В. Феллер

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ И ЕЕ НРИЛОЖЕНИЯ

Том 2

Это второй том учебника по теории вероятностей — первый вышел' двумя изданиями на английском языке и тремя изданиями на русском языке и завоевал заслуженную популярность.

Автор книги — крупный специалист по теории вероятностей. Его учебник написан на высоком научном и методическом уровне и содержит большое число примеров применений теории в физике, биологии и экономике. Данный том посвящен непрерывным распределениям. Вместе с первым томом он составляет прекрасное учебное руководство, в котором очень удачно сочетаются и принципиальные основы, и важнейшие приложения теории вероятностей.

Книга рассчитана на читателей различных уровней — от студентов младших курсов университетов до специалистов-математиков. Она, безусловно, заинтересует также физиков и инженеров различных специальностей, которые в своей работе пользуются вероятностными методами.

ОГЛАВЛЕНИЕ

Предисловие к русскому изданию 5
Предисловие 8
Глава 1. Показательные и равномерные плотности 13
§ 1. Введение 13
§ 2. Плотности. Свертки 16
§ 3. Показательная плотность 21
§ 4. Парадоксы, связанные с временем ожидания. Пуассоновский процесс 24
§ 5. Устойчивость неудач 29
§ 6. Времена ожидания и порядковые статистики 32
§ 7. Равномерное распределение 36
§ 8. Случайные разбиения 40
§ 9. Свертки и теоремы о покрытии 42
§ 10. Случайные направления 46
§ 11. Использование меры Лебега 51
§ 12. Эмппрические распределения 55
§ 13. Задачи 58
Глава II. Специальные плотности. Рандомизация 64
§ 1. Обозначения и определения 64
§ 2. Гамма-распределения 66
§ 3. Распределения математической статистики, связанные с гамма-распределениями 67
§ 4. Пекоторые распространенные плотности 69
§ 5. Рандомизация и смеси 74
§ 6. Дискретные распределения 76
§ 7. Бесселевы фупкции и случайные блуждания 79
§ 8. Распределения на окружности 83
§ 9. Задачи 86
Глава III. Многомерные плотности. Пормальные плотности и процессы 89
§ 1. Плотности 89
§ 2. Условные распределения 95
§ 3. Возвращение к показательному и равномерному распределениям 98
§ 4. Характеризация нормального распределения 102
§ 5. Матричные обозначения. Матрица ковариаций 106

§ 6. Нормальные плотности и распределения 108
§ 7. Стационарные нормальные процессы 114
§ 8. Марковские нормальные плотности 122
§ 9. Задачи 128
Глава IV. Вероятностные меры и пространства 132
§ 1. Бэровские функции 132
§ 2. Функции интервалов и интегралы в R^r 135
§ 3. Вероятностные меры и пространства 142
§ 4. Случайные величины. Математические ожидания 145
§ 5. Теорема о продолжении 149
§ 6. Нроизведения пространств. Носледовательности независимых случайных величин 153
§ 7. Нулевые множества. Нополнение 158
Глава V. Вероятностные распределения в R^r 160
§ 1. Распределения и математические ожидания 161
§ 2. Нредварительные сведения 170
§ 3. Плотности 174
§ 3а. Сингулярные распределения 177
§ 4. Свертки 179
§ 5. Симметризация 186
§ 6. Интегрпрование по частям. Существование моментов 189
§ 7. Неравенство Чебышева 191
§ 8. Дальнейшие неравенства. Выпуклые функции 192
§ 9. Нростые условные распределения. Смеси 196
§ 10. Условные распределения 200
§ 10а. Условные математические ожидания 203
§ 11. Задачи 206
Глава VI. Некоторые важные распределения и процессы 210
§ 1. Устойчивые распределения в R^1 210
§ 2. Нримеры 216
§ 3. Безгранично делимые распределения в R^1 220
§ 4. Нроцессы с независимыми прпращениями 224
§ 5. Обобщенные пуассоновские процессы и задачи о разорении 228
§ 6. Нроцессы восстановления 230
§ 7. Нримеры и задачи 234
§ 8. Случайные блуждания 240
§ 9. Нроцессы массового обслуживания 244
§ 10. Возвратные и невозвратные случайные блуждания 252
§ 11. Общие марковские цепи 258
§ 12. Мартингалы 265
§ 13. Задачи 272
Глава VII. Законы больших чисел. Нрименения в анализе 275
§ 1. Основная лемма. Обозначения 275
§ 2. Нолиномы Бернштейна. Абсолютно монотонные функции 278
§ 3. Нроблемы моментов 280
§ 4. Нрименение к симметрично зависимым, случайным величинам 283
§ 5. Обобщенная формула Тейлора и полугруппы 286
§ 6. Формулы обращения для преобразования Лапласа 288
§ 7. Законы больших чисел для одинаково распределенных случайных величин 290
§ 8. Усиленный закон больших чисел для мартингалов 295
§ 9. Задачи 300

Глава VIII. Основные предельные теоремы 302
§ 1. Сходимость мер 302
§ 2. Специальные свойства 307
§ 3. Распределения как операторы 311
§ 4. Центральная предельная теорема 315
§ 5. Бесконечные свертки 324
§ 6. Теоремы о выборе 325
§ 7. Эргодические теоремы для цепей Маркова 330
§ 8. Правильно меияющиеся функции 334
§ 9. Асимптотические свойства правильно меияющихся функций 339
§ 10. Задачи 344
Глава IX. Безгранично делимые распределения и полугрупны 349
§ 1. Общее знакомство с темой 349
§ 2. Полугруппы со сверткой 352
§ 3. Подготовительные леммы 356
§ 4. Случай конечных дисперсий 358
§ 5. Основная теорема 361
§ 6. Пример: устойчивые полугруппы 366
§ 7. Схемы серий 369
§ 8. Области притяжения 373
§ 9. Различные распределения. Теорема о трех рядах 378
§ 10. Задачи 381
Глава X. Марковские процессы и полугруппы 383
§ 1. Псевдопуассоновский тип 384
§ 2. Вариант: линейные прпращения 387
§ 3. Скачкообразные процессы 389
§ 4. Диффузионные процессы в R^1 394
§ 5. Прямое уравнение. Граничные условия 400
§ 6. Диффузия в многомерном случае 407
§ 7. Подчиненные процессы 408
§ 8. Марковские процессы и полугруппы 413
§ 9. «Показательная формула» в теории полугрупп 417
§ 10. Производящие операторы. Обратное уравнение 420
Глава XI Теория восстановления 423
§ 1. Теорема восстановления 423
§ 2. Уравнение $\zeta = F * \zeta$ 429
§ 3. Устойчивые процессы восстановления 431
§ 4. Уточнения 436
§ 5. Центральная предельная теорема 438
§ 6. Обрывающиеся (невозвратные) процессы 440
§ 7. Применения 444
§ 8 Существование пределов в случайных процессах 446
§ 9. Теория восстановления на всей прямой 448
§ 10. Задачи 453
Глава XII. Случайные блуждания в R^1 456
§ 1. Обозначения и соглашения 457
§ 2 Двойственность 461
§ 3. Распределение лестничных высот Факторизация Винера—Хопфа 466
§ 4 Примеры 472
§ 5. Применения 477
§ 6. Одна комбинаторная лемма 480
§ 7. Распределение лестничных моментов 481
§ 8 Закон арксинуса 484
§ 9. Различные дополнения 489
§ 10 Задачи 491
Глава XIII. Преобразование Лапласа. Тауберовы теоремы. Резольвенты 495

§ 1. Определения. Теорема непрерывности 495
§ 2. Элементарные свойства 500
§ 3. Примеры 502
§ 4 Вполне монотонные функции. Формулы обращения 504
§ 5 Тауберовы теоремы 508
§ 6. Устойчивые распределения 514
§ 7. Безгранично-делимые распределения 516
§ 8 Многомерный случай 519
§ 9. Преобразования Лапласа для полугрупп 520
§ 10. Теорема Хилле—Иосида 526
§ 11 Задачи 530
Глава XIV. Применение преобразования Лапласа 534
§ 1. Уравнение восстановления: теория 534
§ 2. Уравнение типа уравнения восстановления: примеры 536
§ 3. Предельные теоремы, включающие распределения арксинуса 539
§ 4 Периоды заиятости и соответствующие ветвящиеся процессы 542
§ 5. Диффузионные процессы 544
§ 6. Процессы размножения и гибели. Случайные блуждания 549
§ 7. Дифференциальные уравнения Колмогорова 553
§ 8. Пример: чистый процесс размножения 559
§ 9. Вычисление $P(\infty)$ и времен первого прохождения 569
§ 10. Задачи 566
Глава XV. Характеристические функции 569
§ 1. Определение. Основные свойства 569
§ 2. Специальные плотности. Смеси 573
§ 3. Единственность. Формулы обращения 579
§ 4. Свойства регулярности 584
§ 5. Центральная предельная теорема для одинаково распределенных слагаемых 588
§ 6. Условие Линдеберга 592
§ 7. Характеристические функции многомерных распределений 596
§ 8. Две характеризации нормального распределения 600
§ 9. Задачи 603
Глава XVI. Асимптотические разложения, связанные с центральной предельной теоремой 607
§ 1. Обозначения 608
§ 2. Асимптотические разложения для плотностей 609
§ 3. Сглаживание 613
§ 4. Асимптотические разложения для распределений 616
§ 5. Теорема Берри—Эссеена 620
§ 6. Большие отклонения 622
§ 7. Различно распределенные слагаемые 626
§ 8. Задачи 630
Глава XVII. Безгранично делимые распределения 632
§ 1. Теорема о сходимости 632
§ 2. Безгранично делимые распределения 638
§ 3. Примеры. Специальные свойства 644
§ 4. Устойчивые характеристические функции 648
§ 5. Области притяжения 652
§ 6. Устойчивые плотности 657
§ 7. Схема серий 659
§ 8. Класс L 663
§ 9. Частичное притяжение. 666

«Универсальные законы»
§ 10. Бесконечные свертки 669
§ 11. Многомерный случай 670
§ 12. Задачи 671
Глава XVIII. Применение методов Фурье к случайным блужданиям 675
§ 1. Основное тождество 675
§ 2 Конечные интервалы. Вальдовская аппроксимация 678
§ 3 Факторизация Винера—Хопфа 681
§ 4. Обсуждение результатов Применения 684
§ 5. Уточнения 687
§ 6 Возвращения в нуль 689
§ 7. Критерии возвратности 690
§ 8 Задачи 693
Глава XIX Гармонический анализ 6У5
§ 1 Равенство Парсеваля 695
§ 2 Положительно определенные функции 697
§ 3 Стационарные процессы 700
§ 4. Ряды Фурье 703
§ 5 Формула суммпрования Пуассона 707
§ 6. Положнтельно определенные последовательности 710
§ 7. L^2-теория 713
§ 8 Случайные процессы и стохастические интегралы 719
§ 9. Задачи 726
Предметный указатель 736
Именной указатель 744

ПРЕДМЕТНЫЙ УКАЗАТЕЛБ

Абелевы теоремы 511
Абсолютная непрерывность 174, 176
Абсолютно беспристрастная последовательность 265, 295
— монотонные функции 279, 280, 505
Аварии 78, 228
Автобусы (очереди на автобусы, ожидание) 37, 76, 88, 236
Аддитивные функции множества 136
Алгебра множеств 143, 144, 150
— —, порождения 204
Арифметическое распределение 173; см. также Решетчатое распределение
Арксинуса распределения 70
— — в процессах восстановления 539
— — в случайном блуждании 483
— — обобщенное 604
Арцела — Асколи теорема 329
Асимптотическая пренебрежимость 223, 594
Асимптотически несмещенная оценка 277
— плотное (множество) 184
Атомические меры 172
Атомы (мер) 172
— при свертке 183—184
Банаха пространство 313, 414, 557
Башелье процесс; см. Броуновское движение
Безгранично делимые полугрупны 524
— — распределения 220, 351,516,632
— — — в R^r 670, 674
Бернулли испытания и случайный выбор 53
Берштейна полиномы (в R^2 300) 278
Бесконечная дифференцпруемость 312, 353
Бесконечные свертки 324, 379, 669, 674
Бесселевы функции 79—80, 598

— —, безграничная делимость 221, 518, 646
— — в стохастических процессах 81, 385
— —, Лапласа преобразование 503, 504, 550
— —.распределения, связанные с ними 80 и ел., 187, 207, 208
— —, характеристическая функция 573
Бета-интеграл 67
— плотность 70
— — в задаче восстановления 538
Борелевская алгебра 143, 150
Борелевское множество 143
— —, аппроксимация 145, 156
— —, измеримое по 146
— —, соглашения об обозначениях 160
Бохнера интеграл 521—522
Броуновское движение 127, 394, 544
— — в R^r 218, 407
— —, непрерывность траекторий 226, 365, 395
— —, первое прохождение 217, 403, 545
— —, подчиненность 411
— — с упругой силой 397
Бэровские функции 134, 139, 148, 161
Бюффона задача об игле 83
Вальда тождество 469, 492, 679
Вейбула распределение 73
Вейерштрасса теорема об аппроксимации 278
Вероятность вхождения; см. Вероятность достижения
Вероятностные меры и пространства 142, 168
Ветвящиеся процессы 298, 507, 542
Взаимные пары функций 575
Видимость в x-направлении 23
Винера процесс; см. Броуновское движение
Винера—Хопфа интегральное уравнение 470
— — факторизация 456, 465, 494, 681, 690
Вложенные процессы восстановления 241, 446
Водохранилище; см. Хранение запасов
Возвращение в нуль 490, 689
Возраст; см. Длительность
Вполне монотонные функции 282,504, 517
— — — абстрактные 520
— — —.интерполяция 531
Вращения 103, 113
Время возвращения 231
— — максимальное 238, 453
— жизни; см. Длительность, Время возвращения
Выборка, медиана, экстремумы 32
— среднее значение, дисперсия 111
Выпуклые функции 193
— — и мартингалы 271
Вырожденное распределение 108, 113
Вырожденные процессы 118
Гамеля уравнение 358, 366
Гамма-плотность 24, 66
— распределение 23, 66, 500
— —, безграничная делимость 220, 355, 518, 645
— — как предел порядковых статистик 39
— —, подчинение 412
— —, приближение 276
— — рандомизпрованное 80
— функция 66, 86
Гармонические функции 298
Гейгера счетчики 238, 240, 439, 453, 536
Гёльдера неравенство 195

Гильбертовы пространства 715, 720
Гиперболические функции и плотности 87, 574, 603, 645, 670, 710
Гравитационные поля 216, 272
Граничные условия 400, 546
— значения 299
Грина функция 396, 545, 567
Грунппровка данных 17
Гюйгенса принцип 71
Двойные орбиты 51
Двойственность 461
Двустороннее преобразование Лапласа 499
Двусторонняя показательная плотность 69
— — —, факторизация 682
— — —, характеристическая функция 573
Дисперсия 18, 65, 171
— инфинитезимальная 397
— условная 96
Диффузия в генетике 398
— в R^1 394, 532, 544, 567
— в R^r 407
— с упругой силой 397
Длина случайных цепей 262
Длительность, диффузия 404—405
—, мертвый период 238
—, период занятости 452, 553
—, процесс восстановления 234, 273, 441
—, процесс размножения 561
Дробная доля 186, 327
Дробовый эффект 223, 672
Естественный масштаб в диффузии 395
Жордана разложення 173
Задача о разорении 228, 250, 388, 538
— — —, оценки 444, 479
Задержки в движении 237, 445, 454, 544, 567
Законы больших чисел 274, 290, 346, 502, 586
— — — для мартингалов 299
— — — для серий 378, 674
— — — для стационарных последовательностей 301
— — —,обратная теорема 292
— нуля или единицы 157
Заражение 79
Значащие цифры 85
Иенсена неравенство 193, 271
Измеримое пространство 144
Измеримость 146
Изометрия 722
Индикатор множества 132
Интегральное уравнение Абеля 51
Интегрпрование по частям 189
Интервал непрерывности 303
Инфинитезимальная скорость и дисперсия 397
Ионизация 385
Исчезающие на бесконечности функции 303
«Исчезновение», случайные блуждания 490
Каноническая мера 633
Кантора тип распределения 54, 177, 670
— — — (свертки) 186
Квазиустойчивость 215
Квазихарактеристические функции 632
Ковариация 91
—, матрица 107
— процесса 116, 700, 725
Колмогорова дифференциальные уравнения 293, 553
— неравенство 296, 301
— Смпрнова теорема 57, 405
— теорема о трех рядах 379
Компактность в схеме серий 370
Конкордантные (согласованные)

функции 266, 298
Координатные случайные величины 17, 91
Корреляция 91
Коши — Леви теорема 595, 600
— распределение 71, 87, 292, 599
— — в R^r 94, 97, 128, 599, 671
— — и броуновское движение 218, 411
— —, случайное блуждание 259, 693
— —, устойчивость 216, 634, 645
Крамера оценка для вероятности разорения 229, 444, 471, 480
Критерий значимости 201
Кронекера лемма 299
— символ 554
— ядро 260, 554
Круг, распределение на нем 46, 83, 88, 104
— — равномерное на нем 327, 333
—,теорема покрытия 44
Кэмпбелла теорема 224, 347, 672
Лапласа второй закон распределения 70
— преобразование 288, 495, 534
Лапласа преобразование $вЯ^{1}$' 519
— — двустороннее 499
— — для полугрупп 520
— Стильтьеса преобразование 489
Лебега интеграл 135
— мера 51, 142, 159
— Пикодима теорема 176
— пополнение 159
— Стильтьеса интеграл 136, 139, 166
— теорема о разложении 179
Леви каноническое представление 643
— Крамера теорема 600
— метрика 345
— Парето распределение 219
—,примеры 272, 382, 646
Лежандра формула удвоения 86
Лестничные величины 241, 457, 481
— — в процессе ожидания 248
Линдеберга условие для диффузии 395
— условия 320, 346, 592
Линейные операций над стохастическими процессами 702
— прпращения в скачкообразных процессах 387
Линейный функционал 151
Логарифм комплексных чисел 639
Логарифмическое распределение 86
Логистическое распределение и развитие 73
Локально компактные пространства 152, 156, 303
Ляпунова условие 346
Мажорпрованная сходимость 141
Максвелла распределение 49, 68, 103
Максимальная оценка 479
— характеристическая функция 685, 694
— частная сумма 242, 250, 383, 470, 477, 487
Максимальное время возвращения 238, 453
Максимальный член 207, 215, 337, 347
— —, влияние на сумму 532; см. также Порядковые статистики, Рекордные значения
Марковские процессы с дискретным временем 122, 129, 258, 274
— — — — — — и мартингалы 298
— — — — — —, спектральная функция 713, 728
Марковские процессы с дискретным временем, эргодические теоремы 330
— — с непрерывным временем 125, 383, 700
— — — — — — в процессах

восстановления 436
— — — — — в счетных пространствах 553
— — — — — и полугруппы 413, 521
— — — — —, эргодические теоремы 446, 564
Марковское свойство 22; см. Строго марковское свойство
Мартингалы 265, 295
— в процессах восстановления 431
— в случайных блужданиях 466
—, неравенства 295, 296, 301
Масштабные параметры 64, 171
Медиана 32, 172
Медленно меияющиеся фупкции; см. Нравильно меияющиеся функции
Мера скорости (speed measure) 395
«Мертвый» период 240
Метод ограничений 403
Метрики 345; см. также Банахово пространство, Гильбертово пространство
Минимальное решение, Винера — Хопфа уравнение 469
— —, дифференциальное уравнение Колмогорова 555
— —, диффузия 402
— —, полумарковские процессы 568
— —, скачкообразные процессы 393
— —, уравнение восстановления 255
Миттаг-Леффлёра функция 519
Момент регенерации 231
Моменты 18, 170, 189
— восстановления 231, 432
— в процессах восстановления 442
—,неравенство 195
—, проблема однозначности 283, 290, 532, 527
— — — в R^r 606
—, производящие функции 499
—, сходимость 306, 328
—, Хаусдорфа проблема 281
Монотонной сходимости принцип 140
— — свойства 414
Найквиста формула 708
Направляющий процесс 411
Невозвратные случайные блуждания 253
Невозвратные случайные блуждания, свойства 257, 465, 483, 690, 694
— — —, уравнение восстановления и теорема для них 252 449, 450, 494
Невозможность систем игры 270, 271
Независимость случайных величин 19,92, 95, 153, 169
— — —, комплексные величины 570
— — —, критерий для нее 170
Независимые прпращения 123, 224, 352, 365, 380, 724
— —, разрывы 366, 380
Неймана тождество 82
Нелинейное восстановление 454
Непосредственная интегрпруемость по Риману 426
Непрерывность полугрупп 416; см. также Фикспрованные разрывы
Несмещенная оценка 277
Несобственные (дефектные) распределения 162, 164, 260
— — — в процессах восстановления 234
Неудачи 29
Норма 313, 413, 714, 720
—, топология, порождаемая 345
Нормальное распределение 65, 87, 574
— — в R^r 108, 597
— — вырожденное 112
— — двумерное 93, 96, 129
Нормальное распределение марковское 122
— —.область притяжения 377, 652
— — характеристические свойства

— — частное 128, 206
Нормальные полугруппы 360, 368, 381
Нормальный стохастический процесс
Носитель 64
Нулевая схема серий 222, 659
Нулевое множество 158, 175 714 720
Область притяжения 214
— —.критерий для нее 373, 382 514 652
— — нормальная 657
— — частичная 382, 646, 656
Обобщенный пуассоновский процесс 225, 365, 388
— — — и полугруппы 357, 360
— — —, подчинение в нем 412
— — —, разорение в нем 228, 250, 538
Образование колонн 59, 567
Обратное уравнение для диффузии 396, 398 и сл., 407, 545 и сл.
— — — полугрупп 417, 420 и т. д.
— — — полумарковских процессов 568
— —, минимальное решение 392, 555 и т. д.
— —, скачкообразный процесс 390, 554 и т. д.
Обрывающийся (terminating) или невозвратный (transient) процесс 234, 440 и сл.
Обслуживание больных 229
Обслуживающие устройства; см. Очереди
Одновершинные [упимодальные] распределения 197, 603
— —,свертки 208
Оператор перехода 414
— сдвига 722
Операторы, связанные с распределениями 311
Операционное время 227, 409
Орбиты двойные 51
Орнстейна—Уленбека процесс 127, 397
Ортогональные матрицы 114
Остаточные события 157
Остающееся время ожидания 236
— — —, предельная теорема 434, 436, 453, 540
Осциллпрующие случайные блуждания 258, 464, 687
Отношение правдоподобия 267
Отражающий экран 402, 406, 532
Оценка 60
Очереди 75, 76, 87, 88, 248
— параллельные 31, 32, 35, 60; см.также Нериоды занятости
Ошибки округления 37, 84
Нарадокс контроля 235, 436
Нарадоксы 24, 38, 235, 436
Нараметры расположения 64, 171
Нарето распределение 70
Нарсеваля равенство 530, 695, 716, 719
— — как критерий Хинчина 718
Нервое возвращение 459, 491, 566, 689
Нервые прохождения, диффузия 217, 403, 544
— —, марковские цепи 562
— —, процесс размножения и гибели 82, 88, 551, 565
Нериодограмма 101
Нериоды заиятости 246, 251, 553
— — и ветвящийся процесс 542, 517
Нетербургская игра 293
Нпрсона система распределений 67, 70
Планшереля преобразование 715
— теорема 582
Плотность 16, 64, 174
Новторные математические ожидания 203

Поглощающие экраны 402 и т. д., 532, 546 и т. д
Поглощение (физическое) 41, 48, 385
Подчиненные процессы 408, 518, 525, 646, 673
— — и показательная формула 419
Пойа критерий 577, 581
— урновая схема 266, 285, 297
Показательная формула для полугрупп 287, 409, 417
Показательное распределение 21, 58, 98
— — двумерное 128
— — двустороннее 69, 187
— — и равномерное распределение 62, 98
Показательное распределение как предел 39, 62, 434; см. также Гамма-распределения
Положительно определенные матрицы 107
— — последовательности 710
— — функции 697
Полугруппы 286, 413, 520
— со сверткой 353
— устойчивые 366
Полумарковский процесс 553, 568
Полумартингалы 271
Полярные координаты 92
Пополнение мер 159
Популяция, рост 399, 401
—,случайное рассеивание 319
Порядковые статистики 32, 36, 129
— —, предельные теоремы 39, 62, 87
— —, применение к статическому оцениванию 60
Потенциал 558
Потеря энергии 385, 387; см. Столкновение частиц
Потерянные вызовы 239, 566
Правильно меияющиеся функции 334, 348
Приложение к астрономии 51, 115 170—171, 261, 272, 388
Принцип инвариантности 400
— отражения 218, 547, 548
Притяжение; см. Область притяжения
Проекция 90
— случайных векторов 47, 49, 62
Произведение мер и пространств 153
Производящие операторы 354, 420, 522, 545
— фупкции, интерпретация с помощью «исчезновения» 490
— —, характеризация 279
Произвольная остановка 270
Пропуски (пробелы) большие 236, 445, 536
— малые 273
Прохождение света через вещество 41, 47, 62
— — в звездных системах 261, 388
— — и принцип Гюйгенса 71
— — через сферу 47
Процесс авторегрессии 116, 123
— восстановления с запаздыванием 234, 431
Прочность на разрыв 22
Прошедшее время ожидания 236, 453
— — —, предельные теоремы 436, 540
Прямое уравнение, диффузия 400, 544
Прямое уравнение, диффузия, минимальное решение 393, 556
— —,полугруппы 417
— —, скачкообразный процесс 387, 391
— — — — (счетность) 554
Прямоугольная плотность 36, 70
Псевдопуассоновский процесс 409
— —, показательная формула 417
— —, полугрупна 526

— — с линейными прпращениями 387
Пуанкаре задача о рулетке 84
Пуассона формула суммпрования 85, 405, 707, 727
— ядро 704
Пуассоновские ансамбли 28; см. также Гравитационные поля
— распределения, аппроксимация с их помощью 346
— — как крайние точки 644
— —, разность 207, 645
— —, формула Тейлора 286
Пуассоновский процесс 24
— — как предел в процессах восстановления 434
— —, supremum 299
Равномерное распределение 36, 327
Равностепенная непрерывность 307, 329
Радиация звездная 261
Радона—Пикодима производная 174
— — теорема 176
Развитие логистическое 73
Разложение единицы 722
Размножения и гибели процессы 549, 567
— — — — и диффузия 567
— — — —, периоды занятости в них 552; см. также Рандомизованные случайные блуждания
— процесс 60, 325, 559
— — двусторонний 561
Разности (обозначения) 277
Разрывы фикспрованные 227, 380
Рандомизпрованное случайное блуждание 81, 549, 644
Рандомизация и подчиненные процессы 409
—, полугруппы 288, 419
—, симметрично зависимые случайные величины 283
Распределение роста 97
Расслоение 78
Расстояние между распределениями 345
— по вариации 345
Расхождение; см. Эмпприческое распределение
Регистрация 240, 439
Регрессия 94, 112
Регулярное стохастическое ядро 331
Резольвенты 495, 527, 557
— и полная монотонность 529
Рекордные значения 29, 58, 59; см.также Порядковые статистики
Решающая функция 269
Решетчатые
Распределения 173, 184
— —, характеристические функции 572, 583
— —,центральная предельная теорема для 591, 617
Римана — Лебега теорема 586
Рисса теорема о представлении 152
Рулетка 37, 84
Самовосстанавливающиеся
Совокупности 234
Свертка ядер 261
Свертки, бесконечные 324, 379, 669, 674
— —, поведение F^{n*} 347
— на окружности 86, 333
— плотностей 19, 65, 95
— распределений 179
— — (сингулярных распределений) 186
Свободный пробег 22, 29
Сглаживание 115, 181, 613
Сдвиг, оператор 287, 313
—, полугруппы 355, 421
— принцип 499
Сепарабельность 328
Сервостохастический процесс 130

Серии 62; См. также Рекордные значения
Сжатие 414, 521
Сильная непрерывность в нуле 416, 521
— сходимость 314, 416
Симметризация 186
—, неравенства 188
Симметрично зависимые величины 283, 489
— — —, центральная предельная теорема 346
Симметричные события 157
Сингулярные (вырожденные) распределения 54, 137, 177
— —, свертки 186, 670
Скалярное произведение 714
Скачкообразный процесс 389
— — с бесконечным числом скачков 393, 554
Скользящее среднее 115, 723
Скрытая периодичность 101
Слабая сходимость 304
Случайная величина 16, 91, 145, 165
— — комплексная 569
Случайное рассеивание 319
Случайные блуждания в R^1 240, 252, 456, 575
— — — — — и эмпирическое распределение 56, 62
— — — — — простые (Бернулли) 270, 381, 460, 462, 491, 503, 680
— — — — — сопряженные 474
— — — в R^r 50, 598
— — — — —, центральная предельная теорема 319
— импульсы 240
— направления 46, 61, 62
— —, сложение 48, 71, 186, 262, 598
— разбиения 37, 98, 101; см. Теоремы о покрытии
— — (расщепления частиц) 40, 61, 129
— суммы 75, 199, 208
— —, характеристическая функция 576
— —, центральная предельная теорема для них 323, 605
— цепи 261
Случайный вектор 48, 50; см. также Случайные направления
— выбор 14, 36, 32
— — и бросание монеты 53; см. также Теоремы о покрытии, Случайные разбиения
Смеси 74, 98, 196, 208
—, преобразование 503
Снедекора плотность 69, 70
Снос в диффузии 397
— в случайных блужданиях 465, 686
Собственное распределение 164
— —, сходимость 304, 344, 633
Совместное распределение 91
Совпадения 273
Сопряженные случайные блуждания 474
Спектр мощностей 701
Спектральная мера 701, 726 (заданная 703)
— —, представление для унитарных операторов 722
Статистика 32
Стационарность, мера и вероятность 262, 274, 332
—, прпращение 225, 352
—, процесс 114, 129, 700, 710 (закон больших чисел 301) (Ср. Эргодическая теорема)
Стационарный режим 27, 263, 388
— — в процессах восстановления 434
Столкновение частиц 261, 385, 387
Стохастическая ограниченность 310, 369

Стохастические интегралы 724
Стохастическое разностное уравнение 128
— ядро 199, 259, 330
Строго марковское свойство 38
— устойчивые распределения 211
Структура (алгебраическая) 414
Стьюдента плотность 69, 72
Субмартингал 271, 296
Субстохастическое ядро 260
Сумма случайных векторов 48, 186, 262, 598
Суммпруемость по Абелю 704, 706, 728
— — Чезаро 706, 727
Суперпозиция процессов восстановления 434
Схемы серий 222, 272, 369, 659, 674
Сходимость в среднем квадратичном 714
— мер 304, 326, 344
— операторов 314, 353, 414
— по вероятности, 290, 309
— сильная 314, 416
— слабая 305
Счетно-аддитивные функции (σ-аддитивные) 137, 150—151, 163
Счетчики; см. Гейгера счетчики, Очереди
Тауберовы теоремы 508
— —, применение 486, 514, 535, 540, 687
Тейлора формула обобщенная 286
Телефонные вызовы 78, 228
— — потерянные 239, 266; см. также Периоды заиятости
Теорема непрерывности, квазихарактеристические фупкции 634
— —, Лапласа преобразования 496
— —,полугруппы 327
— —, характеристические функции 580, 582, 604
Теорема об аппроксимации в среднем 141
— о продолжении 149, 156
— — среднем значении 138
— — трех рядах 379
Теоремы о выборе 325
— — покрытии 42, 61, 101, 273, 538
Теория надежности 73
— страхования 229; см. также Задача о разорении
Типы распределений 64, 171
— —, сходимость 307
Точка роста 172, 184
Точки первого вхождения в процессах восстановления 236, 434, 435, 492; см. также Лестничные величины
— — — в случайных блужданиях 490, 492, 675
Транспортные задачи 59, 237, 445, 454, 544, 567
Треугольная плотность 43, 70, 573
Треугольника неравенство 313
Тэта-функции 405, 406, 710
Универсальные законы Дёблина 568
Унимодальность по Хинчину 603
Унитарные операторы в гильбертовом пространстве 722
Упругая сила 397
Уравнение восстановления 232, 425, 436, 453
— — для всей прямой 253, 448, 494
— —, процесс 25, 230, 272, 431, 453
Уравнения восстановления, теоремы 424, 427, 441
Условные распределения и математические ожидания 95, 200
Устойчивые полугруппы 366
— распределения в процессах восстановления 440

— — R^2 647, 671
— —, плотности 657
— —, подчиненность 412, 518
— —, положительные 514, 673
— —, произведения 219, 519
— — с показателем 1/2 72, 87, 216, 381, 503
Фату теорема 140
Фиксированные разрывы 380
Фильтр 116, 702
Фишера Z-статистика 69
Фоккера — Планка уравнение 356, 386, 391
Формула обращения и проблема моментов 281
— —, преобразование Лапласа 288 506, 507, 530
— —, характеристические функции 581, 584, 599
— удвоения 86
Фубини теорема 141, 155, 181
Функции интервалов 136, 162
Функционал линейный 151
Фурье коэффициенты 705, 708
— обращение 581, 583
— ряд 703, 719, 727
— преобразование 570, 608; см. также Планшереля преобразование
— Стильтьеса преобразование 570
Характеристические функции 569
— — в R^r 596
— —, логарифм 639
— — периодические 703, 706
— —, производная в нуле 635
— —, факторизация 578, 669, 708; см. также Пуассона формула суммпрования
Хи-квадрат 68
Хилле — Иосида теорема 526, 545
Хинчина критерий 718
— Полячека формула 478, 539, 683, 693
Хольцмарка распределение 215, 217, 272
Хранение и управление запасами 228, 229, 247
Центральная предельная теорема 315, 346, 350, 588, 605
— — —, необходимые и достаточные условия 379, 654, 662
— — —, приложения 257, 264, 319
— — —, процессы восстановления 438
Центрпрование 64, 171, 660
— безгранично делимых распределений 637
Цепи, ирочность 22
— случайные 261
Цифры, распределение 52, 85
Частицы, притяжение 382, 646, 666-
—, расщепление 40, 61, 129
— — быстрыми частицами 385, 387; ср. Столкновения частиц
—, унорядочение 107, 167
Частные (маргинальные) распределения 90, 169, 196
— —, заданные 206
Частотная характеристика фильтра 702
Чебышева неравенство 191, 276
— — обобщение 299
— — — на мартингалы 301
— Эрмита многочлен 609
Чепмена — Колмогорова уравнение, дискретное время 127
— — —, непрерывное время 384, 556
— — —, полугрупны 416, 528
Шаг 174
Шварца неравенство 192
Шлемильха формула 82
Шум 346, см также Дробовой эффект
Эджворта разложение 611, 620

Эквивалентные функции 158, 714
Экраны, см Поглощающие экраны, Граничные условия
Эмпирические расиределения 55
Эндоморфизм 414
Эргодические стохастические ядра 332
— теоремы, дискретное время 263, 274, 330
— —, неирерывное время 446, 562, см Стационарный режим
Эрлангова плотность 66
Эрмита полиномы 609, 630
— разложение 620
Ядра 199, 259, 330
Якоби тэта-функция 405, 406, 710
— σ-аддитивность 137
— алгебра 143
— конечная мера 144

ИМЕННОЙ УКАЗАТЕЛЬ

Амбарцумян В. А. 261, 388
Андре (Andre D.) 2)8
Аннекен (Hennequin P. L.) 131, 729
Ьайес (Bayes T.) 77
Ьакстер (Baxter G.) 472, 481, 490, 647, 681
Ьартлетт (Bartlett M. S.) 730
Ьаруча-Рид (Barucha-Reid A. T.) 399, 730
Ьенеш (Benes E. V.) 446, 730
Ьенфорд (Benford F.) 85
Ьергстрем (Bergstrom П.) 657
Ьернштейн С. П. 103, 505
Ьерри (Berry A. C.) 607, 620, 621, 625, 627, 630
Ьикель (Bickel P. J.) 455
Ьиллингслей (Billingsley P.) 57, 323, 406
Ьлекуэлл (Blackwell D.) 424, 449
Ьлюм (Blum J. R.) 346
Ьоль (Bohl) 327
Ьоттс (Botts T. A.) 131
Ьохнер (Bochner S.) 383, 411, 520, 697, 699, 712, 729
Ьрело (Brelot M.) 299
Ьудро (Boudreau P. E.) 248
Ьурбаки (Bourbaki N.) 131
Ьюльман (Buhlman П.) 285
Вальд (Wald A.) 283, 465, 474
Вебер (Weber П.) 504
Вейль (Weyl П.) 327
Веисс (Weiss B.) 728
Виленкин П. Я. 699, 729
Винер (Wiener N.) 426, 718
Винтнер (Wintrier A.) 209
Волд (Wold П.) 597
Вольфовиц (Wolfowitz J.) 252, 153, 447, 449
Гальтон (Gallon F.) 97
Гейтлер (Ieitler W.) 385
Гельфанд И. М. 699, 729
Герглотц (Herglotz G.) 712
Гири (Geary R. C.) 111
Гнеденко Ь. В. 57, 62, 337, 607, 618, 657, 668, 730
Гренандер (Grenander U.) 102, 729, 730
Гринвуд (Greenwood M.) 79
Гриффин (Griffin J. S.) 248
Гуд (Good D. J.) 542
Гумбел (Gumbel E. J.) 206
Дарлинг (Darling D. A.) 533
Дворецкий А. 333
Дерман (Derman C.) 564
Деблин (Doblin W.) 215, 668
Джильберт (Gilbert E.) 273
Донскер (Donsker M.) 57, 406
Дуб (Doob J. L.) 130, 206, 266 297, 299 431, 730
Дынкин Е. Ь. 383, 395, 540, 729
Зигмунд (Zygmund A.) 635
Золотарев В. М. 659

Иахав (Yahav J. A.) 455
Ибрагимов И. А. 209
Ито (Ito К.) 395, 400, 729
Кантор (Cantor G.) 326
Карамата (Karamata J.) 215, 302, 334, 339 512
Карлеман (Carleman T.) 282, 587
Карлин (Carlin S.) 2S3, 449, 730
Кац (Кас М.) 103, 248, 406
Кельвин (Kelvin) 403
Кемперман (Kemperman J. H. B.) 472, 729
Кендалл (Kendall D. C.) 244, 288, 542
Кенуй (Quenoinlle M. H.) 50
Кимура (Kimura M.) 399
Кингман (Kingman J. F. C.) 231
Кифер (Kiefer J.) 252, 447
Кокс (Cox D. R.) 730
Колмогоров А. Н. 155, 156, 206, 224, 292, 299, 395, 607, 618, 730
Королюк В. С. 56, 57, 62
Коши (Cauchy A.) 215, 581
Крамер (Cramer H.) 597, 607, 620, 628, 625, 730
Крикеберг (Knckeberg К.) 729
Ламперти (Lamperti A) 238
Ландау Л. Д. 385
Ландау (Landau F.) 405, 513
Лаха (Laha R. G.) 729
Лебег (Lebeque H.) 152
Леви Ноль (Levy Paul) 215, 224, 226, 266, 320, 365, 380, 400, 568, 647, 663, 668, 729, 730
Ле Кам (Le Cam L.) 346
Линдеберг (Lindeberg J. W.) 585
Линдли (Lindley D. V.) 244, 456
Литтл (Little J. D. C.) 544
Литтльвуд (Littlewood J. F.) 195, 512
Лоэв (Loeve M.) 131, 132, 266, 285, 297, 323, 383, 729
Лукач (Lukacs F.) 111 729
Лундберг (Lundberg F.) 229
Маккин (McKean H. P.) 395 400, 729
Макшейн (McShane E. J.) 131
Мандельбройт (Mandelbrot B.) 219, 348
Марков А. 283
Маршалл (Marshall A. W.) 301
Мидлтон (Middleton D.) 709
Миллс (Mills H. D.) 130
Моран (Moran P. A. P.) 730
Мюнц (Muntz C.) 300
Неве (Neveu J.) 131, 297, 729
фон Нейман (Neumann J.) 62
Нейман (Neyman J.) 228
Нелсон (Nelson E.) 128 411
Ньюэл (Newell G. F.) 59
Орей (Orey C.) 449
Орнштейн (Ornstem D.) 257
Найк (Pyke R.) 229, 457, 539, 568
Нетерсен (Petersen D. P.) 709
Нетров В. В. 625
Нинкхэм (Pinkham R. S.) 86
Нитман (Pittman E. J.) 635
Нойа (Polya G.) 79, 195, 215, 228
Ноллард (Pollard H.) 424, 449, 520
Ноляк (Pollak H. O.) 273
Нолячек (Pollaczek F.) 250
Норт (Port S. C.) 339
Нрабу (Prabhu H. U.) 730
Нрохоров Ю. В. 57, 406
Райт (Wright S.) 399
Райт (Wright E. M.) 327
Рвачева Е. Л. 57
Реньи (Renyi A.) 406
Риордан (Riordan J.) 730
Рисс (Riesz M.) 288
Рихтер (Richter W.) 625
Роббинс (Robbins H. E.) 128, 323, 423
Розенблат (Rosenblatt M.) 102, 346, 730
Рой (Roy S.) 423
Ройден (Royden H. L.) 283
Рэй (Ray D.) 395

Рэлей (Rayleigh) 50
Севидж (Savage J. L.) 157, 285
Серпинский (Sierpinski W.) 327
Скеллам (Skellam J. G.) 320
Скитович В.Н. 104
Скороход А. В. 57, 730
Смирнов Н. В. 57, 62, 405
Смит (Smith W. L.) 423, 449, 568, 730
Снелл (Snell J. L.) 431
Спарре-Андерсен (Sparre-Andersen E.) 456, 481, 486, 489
Спицер (Spitzer F.) 197, 374, 386, 395, 569, 575, 580, 615
Стадден (Studden W.) 283, 730
Стейн (Stein C.) 678
Стивенс (Stevens W. L.) 44, 102
Стильтьес (Stieltjes) 283
Сьюэлл (Sewall) 399
Такая (Takacs L.) 542, 730
Тамаркин Я. Д. 273, 301
Таннер (Tanner J. C.) 237
Тейчер (Teicher H.) 346
Титчмарш (Titchmarsch E. C.) 688
Тортра (Tortrat A.) 676
Троттер (Trotter H. P.) 320, 411
Угахери (Ugaheri T.) 274
Уилкс (Wilks S. S.) 59
Уильямсон (Williamson R. E.) 431
Уиттекер (Whittaker J. M.) 709
Уокс (Wax N.) 730
Уолл (Woll J. W.) 411
Уоллес (Wallace D. L.) 621
Уолш (Walsh J. W.) 71
Фейер (Fejer L.) 706
Филлипс (Phillips R. S.) 288, 354, 520, 525, 730 де
Финетти (Finetti B.) 61, 224, 283
Фишер (Fisher R. A.) 102, 337, 399
Фреше (Frechet H.) 207
Фукс (Puchs W. H. Y.) 690
Халмош (Halmos P. R.) 131, 270
Хант (Hunt G.) 266
Харди (Hardy G. H.) 195, 327, 512
Харрис (Harris
T. E.) 298
Хаусдорф (Hausdorff F.) 281
Хелли (Helly E.) 326
Хилле (Hille E.) 286, 288, 354, 476 520, 575, 601, 730
Хинчин А. Я. 215, 224, 291, 474, 643, 664, 668, 730
Хобби (Hobby C.) 457
Хопф (Hopf E.) 471, 472
Хьюитт (Hewitt E.) 157, 285
Чандрасекар (Chandrasekhar S.) 50
Чернов (Chernoff H.) 346, 631
Чжоу (Chow S.) 423
Чжун (Chung K. L.) 134, 208, 256, 257, 288, 420, 449, 553, 690, 729
Шапиро (Shapiro J. M.) 647
Шварц (Schwartz L.) 699
Шеллинг (Schelling H.) 264, 285
Шеннон (Shannon C.) 708
Шепп (Shepp L.) 87, 197, 694, 728
Шоке (Choquet G.) 429, 670
Шохат (Shohat J. A.) 283, 301
Эйнштейн А. (Einstein A.) 395
Эйнштейн (Einstein A. Jr.) 228
Эссеен (Esseen G.) 607, 618, 620, 621
Эсшер (Esscher F.) 625
Эрдёш (Erdos P.) 406, 424
Юл (Jule G. U.) 79
Яноши (Janossy L.) 385

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

Второй том книги известного американского математика В. Феллера «Введение в теорию вероятностей и ее приложения» вышел в США в 1966 г. Напомню, что первый том, посвященный дискретным распределениям, дважды издавался в США в 1950 году и (с изменениями и дополнениями) в 1958 году. Оба издания были переведены на русский язык (в 1952 и 1964 годах).

Несмотря на большой промежуток времени между выходом первого и второго томов, оба тома имеют общий замысел и составляют единое целое. Книга дает строгое изложение теории вероятностей как самостоятельного раздела математики и в то же время знакомит читателя с опытными основаниями теории и различными применениями. Последнее достигается включением большого числа примеров и задач.

Книга очень богата содержанием (достаточно сказать, например, что гл. XV, XVI и XVII второго тома полностью поглощают известную монографию Б. В. Гнеденко и А. Н. Колмогорова «Предельные распределения для сумм независимых случайных величин»). Многочисленные отступления от основного текста содержат сведения, интересные и специалистам. Большая работа, проведенная автором при подготовке второго тома, позволила существенно упростить изложение целых разделов (например, вывод асимптотических формул в теории случайных блужданий и задачах о разорении и многое другое).

Изложение тщательно продумано. Оно часто дополняется замечаниями, отражающими отношение автора к приводимым фактам (см., например, замечание в гл. VI, 7 о том, что рассмотрение в теории очередей потоков вызовов с независимыми промежутками между вызовами и с законом распределения длины этих промежутков, отличным от вырожденного или показательного, создает лишь иллюзию общности; трудно, говорит автор, найти примеры таких процессов, кроме разве движения автобуса по кольцевому маршруту без расписания). Напомню, что

в первом томе автор удачно продемонстрировал тот факт, что сравнительно простые модели позволяют хотя бы в первом приближении правильно описать широкий круг практических задач (такими моделями в первом томе являются, например, размещения частиц по ячейкам, урновые схемы и случайные блуждания). Во многих случаях, где интуиция не подсказывает правильного порядка соответствующих вероятностей, автор приводил численные результаты. Подчеркивался ряд свойств случайности, идущих в разрез с интуитивными представлениями (закон арксинуса, пропорциональность времени до n-го возвращения величине n^2 и т. п.). Эти тенденции сохраняются и во втором томе (см., в частности, неоднократное обсуждение парадокса инспекции и сходных тем).

После того, как разобран дискретный случай, элементарная теория непрерывных распределений требует лишь нескольких слов дополнительного объяснения. Поэтому до всякой общей теории, в первых трех главах, автор излагает задачи, связанные с тремя важнейшими распределениями — равномерным, показательным и нормальным. При этом автор затрагивает и ряд глубоких вопросов (случайные разбиения и теоремы о покрытиях, отклонение эмпирического распределения от теоретического, характеризация нормального распределения независимостью статистик, структура некоторых стационарных нормальных последовательностей и т. п.).

Необходимые сведения из теории меры сообщаются в четвертой главе. Автор подчеркивает вспомогательный характер этих сведений (что отличает книгу Феллера от ряда других современных руководств, где до трети объема уходит на теорию меры и интеграла).

Важную роль играет гл. VI, являющаяся, как замечает автор, собранием введений во все последующие главы.

В настоящем томе, по-видимому, впервые излагается так обстоятельно теория преобразований Лапласа в применении к вероятностным проблемам (гл. VII, XIII, XIV). Много места отведено вопросам теории восстановления и случайным блужданиям (гл. VI, XI, XII, XIV и XVIII). Предельные теоремы для сумм независимых величин излагаются дважды: сначала как иллюстрация операторных методов (гл. IX), а затем в общей форме, в гл. XVII.

Можно надеяться, что выход в свет второго тома книги Феллера окажет заметное воздействие на многие стороны развития теории вероятностей, в частности сильно повлияет на характер преподавания теории вероятностей, позволив, наконец, привести его в соответствие с современными требованиями.

Профессор Феллер, узнав о подготовке перевода второго тома, любезно прислал список ряда необходимых исправлений, которые были внесены в текст. Я весьма благодарен ему за эту любезность. В работе над переводом существенную помощь мне оказали А. В. Прохоров и В. В. Сазонов. Они внимательно прочли рукопись перевода и сделали ряд ценных замечаний, которыми я воспользовался. Они также указали ряд неточностей в оригинале. Пользуюсь случаем выразить им свою глубокую благодарность.

9 мая 1967 г. *Ю. Прохоров*

ПРЕДИСЛОВИЕ

В то время когда писался первый том этой книги (между 1941 и 1948 гг.), интерес к теории вероятностей еще не был столь широким. Преподавание велось очень ограниченно, и такие вопросы, как цепи Маркова, которые теперь интенсивно используются в самых различных дисциплинах, рассматривались лишь как весьма специальные главы чистой математики. Поэтому первый том можно уподобить годному на все случаи жизни путеводителю для человека, отправляющегося в чужую страну. Чтобы отразить сущность теории вероятностей, в первом томе нужно было подчеркнуть как математическое содержание этой науки, так и удивительное разнообразие потенциальных применений. Некоторые предсказывали, что неизбежно возникающие из этой задачи колебания степени трудности изложения ограничат полезность книги. На самом деле первым томом книги широко пользуются и сейчас, когда новизна этой книги пропала, а ее материал (включая способ изложения) можно почерпнуть из более новых книг, написанных для специальных целей. Книга, как нам кажется, все еще завоевывает новых друзей. Тот факт, что неспециалисты не застряли в местах, оказавшихся трудными для студентов-математиков, показывает невозможность объективного измерения уровня трудности; этот уровень зависит от типа информации, которую ищет читатель, и от тех деталей, которые он готов пропустить. Так путешественник часто стоит перед выбором: взбираться ли ему на гору самому или использовать подъемник.

Ввиду сказанного второй том написан в том же самом стиле, что и первый. Он включает более трудные разделы математики, но значительная часть текста может восприниматься с различной степенью трудности. Поясним это на примере способа изложения теории меры. Гл. IV содержит неформальное введение в основные идеи теории меры, а также дает основные понятия теории вероятностей. В этой же главе перечисляются некоторые факты теории меры, применяемые в последующих главах для того, чтобы сформулировать аналитические теоремы в их простейшей форме и избежать праздной дискуссии об условиях ре-

гулярности. Основное назначение теории меры в этой связи состоит в том, чтобы оправдать формальные операции и переходы к пределу, чего никогда не потребовал бы нематематик. Поэтому читатели, заинтересованные в первую очередь в практических результатах, никогда не почувствуют какой-либо необходимости в теории меры.

Чтобы облегчить доступ к отдельным темам, изложение во всех главах сделано столь замкнутым, сколь это было возможно. Иногда частные случаи разбираются отдельно от общей теории. Некоторые темы (такие, как устойчивые законы или теория восстановления) обсуждаются в нескольких местах с различных позиций. Во избежание повторений определения и иллюстративные примеры собраны в гл. VI, которую можно описать как собрание введений к последующим главам. Остов книги образуют гл. V, VIII и XV. Читатель сам решит, сколько подготовительных глав нужно ему прочитать и какие экскурсы произвести.

Специалисты найдут в этом томе новые результаты и доказательства, но более важной представляется попытка сделать единой общую методологию. В самом деле, некоторые части теории вероятностей страдают от того, что они недостаточно внутренне согласованы, а также от того, что группировка материала и способ изложения в большой степени определяются случайностями исторического развития. В образующейся путанице тесно связанные между собой проблемы выступают разобщенными, а простые вещи затемняются усложненными методами. Значительные упрощения были достигнуты за счет систематического применения и развития наилучших доступных сейчас методов. Это относится, в частности, к такой беспорядочной области, как предельные теоремы (гл. XVI—XVII). В других местах упрощения были достигнуты за счет трактовки задач в их естественном контексте. Например, элементарное исследование специального случайного блуждания привело к обобщению асимптотической оценки, которая ранее была выведена тяжелыми, трудоемкими методами в математической теории страхования (и независимо, при более ограничительных условиях, — в теории очередей).

Я пытался достичь математической строгости, не впадая в педантизм. Например, утверждение, что $1/(1+\xi^2)$ является характеристической функцией для $\frac{1}{2}e^{-|x|}$, представляется мне желательным и законным сокращением для логически корректного варианта: функция, которая в точке ξ принимает значение $1/(1+\xi^2)$, является характеристической функцией для функции, которая в точке x принимает значение $\frac{1}{2}e^{-|x|}$.

Боюсь, что краткие исторические замечания и ссылки не отдают должного многим авторам, внесшим свой вклад в теорию вероятностей. Однако всюду, где было возможно, я старался сделать это. Первоначальные работы во многих случаях перекрыты более новыми исследованиями, и, как правило, полные ссылки даются только на статьи, к которым читатель может обратиться для получения дальнейшей информации. Например, нет ссылок на мои собственные работы по предельным теоремам, в то же время какая-либо статья, описывающая результаты наблюдений или идеи, лежащие в основе примера, цитируется, даже если она совсем не содержит математики [1]). В этих обстоятельствах авторский указатель не дает никаких сведений о важности тех или иных исследований для теории вероятностей. Другая трудность состояла в справедливой оценке работ, которым мы обязаны новыми направлениями исследований, новыми подходами, новыми методами. Некоторые теоремы, рассматривавшиеся в свое время как поразительно оригинальные и глубокие, теперь получили простые доказательства и выступают в окружении более тонких результатов. Трудно воспринимать такие теоремы в их исторической перспективе и нелегко понять, что здесь, как и в других случаях, первый шаг значит очень многое.

Я благодарен Исследовательскому бюро армии США за поддержку моей работы по теории вероятностей в Принстонском университете, где мне помогали Дж. Голдман, Л. Пит, М. Силверстейн. Они устранили значительное число неточностей и неясных мест. Все главы переписывались много раз, и первоначальные варианты предшествующих глав распространялись среди моих друзей. Таким образом, благодаря замечаниям Дж. Эллиота, Р. С. Пинхема и Л. Дж. Сэвиджа возник ряд улучшений. Я особенно благодарен Дж. Л. Дубу и Дж. Волфовицу за советы и критику. График случайного блуждания Коши подготовлен Г. Тротером. За печатанием наблюдала г-жа Х. Мак-Дугл, и внешний вид книги многим обязан ей.

Вильям Феллер

Октябрь, 1965

[1]) Эта система была принята и в первом томе, однако она была неправильно понята некоторыми последующими авторами; они теперь приписывают методы, использованные в книге, работавшим до этого ученым, которые не могли знать их.

ОБОЗНАЧЕНИЯ

Интервалы обозначаются чертой сверху:
$\overline{a, b}$ — открытый, $\overline{|-|}\, a, b$ — замкнутый интервал; полуоткрытые интервалы обозначаются $\overline{-|}\, a, b$ и $\overline{|-}\, a, b$. *Эти обозначения применяются также и в многомерном случае* Соответствующие соглашения о векторных обозначениях и упорядочении см. в гл. V, 1 (а также в гл. IV, 2). Символ (a, b) резервирован для пар и для точек.

$\mathscr{R}^1$, $\mathscr{R}^2$, $\mathscr{R}^r$ обозначают прямую, плоскость и r-мерное пространство.

1 означает ссылку на первый том; римские цифры указывают номер главы. Так, I; гл. XI, (3.6) указывает на отношение к параграфу 3 главы XI первого тома.

▶ указывает конец доказательства или серии примеров.

$\mathfrak{n}$ и $\mathfrak{N}$ обозначают соответственно нормальную плотность и нормальное распределение с нулевым математическим ожиданием и единичной дисперсией.

O, o и $\sim$ Пусть u и v зависят от параметра x, который стремится, скажем, к a. Предполагая, что v положительно, мы пишем

$$\left.\begin{aligned} u &= O(v) \\ u &= o(v) \\ u &\sim v \end{aligned}\right\} \text{ если } \frac{u}{v} \begin{cases} \text{ограничено} \\ \to 0 \\ \to 1 \end{cases}$$

$f(x)\,U\{dx\}$ об этом сокращении см. гл. V. 3.

Относительно борелевских множеств и бэровских функций см. введение к главе V.

ГЛАВА I

ПОКАЗАТЕЛЬНЫЕ И РАВНОМЕРНЫЕ ПЛОТНОСТИ

§ 1. Введение

В первом томе мы неоднократно имели дело с вероятностями, определяемыми суммами многих малых слагаемых, и мы пользовались приближениями вида

$$\mathbf{P}\{a < \mathbf{X} < b\} \approx \int_a^b f(x)\,dx. \tag{1.1}$$

Основным примером служит нормальное приближение к биномиальному распределению [1]). Приближение этого типа обыкновенно устанавливается в форме предельной теоремы, включающей ряд все более и более «тонких» дискретных вероятностных моделей. Во многих случаях этот переход к пределу приводит, вообще говоря, к новому выборочному пространству, и последнее может быть интуитивно проще, чем первоначальная дискретная модель.

Примеры. а) *Показательные времена ожидания.* Для того чтобы описать времена ожидания дискретной моделью, мы должны сделать время дискретным и условиться, что изменения могут осуществляться только в моменты времени $\delta, 2\delta, \ldots$. Простейшее время ожидания $\mathbf{T}$ есть время ожидания первого успеха в последовательности испытаний Бернулли с вероятностью успеха p_δ. Тогда $\mathbf{P}\{\mathbf{T} > n\delta\} = (1 - p_\delta)^n$ и среднее время ожидания равно $\mathbf{E}(\mathbf{T}) = \delta/p_\delta$. Эту модель можно усовершенствовать, если предположить δ убывающим так, что математическое ожидание $\delta/p_\delta = \alpha$ остается фиксированным. При малом δ и фиксированном t мы имеем

$$\mathbf{P}\{\mathbf{T} > t\} = \left(1 - \frac{\delta}{\alpha}\right)^{t/\delta} \approx e^{-t/\alpha}, \tag{1.2}$$

---

[1]) *Другие примеры*: распределение арксинуса, гл. III, 5; распределение числа возвращений в начальное состояние и времен первого прохождения, гл. III, 6 и гл. III, 8; предельные теоремы для случайных блужданий, гл. XIV; равномерное распределение, задача 20 в гл. XI, 7.

что доказывается переходом к логарифмам. Эта модель описывает время ожидания как геометрически распределенную дискретную случайную величину, а (1.2) утверждает, что «в пределе» получается показательное распределение. Интуитивно представляется более естественным начать с выборочного пространства, точками которого служат вещественные числа, и ввести показательное распределение непосредственно.

б) *Случайный выбор.* «Выбор точки наудачу» в интервале [1]) $\overline{0,1}$ является воображаемым экспериментом с очевидным интуитивным смыслом. Этот эксперимент может быть описан дискретными приближениями, но проще взять в качестве выборочного пространства весь интервал и приписать каждому подинтервалу в качестве вероятности его длину. Воображаемый эксперимент, заключающийся в том, что производятся два независимых выбора точек в $\overline{0,1}$, дает пару вещественных чисел, и поэтому естественным выборочным пространством служит единичный квадрат. В этом выборочном пространстве «вероятность» почти инстинктивно приравнивается «площади». Этого вполне достаточно для некоторых элементарных целей, но рано или поздно возникает вопрос о том, что же в действительности означает слово «площадь». ▶

Как показывают эти примеры, непрерывное выборочное пространство может быть проще для восприятия, чем дискретная модель, но определение вероятностей в нем зависит от такого аппарата, как интегрирование и теория меры. В счетных выборочных пространствах можно было приписать вероятности *всем* вообще событиям, в то время как в пространствах произвольной природы эта простая процедура ведет к логическим противоречиям, и наша интуиция должна приспосабливаться к крайностям формальной логики. Мы вскоре увидим, что наивный подход может привести к затруднениям даже в сравнительно простых задачах. Ради справедливости стоит сказать, что в то же время многие значительные вероятностные задачи не требуют четкого определения понятия вероятности. Иногда они имеют аналитический характер, а вероятностное содержание служит лишь опорой для нашей интуиции. В таких ситуациях использование вероятностных терминов безвредно так же, как обращение к массам и центрам тяжести в абстрактных пространствах. Еще более существен тот факт, что при изучении сложных стохасти-

[1]) Для обозначения интервалов используется черта. Символ (a, b) сохраняется для координатной записи точек на плоскости, *см. обозначения в начале книги.*

ческих процессов (и соответственно усложненных выборочных пространств) могут возникнуть важные и понятные задачи, не связанные с теми тонкими средствами, которые используются при анализе процесса в целом. Типичное рассуждение может иметь следующий вид: если процесс вообще допускает математическое описание, то случайная величина **Z** должна иметь такие-то и такие-то свойства, а ее распределение должно в силу этого удовлетворять такому-то уравнению. Хотя вероятностные соображения могут сильно влиять на ход исследования этого уравнения, последнее в принципе не зависит от аксиом теории вероятностей. Специалисты различных прикладных областей настолько привыкли иметь дело с подобными проблемами, что они отрицают необходимость теории меры: они не знакомы с задачами другого типа и с ситуациями, где нечеткие рассуждения приводили к неверным результатам [1]).

Эти замечания постепенно разъяснятся на протяжении этой главы, которая предназначена служить неформальным введением в теорию. В ней описываются некоторые аналитические свойства двух важных распределений, которые будут постоянно использоваться в этой книге. Специальные темы затрагиваются отчасти по причине важных применений, отчасти для демонстрации новых задач, требующих привлечения новых методов. Нет необходимости изучать их систематически или в том порядке, как они представлены. Всюду в этой главе вероятности понимаются как элементарные интегралы со всеми вытекающими отсюда ограничениями. Использование вероятностного «жаргона» и таких терминов, как случайные величины и математические ожидания, может быть оправдано с двух точек зрения. Его можно рассматривать как техническое вспомогательное средство для нашей интуиции, имеющее в основе формальную аналогию с изложенным в первом томе. С другой стороны, все сказанное в этой главе может быть формально безупречно обосновано предельным переходом от дискретной модели, описанной в примере (2, а). Последняя процедура, хотя и не является ни необходимой, ни желательной, может быть хорошим упражнением для начинающих.

---

[1]) Роли строгости и интуиции часто понимаются неправильно. Как уже отмечалось в первом томе, природная интуиция и естественный способ мышления дают мало, но они становятся сильнее по мере развития математической теории. Сегодняшняя интуиция и применения теории опираются на сложные теории вчерашнего дня. Сверх того, строгая теория не роскошь, а способ экономии мышления. В самом деле, наблюдения показывают, что в применениях многие авторы полагаются на длинные вычисления, а не на простые рассуждения, которые им кажутся рискованными [ближайшая иллюстрация — пример (5, а)].

§ 2. Плотности. Свертки

Плотностью вероятностей на прямой (или в $\mathscr{R}^1$) называется функция f, такая, что

$$f(x) \geqslant 0, \qquad \int_{-\infty}^{+\infty} f(x)\,dx = 1. \tag{2.1}$$

Для начала мы рассмотрим только кусочно-непрерывные плотности (общий случай см. в гл. V, 3). Каждой плотности f мы поставим в соответствие *функцию распределения*[1]) F, определенную равенством

$$F(x) = \int_{-\infty}^{x} f(y)\,dy. \tag{2.2}$$

Это есть монотонная непрерывная функция, возрастающая от 0 до 1. Мы скажем, что f и F *сосредоточены на интервале* $a \leqslant x \leqslant b$, если f равна нулю вне этого интервала. Плотность f будет рассматриваться как функция, задающая вероятности на интервалах прямой; интервалу[2]) $\overline{a, b} = \{a < x < b\}$ соответствует вероятность

$$F(b) - F(a) = \int_{a}^{b} f(x)\,dx. \tag{2.3}$$

Иногда эта вероятность будет обозначаться $\mathbf{P}\{\overline{a, b}\}$. При таком распределении отдельной точке соответствует вероятность 0, а замкнутому интервалу $\overline{|a, b|}$ — та же вероятность, что и $\overline{a, b}$.

В простейшей ситуации вещественная прямая служит *«выборочным пространством»*, т. е. результат воображаемого эксперимента представляется числом. (Точно так же, как и в первом томе, это лишь первый шаг в построении выборочных пространств, представляющих последовательности экспериментов.) *Случайные величины* являются функциями, определенными на выборочном пространстве. Для простоты мы договоримся на некоторое время считать функцию $\mathbf{U}$ случайной величиной, если

---

[1]) Мы напоминаем, что под «функцией распределения» понимается непрерывная справа неотрицательная функция с пределами 0 и 1 на $\pm\infty$. В первом томе мы преимущественно имели дело со ступенчатыми функциями. Теперь мы сосредоточим наше внимание на функциях распределения, задаваемых интегралами. Более подробно мы будем изучать функции распределения в гл. V.

[2]) Обозначения см. в списке обозначений в начале книги.

для каждого t событие $\{\mathbf{U}\leqslant t\}$ содержит лишь конечное множество интервалов. Тогда функция

$$G(t)=\mathbf{P}\{\mathbf{U}\leqslant t\} \qquad (2.4)$$

может быть определена как интеграл от f по объединению этих интервалов. Функция G, заданная равенством (2.4), называется *функцией распределения* величины $\mathbf{U}$. Если G является интегралом от некоторой функции g, то g называется *плотностью* распределения G, или (что равнозначно) плотностью величины $\mathbf{U}$.

Основная случайная величина — это, конечно, координатная величина [1]) $\mathbf{X}$ сама по себе, и все другие случайные величины являются функциями $\mathbf{X}$. Функция распределения $\mathbf{X}$ тождественна функции F, посредством которой задаются вероятности. Необходимо сказать, что любая случайная величина $\mathbf{Y}=g(\mathbf{X})$ может рассматриваться как координатная величина на новой прямой.

Как сказано выше, употребление этих терминов может быть оправдано аналогией с рассмотренным в первом томе. Следующий пример показывает, что наша модель может быть получена из дискретных моделей предельным переходом.

Примеры. а) *Группировка данных.* Пусть F — заданная функция распределения. Выберем фиксированное $\delta>0$ и рассмотрим дискретную случайную величину $\mathbf{X}_\delta$, которая в интервале $(n-1)\delta<\mathbf{X}<n\delta$ принимает постоянное значение $n\delta$. Здесь $n=0,\ \pm1,\ \pm2,\ \ldots$. В первом томе мы использовали совокупность чисел, кратных δ, как выборочное пространство, и описывали распределение вероятностей $\mathbf{X}_\delta$ равенством

$$\mathbf{P}\{\mathbf{X}_\delta=n\delta\}=F(n\delta)-F((n-1)\delta). \qquad (2.5)$$

Теперь $\mathbf{X}_\delta$ — случайная величина в расширенном выборочном пространстве, и ее функция распределения — это функция, которая в интервале $(n-1)\delta<x\leqslant n\delta$ равна $F(n\delta)$. В непрерывной модели $\mathbf{X}_\delta$ служит приближением к $\mathbf{X}$, получаемым отождествлением наших интервалов с их концами (процедура, известная статистикам как группировка данных). В духе первого тома мы будем рассматривать $\mathbf{X}_\delta$ как основную случайную величину, и δ — как свободный параметр. Полагая $\delta\to0$, мы получим предельные теоремы, устанавливающие, например, что F служит предельным распределением для $\mathbf{X}_\delta$.

---

[1]) Всюду, где это возможно, мы будем обозначать случайные величины (т. е. функции на выборочном пространстве) заглавными полужирными буквами, оставляя малые буквы для обозначения чисел или масштабных параметров. Это справедливо, в частности, для координатной величины $\mathbf{X}$, а именно функции, определенной равенством $\mathbf{X}(x)=x$.

б) При $x>0$ событие $\{\mathbf{X}^2 \leqslant x\}$ равносильно событию $\{-\sqrt{x} \leqslant \mathbf{X} \leqslant \sqrt{x}\}$; случайная величина $\mathbf{X}^2$ имеет распределение, сосредоточенное на $\overline{0, \infty}$ с функцией распределения $F(\sqrt{x}) - F(-\sqrt{x})$. После дифференцирования видно, что *плотность* g величины $\mathbf{X}^2$ равна $g(x) = {}^1/_2[f(\sqrt{x}) + f(-\sqrt{x})]/\sqrt{x}$ при $x>0$ и $g(x)=0$ при $x<0$. Аналогично функция распределения $\mathbf{X}^3$ равна $F(\sqrt[3]{x})$ и имеет плотность ${}^1/_3 f(\sqrt[3]{x})/\sqrt[3]{x^2}$. ▶

Математическое ожидание $\mathbf{X}$ определяется формулой

$$\mathbf{E}(X) = \int_{-\infty}^{+\infty} x f(x)\, dx \tag{2.6}$$

при условии, что интеграл сходится абсолютно. Математические ожидания аппроксимирующих дискретных величин $\mathbf{X}_\delta$ из примера (а) совпадают с римановыми суммами этого интеграла, и поэтому $\mathbf{E}(\mathbf{X}_\delta) \to \mathbf{E}(\mathbf{X})$. Если u есть ограниченная непрерывная функция, то те же самые рассуждения применимы к случайной величине $u(\mathbf{X})$, и соотношение $\mathbf{E}(u(\mathbf{X}_\delta)) \to \mathbf{E}(u(\mathbf{X}))$ влечет

$$\mathbf{E}(u(\mathbf{X})) = \int_{-\infty}^{+\infty} u(x) f(x)\, dx. \tag{2.7}$$

Здесь следует отметить, что математическое ожидание может быть вычислено без явного использования распределения $u(\mathbf{X})$ (см. 1, гл. IX, 2). Те же доводы применимы к неограниченным величинам, как, например, $\mathbf{X}^2$.

Второй момент $\mathbf{X}$ определяется равенством

$$\mathbf{E}(\mathbf{X}^2) = \int_{-\infty}^{+\infty} x^2 f(x)\, dx \tag{2.8}$$

при условии, что интеграл сходится. Обозначим $\mu = \mathbf{E}(\mathbf{X})$. Тогда *дисперсия* $\mathbf{X}$ определяется как

$$\operatorname{Var}(\mathbf{X}) = \mathbf{E}((\mathbf{X}-\mu)^2) = \mathbf{E}(\mathbf{X}^2) - \mu^2. \tag{2.9}$$

Замечание. Если величина $\mathbf{X}$ *положительна* (т. е. если плотность f сосредоточена на $\overline{0, \infty)}$, и если интеграл в выражении (2.6) расходится, то безвредно и даже удобно говорить, что $\mathbf{X}$ *имеет бесконечное математическое ожидание* и писать $\mathbf{E}(\mathbf{X}) = \infty$. Точно так же говорят, что $\mathbf{X}$ имеет бесконечную дисперсию, когда интеграл в (2.8) расходится. Для величин, принимающих положительные и отрицательные значения, математическое ожи-

дание остается неопределенным, когда интеграл (2.6) расходится. Типичный пример доставляет плотность $\pi^{-1}(1+x^2)^{-1}$.

Понятие плотности переносится на размерности более высоких порядков, но общее обсуждение мы отложим до гл. III. Пока же мы будем рассматривать только аналог произведения вероятностей, введенного определением 2 в 1, гл. V, 4 для описания комбинаций независимых экспериментов. Другими словами, в этой главе мы будем иметь дело только с плотностями, представляемыми произведениями вида $f(x)g(y)$, $f(x)g(y)h(z)$ и т. д., где f, g, ... — плотности на прямой. Задание плотности вида $f(x)g(y)$ на плоскости $\mathscr{R}^2$ означает отождествление «вероятностей» с интегралами

$$\mathbf{P}\{A\} = \int\int_A f(x)\,g(y)\,dx\,dy. \tag{2.10}$$

Когда мы говорим *«две независимые случайные величины* $\mathbf{X}$ и $\mathbf{Y}$ *с плотностями* f *и* g», то это есть сокращение высказывания о том, что вероятности на $(\mathbf{X}, \mathbf{Y})$-плоскости определяются в соответствии с формулой (2.10). Отсюда следует правило умножения для интервалов, например $\mathbf{P}\{\mathbf{X}>a,\ \mathbf{Y}>b\} = \mathbf{P}\{\mathbf{X}>a\}\mathbf{P}\{\mathbf{Y}>b\}$. Аналогия с дискретным случаем настолько очевидна, что дальнейшие разъяснения не требуются.

Можно ввести много новых случайных величин, рассматривая функции $\mathbf{X}$ и $\mathbf{Y}$, но наиболее важную роль играет сумма $\mathbf{S}=\mathbf{X}+\mathbf{Y}$. Событие $A=\{\mathbf{S}\leqslant s\}$ изображается полуплоскостью точек (x, y), таких, что $x+y\leqslant s$. Обозначим функцию распределения $\mathbf{Y}$ через G, так что $g(y)=G'(y)$. Чтобы найти *функцию распределения* $\mathbf{X}+\mathbf{Y}$, мы интегрируем в (2.10) по области $y\leqslant s-x$. В результате получаем

$$\mathbf{P}\{\mathbf{X}+\mathbf{Y}\leqslant s\} = \int_{-\infty}^{+\infty} G(s-x)\,f(x)\,dx. \tag{2.11}$$

Ввиду симметрии F и G можно поменять ролями без влияния на результат. После дифференцирования видно, что *плотность* $\mathbf{X}+\mathbf{Y}$ *равна любому из двух интегралов*

$$\int_{-\infty}^{+\infty} f(s-y)\,g(y)\,dy = \int_{-\infty}^{+\infty} f(y)\,g(s-y)\,dy. \tag{2.12}$$

Операция, определенная формулой (2.12), является частным случаем свертки, которая будет введена в гл. V, 4. До этого времени мы будем использовать термин «свертка» только для плотностей: *свертка двух плотностей* f *и* g *есть функция, определенная в* (2.12). *Свертка будет обозначаться* $f*g$.

Повсюду в первом томе мы имели дело со свертками дискретных распределений. Соответствующие правила справедливы и здесь. В соответствии с (2.12) мы имеем $f*g=g*f$. Задавая третью плотность h, мы можем составить свертку $(f*g)*h$. Это есть плотность суммы $\mathbf{X}+\mathbf{Y}+\mathbf{Z}$ трех независимых величин с плотностями f, g, h. Тот факт, что суммирование коммутативно и ассоциативно, влечет те же свойства для сверток, и поэтому результат $f*g*h$ не зависит от порядка операций.

Положительные случайные величины играют важнейшую роль, и потому полезно отметить, что *если плотности* f и g *сосредоточены* на $\overline{0, \infty}$, *то свертка* $f*g$ *равна*

$$f*g(s)=\int_0^s f(s-y)\,g(y)\,dy=\int_0^s f(x)\,g(s-x)\,dx. \qquad (2.13)$$

Замечание *о понятии случайной величины.* Использование прямой линии или евклидовых пространств $\mathscr{R}^n$ как выборочных пространств иногда затемняет различие между случайными величинами и «обычными» функциями одной или большего числа переменных. В первом томе случайная величина $\mathbf{X}$ могла принимать лишь счетное множество значений, и в этом случае было ясно, говорили ли мы о функции (например, квадрате или экспоненте), определенной на прямой, или о случайной величине $\mathbf{X}^2$ или $e^{\mathbf{X}}$, определенной на выборочном пространстве. Даже по внешнему виду эти функции совершенно различны, поскольку «обычная» показательная функция принимает все положительные значения, тогда как $e^{\mathbf{X}}$, имеет счетную область значений. Для того чтобы увидеть изменение в этой ситуации, рассмотрим теперь «две независимые случайные величины $\mathbf{X}$ и $\mathbf{Y}$ с одинаковой плотностью f». Другими словами, выборочным пространством служит плоскость и вероятности определяются как интегралы от $f(x)f(y)$. Теперь любая функция двух переменных *может* быть определена *на* выборочном пространстве, и тогда она становится случайной величиной, однако следует иметь в виду, что функция двух переменных может быть также определена независимо от нашего выборочного пространства. Например, некоторые статистические задачи требуют введения случайной величины $f(\mathbf{X})f(\mathbf{Y})$ [см. пример (11, г) гл. VI]. С другой стороны, вводя наше выборочное пространство $\mathscr{R}^2$, мы очевидным образом упоминали «обычную» функцию f, определенную независимо от выборочного пространства. Эта «обычная» функция задает множество случайных величин, например $f(\mathbf{X})$, $f(\mathbf{Y})$, $f(\mathbf{X}\pm\mathbf{Y})$ и т. д. Таким образом, одна и та же функция f может служить или случайной величиной, или «обычной» функцией.

Как правило (и в каждом отдельном случае), будет ясно, имеем ли мы дело со случайной величиной или нет. Тем не менее в общей теории возникают ситуации, при которых функции (такие, как условные вероятности и условные математические ожидания) могут рассматриваться как «свободные» функции или как случайные величины, и это отчасти приводит к путанице, если свобода выбора не понимается должным образом.

Замечание *о терминологии и обозначениях.* Чтобы избежать громоздких формулировок, условимся называть $\mathbf{E}(\mathbf{X})$ математическим ожиданием величины $\mathbf{X}$ или плотности f или распределения F Подобные вольности будут введены и для других терминов Например, «свертка» фактически означает операцию, но этот термин применяют и к результату операции, и функция $f*g$ называется «сверткой».

В старой литературе термины «распределение» и «функция частот» применялись к тому, что мы теперь называем плотностями, наши функции распределения описывались как «кумулятивные», а сокращение с. d f.[1]) употребительно до сих пор.

§ 3. Показательная плотность

Для произвольного, но фиксированного $\alpha>0$ положим

$$f(x)=\alpha e^{-\alpha x},\ F(x)=1-e^{-\alpha x} \text{ при } x\geqslant 0 \qquad (3.1)$$

и $F(x)=f(x)=0$ при $x<0$. Тогда f — показательная плотность, F — ее функция распределения. Простые вычисления показывают, что *математическое ожидание равно* α^{-1}, *дисперсия* — α^{-2}.

В примере (1, а) показательное распределение было выведено как предел геометрических распределений. Метод примера (2, а) приводит к тому же результату. Напомним, что в стохастических процессах геометрическое распределение часто имеют времена ожидания или продолжительность существования и что это обусловлено «отсутствием последействия», описанным в **1**, гл. XIII, 9: *каков бы ни был настоящий возраст, оставшаяся продолжительность существования не зависит от прошлого и имеет то же самое распределение, что и сама продолжительность существования.* Теперь будет показано, что это свойство переносится на показательное распределение и только на показательное распределение.

Пусть $\mathbf{T}$ — произвольная положительная величина, которую мы будем интерпретировать как время жизни или как время ожидания. Удобно заменить функцию распределения $\mathbf{T}$ «хвостом распределения»

$$U(t)=\mathbf{P}\{\mathbf{T}>t\}. \qquad (3.2)$$

[1]) Cumulative distribution function — кумулятивная функция распределения. — *Прим. перев.*

С наглядной точки зрения $U(t)$ есть «вероятность того, что время жизни (от момента рождения) превзойдет t». При данном возрасте s событие, состоящее в том, что оставшееся время жизни превосходит t, записывается как $\{\mathbf{T}>s+t\}$, и условная вероятность этого события (при данном возрасте s) равна отношению $U(s+t)/U(s)$. Это есть распределение оставшегося времени жизни. Оно совпадает с распределением общего времени жизни тогда и только тогда, когда

$$U(s+t)=U(s)\,U(t) \qquad s,t>0. \tag{3.3}$$

В 1, гл. XVII, 6 было показано, что положительное решение этого уравнения должно иметь вид $U(t)=e^{at}$, и, следовательно, *отсутствие старения, описанное выше курсивом, имеет место тогда и только тогда, когда распределение времени жизни показательно.*

Мы будем называть это свойство отсутствия последействия (lack of memory) *марковским свойством* показательного распределения. Аналитически марковское свойство сводится к утверждению, что только для показательного распределения F хвосты $U=1-F$ удовлетворяют равенству (3.3). Этим объясняется постоянное присутствие показательного распределения в марковских процессах. (Усиленный вариант марковского свойства будет описан в § 6) Наше описание относится к временны́м процессам, но наши рассуждения имеют общий характер и марковское свойство остается осмысленным, когда время заменяется каким-нибудь другим параметром.

Примеры. а) *Прочность на разрыв.* Чтобы получить непрерывный аналог общеизвестной конечной цепи, прочность которой равна прочности ее самого слабого звена, обозначим $U(t)$ вероятность того, что нить *длины* t (данного материала) может выдержать некоторый фиксированный груз. Нить длины $s+t$ не разрывается, если два отрезка по отдельности выдерживают данный груз. В предположении, что взаимодействие отсутствует, эти два события должны рассматриваться как независимые и U должно удовлетворять (3.3). Здесь длина нити выполняет роль временно́го параметра, а длина, при которой нить порвется, является показательно распределенной случайной величиной.

б) *Задача свободного пробега.* Будем представлять себе звезды как шары фиксированного радиуса $\rho>0$. Рассмотрим некоторое множество звезд в пространстве и выберем за начало отсчета точку, не содержащуюся ни в какой звезде. Возьмем ось x произвольно направленной. Нас интересует *длина самого длинного интервала* $\overline{0,\ x}$, *не пересекающего никакую звезду.*

Она характеризует *видимость в x-направлении*. То же самое описание применимо в двумерном случае к *видимости в лесу*, когда деревья предполагаются круговыми цилиндрами.

Видимость зависит от строения звездного поля, для которого мыслимы многие модели. Мы описываем этот ансамбль пуассоновскими распределениями, которые служат моделью «чистой случайности» (perfect randomness) в астрономии, физике и статистике; более подробно мы поговорим об этом в следующем параграфе. На данном этапе понятие чистой случайности не определено, и мы свободны ввести его произвольными *постулатами*; при этом возникает сомнение в отношении существования фактических моделей, подчиняющихся нашим постулатам (другими словами, являются ли постулаты противоречивыми или нет). Интуитивно ясно, что первым свойством, характеризующим чистую случайность, должно быть отсутствие взаимодействия между различными областями: из того, что происходит внутри области A, невозможно получить вывод о звездах в области B. *Если* ансамбль звезд обладает этим свойством, мы можем повторить аргументы последнего примера: вероятность, что сегмент длины $s+t$ не пересекает ни одной звезды, должна быть произведением соответствующих вероятностей для двух сегментов длины s и t. При этих условиях *видимость в данном направлении должна быть показательно распределенной случайной величиной*. [См. пример (4, б)].

Важное «если», на котором построены наши рассуждения, отнюдь не травиально, так как оно заставляет нас допустить, что звезды перекрываются (в противном случае взаимные расстояния между центрами превышали бы 2ρ; это представляло бы собой форму взаимодействия). Это обстоятельство практически несущественно, если радиус ρ мал в сравнении со средним расстоянием между ближайшими друг к другу звездами. Поэтому астрономы принимают эту модель как разумное приближение.

Следующая теорема будет использоваться неоднократно.

Теорема. *Если* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — *взаимно независимые случайные величины с показательным распределением* (3.1), *то сумма* $\mathbf{X}_1 + \ldots + \mathbf{X}_n$ *имеет плотность* g_n *и функцию распределения* G_n, *задаваемые формулами*

$$g_n(x) = \alpha \frac{(\alpha x)^{n-1}}{(n-1)!} e^{-\alpha x}, \qquad x > 0, \tag{3.4}$$

$$G_n(x) = 1 - e^{-\alpha x}\left(1 + \frac{\alpha x}{1!} + \ldots + \frac{(\alpha x)^{n-1}}{(n-1)!}\right), \qquad x > 0. \tag{3.5}$$

Доказательство. При $n=1$ утверждение сводится к определению (3.1). Плотность g_{n+1} определяется сверткой

$$g_{n+1}(t)=\int_0^t g_n(t-x)\,g_0(x)\,dx, \tag{3.6}$$

и в предположении справедливости (3.4) это сводится к

$$g_{n+1}(t)=\frac{\alpha^{n+1}}{(n-1)!}\,e^{-\alpha t}\int_0^t x^n\,dx=\alpha\,\frac{(\alpha t)^n}{n!}\,e^{-\alpha t}. \tag{3.7}$$

Таким образом, (3.4) выполняется по индукции для всех n. Справедливость равенства (3.5) доказывается дифференцированием. ▶

Плотности g_n входят в семейство *гамма-плотностей*, которое будет введено в гл. II, 2. Они представляют собой непрерывный аналог отрицательного биномиального распределения, определенного в **1**, гл. VI, 8 как распределение суммы n случайных величин с одинаковым геометрическим распределением. (См. задачу 7.)

§ 4. Парадоксы, связанные с временем ожидания. Пуассоновский процесс

Обозначим $\mathbf{X}_1$, $\mathbf{X}_2$, ... взаимно независимые случайные величины с одинаковым показательным распределением (3.1). Пусть $\mathbf{S}_0=0$,

$$\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n, n=1, 2, \ldots. \tag{4.1}$$

Мы вводим семейство новых случайных величин $\mathbf{N}(t)$ следующим образом: $\mathbf{N}(t)$ *есть число индексов* $k\geqslant 1$, *таких, что* $\mathbf{S}_k\leqslant t$. Событие $\{\mathbf{N}(t)=n\}$ происходит, когда $\mathbf{S}_n\leqslant t$, но $\mathbf{S}_{n+1}>t$. Так как $\mathbf{S}_n$ имеет распределение G_n, то вероятность этого события равна $G_n(t)-G_{n+1}(t)$ или

$$\mathbf{P}\{\mathbf{N}(t)=n\}=e^{-\alpha t}\,\frac{(\alpha t)^n}{n!}. \tag{4.2}$$

Иначе говоря, *случайная величина* $\mathbf{N}(t)$ *имеет пуассоновское распределение с математическим ожиданием* αt.

Эта аргументация выглядит как новый вывод пуассоновского распределения, но на самом деле это только перефразировка первоначального вывода из **1**, гл. VI, 6 в терминах случайных величин. Для наглядного описания рассмотрим случайно наступающие события (такие, как вспышки космических лучей или телефонные вызовы), которые мы называем «поступлениями».

Допустим, что последействие отсутствует, т. е. что прошлое не дает возможности заключений в отношении будущего. Как мы видели, это условие требует, чтобы время ожидания $\mathbf{X}_1$ первого поступления было распределено показательно. Но при каждом поступлении процесс начинается снова, как вероятностная копия всего процесса: времена ожидания $\mathbf{X}_k$ между последовательными поступлениями должны быть независимыми и должны иметь одно и то же распределение. Сумма $\mathbf{S}_n$ изображает момент n-го поступления, а $\mathbf{N}(t)$ — *число поступлений* в интервале $\overline{0, t}$. В такой форме доводы отличаются от первоначального вывода пуассоновского распределения лишь использованием лучших специальных терминов.

(По терминологии стохастических процессов последовательность $\{\mathbf{S}_n\}$ есть *процесс восстановления* с показательными *временами* $\mathbf{X}_k$ *между поступлениями*; общее понятие процесса восстановления см. в гл. VI, 6.)

Даже эта простая ситуация приводит к кажущимся противоречиям, которые иллюстрируют необходимость более сложного подхода. Мы начнем с несколько наивной формулировки.

Пример. *Парадокс времени ожидания* Автобусы прибывают в согласии с пуассоновским процессом, причем среднее время между последовательными автобусами равно α^{-1}. Я прихожу в момент t. Каково математическое ожидание $\mathbf{E}(\mathbf{W}_t)$ времени $\mathbf{W}_t$, в течение которого я жду следующий автобус? (Понятно, что момент t моего прибытия не зависит от автобусов — скажем я прихожу ровно в полдень.) Очевидно, имеются два противоречивых ответа.

а) Отсутствие последействия, свойственное пуассоновскому процессу, влечет независимость $\mathbf{E}(\mathbf{W}_t)$ от t, т. е.

$$\mathbf{E}(\mathbf{W}_t) = \mathbf{E}(\mathbf{W}_0) = \alpha^{-1}.$$

б) Момент моего прибытия «выбран наудачу» в интервале между двумя последовательными автобусами и по соображениям симметрии:

$$\mathbf{E}(\mathbf{W}_t) = \frac{1}{2}\alpha^{-1}.$$

Оба рассуждения кажутся приемлемыми, и оба использовались на практике. Что же делать с противоречиями? Легчайший выход у формалиста, который отказывается рассматривать задачу, если она не сформулирована строго. Но задачи не решаются их игнорированием.

Теперь мы покажем, что *оба* рассуждения по существу, если не формально, *корректны*. Ошибка лежит в непредвиденном пункте. Она связана с явлением, известным в общей теории

восстановления, где оно вызвало серьезные затруднения, прежде чем было правильно понято. ▶

Мы рассмотрим последовательность времен между прибытиями

$$\mathbf{X}_1=\mathbf{S}_1,\ \mathbf{X}_2=\mathbf{S}_2-\mathbf{S}_1,\ldots.$$

По предположению $\mathbf{X}_k$ имеют одинаковое показательное распределение с математическим ожиданием α^{-1}. Выбирая «любое» отдельное $\mathbf{X}_k$, мы получаем случайную величину, и интуитивно ожидается, что ее математическое ожидание будет равно α^{-1} при условии, что выбор сделан без знания о последовательности $\mathbf{X}_1, \mathbf{X}_2, \ldots$. Однако это не верно. В нашем примере мы взяли тот элемент $\mathbf{X}_k$, для которого $\mathbf{S}_{k-1}<t\leqslant\mathbf{S}_k$, где t фиксировано. Этот выбор сделан без учета действительного процесса, но оказывается, что так выбранное $\mathbf{X}_k$ имеет двойное математическое ожидание $2\alpha^{-1}$. С учетом этого факта противоречие исчезает, так как в пункте (б) нашего примера постулируется, что среднее время ожидания равно α^{-1}.

Это разрешение парадокса вызвало шок у опытных работников, однако оно становится интуитивно ясным, когда должным образом подбирается способ рассуждения. Грубо говоря, длинный интервал имеет больше шансов накрыть точку t, нежели короткий. Это приблизительное утверждение подтверждает следующее

Предложение. *Пусть* $\mathbf{X}_1, \mathbf{X}_2, \ldots$ *взаимно независимы и одинаково показательно распределены с математическим ожиданием* α^{-1}. *Пусть* $t>0$ *фиксировано, но произвольно. Элемент* $\mathbf{X}_k$, *удовлетворяющий условию* $\mathbf{S}_{k-1}<t\leqslant\mathbf{S}_k$, *имеет плотность*

$$v_t(x)=\begin{cases}\alpha^2 x e^{-\alpha x} & \text{для } 0<x\leqslant t,\\ \alpha(1+\alpha t)\,e^{-\alpha x} & \text{для } x>t.\end{cases} \tag{4.3}$$

Здесь главное в том, что плотность (4.3) *не является общей плотностью для* $\mathbf{X}_k$. Явный вид этой плотности представляет второстепенный интерес.

Доказательство. Пусть k — такой индекс, что $\mathbf{S}_{k-1}<t\leqslant\mathbf{S}_k$, и пусть $\mathbf{L}_t$ равно $\mathbf{S}_k-\mathbf{S}_{k-1}$. Мы должны доказать, что $\mathbf{L}_t$ имеет плотность (4.3). Допустим сначала, что $x<t$. Событие $\{\mathbf{L}_t\leqslant x\}$ происходит, если и только если $\mathbf{S}_n=y$ и $t-y\leqslant\mathbf{X}_{n+1}\leqslant x$ при некоторой комбинации n, y. Отсюда следует неравенство $t-x\leqslant y<t$. Суммируя по всем возможным n и y, получаем

$$\mathbf{P}\{\mathbf{L}_t\leqslant x\}=\sum_{n=0}^{\infty}\int_{t-x}^{t} g_n(y)\cdot[e^{-\alpha(t-y)}-e^{-\alpha x}]\,dy. \tag{4.4}$$

Но $g_0(y)+g_1(y)+\ldots=\alpha$ тождественно, и поэтому

$$\mathbf{P}\{\mathbf{L}_t \leqslant x\} = 1 - e^{-\alpha x} - \alpha x e^{-\alpha x}. \tag{4.5}$$

Дифференцируя, мы получим (4.3) для $x<t$. Для $x>t$ применимо это же доказательство с той разницей, что y меняется от 0 до x, и мы должны прибавить к правой части равенства (4.4) вероятность $e^{-\alpha t}-e^{-\alpha x}$ того, что $0<t<\mathbf{S}_1<x$. Этим доказательство завершается. ▶

Скачок функции (4.3) при $x=t$ обусловлен специальной ролью начала как исходной точки процесса. Очевидно,

$$\lim_{t\to\infty} v_t(x) = \alpha^2 x e^{-\alpha x}, \tag{4.6}$$

откуда видно, что со временем роль начала исчезает и для «старого» процесса распределение $\mathbf{L}_t$ почти не зависит от t. Удобнее выразить это, сказав, что правая часть (4.6) задает *«стационарную» плотность* $\mathbf{L}_t$.

В обозначениях, принятых в доказательстве, время ожидания $\mathbf{W}_t$, рассмотренное в примере, есть случайная величина $\mathbf{W}_t = \mathbf{S}_k - t$. Из доказательства теоремы следует также, что

$$\mathbf{P}\{\mathbf{W}_t \leqslant x\} = e^{-\alpha t} - e^{-\alpha(x+t)} + \sum_{n=0}^{\infty} \int_0^t g_n(y)\left[e^{-\alpha(t-y)} - e^{-\alpha(x+t-y)}\right] dy = 1 - e^{-\alpha x}. \tag{4.7}$$

Таким образом, $\mathbf{W}_t$ *имеет то же самое показательное распределение, что и* $\mathbf{X}_k$, в согласии с доводами пункта (а). (См. задачу 8.)

Наконец несколько слов о *пуассоновском процессе*. Пуассоновские величины $\mathbf{N}(t)$ были введены как функции, определенные на выборочном пространстве бесконечной последовательности случайных величин $\mathbf{X}_1, \mathbf{X}_2, \ldots$. Эта процедура удовлетворительна для многих целей, но более естественно другое выборочное пространство. Мысленный эксперимент «регистрации числа поступивших вызовов вплоть до момента t» дает при каждом положительном t целое число, и результатом поэтому является ступенчатая функция с единичными скачками. Эти ступенчатые функции служат точками выборочного пространства; выборочное пространство является функциональным пространством — пространством всех возможных «траекторий». В этом пространстве $\mathbf{N}(t)$ определяется как значение ординаты в момент t, а $\mathbf{S}_n$ — как координата n-го скачка и т. д. Теперь могут быть рассмотрены события, которые не выразимы в терминах первоначальных случайных величин. Типичный пример, имею-

щий практический интерес (см. задачу о разорении в гл. VI, 5), доставляет событие, заключающееся в том, что $\mathbf{N}(t) > a + bt$ при некотором t. Отдельная траектория (точно так же, как отдельная бесконечная последовательность из ± 1 в биномиальных испытаниях) представляет собой естественный и неизбежный объект вероятностного исследования. Стоит только привыкнуть к новой терминологии, как пространство всех траекторий становится наиболее наглядным выборочным пространством.

К сожалению, введение вероятностей в пространстве выборочных траекторий — дело далеко не простое. Для сравнения заметим, что переход от дискретных выборочных пространств к прямой, плоскости и т. д. и даже к бесконечным последовательностям случайных величин ни мысленно, ни технически не труден. В связи с функциональными пространствами возникают проблемы нового типа и мы предупреждаем читателя, что не будем иметь с ними дело в этом томе. Мы удовлетворимся частным рассмотрением выборочных пространств последовательностей (счетного множества координатных величин). Упоминание о стохастических процессах вообще и о пуассоновском процессе в частности мы будем делать свободно, но только для того, чтобы подготовить интуитивную основу или привлечь интерес к нашим проблемам.

Пуассоновские ансамбли точек

Как показано в **1**, гл. VI, 6, пуассоновский закон управляет не только «точками, распределенными случайно по оси времени», но также ансамблями точек (такими, как дефекты в изделиях или изюминки в булках), распределенных случайно на плоскости или в пространстве при условии, что t интерпретируется как площадь или объем. Основное предположение состоит в том, что вероятность нахождения k точек в заданной области зависит только от площади или объема области, но не от ее формы и что явления в неперекрывающихся областях независимы. Путем несложных формальных выкладок можно прийти к интересным результатам, касающимся таких случайных ансамблей точек, но замечания относительно пуассоновского процесса одинаково применимы и к пуассоновским ансамблям; полное вероятностное описание сложно и лежит вне рамок настоящего тома.

Примеры. а) *Ближайшие соседи.* Времена между поступлениями в пуассоновском процессе могут быть описаны как расстояния между ближайшими соседями. Это понятие переносится на плоскость. Высказывание, что расстояние ближайшего к 0 соседа не превосходит r, сводится к высказыванию, что круг

радиуса r содержит по крайней мере одну точку ансамбля. Для пуассоновского ансамбля вероятность не обнаружить точку внутри любой области с площадью v равна e^{-av}, и отсюда видно, что *распределение ближайших соседей в пуассоновском ансамбле задается в* $\mathscr{R}^2$ *функцией* $1-e^{-\alpha\pi r^2}$; в $\mathscr{R}^3$ мы получим $1-e^{-\beta r^3}$, где $\beta=4/3\pi\alpha$.

б) *Свободная видимость.* Мы возвращаемся к задаче свободного пробега (или видимости) из примера (3, *б*). В свете терминологии этого параграфа предполагается, что центры звезд представляют собой выборку из пуассоновского ансамбля и что каждая звезда есть шар радиуса ρ. (Это влечет возможность перекрытий.) Предполагается, что начало не содержится ни в какой звезде. Мы интересуемся расстоянием $\mathbf{L}$ *до ближайшего соседа в* x*-направлении* (т. е. наибольшим сегментом $\overline{0, a}$ оси x, не пересекающим ни одну звезду).

Мы уже вывели из априорных соображений, что $\mathbf{L}$ должно иметь показательное распределение. Теперь этот результат будет выведен из предполагаемых свойств нашей звездной системы. Событие $\{\mathbf{L}>x\}$ означает, что ни один звездный *центр* не содержится внутри сферы радиуса ρ с центром в точке из $\overline{0, x}$, и известно заранее, что сфера вокруг начала не содержит центров звезд. Объединение оставшихся сфер есть область Ω, ограниченная круговым цилиндром и двумя полусферами с центрами в 0 и в x соответственно. Объем области Ω, очевидно, равен $\pi\rho^2 x$, и, следовательно, $\mathbf{P}\{\mathbf{L}>x\}=e^{-\alpha\pi\rho^2 x}$. Таким образом, мы не только показали, что распределение показательно, но и нашли связь между параметром распределения и ансамблем звезд (в $\mathscr{R}^2$ площадь Ω, очевидно, равна $2\rho x$, и поэтому $\mathbf{P}\{\mathbf{L}>x\}=e^{-2\alpha\rho x}$). ▶

§ 5. Устойчивость неудач

Общеизвестно, что тот, кто встает в очередь, может быть вынужден ждать неопределенно долгое время. Подобные неудачи преследуют нас во многих положениях. Как может способствовать объяснению этого теория вероятностей? Для ответа на этот вопрос мы рассмотрим три примера, типичных для множества ситуаций. Они иллюстрируют неожиданные общие свойства случайных флуктуаций.

Примеры. а) *Рекордные значения.* Обозначим $\mathbf{X}_0$ мое время ожидания (или размер финансовых потерь) некоторого случайного события. Предположим, что мои друзья подвергли себя того же самого типа опыту. Обозначим результаты $\mathbf{X}_1, \mathbf{X}_2, \ldots,$

Чтобы отразить «беспристрастность», предположим, что $\mathbf{X}_0, \mathbf{X}_1 \ldots$ взаимно независимые случайные величины с одним и тем же распределением. Природа последнего в действительности не имеет значения, но, так как показательное распределение служит моделью для случайности появления, мы предположим, что $\mathbf{X}_j$ показательно распределены в соответствии с (3.1). Для простоты описания мы предполагаем последовательность $\{\mathbf{X}_j\}$ неограниченной.

Чтобы измерить размер моей неудачи, я спрашиваю, как много времени должно пройти, прежде чем один из моих друзей испытает бо́льшую неудачу (мы пренебрегаем событием $\mathbf{X}_k = \mathbf{X}_0$, вероятность которого равна нулю). Более формально, мы вводим время ожидания $\mathbf{N}$ *как значение первого индекса n, такого, что* $\mathbf{X}_n > \mathbf{X}_0$. Событие $\{\mathbf{N} > n-1\}$ происходит, если и только если максимальный член строки $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{n-1}$ является начальным; по соображениям симметрии вероятность этого события равна n^{-1}. Событие $\{\mathbf{N}=n\}$ — это то же, что $\{\mathbf{N}>n-1\} - \{\mathbf{N}>n\}$, и, следовательно, при $n=1, 2, \ldots$

$$\mathbf{P}\{\mathbf{N}=n\} = \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)}. \tag{5.1}$$

Этот результат полностью подтверждает, что мне действительно ужасно не везет: *случайная величина* $\mathbf{N}$ *имеет бесконечное математическое ожидание*! Было бы достаточно плохо, если бы нужно было провести в среднем 1000 испытаний, чтобы побить «рекорд» моих неудач, но действительное время ожидания имеет бесконечное математическое ожидание.

Отметим, что наше рассуждение не зависит от условия, что $\mathbf{X}_k$ показательно распределены. В действительности всякий раз, когда величины $\mathbf{X}_j$ независимы и имеют одну и ту же непрерывную функцию распределения, то первое «рекордное» значение имеет распределение (5.1). Тот факт, что это распределение не зависит от F, используется статистиками для проверки независимости. (См. также задачи 10 и 11.)

Замечательный и общий характер результата (5.1) в сочетании с простотой доказательства способны внушить подозрение. На самом деле доказательство безупречно (за исключением неформальных рассуждений), но те, кто предпочитает полагаться на надежные вычисления, могут легко проверить правильность (5.1), используя непосредственное определение рассматриваемой вероятности, как $(n+1)$-кратного интеграла от $\alpha^{n+1}e^{-\alpha(x_0+\cdots+x_n)}$ по области, определенной неравенствами $0<x_0<x_n$ и $0<x_j<x_0$ при $j=1, \ldots, n-1$.

Дадим *другой вывод* формулы (5.1), представляющий собой поучительное упражнение с условными вероятностями; он менее прост, но приводит к до-

полнительным результатам (задача 9). При условии что $\mathbf{X}_0=x$, вероятность большого значения в более поздних испытаниях равна $p=e^{-\alpha x}$, и мы имеем дело с временем ожидания первого «успеха» в испытаниях Бернулли с вероятностью успеха p. Условная вероятность, что $\mathbf{N}=n$ при условии $\mathbf{X}_0=x$, равна поэтому $p(1-p)^{n-1}$. Чтобы получить $\mathbf{P}\{\mathbf{N}=n\}$, мы должны умножить на плотность $\alpha e^{-\alpha x}$ и проинтегрировать по x. Подстановка $1-e^{-\alpha x}=t$ сводит подинтегральную функцию к функции $t^{n-1}(1-t)$, интеграл от которой равен $n^{-1}-(n+1)^{-1}$, что соответствует (5.1).

б) *Отношения.* Если $\mathbf{X}$ и $\mathbf{Y}$ — две независимые случайные величины с одинаковым показательным распределением, то отношение $\mathbf{Y}/\mathbf{X}$ является новой случайной величиной. Ее функция распределения получается интегрированием выражения $\alpha^2 e^{-\alpha(x+y)}$ по области $0<y<tx$, $0<x<\infty$. Интегрирование по y приводит к результату

$$\mathbf{P}\left\{\frac{\mathbf{Y}}{\mathbf{X}}\leqslant t\right\}=\int_0^\infty \alpha e^{-\alpha x}(1-e^{-\alpha t x})\,dx=\frac{t}{1+t}. \tag{5.2}$$

Соответствующая *плотность* равна $(1+t)^{-2}$. Заслуживает внимания то, что *величина* $\mathbf{Y}/\mathbf{X}$ *имеет бесконечное математическое ожидание.*

Мы находим здесь новое подтверждение устойчивости неудач. Конечно, Петр имеет основание для недовольства, если ему пришлось ждать в три раза больше, чем ждал Павел. Однако распределение (5.2) приписывает этому событию вероятность $^1/_4$. Отсюда следует, что в среднем в одном из двух случаев или Павел, или Петр имеют основания для недовольства. Наблюденная частота практически возрастает, так как очень короткие времена ожидания, естественно, проходят незамеченными.

в) *Параллельные очереди.* Я приезжаю в своем автомобиле на автомобильную инспекционную станцию (или ко въезду в туннель, или на автомобильный паром и т. д.). Имеются на выбор две очереди, но раз уж я встал в одну очередь, я должен стоять в ней. Мистер Смит, который приехал вслед за мной, занимает место, которое я мог бы выбрать, и я теперь слежу за тем, продвигается ли он быстрее или медленнее меня. Большее время мы стоим неподвижно, но время от времени то одна, то другая очередь передвигается на один автомобиль вперед. Чтобы увеличить влияние чистого случая, мы предположим две очереди стохастически независимыми; интервалы времени между последовательными продвижениями также являются независимыми величинами с одинаковым показательным распределением. При данных обстоятельствах последовательные продвижения представляют собой испытания Бернулли, где «успех» — это то,

что продвинулся вперед я, «неудача» — что продвинулся мистер Смит. Так как вероятности успеха равны $^1/_2$, мы в сущности имеем дело с симметричным случайным блужданием, и любопытные свойства флуктуаций при случайных блужданиях находят неожиданную интерпретацию. (Для простоты описания мы отвлечемся от того, что имеется только ограниченное число автомобилей.) Смогу ли я когда-нибудь обогнать мистера Смита? В переводе на язык случайных блужданий этот вопрос звучит так: будет ли когда-нибудь иметь место первое прохождение через $+1$. Как известно, это событие имеет вероятность, равную единицу. Однако среднее время ожидания для него бесконечно. Такое ожидание дает достаточные возможности для оплакивания моих неудач, причем раздражение мое только возрастает оттого, что мистер Смит приводит те же самые доводы. ▶

§ 6. Времена ожидания и порядковые статистики

Пусть дана упорядоченная строка $(x_1, \ldots, x_n)$ из n вещественных чисел. Переставляя их в порядке возрастания величины, получим новую строку

$$(x_{(1)}, x_{(2)}, \ldots, x_{(n)}), \text{ где } x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}.$$

Эта операция, примененная ко всем точкам пространства $\mathscr{R}^n$, дает нам n вполне определенных функций, которые будут обозначаться $\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n)}$. Если в $\mathscr{R}^n$ определены вероятности, то эти функции становятся случайными величинами. Мы говорим, что $(\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n)})$ получается перестановкой в соответствии с возрастанием величины из $\mathbf{X}_1, \ldots, \mathbf{X}_n$. Величина $\mathbf{X}_{(k)}$ называется *k-й порядковой статистикой* [1]) данной выборки $\mathbf{X}_1, \ldots, \mathbf{X}_n$, в частности $\mathbf{X}_{(1)}$ и $\mathbf{X}_{(n)}$ — это *выборочные экстремумы*; если $n = 2\nu + 1$ нечетно, то $\mathbf{X}_{(\nu+1)}$ есть *выборочная медиана*.

Мы применим эти понятия к частному случаю независимых случайных величин $\mathbf{X}_1, \ldots, \mathbf{X}_n$ с одинаковым показательным распределением с плотностью $\alpha e^{-\alpha x}$.

Пример. а) *Параллельные очереди.* Будем интерпретировать $\mathbf{X}_1, \ldots, \mathbf{X}_n$ как длительности n времен обслуживания, начиная с

[1]) Строго говоря, термин «выборочная статистика» является синонимом термина «функция от выборочных величин», т. е. термина «случайная величина». Он используется для лингвистического выделения различной роли, играемой в *данном контексте* первоначальными величинами (выборкой) и некоторыми полученными из них величинами. Например, «выборочное среднее» $(\mathbf{X}_1 + \ldots + \mathbf{X}_n)/n$ называется статистикой. Порядковые статистики часто встречаются в статистической литературе. Мы придерживаемся стандартной терминологии, за исключением того, что экстремумы обычно называются экстремальными *«значениями»*.

момента 0, в почтовом отделении с n окошками. Порядковые статистики изображают последовательные моменты окончаний, или, как говорится, *моменты последовательных разгрузок* («выходной процесс»). В частности, $\mathbf{X}_{(1)}$ представляет собой время ожидания первой разгрузки. Теперь если предположение об отсутствии последействия осмысленно, то время ожидания $\mathbf{X}_{(1)}$ должно обладать марковским свойством, т. е. $\mathbf{X}_{(1)}$ должно быть распределено показательно. Фактически событие $\{\mathbf{X}_{(1)}>t\}$ есть одновременная реализация n событий $\{\mathbf{X}_k>t\}$, каждое из которых имеет вероятность $e^{-\alpha t}$; ввиду предположенной независимости вероятности перемножаются, и мы имеем

$$\mathbf{P}\{\mathbf{X}_{(1)} > t\} = e^{-n\alpha t}. \tag{6.1}$$

Сделаем теперь следующий шаг и рассмотрим положение в момент $\mathbf{X}_{(1)}$. Предположенное отсутствие памяти, по-видимому, подразумевает, что восстанавливается первоначальное положение с той разницей, что теперь в операциях участвует $n-1$ окошко; продолжение процесса не должно зависеть от $\mathbf{X}_{(1)}$ и должно быть копией всего процесса. В частности, время ожидания следующей разгрузки, а именно $\mathbf{X}_{(2)}-\mathbf{X}_{(1)}$, должно иметь распределение

$$\mathbf{P}\{\mathbf{X}_{(2)} - \mathbf{X}_{(1)} > t\} = e^{-(n-1)\alpha t}, \tag{6.2}$$

аналогичное (6.1). Это рассуждение приводит к следующей общей теореме, касающейся порядковых статистик для независимых величин с одинаковым показательным распределением:

Предложение[1]). *Случайные величины* $\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)}-\mathbf{X}_{(1)}\ldots$ $\ldots\mathbf{X}_{(n)}-\mathbf{X}_{(n-1)}$ *являются независимыми, и плотность* $\mathbf{X}_{(k+1)}-\mathbf{X}_{(k)}$ *равна* $(n-k)\alpha e^{-(n-k)\alpha t}$.

Прежде чем проверить это утверждение формально, рассмотрим его следствия. Если $n=2$, то разность $\mathbf{X}_{(2)}-\mathbf{X}_{(1)}$ означает «остающееся время ожидания» после истечения более короткого из двух времен. Теорема утверждает, что это остаточное время ожидания имеет то же самое показательное распределение, что и первоначальное время ожидания, и не зависит от $\mathbf{X}_{(1)}$. Это есть расширение марковского свойства, сформулированного для *фиксированных моментов* t, на зависящее от случая время остановки $\mathbf{X}_{(1)}$. Это свойство называется *строго марковским свойством*. (Так как мы имеем дело лишь с конечным множеством величин, мы в состоянии *вывести* строго марковское свойство из

[1]) Эту теорему неоднократно открывали в связи с задачами статистического оценивания, но ее характер остается неясным без обращения к марковскому свойству. См. также задачу 14.

обычного, но в более сложных стохастических процессах различие существенно.)

Доказательство теоремы служит примером формальных действий с интегралами. Для упрощения типографского набора формул положим $n=3$. Мы начинаем с замечания [1]), что

$$\mathbf{P}\{\mathbf{X}_{(1)} > t_1,\ \mathbf{X}_{(2)} - \mathbf{X}_{(1)} > t_2,\ \mathbf{X}_{(3)} - \mathbf{X}_{(2)} > t_3\} = \\ = 3!\,\mathbf{P}\{\mathbf{X}_1 > t_1,\ \mathbf{X}_2 - \mathbf{X}_1 > t_2,\ \mathbf{X}_3 - \mathbf{X}_2 > t_3\}. \quad (6.3)$$

Это следует из того факта, что 3! перестановок из $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ имеют одно и то же распределение вероятностей. (Аналитически: пространство $\mathscr{R}^3$ разбивается на шесть частей, конгруэнтных области, определенной неравенствами $x_1<x_2<x_3$. Каждая из них вносит одну и ту же величину в соответствующий интеграл. Границы, где две или более координат равны, имеют вероятность нуль и не играют роли.) Чтобы получить вероятность (6.3), мы должны проинтегрировать $\alpha^3 e^{-\alpha(x_1+x_2+x_3)}$ по области, определенной неравенствами

$$x_1 > t_1,\ x_2 - x_1 > t_2,\ x_3 - x_2 > t_3.$$

Интегрирование по x_3 приводит к

$$3!\, e^{-\alpha t_3} \int_{t_1}^{\infty} \alpha e^{-\alpha x_1}\, dx_1 \int_{x_1+t_2}^{\infty} \alpha e^{-2\alpha x_2}\, dx_2 = \\ = 3e^{-\alpha t_3 - 2\alpha t_2} \int_{t_1}^{\infty} \alpha e^{-3\alpha x_1}\, dx_1 = e^{-\alpha t_3 - 2\alpha t_2 - 3\alpha t_1}. \quad (6.4)$$

Таким образом, совместное распределение трех величин $\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)} - \mathbf{X}_{(1)}$, $\mathbf{X}_{(3)} - \mathbf{X}_{(2)}$ есть произведение трех показательных распределений, и это доказывает наше утверждение.

Отсюда, в частности, следует, что $\mathbf{E}(\mathbf{X}_{(k+1)} - \mathbf{X}_{(k)}) = 1/(n-k)\alpha$. Суммируя по $k=0, 1, \ldots, \nu-1$, мы получаем

$$\mathbf{E}(\mathbf{X}_{(\nu)}) = \frac{1}{\alpha}\left(\frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{n-\nu+1}\right). \quad (6.5)$$

Отметим, что это математическое ожидание было вычислено без знания распределения $\mathbf{X}_{(\nu)}$. Итак, мы имеем еще один пример преимущества представления случайной величины как суммы других величин. (См. **1**, гл. IX, 3.)

б) *Использование строго марковского свойства.* В качестве образного примера рассмотрим следующую ситуацию: в мо-

[1]) Подобные вероятности большей частью вычисляются по тому же самому принципу.

мент 0 три лица A, B и C прибывают в почтовое отделение и находят два окна свободными. Три времени обслуживания представляют собой независимые случайные величины $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ с одинаковым показательным распределением. Обслуживание A и B начинается сразу же, а обслуживание C начинается в момент $\mathbf{X}_{(1)}$, когда или A, или B уходит. Мы покажем, что марковское свойство приводит к простым ответам на различные вопросы.

(i) Какова вероятность, что C не последним уйдет из почтового отделения? Ответ: $1/2$, так как момент $\mathbf{X}_{(1)}$ первого ухода устанавливает симметрию между C и другим обслуживаемым лицом.

(ii) Каково распределение времени $\mathbf{T}$, проведенного лицом C в почтовом отделении? Ясно, что величина $\mathbf{T}=\mathbf{X}_{(1)}+\mathbf{Z}$ есть сумма двух независимых величин, которые распределены показательно с параметрами 2α и α. Свертка любых двух показательных распределений легко находится (задача 6), и мы получаем, что $\mathbf{T}$ имеет плотность $u(t)=2\alpha(e^{-\alpha t}-e^{-2\alpha t})$ и что $\mathbf{E}(\mathbf{T})=\frac{3}{2\alpha}$.

(iii) Каково распределение момента последнего ухода? Обозначим моменты последовательных уходов $\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)}$, $\mathbf{X}_{(3)}$. Разность $\mathbf{X}_{(3)}-\mathbf{X}_{(1)}$ представляет собой максимум двух показательных времен обслуживания, и поэтому

$$\mathbf{P}\{\mathbf{X}_{(3)}-\mathbf{X}_{(1)}\leqslant t\}=(1-e^{-\alpha t})^2.$$

Это распределение снова имеет плотность u. Теперь $\mathbf{X}_{(1)}$ не зависит от $\mathbf{X}_{(3)}-\mathbf{X}_{(1)}$ и имеет плотность $2\alpha e^{-2\alpha t}$. Формула свертки, использовавшаяся в (ii), показывает, что $\mathbf{X}_{(3)}$ имеет плотность $4\alpha[e^{-\alpha t}-e^{-2\alpha t}-\alpha t e^{-2\alpha t}]$ и что $\mathbf{E}(\mathbf{X}_{(3)})=2/\alpha$.

Преимущества этого метода становятся очевидными при сравнении с непосредственным вычислением, но последнее применимо и в случае других распределений времени обслуживания (задача 18).

в) *Распределение порядковых статистик.* В качестве заключительного упражнения мы выведем распределение $\mathbf{X}_{(k)}$. Событие $\{\mathbf{X}_{(k)}\leqslant t\}$ означает, что по крайней мере k из n величин $\mathbf{X}_j$ не превосходят t. Это равносильно наступлению по крайней мере k «успехов» в n независимых испытаниях, и поэтому

$$\mathbf{P}\{\mathbf{X}_{(k)}\leqslant t\}=\sum_{j=k}^{n}\binom{n}{j}(1-e^{-\alpha t})^j e^{-(n-j)\alpha t}. \tag{6.6}$$

После дифференцирования видно, что *плотность* $\mathbf{X}_{(k)}$ равна

$$n\binom{n-1}{k-1}(1-e^{-\alpha t})^{k-1}e^{-(n-k)\alpha t}\cdot\alpha e^{-\alpha t}. \tag{6.7}$$

Этот результат может быть получен непосредственно следующими нестрогими рассуждениями. Мы разыскиваем (вплоть до членов пренебрежимо малых в пределе при $h \to 0$) вероятность события, заключающегося в том, что одна из величин $\mathbf{X}_j$ лежит между t и $t+h$, а $k-1$ величин из оставшихся $n-1$ величин не превосходят t, в то время как другие $n-k$ величин строго больше $t+h$. Перемножая число наборов и соответствующие вероятности, приходим к (6.7). Начинающим рекомендуется формализовать это рассуждение, а также вывести формулу (6.7) из дискретной модели. ➤

§ 7. Равномерное распределение

Случайная величина $\mathbf{X}$ *распределена равномерно в интервале* $\overline{a, b}$, если ее плотность постоянна — равна $(b-a)^{-1}$ при $a<x<b$ и равна 0 вне этого интервала. В этом случае величина $(\mathbf{X}-a)(b-a)^{-1}$ распределена равномерно в $\overline{0, 1}$, и мы будем обычно использовать этот интервал как стандартный. Из-за внешнего вида их графиков плотности равномерного распределения называются *«прямоугольными»*.

При равномерном распределении интервал $\overline{0, 1}$ становится выборочным пространством, в котором вероятности интервалов тождественны их длинам. Выборочное пространство, соответствующее двум независимым величинам $\mathbf{X}$ и $\mathbf{Y}$, которые равномерно распределены на $\overline{0, 1}$, представляет собой единичный квадрат в $\mathscr{R}^2$, а вероятности в нем совпадают с площадью. Те же идеи применимы к трем и большему числу величин.

Про равномерно распределенную случайную величину часто говорят: *«точка* $\mathbf{X}$, *выбранная наудачу»*. Результат мысленного эксперимента «n независимых случайных выборов точки в $\overline{0, 1}$» требует для своего вероятностного описания n-мерного гиперкуба, но эксперимент сам по себе дает n точек $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *в том же самом интервале*. С вероятностью единица никакие две из этих точек не совпадают, и, следовательно, они разбивают интервал $\overline{0, 1}$ на $n+1$ подинтервалов. Переставляя n точек $\mathbf{X}_1, \ldots, \mathbf{X}_n$ в их естественном порядке слева направо, мы получаем n новых случайных величин, которые обозначим $\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n)}$. Это есть *порядковые статистики*, определенные в предыдущем параграфе. Подинтервалы разбиения теперь имеют вид $\overline{0, \mathbf{X}_{(1)}}$, $\overline{\mathbf{X}_{(1)}, \mathbf{X}_{(2)}}$ и т. д.

Понятие *точки, выбранной наудачу на окружности*, говорит само за себя. Чтобы отчетливо представить себе исход n независимых выборов точки на окружности, мы вообразим *окружность, ориентированную* против часовой стрелки, так что интер-

валы имеют левый и правый концы и могут быть изображены в форме $\overline{a, b}$. Две точки $\mathbf{X}_1$ и $\mathbf{X}_2$, выбранные независимо и наудачу, делят окружность на два интервала $\overline{\mathbf{X}_1, \mathbf{X}_2}$ и $\overline{\mathbf{X}_2, \mathbf{X}_1}$. (Мы снова пренебрегаем событием $\mathbf{X}_1 = \mathbf{X}_2$, имеющим вероятность 0.)

Примеры. а) *Опытные интерпретации. Колесо рулетки* обычно приводится как средство осуществления «случайного выбора» на окружности. *Ошибка округления* при вычислениях с шестью десятичными знаками обычно рассматривается как случайная величина, распределенная равномерно в интервале длиной 10^{-6}. (Для ошибок, совершенных при отбрасывании двух последних десятичных знаков, дискретная модель со 100 возможными значениями более уместна, хотя менее практична.) Время ожидания пассажира, прибывшего на *автобусную станцию*, не зная расписания, может рассматриваться как случайная величина, равномерно распределенная в интервале между последовательными отъездами автобусов. Большой теоретический интерес представляют применения к *случайным разбиениям*, обсуждаемым в § 8. Во многих задачах математической статистики (таких, как непараметрические критерии) равномерное распределение входит косвенным образом: при условии, что произвольная случайная величина $\mathbf{X}$ имеет непрерывную функцию распределения F, случайная величина $F(\mathbf{X})$ *распределена равномерно на* $\overline{0, 1}$ (см. § 12).

б) *Индуцированное разбиение.* Мы докажем следующее утверждение: *n независимо и случайно выбранных на* $\overline{0, 1}$ *точек* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *разбивают* $\overline{0, 1}$ *на* $n+1$ *интервалов, длины которых имеют одно и то же распределение*

$$\mathbf{P}\{\mathbf{L} > t\} = (1-t)^n, \quad 0 < t < 1. \tag{7.1}$$

Отметим, что (7.1) относится к *хвосту* распределения.

Интуитивно можно было бы ожидать, что по крайней мере два конечных интервала должны иметь другое распределение. То, что все интервалы будут иметь одинаковое распределение, становится ясным при рассмотрении эквивалентной ситуации на (ориентированной) окружности единичной длины. Здесь $n+1$ точек $\mathbf{X}_1, \ldots, \mathbf{X}_{n+1}$, выбранные независимо и наудачу, разбивают окружность на $n+1$ интервалов, и по соображениям симметрии эти интервалы должны иметь одинаковое распределение. Представим теперь, что окружность разрезана в точке $\mathbf{X}_{n+1}$ для того, чтобы получить интервал, в котором $\mathbf{X}_1, \ldots, \mathbf{X}_n$ выбраны независимо и наудачу. Длины всех $n+1$ интервалов индуцированного разбиения имеют одно и то же распределение. Рассматривая самый левый интервал $\overline{0, \mathbf{X}_{(1)}}$, можно видеть, что это

распределение задается формулой (7.1). Длина этого интервала превосходит t, если все n точек $\mathbf{X}_1, \ldots, \mathbf{X}_n$ лежат в $\overline{t, 1}$. Вероятность этого равна $(1-t)^n$.

Хорошим упражнением является проверка утверждения в частном случае $n=2$ путем рассмотрения трех событий в единичном квадрате, изображающем выборочное пространство. [Сделаем замечание, необходимое при численной проверке. Вероятность события $\{\mathbf{X}_{(k+1)} - \mathbf{X}_{(k)} > t\}$ равна интегралу от постоянной 1 по объединению $n!$ конгруэнтных областей, определенных или цепью неравенств $x_1 < \ldots < x_j < x_k + t < x_{k+1} < \ldots < x_n$, или аналогичными цепями, полученными перестановкой индексов. Другой способ вычисления, приводящий к более сильному результату, содержится в примере (3, в) гл. III.]

в) *Парадокс* [1]). Пусть две точки $\mathbf{X}_1$ и $\mathbf{X}_2$ выбраны независимо и наудачу на окружности единичной длины. Тогда *длины двух интервалов* $\overline{\mathbf{X}_1, \mathbf{X}_2}$ *и* $\overline{\mathbf{X}_2, \mathbf{X}_1}$ *распределены равномерно, а длина* λ *интервала, содержащего произвольную точку* P, *имеет другое распределение* (*с плотностью* $2x$).

В частности, средняя длина каждого из двух интервалов равна $1/2$, а средняя длина интервала, содержащего P, равна $2/3$. Точка P берется фиксированной, но произвольной. Поэтому можно ожидать, что интервал, покрывающий P, выбран (заимствуя фразу у философов вероятности) «без предварительного знания его свойств». Наивная интуиция не подготовлена, конечно, воспринять большое различение между покрытием и непокрытием произвольной точки, но после должного размышления это различие становится «интуитивно очевидным». На самом деле, однако, даже довольно опытные авторы попадали в ловушку.

Для доказательства представим себе, что окружность разъединена в точке P. Нам остаются две точки, выбранные независимо и наудачу в $\overline{0, 1}$. Используя те же обозначения, что и раньше, мы находим, что событие $\{\lambda < t\}$ происходит тогда и только тогда, когда $\mathbf{X}_{(2)} - \mathbf{X}_{(1)} > 1 - t$. По формуле (7.1) вероятность этого равна t^2. Величина λ имеет потому плотность $2t$, как утверждалось. (Начинающим рекомендуется провести непосредственную вычислительную проверку.)

г) *Распределение порядковых статистик.* Если $\mathbf{X}_1, \ldots, \mathbf{X}_n$ независимы и распределены равномерно в интервале $\overline{0, 1}$, то число величин, удовлетворяющих неравенству $0 < \mathbf{X}_j \leqslant t < 1$, имеет биномиальное распределение с вероятностью «успеха», равной t. Теперь событие $\{\mathbf{X}_{(k)} \leqslant t\}$ происходит, если и только если

[1]) Тесно связан с парадоксом времени ожидания § 4.

по крайней мере k из наших величин не превосходят t, и, следовательно,

$$\mathbf{P}\{\mathbf{X}_{(k)} \leqslant t\} = \sum_{j=k}^{n} \binom{n}{j} t^j (1-t)^{n-j}. \tag{7.2}$$

Это дает нам функцию распределения k-й порядковой статистики. После дифференцирования находим, что *плотность* $\mathbf{X}_{(k)}$ *равна*

$$n\binom{n-1}{k-1} t^{k-1}(1-t)^{n-k}. \tag{7.3}$$

Это можно увидеть непосредственно из следующего: вероятность того, что одна из величин $\mathbf{X}_j$ лежит между t и $t+h$ и что $k-1$ из оставшихся величин меньше t, в то время как $n-k$ величин больше $t+h$, равна

$$n\binom{n-1}{k-1} t^{k-1}(1-t-h)^{n-k} h.$$

Деля последнее выражение на h и устремляя h к 0, получим (7.3).

д) *Предельные теоремы.* Чтобы понять характер распределения $\mathbf{X}_{(1)}$, когда n велико, лучше ввести $\mathbf{E}(\mathbf{X}_{(1)}) = (n+1)^{-1}$ как новую единицу измерения. Тогда при $n \to \infty$ мы получим для хвоста функции распределения

$$\mathbf{P}\{n\mathbf{X}_{(1)} > t\} = \left(1 - \frac{t}{n}\right)^n \to e^{-t}. \tag{7.4}$$

Словами это соотношение обычно описывают так: *в пределе* $\mathbf{X}_{(1)}$ *показательно распределена с математическим ожиданием* n^{-1}. Аналогично

$$\mathbf{P}\{n\mathbf{X}_{(2)} > t\} = \left(1 - \frac{t}{n}\right)^n + \binom{n}{1}\frac{t}{n}\left(1 - \frac{t}{n}\right)^{n-1} \to e^{-t} + te^{-t}, \tag{7.5}$$

и в правой части мы узнаем хвост гамма-распределения G_2 из формулы (3.5). Подобным способом легко проверить, что при каждом фиксированном k при $n \to \infty$ *распределение* $n\mathbf{X}_{(k)}$ *стремится к гамма-распределению* G_k (см. задачу 31).

Теперь G_k есть распределение суммы k независимых показательно распределенных величин, тогда как $\mathbf{X}_{(k)}$ есть сумма первых k интервалов, рассмотренных в примере (б). Мы можем поэтому сказать, что длины последовательных интервалов нашего разбиения ведут себя в пределе так, как если бы они были взаимно независимыми показательно распределенными величинами.

Ввиду очевидной связи (7.2) с биномиальным распределением для получения приближений к распределению $\mathbf{X}_{(k)}$, когда

и n и k велики, может быть использована центральная предельная теорема.

е) *Отношения.* Пусть $\mathbf{X}$ выбрано наудачу в $\overline{0, 1}$. Обозначим через $\mathbf{U}$ длину *более короткого* из интервалов $\overline{0, \mathbf{X}}$ и $\overline{\mathbf{X}, 1}$ и через $\mathbf{V}=1-\mathbf{U}$ длину более длинного. Случайная величина $\mathbf{U}$ равномерно распределена между 0 и $1/2$, так как событие $\{\mathbf{U}<t<1/2\}$ происходит тогда и только тогда, когда или $\mathbf{X}<t$, или $1-\mathbf{X}<t$, и потому имеет вероятность $2t$. По соображениям симметрии величина $\mathbf{V}$ распределена равномерно между $1/2$ и 1, и отсюда $\mathbf{E}(\mathbf{U})=1/4$, $\mathbf{E}(\mathbf{V})=3/4$. Что мы можем сказать об отношении $\mathbf{V}/\mathbf{U}$? Оно необходимо превосходит 1 и лежит между 1 и t тогда и только тогда, когда или

$$\frac{1}{1+t} \leqslant \mathbf{X} \leqslant \frac{1}{2},$$

или

$$\frac{1}{2} \leqslant \mathbf{X} \leqslant \frac{t}{1+t}.$$

При $t>1$ отсюда следует, что

$$\mathbf{P}\left\{\frac{\mathbf{V}}{\mathbf{U}} \leqslant t\right\} = \frac{t-1}{t+1}, \tag{7.6}$$

и плотность этого распределения равна $2(t+1)^{-2}$. Мы видим, что отношение $\mathbf{V}/\mathbf{U}$ *имеет бесконечное математическое ожидание.* Этот пример показывает, сколь малая информация содержится в наблюдении $\mathbf{E}(\mathbf{V})/\mathbf{E}(\mathbf{U})=3$. ▶

§ 8. Случайные разбиения

Задача этого параграфа завершает предыдущую цепь примеров и выделена из них отчасти ввиду ее важности в физике, отчасти ввиду того, что она может служить прототипом обычных марковских цепей.

Формально мы будем иметь дело с произведениями вида $\mathbf{Z}_n=\mathbf{X}_1\mathbf{X}_2\ldots\mathbf{X}_n$, где $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — взаимно независимые величины, равномерно распределенные в интервале $\overline{0, 1}$.

Примеры *для применений.* В некоторых процессах столкновения физическая *частица* расщепляется на две и ее масса m делится между ними. Различным процессам могут соответствовать различные законы разбиения, но чаще всего предполагается, что часть исходной массы, полученная каждой дочерней частицей, распределена равномерно в $\overline{0, 1}$. Если одна из двух частиц выбрана наудачу и подвергается новому столкновению,

то (в предположении отсутствия взаимодействия, т. е. независимости столкновений) массы двух частиц второго поколения задаются произведениями $m\mathbf{X}_1\mathbf{X}_2$ и т. д. (см. задачу 20). С небольшими терминологическими изменениями эта модель применима также к разбиениям зерен минералов или галек и т. п. Вместо масс можно также рассматривать *потери энергии* при столкновениях, и описание несколько упрощается, если заниматься изменениями энергии *одной и той же* частицы при последовательных столкновениях. В качестве последнего примера рассмотрим изменения в *интенсивности света*, проходящего через вещество. Пример (10, а) показывает, что, когда световой луч проходит через сферу радиуса R «в случайном направлении», расстояние, пробегаемое лучом в сфере, распределено равномерно между 0 и $2R$. При наличии равномерного поглощения такое прохождение будет уменьшать интенсивность случайного луча на величину, которая распределена равномерно в интервале $\overline{0, a}$ (где $a<1$ зависит от степени поглощения). Масштабный параметр не влияет серьезно на нашу модель, и видно, что n независимых прохождений будут уменьшать интенсивность света на величину вида $\mathbf{Z}_n$. ▶

Для того чтобы найти распределение $\mathbf{Z}_n$, мы можем действовать двумя способами.

(i) *Редукция к показательным распределениям.* Так как суммы обычно предпочтительнее произведений, мы перейдем к логарифмам, полагая $\mathbf{Y}_k = -\log \mathbf{X}_k$. Величины $\mathbf{Y}_k$ взаимно независимы, и при $t>0$

$$\mathbf{P}\{\mathbf{Y}_k \geqslant t\} = \mathbf{P}\{\mathbf{X}_k \leqslant e^{-t}\} = e^{-t}. \tag{8.1}$$

Функция распределения G_n суммы $\mathbf{S}_n = \mathbf{Y}_1 + \ldots + \mathbf{Y}_n$ n независимых показательно распределенных величин была получена в формуле (3.5), и тогда функция распределения величины $\mathbf{Z}_n = e^{-\mathbf{S}_n}$ задается выражением $1 - G_{n-1}(\log t^{-1})$, где $0 < t < 1$. Плотность этого распределения равна $t^{-1}g_{n-1}(\log t^{-1})$ или

$$f_n(t) = \frac{1}{n-1!}\left(\log \frac{1}{t}\right)^{n-1}, \quad 0 < t < 1. \tag{8.2}$$

Подстановка $t = e^{-x}$ показывает, что f_n действительно есть плотность с математическим ожиданием 2^{-n}.

Наша задача решается в явном виде. Этот метод показывает преимущества вывода с помощью соответствующего преобразования, но успех зависит от случайной эквивалентности нашей задачи и задачи предварительно решенной.

(ii) *Рекуррентная процедура* имеет преимущество в том, что она применима к родственным задачам вообще. Пусть $F_n(t) =$

$= \mathbf{P}\{\mathbf{Z}_n \leqslant t\}$ и $0<t<1$. По определению $F_1(t)=t$. Допустим, что F_{n-1} известно, и рассмотрим $\mathbf{Z}_n = \mathbf{Z}_{n-1}\mathbf{X}_n$ как произведение двух независимых величин. При условии $\mathbf{X}_n = x$ событие $\{\mathbf{Z}_n \leqslant t\}$ происходит тогда и только тогда, когда $\mathbf{Z}_{n-1} \leqslant t/x$ и имеет вероятность $F_{n-1}(t/x)$. Суммируя по всем возможным x, мы получаем

$$F_n(t) = \int_0^1 F_{n-1}\left(\frac{t}{x}\right) dx. \tag{8.3}$$

Эта формула в принципе позволяет нам вычислить последовательно $F_2, F_3, \ldots$. На практике удобнее иметь дело с соответствующими плотностями f_n. По предположению f_1 существует. Предположим по индукции существование f_{n-1}. Вспоминая, что $f_{n-1}(s) = 0$ при $s > 1$, мы получаем после дифференцирования (8.3)

$$f_n(t) = \int_t^1 f_{n-1}\left(\frac{t}{x}\right)\frac{dx}{x}, \quad 0 < t < 1, \tag{8.4}$$

и элементарные вычисления показывают, что f_n действительно задается формулой (8.2).

§ 9. Свертки и теоремы о покрытии

Результаты этого параграфа в какой-то мере занятны и сами по себе и в связи с некоторыми очевидными приложениями. Кроме того, они несколько неожиданно возникают в совсем, казалось бы, удаленных темах, таких, как, например, критерии значимости в гармоническом анализе [пример (3, е), гл. III], пуассоновские процессы [гл. XIV, (2, а)] и случайные сдвиги [пример (10, д)]. Поэтому не удивительно, что все формулы, так же как и их варианты, выводились несколько раз различными способами. Способ, используемый в последующем, выделяется простотой и применимостью к близким проблемам.

Пусть $a>0$ фиксировано и пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ взаимно независимые случайные величины, равномерно распределенные на $\overline{0, a}$. Пусть $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. Наша первая задача заключается в нахождении распределения U_n суммы $\mathbf{S}_n$ и ее плотности $u_n = U_n'$.

По определению $u_1(x) = a^{-1}$ при $0<x<a$ и $u_1(x) = 0$ в других точках (прямоугольная плотность). Для $n>1$ плотности u_n определяются формулой свертки (2.13), которая в данном случае имеет вид

$$u_{n+1}(x) = \frac{1}{a}\int_0^a u_n(x-y)\,dy = \frac{1}{a}\left[U_n(x) - U_n(x-a)\right]. \tag{9.1}$$

Легко видеть, что

$$u_2(x)=\begin{cases} xa^{-2} & 0\leqslant x\leqslant a \\ (2a-x)\,a^{-2}, & a\leqslant x\leqslant 2a \end{cases} \tag{9.2}$$

и, конечно, $u_2(x)=0$ для всех других x. График представляет собой треугольник, и u_2 называется *треугольной плотностью*. Аналогично u_3 объединяет три квадратных полинома, определенные соответственно в интервалах $\overline{0, a}$, $\overline{a, 2a}$ и $\overline{2a, 3a}$. Продолжая эти рассуждения, можно вскоре обнаружить общее правило для вычисления u_n. Чтобы выразить его символически, мы введем следующее

Обозначение. *Мы записываем положительную часть x в виде*

$$x_+=\frac{x+|x|}{2}. \tag{9.3}$$

Для удобства набора мы будем писать $f_+(x)$ [вместо $f(x)_+$] для положительной части $f(x)$ и, в частности, x_+^n для обозначения положительной части x^n [1]).

Другими словами, f_+ есть такая функция, что $f_+(x)=0$ всюду, где $f(x)\leqslant 0$ (обычно это обозначается $f_+=f\cap 0$). Заметим, что $(x-a)_+$ есть нуль при $x<a$ и линейная функция, когда $x>a$. С помощью этого обозначения равномерное распределение можно записать в форме

$$U_1(x)=(x_+-(x-a)_+)\,a^{-1}. \tag{9.4}$$

Теорема 1. *Пусть $U_n(x)=\mathbf{P}\{\mathbf{S}_n\leqslant x\}$. Обозначим через $u_n=U_n'$ плотность этого распределения. Тогда при $n=1, 2, \ldots$ и для всех x*

$$U_n(x)=\frac{1}{a^n n!}\sum_{\nu=0}^{n}(-1)^\nu\binom{n}{\nu}(x-\nu a)_+^n. \tag{9.5}$$

При [2]) $n=1, 2, 3, \ldots$

$$u_n(x)=\frac{1}{a^n(n-1)!}\sum_{\nu=0}^{n}(-1)^\nu\binom{n}{\nu}(x-\nu a)_+^{n-1}. \tag{9.6}$$

Заметим, что если точка x расположена между $(k-1)a$ и ka, то только k членов в формулах отличны от 0. При $x<0$ обе

[1]) Несмотря на это замечание, в последующих теоремах надо считать x_+^n равным $(x_+)^n$.

[2]) Формула (9.6) остается верной для $n=1$ при условии, что x_+^0 равно 0, если $x<0$, и 1, если $x>0$. Естественно, что U_1 имеет только односторонние производные в 0 и a.

суммы равны 0. Справедливо также то, что $u_n(x)$ равняется 0 при $x>na$, хотя из формулы (9.6) это не очевидно.

Доказательство. При $n=1$ справедливость (9.6) очевидна, а (9.5) сводится к (9.4). Допустим, что формула (9.5) верна. Формула (9.1) представляет u_{n+1} в виде разности двух известных сумм. Заменяя индекс суммирования ν в первой сумме на $\nu-1$ и используя тождество

$$\binom{n}{\nu}+\binom{n}{\nu-1}=\binom{n+1}{\nu},$$

мы получаем для $u_{n+1}(x)$ выражение, совпадающее с правой частью (9.6), где n заменено на $n+1$. Интегрирование доказывает справедливость (9.5) для $n+1$, и этим завершается доказательство.

(*Другое доказательство*, использующее переход к пределу в дискретной модели, содержится в [1], гл. XI, 7, задача 20.) ▶

Теорема 1 применима также к величинам, которые равномерно распределены в интервале $-b, b$, а не в $0, a$. В самом деле, если $a=2b$, то величины $\mathbf{X}_k-b$ принадлежат к такому типу, и, чтобы получить распределение $\sum_1^n(\mathbf{X}_k-b)$, мы должны просто заменить x на $x+nb$ в (9.5) и (9.6). Таким образом, *плотность суммы n независимых величин, распределенных равномерно в $-b, b$, задается выражением*

$$u_n(x+nb)=\frac{1}{(2b)^n(n-1)!}\sum_{\nu=0}^{n}(-1)^\nu\binom{n}{\nu}(x+(n-2\nu)b)_+^{n-1}. \quad (9.7)$$

По счастливому совпадению в (9.6) находит свой ответ задача, на вид не связанная с этой. Мы доказываем это вычислениями, но последующее комбинаторное доказательство лучше объясняет связь между двумя задачами.

Теорема 2[1]). *На окружности длины t заданы $n\geqslant 2$ дуг длины a, центры которых выбраны независимо и наудачу. Вероятность $\varphi_n(t)$ того, что эти n дуг покрывают всю окружность, равна*

$$\varphi_n(t)=a^n(n-1)!\,u_n(t)\frac{1}{t^{n-1}}, \quad (9.8)$$

[1]) Красивое геометрическое доказательство см. в работе Stevens W. L., Solution to a geometrical problem in probability, *Ann. Eugenics*, 2 (1939), 315—320. Стивенс также получил вероятность того, что имеется ровно k пробелов длины $\geqslant x$. С соглашением, содержащимся в предыдущей сноске, формула (9.9) также верна для $n=1$.

или, что то же самое,

$$\varphi_n(t)=\sum_{\nu=0}^{n}(-1)^{\nu}\binom{n}{\nu}\Big(1-\nu\frac{a}{t}\Big)_{+}^{n-1}. \tag{9.9}$$

Прежде чем доказывать это, придадим теореме форму, в которой она будет использована позже. Выберем один из n центров в качестве начала и развернем окружность в интервал длины t. Оставшиеся $n-1$ центров случайно распределятся в интервале $\overline{0,t}$, и теорема 2 выражает то же самое, что и

Теорема 3. *Пусть интервал $\overline{0,t}$ разбит на n подинтервалов посредством независимого и случайного выбора $n-1$ точек $\mathbf{X}_1,\ldots,\mathbf{X}_{n-1}$ деления. Вероятность $\varphi_n(t)$ того, что ни один из этих интервалов не превосходит a, равна* (9.9).

Отметим, что $\varphi_n(t)$, рассматриваемая при фиксированном t как функция a, представляет собой *функцию распределения максимальной длины среди длин n интервалов*, на которые разбит интервал $\overline{0,t}$. В связи с этим см. задачи 21—24.

Доказательство. Достаточно доказать теорему 3. Мы докажем рекуррентную формулу

$$\varphi_n(t)=(n-1)!\int_0^a\varphi_{n-1}(t-x)\Big(\frac{t-x}{t}\Big)^{n-2}\frac{dx}{t}. \tag{9.10}$$

Ее справедливость прямо следует из определения φ_n как $(n-1)$-кратного интеграла, но предпочтительнее толковать (9.10) вероятностно следующим образом. Наименьшее значение среди $\mathbf{X}_1,\ldots,\mathbf{X}_{n-1}$ должно быть меньше, чем a, причем существует $n-1$ возможностей выбрать это значение. При условии, что $\mathbf{X}_1=x$, вероятность того, что точка $\mathbf{X}_1$ самая левая, равна $\Big[\frac{t-x}{t}\Big]^{n-2}$; тогда $\mathbf{X}_2,\ldots,\mathbf{X}_{n-1}$ распределены равномерно в интервале $\overline{x,t}$ и (условная) вероятность того, что они удовлетворяют условиям теоремы, равна $\varphi_{n-1}(t-x)$. Суммируя все возможности, мы получаем (9.10). (Недоверчивые читатели могут повторить доказательство, заменяя предположение $\mathbf{X}_1=x$ на $x-h<\mathbf{X}_1<x$.)

Для упрощения выражения (9.10) мы введем новые функции, задаваемые формулой

$$\varphi_n(t)=(n-1)!\,a^n w_n(t)\frac{1}{t^{n-1}}. \tag{9.11}$$

Легко проверить, что $w_2 = u_2$ и

$$w_n(t) = \frac{1}{a}\int_0^a w_{n-1}(t-x)\,dx. \tag{9.12}$$

Это есть в точности рекуррентная формула (9.1), с которой мы начинали, и поэтому $w_n = u_n$ для всех n. ▶

Хотя это утверждение доказывает теорему, оно плохо объясняет таинственную связь между двумя вероятностями, входящими в формулу (9.8). Чтобы лучше разобраться в этом, рассмотрим теперь *дискретный аналог* (9.8). Эта модель интуитивно понятна, и *новое доказательство* двух теорем может быть получено простым переходом к пределу.

В целях сохранения обозначений допустим, что a и t — положительные целые числа, и пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ независимые величины, принимающие значения $1, 2, \ldots, a$ с вероятностью a^{-1} каждое. Обозначим через N_n число представлений $t = y_1 + y_2 + \ldots + y_n$ в виде суммы целых чисел $y_j \leqslant a$. По самому определению $N_n = a^n \mathbf{P}\{\mathbf{S}_n = t\}$. С другой стороны, пусть $\varphi_n(t)$ будет вероятностью того, что $n-1$ целых чисел, выбранных независимо и наудачу между 1 и $t-1$, разбивают интервал $\overline{0, t}$ на n подинтервалов длины $\leqslant a$. Так как число возможных выборов равно $\binom{t-1}{n-1}$, мы имеем $N_n = \binom{t-1}{n-1}\varphi_n(t)$, и, следовательно,

$$\binom{t-1}{n-1}\varphi_n(t) = a^n \mathbf{P}\{\mathbf{S}_n = t\}. \tag{9.13}$$

Это есть дискретный аналог формулы (9.8). При $t \to \infty$ биномиальный коэффициент эквивалентен $t^{n-1}/(n-1)!$ и (9.13) принимает вид (9.8) [Чтобы получить (9.8) формальным переходом к пределу, допустим, что $\mathbf{X}_j$ принимают значения $h, 2h, \ldots, \alpha h$, и рассмотрим интервал длины τh; положим затем $h \to 0$ и $\alpha h \to a$, $\tau h \to t$.]

§ 10. Случайные направления

Выбор случайного направления на плоскости $\mathscr{R}^2$ — это то же самое, что выбор наудачу точки на окружности. Если желательно задать направление его углом с правой частью x-оси, то на окружности следует ввести координату θ с $0 \leqslant \theta < 2\pi$. Для случайных направлений в пространстве $\mathscr{R}^3$ в качестве выборочного пространства служит единичная сфера; каждая область имеет вероятность, равную ее площади, деленной на 4π. Выбор случайного направления в $\mathscr{R}^3$ эквивалентен выбору наудачу точки на этой единичной сфере. Так как это включает пару случайных

величин (долготу и широту), то следовало бы отложить обсуждение трехмерного случая до гл. III, но в данном контексте он появляется более естественно.

Предложения. (i) *Обозначим через* **L** *длину проекции единичного вектора со случайным направлением в* $\mathscr{R}^3$ *на фиксированную прямую, скажем x-ось. Тогда величина* **L** *распределена равномерно на* $\overline{0, 1}$ *и* $\mathbf{E}(\mathbf{L}) = 1/2$.

(ii) *Пусть* **U** *— длина проекции того же вектора на фиксированную плоскость, скажем x, y-плоскость. Тогда* **U** *имеет плотность* $t/\sqrt{1-t^2}$ *при* $0<t<1$, *и* $\mathbf{E}(\mathbf{U}) = \pi/4$.

Важно то, что две проекции имеют различные распределения. То, что первое распределение равномерно, не есть свойство случайности, но зависит от числа измерений. Аналог теоремы (i) в $\mathscr{R}^2$ доставляет

Предложение (iii). *Пусть* **L** *— длина проекции случайного единичного вектора в* $\mathscr{R}^2$ *на x-ось. Тогда* **L** *имеет плотность* $2/(\pi\sqrt{1-x^2})$, *и* $\mathbf{E}(\mathbf{L}) = 2/\pi$.

Доказательства (iii). Если θ — угол между нашим случайным направлением и y-осью, то $\mathbf{L} = |\sin\theta|$ и, следовательно, при $0<x<1$

$$\mathbf{P}\{\mathbf{L} \leqslant x\} = \frac{2}{\pi} \arcsin x. \tag{10.1}$$

Предложение (iii) теперь доказывается дифференцированием.

(i), (ii). Вспомним элементарную теорему о том, что площадь сферического пояса между двумя параллельными плоскостями пропорциональна высоте пояса. При $0<t<1$ событие $\{\mathbf{L} \leqslant t\}$ представляется поясом $|x_1| \leqslant t$ высоты $2t$, тогда как событие $\{\mathbf{U} \leqslant t\}$ соответствует поясам $|x_3| \geqslant \sqrt{1-t^2}$ общей высоты $2 - 2\sqrt{1-t^2}$. Этим обе функции распределения определяются с точностью до числовых множителей. Их значения просто находятся из условия, что обе функции распределения равны 1 при $t=1$. ▶

Примеры. а) *Прохождение через сферу.* Пусть Σ — сфера радиуса r и N — точка на ней. Линия, проведенная через N в случайном направлении, пересекает Σ в точке P. Тогда: *Длина сегмента NP является случайной величиной, равномерно распределенной между* 0 *и* $2r$.

Чтобы увидеть это, рассмотрим ось NS сферы и треугольник NPS, который имеет прямой угол в точке P и угол Θ в точке N.

Длина NP равна тогда $2r\cos\Theta$. Но $\cos\Theta$ является в то же время проекцией единичного вектора прямой NP на диаметр NS, и поэтому величина $\cos\Theta$ равномерно распределена на $\overline{0,\ 1}$.

В физике эта модель употребляется для описания прохождения света через «случайно распределенные сферы». Возникающее при этом *поглощение света* использовалось в качестве одного из примеров процессов случайных разбиений в предыдущем параграфе (см. задачу 26).

б) *Круговые объекты под микроскопом.* Через микроскоп наблюдается проекция клетки на x_1, x_2-плоскость, а не ее действительная форма. В некоторых биологических экспериментах клетки имеют линзообразную форму и могут рассматриваться как круговые диски. Один лишь горизонтальный диаметр проектируется в свою натуральную величину, а весь диск проектируется в эллипс, малая ось которого является проекцией наиболее круто наклоненного радиуса. Обычно предполагается, что ориентация диска случайна; это означает, что направление его нормали выбрано наудачу. В этом случае проекция единичной нормали на x_3-ось распределена равномерно в $\overline{0,\ 1}$. Но угол между этой нормалью и x_3-осью равен углу между наиболее круто наклоненным радиусом и x_1, x_2-плоскостью. Следовательно, отношение малой и большой осей распределено равномерно в $\overline{0,\ 1}$. Иногда обработка экспериментов основывалась на ошибочном мнении, что угол между наиболее круто наклоненным радиусом и x_1, x_2-плоскостью должен быть распределен равномерно.

в) *Почему две скрипки звучат вдвое громче, чем одна?* (Вопрос серьезен, так как громкость пропорциональна *квадрату* амплитуды колебаний.) Поступающие волны могут быть изображены случайными единичными векторами, и эффект наложения звука двух скрипок соответствует сложению двух независимых случайных векторов. По теореме косинусов квадрат длины результирующего вектора равен $2+2\cos\Theta$. Здесь Θ — угол между двумя случайными векторами, и, следовательно, $\cos\Theta$ распределен равномерно в $\overline{-1,\ 1}$ и имеет нулевое математическое ожидание. Математическое ожидание квадрата длины суммарного вектора поэтому действительно равно 2. ▶

Под *случайным вектором в* $\mathscr{R}^3$ понимается вектор, проведенный в случайном направлении, длина которого $\mathbf{L}$ является случайной величиной, не зависящей от его направления. Вероятностные свойства случайного вектора полностью определяются свойствами его проекции на x-ось, и, используя последнюю, час-

то возможно избежать анализа в трех измерениях. Для этого важно знать связь между функцией распределения V истинной длины $\mathbf{L}$ и распределением F длины $\mathbf{L}_x$ проекции на x-ось. Имеем $\mathbf{L}_x = \mathbf{X}\mathbf{L}$, где $\mathbf{X}$ — длина проекции *единичного* вектора с данным направлением. Соответственно величина $\mathbf{X}$ равномерно распределена на $\overline{0, 1}$ и не зависит от $\mathbf{L}$. При условии $\mathbf{X} = x$ событие $\{\mathbf{L}_x \leqslant t\}$ происходит тогда и только тогда, когда $\mathbf{L} \leqslant t/x$, и поэтому [1])

$$F(t) = \int_0^1 V\left(\frac{t}{x}\right) dx, \qquad t > 0, \tag{10.2}$$

Соответствующие *плотности* мы получаем дифференцированием

$$f(t) = \int_0^1 v\left(\frac{t}{x}\right) \frac{dx}{x} = \int_t^\infty v(y) \frac{dy}{y}. \tag{10.3}$$

Повторное дифференцирование приводит к

$$v(t) = -tf'(t), \qquad t > 0, \tag{10.4}$$

Мы нашли таким образом *аналитическую связь между плотностью v длины случайного вектора в $\mathscr{R}^3$ и плотностью f длины его проекции на фиксированное направление.* Соотношение (10.3) используется для отыскания f, когда v известно, а (10.4) в противоположном случае. (Асимметрия между двумя формулами обусловлена тем фактом, что направление не является независимым от длины проекции.)

Примеры. г) *Распределение Максвелла для скоростей.* Рассмотрим случайные векторы в пространстве, проекции которых на x-ось имеют нормальную плотность с нулевым математическим ожиданием и единичной дисперсией. Длина проекции равна ее абсолютному значению, и поэтому

$$f(t) = \sqrt{\frac{2}{\pi}}\, e^{-\frac{1}{2}t^2}, \qquad t > 0. \tag{10.5}$$

Тогда из (10.4) выводим

$$v(t) = \sqrt{\frac{2}{\pi}}\, t^2 e^{-\frac{1}{2}t^2}, \qquad t > 0. \tag{10.6}$$

Это есть плотность Максвелла для скоростей в статистической механике. Обычный вывод объединяет предыдущее рассуждение

[1]) Это рассуждение повторяет доказательство формулы (8.3).

с доказательством того, что плотность f должна иметь вид (10.5). (Другой вывод см. в гл. III, 4.)

д) *Случайные сдвиги в* $\mathscr{R}^3$ (*Рэлей*). Рассмотрим n *единичных* векторов, направления которых выбираются независимо друг от друга и наудачу. Мы разыскиваем распределение *длины* $\mathbf{L}_n$ их векторной суммы. Вместо непосредственного изучения этой суммы мы рассмотрим ее проекцию на x-ось. Эта проекция, очевидно, равна сумме n независимых случайных величин, распределенных равномерно на $-1, 1$. Функция распределения этой проекции дается формулой (9.7) при $b=1$. *Длина* этой проекции равна ее абсолютному значению, и, используя (10.4), мы получаем после простого дифференцирования

$$v_n(x) = \frac{-x}{2^{n-1}(n-2)!} \sum_{\nu=0}^{n} (-1)^\nu \binom{n}{\nu} (x-2\nu)^{n-2} \qquad (10.7)$$

при $x>0$, и $v_n(x)=0$ при $x<0$. Таким образом, это есть плотность [1]) длины $\mathbf{L}_n$.

Эта задача возникает в физике и химии (векторы изображают, например, плоские волны или молекулярные связи). Сведéние к одному измерению, по-видимому, делает эту известную задачу тривиальной.

Тот же метод применим к случайным векторам произвольной длины, и таким образом (10.4) позволяет нам *свести задачи случайного блуждания в* $\mathscr{R}^3$ *к более простым задачам в* $\mathscr{R}^1$. Даже если трудно получить точное решение, ценную информацию дает центральная предельная теорема [см. пример (4, б) гл. VIII]. ▶

Случайные векторы в $\mathscr{R}^2$ определяются тем же способом, но соотношение, соответствующее (10.4), не так просто. Аналогом (10.2) служит соотношение

$$F(x) = \frac{2}{\pi} \int_0^{\pi/2} V\left(\frac{x}{\sin\theta}\right) d\theta, \qquad (10.8)$$

[1]) Обычно ссылаются на статью С. Чандрасекара, который вычислил v_3, v_4, v_6 и преобразование Фурье для v_n. Так как он использовал полярные координаты, его $W_n(x)$ нужно умножить на $4\pi x^2$, чтобы получить нашу плотность v_n. Непосредственный вывод формулы (10.7) (другими методами) принадлежит Кенуню. Статья Чандрасекара перепечатана в книге Вокса (1954).

но чтобы выразить V через F, мы должны опираться на относительно глубокую теорию интегрального уравнения Абеля [1]). Мы констатируем здесь без доказательства, что если F имеет непрерывную плотность f, то

$$1-V(x)=x\int_0^{\pi/2} f\left(\frac{x}{\sin\theta}\right)\frac{d\theta}{\sin^2\theta}. \tag{10.9}$$

(См. задачи 27—28.)

е) *Двойные орбиты.* При наблюдении спектрально двойных орбит астрономы могут измерить только проекции векторов на плоскость перпендикулярно к линии зрения. Эллипс в пространстве проектируется в эллипс в этой плоскости. Большая ось исходного эллипса лежит в плоскости, определяемой лучом зрения и проекцией этой оси, и поэтому разумно предположить, что угол между большой осью и ее проекцией равномерно распределен. Распределение проекции определяется (в принципе) измерениями. Распределение исходной большой оси дается тогда решением (10.9) интегрального уравнения Абеля. ▶

§ 11. Использование меры Лебега

Если множество A в $\overline{0,1}$ представляет собой объединение конечного множества непересекающихся интервалов I_1, I_2, ... с длинами λ_1, λ_2, ..., то равномерное распределение приписывает ему вероятность

$$\mathbf{P}\{A\}=\lambda_1+\lambda_2+\ldots \tag{11.1}$$

Следующие примеры покажут, что некоторые простые, но важные задачи приводят к объединениям бесконечного числа непересекающихся интервалов. Определение (11.1) будет применимо и отождествляет $\mathbf{P}\{A\}$ с лебеговой мерой A. Это согласуется с нашей программой отождествления вероятностей с интегралом от плотности $f(x)=1$, за исключением того, что мы используем интеграл Лебега, а не интеграл Римана (который может не существовать). Из теории Лебега нам требуется только тот факт, что если A представляет собой объединение возможно пересекающихся интервалов I_1, I_2, ..., то мера $\mathbf{P}\{A\}$ существует и не превосходит суммы $\lambda_1+\lambda_2+\ldots$ длин. Для непересекающихся интервалов выполняется равенство (11.1). Использование меры

[1]) Преобразование в интегральное уравнение Абеля достигается посредством замены переменных

$$F_1(x)=F\left(\frac{1}{\sqrt{x}}\right),\quad V_1(x)=V\left(\frac{1}{\sqrt{x}}\right)\quad\text{и}\quad x\sin^2\theta=y.$$

Тогда (10.8) принимает вид

$$F_1(t)=\frac{1}{\pi}\int_0^t\frac{V_1(y)}{\sqrt{y\,(t-y)}}\,dy.$$

Лебега соответствует свободной интуиции, и упрощает дело, так как обосновываются все формальные переходы к пределу. Множество N называется *нулевым множеством*, если оно содержится в множествах произвольно малой меры, т. е. если для каждого ε существует множество $A \supset N$, такое, что $\mathbf{P}\{A\}<\varepsilon$. В этом случае $\mathbf{P}\{N\}=0$.

В дальнейшем $\mathbf{X}$ обозначает случайную величину, распределенную равномерно в $\overline{0,1}$.

Примеры. а) *Какова вероятность того, что величина* $\mathbf{X}$ *рациональна?* Последовательность $\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \ldots$ содержит *все* рациональные числа интервала $\overline{0,1}$ (расположенные в порядке возрастания знаменателей). Выберем $\varepsilon<\frac{1}{2}$ и обозначим через J_k интервал длины ε^{k+1} с центром в k-й точке последовательности. Сумма длин интервалов J_k равна $\varepsilon^2+\varepsilon^3+\ldots<\varepsilon$, и их объединение покрывает рациональные числа. Поэтому по нашему определению *множество всех рациональных чисел имеет вероятность нуль*, и, следовательно, $\mathbf{X}$ иррационально с вероятностью единица.

Здесь уместно спросить, зачем такие множества рассматриваются в теории вероятностей. Один из ответов состоит в том, что их исключением ничего нельзя достичь и что использование теории Лебега действительно упрощает дело, не требуя новых методов. Второй ответ может быть более убедителен для начинающих и для нематематиков; ниже следующие варианты приводят к задачам, несомненно, вероятностной природы.

б) *С какой вероятностью цифра 7 встречается в десятичном разложении величины* $\mathbf{X}$? В десятичном разложении каждой точки x из открытого интервала между 0,7 и 0,8 цифра 7 появляется на первом месте. Для каждого n существуют 9^{n-1} интервалов длины 10^{-n}, содержащих только такие числа, что цифра 7 появляется на n-м месте, но не раньше (при $n=2$ их концами служат 0,07 и 0,08, затем 0,17 и 0,18 и т. д.). Эти интервалы не пересекаются, и их общая длина равна $\frac{1}{10}\left(1+\frac{9}{10}+\left(\frac{9}{10}\right)^2+\ldots\right)=1$. Таким образом, наше событие имеет *вероятность* 1.

Заметим, что некоторые числа имеют два представления, например $0{,}7=0{,}6999\ldots$. Чтобы сделать обсуждаемый нами вопрос определенным, мы должны поэтому точно установить, должна или может цифра 7 появиться в представлении, но наше рассуждение не зависит от этого различия. Причина в том, что только рациональные числа могут иметь два представления, а множество всех рациональных чисел имеет вероятность нуль.

в) *Бросание монеты и случайный выбор.* Посмотрим теперь, как «случайный выбор точки **X** между 0 и 1» может быть описан в терминах дискретных случайных величин. Обозначим через $\mathbf{X}_k(x)$ k-й десятичный знак x (чтобы избежать двусмысленностей, мы будем использовать, когда это возможно, обрывающиеся разложения). Случайная величина $\mathbf{X}_k$ принимает значения 0, 1, ..., 9, каждое с вероятностью $1/10$, и величины $\mathbf{X}_k$ взаимно независимы. Кроме того, мы имеем тождество

$$\mathbf{X} = \sum_{k=1}^{\infty} \frac{1}{10^k} \mathbf{X}_k. \tag{11.2}$$

Эта формула сводит случайный выбор точки **X** к последовательным выборам ее десятичных знаков.

В последующих рассуждениях мы перейдем от *десятичного* разложения к *двоичному*, т. е. мы заменим основание 10 на основание 2. Вместо (11.2) мы теперь имеем

$$\mathbf{X} = \sum_{k=1}^{\infty} \frac{1}{2^k} \mathbf{X}_k, \tag{11.3}$$

где $\mathbf{X}_k$ — взаимно независимые случайные величины, принимающие значения 0 и 1 с вероятностью $1/2$. Эти величины определены на интервале $\overline{0, 1}$, для которого вероятность отождествляется с мерой Лебега (длиной). Эта постановка задачи напоминает игру с бросанием монеты из первого тома, в которой выборочное пространство состоит из бесконечных последовательностей гербов и решеток или нулей и единиц. В этом выборочном пространстве возможно теперь заново интерпретировать (11.3). В этом пространстве $\mathbf{X}_k$ суть координатные величины, а **X** есть случайная величина, ими определяемая; ее функция распределения, конечно, равномерна. Заметим, что второе выборочное пространство содержит две различные выборочные точки 0111111... и 1000000..., хотя соответствующие двоичные разложения изображают одну и ту же точку $1/2$. Однако понятие нулевой вероятности дает нам возможность *отождествить два выборочных пространства.* Другими словами, при пренебрежении событием нулевой вероятности случайный выбор точки **X** между 0 и 1 может быть осуществлен последовательностью бросаний монеты, и наоборот, результат бесконечных бросаний монеты может быть изображен точкой x из интервала $\overline{0, 1}$. Каждая случайная величина, связанная с игрой в монету, может быть представлена функцией на $\overline{0, 1}$ и т. д. Этот удобный и наглядный прием

использовался в теории вероятностей издавна, но он связан с пренебрежением событиями, имеющими нулевую вероятность.

г) *Распределения канторовского типа.* Если мы будем рассматривать в (11.3) сумму слагаемых с четными номерами, или, что то же самое, рассматривать случайную величину

$$\mathbf{Y} = 3\sum_{\nu=1}^{\infty} \frac{1}{4^{\nu}} \mathbf{X}_{\nu}, \tag{11.4}$$

то мы придем к распределению с неожиданными свойствами. (Множитель 3 введен для упрощения рассуждений. Сумма слагаемых с нечетными номерами распределена так же, как $^2/_3\mathbf{Y}$.) Функция распределения $F(x) = \mathbf{P}\{\mathbf{Y} \leqslant x\}$ служит примером так называемых сингулярных распределений.

При подсчете мы рассматриваем $\mathbf{Y}$ как выигрыш игрока, который получает величину $3 \cdot 4^{-k}$, когда k-е бросание симметричной монеты дает в результате решетку. Полный выигрыш лежит между 0 и $3(4^{-1}+4^{-2}+\ldots) = 1$. Если первое испытание приводит к 1, то выигрыш $\geqslant {}^3/_4$, тогда как в противоположном случае $\mathbf{Y} \leqslant 3\ (4^{-2}+4^{-3}+\ldots) = 4^{-1}$. Таким образом, неравенство $^1/_4 < \mathbf{Y} < {}^3/_4$ не может быть осуществлено ни при каких обстоятельствах, и поэтому $F(x) = {}^1/_2$ в этом интервале длины $^1/_2$. Отсюда следует, что F не может иметь скачка, превышающего $^1/_2$.

Отметим теперь, что с точностью до множителя $^1/_4$ испытания с номерами 2, 3, ... представляют собой точную копию всей последовательности, и поэтому график $F(x)$ в интервале $0, {}^1/_4$ отличается от всего графика только преобразованием подобия

$$F(x) = \frac{1}{2} F(4x), \qquad 0 < x < {}^1/_4. \tag{11.5}$$

Отсюда следует, что $F(x) = {}^1/_4$ всюду в интервале длины $^1/_8$ с центром в точке $^1/_8$. По соображениям симметрии $F(x) = {}^3/_4$ всюду в интервале длины $^1/_8$ с центром в $^7/_8$. Мы нашли сейчас три интервала общей длины $^1/_2 + {}^2/_8 = {}^3/_4$, на каждом из которых F принимает постоянное значение, а именно $^1/_4$, $^1/_2$ или $^3/_4$. Следовательно, F не может иметь скачка, превышающего $^1/_4$. Остаются четыре интервала длиной $^1/_{16}$ каждый, и в каждом из них график F отличается от всего графика только преобразованием подобия. Каждый из четырех интервалов поэтому содержит подинтервал половины его длины, на котором F принимает постоянное значение (а именно $^1/_8$, $^3/_8$, $^5/_8$, $^7/_8$ соответственно). Продолжая в том же духе, мы найдем за n шагов $1+2+2^2+\ldots+2^{n-1}$ интервалов общей длины $2^{-1}+2^{-2}+2^{-3}+\ldots+2^{-n} = 1-2^{-n}$, в каждом из которых F принимает постоянное значение.

Таким образом, F *представляет собой непрерывную функцию, возрастающую от* $F(0)=0$ *до* $F(1)=1$ *так, что сумма длин интервалов постоянства равна* 1. Грубо говоря, все возрастание F происходит на множестве меры 0. Мы имеем здесь непрерывную функцию распределения F без плотности f. ▶

§ 12. Эмпирические распределения

«*Эмпирическая функция распределения*» F_n для n точек $a_1, \ldots, a_n$ на прямой есть ступенчатая функция со скачками $1/n$ в точках $a_1, \ldots, a_n$. Другими словами, $nF_n(x)$ равно числу точек a_k в $-\infty, x$, и F_n — функция распределения. Если даны n случайных величин $\mathbf{X}_1, \ldots, \mathbf{X}_n$, то их значения в некоторой точке выборочного пространства образуют строку из n чисел и ее эмпирическая функция распределения называется эмпирическим распределением выборки. При каждом x значение $\mathbf{F}_n(x)$ эмпирического распределения выборки определяет новую случайную величину, а *эмпирическое распределение* $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ — это все семейство случайных величин, зависящих от параметра x (формально мы имеем дело со случайным процессом, где x — временной параметр). Мы не будем пытаться развить здесь теорию эмпирических распределений, но это понятие можно использовать для иллюстрации возникновения сложных случайных величин в простых приложениях. Сверх того, мы в новом свете увидим равномерное распределение.

Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ обозначают взаимно независимые случайные величины с одинаковым непрерывным распределением F. Вероятность того, что любые две величины принимают одно и то же значение, равна нулю, и поэтому мы можем ограничить наше внимание выборками из n различных значений. При фиксированном x число величин $\mathbf{X}_k$, таких, что $\mathbf{X}_k \leqslant x$, имеет биномиальное распределение с вероятностью «успеха» $p=F(x)$, и поэтому *случайная величина* $\mathbf{F}_n(x)$ *имеет биномиальное распределение* с возможными значениями $0, 1/n, \ldots, 1$. Следовательно, при большом n и фиксированном x значение $\mathbf{F}_n(x)$ с большой вероятностью будет довольно близким к $F(x)$, и центральная предельная теорема информирует нас о возможных отклонениях. Более интересен (зависящий от случая) график $\mathbf{F}_n$ в целом и то, насколько он близок к F. Мерой этой близости может быть *максимум расхождения*, т. е.

$$\mathbf{D}_n = \sup_{-\infty < x < \infty} |\mathbf{F}_n(x) - F(x)|. \tag{12.1}$$

Эта новая случайная величина представляет большой интерес для статистиков в силу следующего свойства. *Распределение*

вероятностей случайной величины $\mathbf{D}_n$ *не зависит от* F (конечно, при условии, что F непрерывна).

Для доказательства этого свойства достаточно проверить, что распределение $\mathbf{D}_n$ остается неизменным, когда F заменяется равномерным распределением. Мы сначала покажем, что *величины* $\mathbf{Y}_k = F(\mathbf{X}_k)$ *распределены равномерно в* $\overline{0, 1}$. Для этой цели мы ограничим t интервалом $\overline{0, 1}$, и в этом интервале мы определим v как функцию, обратную F. Событие $\{F(\mathbf{X}_k) \leqslant t\}$ равносильно тогда событию $\{\mathbf{X}_k \leqslant v(t)\}$, которое имеет вероятность $F(v(t)) = t$. Таким образом, $\mathbf{P}\{\mathbf{Y}_k \leqslant t\} = t$, как и утверждалось.

Величины $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ взаимно независимы, и мы обозначим их эмпирическое распределение через $\mathbf{G}_n$. Только что приведенные рассуждения показывают также, что при фиксированном t случайная величина $\mathbf{G}_n(t)$ тождественна с $\mathbf{F}_n(v(t))$. Так как $t = F(v(t))$, то отсюда следует, что в каждой точке выборочного пространства $\mathscr{R}^n$

$$\sup |\mathbf{G}_n(t) - t| = \sup |\mathbf{F}_n(v(t)) - F(v(t))| = \mathbf{D}_n.$$

Этим доказательство завершается.

Тот факт, что распределение $\mathbf{D}_n$ не зависит от исходного распределения F, дает статистикам возможность построить критерии и процедуры оценивания, применимые в ситуации, когда исходное распределение неизвестно. В этой связи другие величины, связанные с $\mathbf{D}_n$, находят даже большее практическое использование.

Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n, \mathbf{X}_1^\sharp, \ldots, \mathbf{X}_n^\sharp$ — $2n$ взаимно независимых случайных величин с одинаковым непрерывным распределением F. Обозначим эмпирические распределения $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ и $(\mathbf{X}_1^\sharp, \ldots, \mathbf{X}_n^\sharp)$ через $\mathbf{F}_n$ и $\mathbf{F}_n^\sharp$ соответственно. Положим

$$\mathbf{D}_{n,n} = \sup_x |(\mathbf{F}_n(x) - \mathbf{F}_n^\sharp(x))|. \tag{12.2}$$

Это есть *максимальное расхождение между двумя эмпирическими распределениями*. Оно обладает тем же свойством, что и $\mathbf{D}_n$, а именно не зависит от распределения F. По этой причине $\mathbf{D}_{n,n}$ используется как статистический критерий проверки «гипотезы о том, что $\mathbf{X}_1, \ldots, \mathbf{X}_n$ и $\mathbf{X}_1^\sharp, \ldots, \mathbf{X}_n^\sharp$ являются случайными выборками из одной и той же популяции».

Распределение $\mathbf{D}_{n,n}$ было найдено с помощью громоздких вычислений и исследований, но в 1951 г. Гнеденко и Королюк показали, что вся проблема сводится к задаче случайного блуждания с хорошо известным решением. Их доказательство отличается изяществом, и мы приведем его для иллюстрации возможностей простых комбинаторных методов.

Теорема. *Вероятность* $\mathbf{P}\{\mathbf{D}_{n,n} \leqslant r/n\}$ *равна вероятности того, что в симметричном случайном блуждании траектория длины* $2n$, *начинающаяся и заканчивающаяся в нуле, содержится в замкнутом интервале* $\overline{-r, r}$.

Доказательство. Достаточно рассмотреть целые r. Расположим $2n$ величин $\mathbf{X}_1, \ldots, \mathbf{X}_n^{\#}$ в порядке возрастания величины и положим $\varepsilon_k = 1$ или $\varepsilon_k = -1$ в зависимости то того, какая величина занимает k-е место, $\mathbf{X}_j$ или $\mathbf{X}_j^{\#}$. Окончательная запись содержит n плюс единиц и n минус единиц, и все $\binom{2n}{n}$ расположений равновероятны. Поэтому возникающие при этом строки $(\varepsilon_1, \ldots, \varepsilon_{2n})$ находятся во взаимно однозначном соответствии с траекториями длины $2n$, начинающимися и заканчивающимися вначале. Если теперь $\varepsilon_1 + \ldots + \varepsilon_j = k$, то первые j мест заняты $(j+k)/2$ неотмеченными и $(j-k)/2$ отмеченными величинами, и поэтому существует точка x, такая, что $\mathbf{F}_n(x) = (j+k)/2n$ и $\mathbf{F}_n^{\#}(x) = (j-k)/2n$. Но тогда $|\mathbf{F}_n(x) - \mathbf{F}_n^{\#}(x)| = |k|/n$. То же рассуждение, проведенное в обратном порядке, завершает доказательство. ▶

Точное выражение для рассматриваемой вероятности содержится в **1**, гл. XIV, (9.1). В самом деле, выражение

$$\binom{2n}{n} \mathbf{P}\left\{\mathbf{D}_{n,n} \leqslant \frac{r}{n}\right\} = w_{r,n}$$

есть вероятность того, что частица, выходящая из начала отсчета, вернется в момент $2n$ в начало, не покидая интервал $\overline{-r, r}$. Последнее условие можно реализовать, если установить в точках $\pm r$ поглощающие экраны, и поэтому $w_{r,n}$ есть вероятность возвращения в начало координат в момент $2n$, когда точки $\pm r$ являются поглощающими экранами [В **1**, гл. XIV, (9.1) рассматривается интервал $\overline{0, a}$, а не $\overline{-r, r}$. Наша вероятность $w_{r,n}$ тождественна $u_{r,2n}(r)$.]

В **1**, гл. XIV было показано, что предельный переход приводит от случайных блужданий к диффузионным процессам, и таким образом легко видеть, что распределение $\sqrt{n}\,\mathbf{D}_{n,n}$ стремится к пределу. Фактически этот предел был найден Смирновым еще в 1939 г., а аналогичный предел для $\sqrt{n}\,\mathbf{D}_n$ — Колмогоровым в 1933 г. Эти вычисления очень сложны и не объясняют связи с диффузионными процессами, что, наоборот, свойственно подходу Гнеденко — Королюка. С другой стороны, они стимулировали плодотворную работу по сходимости стохастических процессов (Биллингслей, Донскер, Прохоров, Скороход и другие).

Можно упомянуть, что теоремы Смирнова в равной степени применимы и к расхождениям $\mathbf{D}_{m,n}$ между эмпирическими распределениями выборок, имеющих различные объемы m и n. Подход, связанный со случайным блужданием, применим к этой проблеме, но он во многом теряет свою изящность и простоту (Гнеденко, Рвачева). Статистиками было исследовано множество вариантов $\mathbf{D}_{m,n}$ (см. задачу 32).

§ 13. Задачи

Во всех задачах предполагается, что заданные величины *взаимно независимы*.

1. Пусть $\mathbf{X}$ и $\mathbf{Y}$ имеют плотности $\alpha e^{-\alpha x}$, сосредоточенные на $\overline{0, \infty}$. Найдите плотности величин

(i) $\mathbf{X}^3$ (ii) $3+2\mathbf{X}$

(iii) $\mathbf{X}-\mathbf{Y}$ (iv) $|\mathbf{X}-\mathbf{Y}|$

(v) Наименьшей из величин $\mathbf{X}$ и $\mathbf{Y}^3$ (vi) Наибольшей из величин $\mathbf{X}$ и $\mathbf{Y}^3$

2. Решите ту же задачу, если плотности $\mathbf{X}$ и $\mathbf{Y}$ равны $1/2$ в интервале $\overline{-1, 1}$ и 0 вне его.

3. Найдите плотности для $\mathbf{X}+\mathbf{Y}$ и $\mathbf{X}-\mathbf{Y}$, если $\mathbf{X}$ имеет плотность $\alpha e^{-\alpha x}$ $(x>0)$, а плотность $\mathbf{Y}$ равна h^{-1} при $0<x<h$.

4. Найдите вероятность того, что $\lambda^2-2a\lambda+b$ имеет комплексные корни, если коэффициенты a и b являются одинаково распределенными случайными величинами, плотность которых

(i) равномерна, т. е. равна h^{-1} при $0<x<h$;

(ii) показательна, т. е. равна $\alpha e^{-\alpha x}$ при $x>0$.

5. Найдите функции распределения величин $\dfrac{\mathbf{X}+\mathbf{Y}}{\mathbf{X}}$ и $\dfrac{\mathbf{X}+\mathbf{Y}}{\mathbf{Z}}$, если величины $\mathbf{X}$, $\mathbf{Y}$ и $\mathbf{Z}$ имеют одинаковое показательное распределение.

6. Если $\mathbf{X}$ и $\mathbf{Y}$ имеют показательные плотности $\alpha e^{-\alpha x}$ и $\beta e^{-\beta x}$, где $\alpha \neq \beta$ то $\mathbf{X}+\mathbf{Y}$ имеет плотность

$$\alpha\beta \frac{e^{-\alpha x} - e^{-\beta x}}{\beta - \alpha}, \quad x > 0.$$

7. Выведите формулу свертки (3.6) для показательного распределения прямым переходом к пределу из формулы свертки для «отрицательного биномиального» распределения из **1**, гл. VI, (8.1).

8. Для пуассоновского процесса § 4 обозначим через $\mathbf{Z}$ время между моментом t и последним предшествующим прибытием или 0 («возраст» текущего времени между прибытиями). Найдите распределение $\mathbf{Z}$ и покажите, что оно стремится к показательному распределению при $t \to \infty$.

9. В примере (5, а) покажите, что вероятность появления первого рекордного значения на n-м месте и при этом не превосходящего x равна

$$\frac{1}{n(n+1)}(1-e^{-\alpha x})^{n+1}.$$

Докажите, что распределение вероятностей первого рекордного значения есть $1-(1+\alpha x)e^{-\alpha x}$.

[Вообще, если величины $\mathbf{X}_j$ положительны и подчинены произвольному непрерывному распределению F, то первая вероятность равна $[n(n+1)]^{-1}F^{n+1}(x)$, а распределение первого рекорда равно $F-(1-F)\log(1-F)^{-1}$.]

10. *Образование колонн в транспорте* [1]). Автомобили отправляются друг за другом из начала координат и двигаются вдоль правой x-оси без обгона. Каждый автомобиль сохраняет постоянную скорость, но мы рассматриваем скорости $v_1, v_2, \ldots$ как выборку из последовательности независимых случайных величин с одинаковой непрерывной функцией распределения F. Когда один автомобиль догоняет другой, медленнее едущий автомобиль, ему приходится тащиться с той же скоростью. Таким образом будет сформирована колонна.

Покажите, что *вероятность того, что данный автомобиль тащится ровно за* $(n-1)$ *автомобилями, стремится к* $[n(n+1)]^{-1}$ независимо от распределения F и времени отъезда. Математическое ожидание числа автомобилей в колонне бесконечно.

11. *Обобщение* [2]) *примера* (5, а) *о рекордном значении.* Вместо того чтобы рассматривать одно предварительное наблюдение $\mathbf{X}_0$, мы начинаем с выборки $(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ с порядковыми статистиками $(\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(m)})$. (Общее для всех величин распределение F не играет роли, коль скоро оно непрерывно.)

а) Пусть $\mathbf{N}$ — первый индекс, такой, что $\mathbf{X}_{m+n} \geqslant \mathbf{X}_{(m)}$. Покажите, что $\mathbf{P}\{\mathbf{N}>n\}=m/(m+n)$. [В примере (5, а) мы имели $m=1$.]

б) Пусть $\mathbf{N}$ — первый индекс n, такой, что $\mathbf{X}_{m+n} \geqslant \mathbf{X}_{(m-r+1)}$; покажите, что

$$\mathbf{P}\{\mathbf{N} > n\} = \frac{\binom{m}{r}}{\binom{m+n}{r}}.$$

При $r \geqslant 2$ мы имеем $\mathbf{E}(\mathbf{N}) = m/(r-1)$ и

$$\mathbf{P}\{\mathbf{N} \leqslant mx\} \to 1 - \frac{1}{(1+x)^r}, \quad m \to \infty.$$

в) Пусть $\mathbf{N}$ — первый индекс, такой, что $\mathbf{X}_{m+n}$ находится вне интервала между $\mathbf{X}_{(1)}$ и $\mathbf{X}_{(m)}$. Тогда

$$\mathbf{P}\{\mathbf{N} > n\} = \frac{m(m-1)}{(m+n)(m+n-1)} \quad \text{и} \quad \mathbf{E}(\mathbf{N}) < \infty.$$

12. Пусть $\mathbf{X}_j$ при $j=0, \ldots, n$ имеет плотность $\lambda_j e^{-\lambda_j x}$ для $x>0$. Здесь $\lambda_j \neq \lambda_k$, если только j не равно k. Положим

$$\frac{1}{\Psi_{k,n}} = (\lambda_0 - \lambda_k) \ldots (\lambda_{k-1} - \lambda_k)(\lambda_{k+1} - \lambda_k) \ldots (\lambda_n - \lambda_k).$$

Покажите, что $\mathbf{X}_0 + \ldots + \mathbf{X}_n$ имеет плотность

$$\lambda_0 \ldots \lambda_n \left[\Psi_{0n} e^{-\lambda_0 x} + \ldots + \Psi_{n,n} e^{-\lambda_n x}\right]$$

Указание. Используйте индукцию, симметрию и задачу 6. Вычислений можно избежать.

Частный случай:

$$f_n(x) = n \sum_{k=1}^{n} (-1)^{k-1} \binom{n-1}{k-1} e^{-kx}, \quad x > 0$$

---

[1]) Newell G. F., *Opns Res.* 7 (1959), 589—598.

[2]) Wilks S. S., *J. Austr. Mat. Soc.*, 1 (1959), 106—112.

представляет собой свертку f_{n-1} с плотностью ne^{-nx}. Выведите, что f_{n-1} есть плотность размаха $\mathbf{X}_{(n)}-\mathbf{X}_{(1)}$ выборки $\mathbf{X}_1, \ldots, \mathbf{X}_n$, где все $\mathbf{X}_j$ имеют одинаковую плотность e^{-x}.

13. *Процессы чистого размножения.* В процессе чистого размножения из 1, гл. XVII, 3 система проходит через последовательность состояний $E_0 \to E_1 \to \ldots$, причем время пребывания в состоянии E_k имеет плотность $\lambda_k e^{-\lambda_k x}$. Таким образом, $\mathbf{S}_n = \mathbf{X}_0 + \ldots + \mathbf{X}_n$ есть момент перехода $E_n \to E_{n+1}$. Обозначьте через $P_n(t)$ вероятность E_n в момент t. Покажите, что $P_n(t) = \mathbf{P}\{\mathbf{S}_n > t\} - \mathbf{P}\{\mathbf{S}_{n-1} > t\}$ и, следовательно (используя последнюю задачу), что

$$P_n(t) = \lambda_0 \ldots \lambda_{n-1}\left[\Psi_{0,n} e^{-\lambda_0 t} + \ldots + \Psi_{n,n} e^{-\lambda_n t}\right]. \tag{1}$$

Дифференциальные уравнения $P_0'(t) = -\lambda_0 P_0(t)$

$$P_n'(t) = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t), \qquad n \geqslant 1 \tag{2}$$

можно вывести (а) из (1) и (б) из свойств сумм $\mathbf{S}_n$.

Указание. По соображениям симметрии достаточно рассмотреть коэффициент при $e^{-\lambda_0 t}$.

Замечание. Дифференциальные уравнения 2 совпадают с уравнениями 1, гл. XVII, (3.2), а (1) дает общее решение с $P_0(0) = 1$.

14. В примере (6, а), где рассматриваются параллельные очереди, мы скажем, что система находится в состоянии k, если k окошек свободны. Покажите, что здесь применима модель процесса чистого размножения (задача 13) с $\lambda_k = (n-k)\alpha$. Докажите, что

$$P_k(t) = \binom{n}{k}(1 - e^{-\alpha t})^k e^{-(n-k)\alpha t}.$$

Выведите отсюда распределение $\mathbf{X}_{(k)}$.

15. Рассмотрите две независимые очереди из m и $n > m$ лиц соответственно, предполагая, что времена обслуживания имеют одно и то же показательное распределение. Покажите, что вероятность более длинной очереди окончиться первой равна вероятности получить n гербов раньше, чем m решеток при бросаниях симметричной монеты. Найдите ту же вероятность, рассматривая отношение $\mathbf{X}/\mathbf{Y}$ двух величин, с гамма-распределениями G_m и G_n, заданными формулой (3.5).

16. *Пример статистического оценивания.* Предполагается, что продолжительности существования электрических ламп имеют показательное распределение с неизвестным математическим ожиданием α^{-1}. Чтобы оценить α, берется выборка из n ламп и наблюдаются продолжительности существования

$$\mathbf{X}_{(1)} < \mathbf{X}_{(2)} < \ldots < \mathbf{X}_{(r)}$$

первых r портящихся ламп. «Наилучшая несмещенная оценка» α^{-1} дается линейной комбинацией $\mathbf{U} = \lambda_1 \mathbf{X}_{(1)} + \cdots + \lambda_r \mathbf{X}_{(r)}$, такой, что $E(\mathbf{U}) = \alpha^{-1}$ и $\mathrm{Var}(\mathbf{U})$ имеет наименьшее возможное значение. Покажите, что

$$\mathbf{U} = (\mathbf{X}_{(1)} + \ldots + \mathbf{X}_{(r)})\frac{1}{r} + \mathbf{X}_{(r)}(n-r)\frac{1}{r},$$

и тогда $\mathrm{Var}(\mathbf{U}) = \dfrac{1}{r}\alpha^{-2}$.

Указание. Проделайте вычисления в терминах *независимых* величин $\mathbf{X}_{(k)} - \mathbf{X}_{(k-1)}$.

17. Пусть величины $\mathbf{X}_1, \ldots, \mathbf{X}_n$ распределены равномерно на $\overline{0, 1}$. Покажите, что размах $\mathbf{X}_{(n)} - \mathbf{X}_{(1)}$ имеет плотность $n(n-1)x^{n-2}(1-x)$ и математическое ожидание $(n-1)/(n+1)$. Какова вероятность, что все n точек лежат внутри некоторого интервала длины t?

18. Ответьте на вопросы примера (6, б), когда три времени ожидания равномерно распределены в интервале $\overline{0, 1}$.

19. На окружности независимо и наудачу выбраны четыре точки. Найдите вероятность того, что хорды $\mathbf{X}_1\mathbf{X}_2$ и $\mathbf{X}_3\mathbf{X}_4$ пересекаются а) без вычислений, используя лишь соображения симметрии; б) из определения через интеграл.

20. В процессе случайного разбиения из § 8 обозначьте через $\mathbf{X}_{11}$, $\mathbf{X}_{12}$, $\mathbf{X}_{21}$, $\mathbf{X}_{22}$ массы четырех осколков второго поколения, индекс 1 относится к наименьшей, а 2 к наибольшей части. Найдите плотности и математические ожидания этих величин.

21. Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ распределены наудачу в $\overline{0, t}$. Обозначим через $p_n(t)$ вероятность того, что все $n+1$ подинтервалов разбиения больше, чем h. Докажите, что

$$p_n(t) = \frac{n}{t^n} \int_0^{t-n} x^{n-1} p_{n-1}(x)\, dx,$$

откуда $p_n(t) = t^{-n}(t - (n+1)h)_+^n$.

[Функция распределения *наиболее короткого подинтервала равна* $1 - p_n(t)$, где t фиксировано, и $h > 0$ является независимой переменной.]

Указание [1]). Рассмотрите событие $\{\mathbf{X}_{(n)} = x\}$.

22. *Продолжение.* Найдите аналогичные выражения для вероятности $g_n(t)$ того, что все расстояния между $\mathbf{X}_k$ превышают $h > 0$ (при отсутствии условий в отношении концов).

23. *Продолжение.* Не используя решение предыдущих задач, найдите, что $p_n(t) = (t-2h)^n t^{-n} g_n(t-2h)$.

24. Сформулируйте аналог задачи 22 для окружности и покажите, что задача 21 дает ее решение.

25. Равнобедренный треугольник образован единичным вектором в x-направлении и единичным вектором в случайном направлении. Найдите распределение длины третьей стороны (i) в $\mathscr{R}^2$ и (ii) в $\mathscr{R}^3$.

---

[1]) *Добавление при корректуре.* То же самое простое рассуждение непосредственно приводит к следующему изящному общему утверждению, полученному недавно Б. де Финетти из геометрических соображений. Обозначим длины последовательных подинтервалов нашего разбиения через $\mathbf{L}_1, \ldots, \mathbf{L}_{n+1}$. Тогда для *произвольных* $x_i \geqslant 0, \ldots, x_{n+1} \geqslant 0$

$$\mathbf{P}\{\mathbf{L}_1 > x_1, \ldots, \mathbf{L}_{n+1} > x_{n+1}\} = t^{-n}(t - x_1 - \ldots - x_{n+1})_+^n. \qquad (*)$$

Задача 21 соответствует частному случаю, когда все x_j равны h, тогда как в примере (7, б) только одна из величин x_j положительна. Теорема о покрытиях 9.3 может быть получена из (*) с использованием формулы 1, гл. IV, (1.5) появления по крайней мере одного из событий.

[Статья де Финетти появилась в *Giornale Ist. Ital. degli Attuari*, 27 (1964), 151—173.]

26. Единичная окружность (сфера) с центром в 0 имеет на положительной x-оси северный полюс. Из полюса выходит луч, и его угол с x-осью распределен равномерно на $\overline{-1/2\pi, 1/2\pi}$. Найдите распределение длины хорды внутри окружности (сферы).

Замечание. В $\mathscr{R}^2$ луч имеет случайное направление, и мы имеем дело с аналогом примера (10, а). В $\mathscr{R}^3$ же задача другая.

27. Отношение *средних длин* случайного вектора и его проекции на x-ось равно 2 в $\mathscr{R}^3$ и $\pi/2$ в $\mathscr{R}^2$.

Указание. Используйте (10.4) и (10.9).

28. Длина случайного вектора распределена равномерно на $\overline{0, 1}$. Найдите плотность длины его проекции на x-ось а) в $\mathscr{R}^3$ и б) в $\mathscr{R}^2$.

Указание. Используйте (10.4) и (10.9).

29. Найдите функцию распределения угла между x-осью и случайно выбранным направлением в $\mathscr{R}^4$.

30. Найдите в $\mathscr{R}^4$ аналог соотношения (10.2) между распределением длин случайного вектора и распределением длины его проекции на x-ось. Проверьте результат задачи 29, ограничившись единичным вектором.

31. *Предельная теорема для порядковых статистик.* а) Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ распределены равномерно в $\overline{0, 1}$. Докажите, что при фиксированном k и $n \to \infty$

$$\mathbf{P}\left\{\mathbf{X}_{(k)} \leqslant \frac{x}{n}\right\} \to G_{k+1}(x), \qquad x > 0,$$

где G_{k+1} есть гамма-распределение (3.5) [см. пример (7, д)].

б) Если $\mathbf{X}_k$ имеют произвольную непрерывную функцию распределения F, то тот же предел существует и для $\mathbf{P}\{\mathbf{X}_{(k)} < \Phi(x/n)\}$, где Φ есть функция, обратная к F (Смирнов).

32. Докажите следующий вариант теоремы Гнеденко — Королюка в § 12:

$$\mathbf{P}\left\{\sup_x \left[\mathbf{F}_n(x) - \mathbf{F}_n^{\sharp}(x)\right] \geqslant \frac{r}{n}\right\} = \frac{\binom{2n}{n-r}}{\binom{2n}{n}},$$

где $r = 1, 2, \ldots, n$. [В отличие от первоначальной формулировки знаки абсолютных величин слева опущены и поэтому в соответствующем случайном блуждании появляется только один поглощающий барьер в точке r.]

33. *Нисходящие серии.* Случайная величина $\mathbf{N}$ определяется как единственный индекс, такой, что $\mathbf{X}_1 \geqslant \mathbf{X}_2 \geqslant \ldots \geqslant \mathbf{X}_{\mathbf{N}-1} < \mathbf{X}_{\mathbf{N}}$. Если величины $\mathbf{X}_j$ имеют одинаковое непрерывное распределение F, докажите, что $\mathbf{P}\{\mathbf{N} = n\} = (n-1)/n!$ и $\mathbf{E}(\mathbf{N}) = e$.

34. *Образование показательно распределенных случайных величин из равномерно распределенных величин*[1]. а) Пусть в предыдущей задаче F будет

[1]) von Neumann J., National Bureau of Standarts, *Appl. Math. Series*, **12** (1951), 36—38.

равномерным распределением в $\overline{0, 1}$. Докажите, что

$$\mathbf{P}\{\mathbf{X}_1 \leqslant x, \mathbf{N} = n\} = \frac{x^{n-1}}{(n-1)!} - \frac{x^n}{n!},$$

откуда $\mathbf{P}\{\mathbf{X}_1 \leqslant x, \mathbf{N}$ четно$\} = 1 - e^{-x}$.

б) Определим $\mathbf{Y}$ следующим образом: «Испытанием» служит последовательность $\mathbf{X}_1, \ldots, \mathbf{X}_N$; мы говорим о «неудаче», если $\mathbf{N}$ нечетно. Мы повторяем независимые испытания до появления первого «успеха». Пусть $\mathbf{Y}$ равно числу неудач плюс первая величина в успешном испытании. Докажите, что

$$\mathbf{P}\{\mathbf{Y} < x\} = 1 - e^{-x}.$$

ГЛАВА II

СПЕЦИАЛЬНЫЕ ПЛОТНОСТИ. РАНДОМИЗАЦИЯ

Основная задача этой главы состоит в том, чтобы перечислить для облегчения ссылок те плотности, которые чаще всего будут появляться в дальнейшем. Рандомизация, описанная во второй части, полезна для многих целей. В качестве иллюстрации дается вывод формул для некоторых распределений, связанных с функциями Бесселя и встречающихся в различных применениях. Мы увидим, что этот простой вероятностный прием заменяет сложные вычисления и трудоемкий анализ.

§ 1. Обозначения и определения

Мы говорим, что плотность f и соответствующее ей распределение F *сосредоточены* (или *сконцентрированы*) [1]) *на интервале* $I=a, b$, если $f(x)=0$ для всех x, не принадлежащих I. Тогда $F(x)=0$ для $x<a$ и $F(x)=1$ для $x>b$. Говорят, что два распределения F и G, а также их плотности f и g принадлежат *одному и тому же типу* или отличаются только «*параметрами расположения*», если они связаны соотношениями

$$G(x)=F(ax+b), \quad g(x)=af(ax+b), \tag{1.1}$$

где $a>0$. Мы будем часто называть b *центрирующим* параметром и a — *масштабным* параметром. Эти термины становятся хорошо понятными после следующего разъяснения: если F служит функцией распределения случайной величины $\mathbf{X}$, то G есть функция распределения величины

$$\mathbf{Y}=\frac{\mathbf{X}-b}{a}. \tag{1.2}$$

В действительности важен только тип распределения и мы, как правило, будем приводить только один образец каждого типа.

[1]) В соответствии с обычной терминологией наименьший замкнутый интервал I с таким свойством следовало бы называть *носителем* f. Новый термин введен в связи с тем, что мы будем употреблять его в более общем смысле: так, распределение может быть сосредоточено на множестве всех целых или всех рациональных чисел.

С помощью соответствующего выбора параметров расположения любой конечный интервал сводится к $\overline{0, 1}$, и мы будем рассматривать только плотности, сосредоточенные либо на всей прямой, либо на правой полупрямой, либо на $\overline{0, 1}$. *Указывая в формулах для плотностей* $0<x<1$ *или* $x>0$, *мы подразумеваем, что* $f(x)=0$ *для всех других* x.

Математическое ожидание m *и дисперсия* σ^2, соответствующие плотности f (или F), определяются равенствами

$$m=\int_{-\infty}^{+\infty} xf(x)\,dx, \quad \sigma^2=\int_{-\infty}^{+\infty}(x-m)^2 f(x)\,dx=\int_{-\infty}^{+\infty} x^2 f(x)\,dx-m^2, \tag{1.3}$$

если только интегралы сходятся абсолютно. Из формул (1.2) следует, что плотности g в этом случае соответствует математическое ожидание $(m-b)/a$ и дисперсия σ^2/a^2. Отсюда ясно, что для каждого типа существует самое большее одна плотность с нулевым математическим ожиданием и единичной дисперсией.

Вспомним [см. гл. I, (2.12)], что *свертка* $f=f_1*f_2$ двух плотностей f_1 и f_2 есть плотность вероятности, определяемая формулой

$$f(x)=\int_{-\infty}^{+\infty} f_1(x-y)f_2(y)\,dy. \tag{1.4}$$

Если f_1 и f_2 сосредоточены на $\overline{0, \infty}$, эта формула сводится к

$$f(x)=\int_0^x f_1(x-y)f_2(y)\,dy, \qquad x>0, \tag{1.5}$$

Функция f представляет собой плотность суммы двух независимых случайных величин с плотностями f_1 и f_2. Заметим, что для $g_i(x)=f_i(x+b_i)$ свертка $g=g_1*g_2$ дается равенством $g(x)=f(x+b_1+b_2)$, как это видно из (1.2).

Мы заканчиваем этот параграф введением обозначений [1])

$$\mathfrak{n}(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}, \quad \mathfrak{N}(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}y^2}\,dy. \tag{1.6}$$

Давно знакомая нам *нормальная плотность с математическим ожиданием* m *и дисперсией* σ^2 определяется как

$$\frac{1}{\sigma}\mathfrak{n}\left(\frac{x-m}{\sigma}\right), \qquad \sigma>0$$

[1]) В первом томе нормальная плотность $\mathfrak{n}$ обозначалась через φ, а функция распределения $\mathfrak{N}$ — через Φ.

В центральной предельной теореме неявно содержится существенный факт, что *семейство нормальных плотностей замкнуто относительно операции свертки*; другими словами свертка двух нормальных плотностей с математическими ожиданиями m_1, m_2 и дисперсиями σ_1^2, σ_2^2 есть нормальная плотность с математическим ожиданием m_1+m_2 и дисперсией $\sigma_1^2+\sigma_2^2$. Ввиду ранее сказанного достаточно доказать это для $m_1=m_2=0$. Утверждается, что

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{x^2}{2\sigma^2}\right]=\frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty}\exp\left[-\frac{(x-y)^2}{2\sigma_1^2}-\frac{y^2}{2\sigma_2^2}\right]dy. \tag{1.7}$$

Справедливость этого утверждения становится очевидной, если произвести замену переменных $z=y\frac{\sigma}{\sigma_1\sigma_2}-x\frac{\sigma_2}{\sigma\sigma_1}$, где x фиксировано. (См. задачу 1.)

§ 2. Гамма-распределения

Гамма-функция Γ определяется как

$$\Gamma(t)=\int_0^\infty x^{t-1}e^{-x}\,dx, \qquad t>0. \tag{2.1}$$

[См. 1, гл. II, (12.23).] Гамма-функция интерполирует факториалы в том смысле, что

$$\Gamma(n+1)=n! \quad \text{для} \quad n=0,\ 1,\ \ldots\ .$$

Интегрирование по частям показывает, что $\Gamma(t)=(t-1)\Gamma(t-1)$ при всех $t>0$.

Гамма-плотности, сосредоточенные на $\overline{0,\infty}$, определяются формулой

$$f_{\alpha,\nu}(x)=\frac{1}{\Gamma(\nu)}\alpha^\nu x^{\nu-1}e^{-\alpha x}, \qquad x>0. \tag{2.2}$$

Здесь $\alpha>0$ — тривиальный масштабный параметр, но $\nu>0$ — существенный параметр. В частном случае $f_{\alpha,1}$ мы имеем дело с *показательной* плотностью, а плотности $f_{\alpha,n+1}$ $(n=0,1,\ldots)$ совпадают с плотностями g_n из гл. I, (3.4). Простые вычисления показывают, что *математическое ожидание* $f_{\alpha,\nu}$ *равно* ν/α, *а дисперсия равна* ν/α^2.

Семейство гамма-плотностей замкнуто относительно операции свертки

$$f_{\alpha,\mu}*f_{\alpha,\nu}=f_{\alpha,\mu+\nu}, \qquad \mu>0,\ \nu>0. \tag{2.3}$$

Это важное свойство обобщает теорему из гл. I, 3 и будет постоянно использоваться; *доказательство* чрезвычайно просто. Согласно (1.5), левая часть равна

$$\frac{\alpha^{\mu+\nu}}{\Gamma(\mu)\,\Gamma(\nu)}\,e^{-\alpha x}\int_0^x (x-y)^{\mu-1}\,y^{\nu-1}\,dy. \tag{2.4}$$

После подстановки $y=xt$ это выражение будет отличаться от $f_{\alpha,\,\mu+\nu}$ только численным множителем, а он равен единице, так как и $f_{\alpha,\,\mu+\nu}$ и (2.4) являются плотностями вероятности.

Последний интеграл при $x=1$ — это так называемый *бета-интеграл* $B(\mu, \nu)$, и как побочный результат доказательства мы получаем, что

$$B(\mu, \nu)=\int_0^1 (1-y)^{\mu-1}\,y^{\nu-1}\,dy=\frac{\Gamma(\mu)\,\Gamma(\nu)}{\Gamma(\mu+\nu)} \tag{2.5}$$

для всех $\mu>0$, $\nu>0$. [С целыми μ и ν эта формула используется в **1**, гл. VI, (10.8) и (10.9). См. также задачу 3 этой главы.]

Что касается *графика* $f_{1,\,\nu}$, то он, очевидно, монотонен, если $\nu\leqslant 1$, и неограничен вблизи нуля, когда $\nu<1$. При $\nu>1$ график $f_{1,\,\nu}$ колоколообразен, $f_{1,\,\nu}$ достигает при $x=\nu-1$ максимума, равного $(\nu-1)^{\nu-1}\,e^{-(\nu-1)}/\Gamma(\nu)$, что близко к $[2\pi(\nu-1)]^{-1/2}$ [формула Стирлинга, **1**, гл. II, задача 12.22]. Из центральной предельной теоремы следует, что $\frac{\sqrt{\nu}}{\alpha}\,f_{\alpha,\,\nu}\left(\frac{\nu}{\alpha}+y\,\frac{\sqrt{\nu}}{\alpha}\right)\to n(x)$, когда $\nu\to\infty$.

§ 3 *. Распределения математической статистики, связанные с гамма-распределениями

Гамма-плотности играют значительную, хотя не всегда явную роль в математической статистике. Прежде всего они фигурируют как «тип III» в классической (теперь до некоторой степени устаревшей) системе плотностей, введенной К. Пирсоном (1894 г.). Более часто они встречаются благодаря тому, что для случайной величины $\mathbf{X}$ с нормальной плотностью n квадрат $\mathbf{X}^2$ имеет плотность $x^{-1/2}\,n(x^{1/2})=f_{1/2,\,1/2}(x)$. Ввиду свойства свертки (2.3) отсюда следует, что

Если $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *— взаимно независимые нормально распределенные величины с математическим ожиданием* 0 *и дисперсией* σ^2, *то величина* $\mathbf{X}_1^2+\ldots+\mathbf{X}_n^2$ *имеет плотность* $f_{1/2\sigma^2,\,n/2}$.

*) Этот параграф затрагивает специальные вопросы и не используется в дальнейшем.

Для статистиков величина $\chi^2 = \mathbf{X}_1^2 + \ldots + \mathbf{X}_n^2$ совпадает с «выборочной дисперсией для нормальной популяции», и ее распределение постоянно используется. В этой связи $f_{1/2,\,1/2n}$ по традиции (восходящей к К. Пирсону) называется *хи-квадрат плотностью с n степенями свободы.*

В статистической механике величина $\mathbf{X}_1^2 + \mathbf{X}_2^2 + \mathbf{X}_3^2$ появляется как квадрат скорости частиц. Сама скорость поэтому имеет плотность

$$v(x) = 2x f_{1/2,\,3/2}(x^2).$$

Это не что иное, как *плотность Максвелла*, найденная другим способом в гл. I, (10.6). (См. также пример в гл. III, 4.)

В теории очередей гамма-распределение иногда называют распределением Эрланга.

Некоторые случайные величины (или «статистики»), важные для статистиков, имеют форму $\mathbf{T} = \mathbf{X}/\mathbf{Y}$, где $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины, $\mathbf{Y} > 0$. Обозначим их функции распределения через F и G, а их плотности через f и g соответственно. Так как величина $\mathbf{Y}$ предполагается положительной, то g сосредоточено на $\overline{0, \infty}$, и поэтому

$$\mathbf{P}\{\mathbf{T} \leqslant t\} = \mathbf{P}\{\mathbf{X} \leqslant t\mathbf{Y}\} = \int_0^\infty F(ty)\, g(y)\, dy. \tag{3.1}$$

Дифференцируя, находим, что *отношение* $\mathbf{T} = \mathbf{X}/\mathbf{Y}$ *обладает плотностью*

$$w(t) = \int_0^\infty f(ty)\, y g(y)\, dy. \tag{3.2}$$

Примеры. а) *Если* $\mathbf{X}$ *и* $\mathbf{Y}$ *имеют плотности* $f_{1/2,\,1/2m}$ и $f_{1/2,\,1/2n}$, то $\mathbf{X}/\mathbf{Y}$ *имеет плотность*

$$w(t) = \frac{\Gamma\left(\frac{1}{2}(m+n)\right)}{\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)} \frac{t^{\frac{1}{2}m-1}}{(1+t)^{\frac{1}{2}(m+n)}} \qquad t > 0. \tag{3.3}$$

Действительно, интеграл (3.2) равен

$$\frac{t^{\frac{1}{2}m-1}}{\Gamma^{\frac{1}{2}(m+n)}\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)} \int_0^\infty y^{\frac{1}{2}(m+n)-1} e^{-\frac{1}{2}(1+t)y} dy, \tag{3.4}$$

а подстановкой $\frac{1}{2}(1+t)y = s$ он сводится к (3.3). В дисперсионном анализе рассматривают частный случай $\mathbf{X} = \mathbf{X}_1^2 + \ldots +$

$+\mathbf{X}_m^2$ и $\mathbf{Y}=\mathbf{Y}_1^2+\ldots+\mathbf{Y}_n^2$, где $\mathbf{X}_1, \ldots, \mathbf{X}_m, \mathbf{Y}_1, \ldots, \mathbf{Y}_n$ — взаимно независимые нормально распределенные величины с плотностью u. Случайная величина $\mathbf{F}=\frac{n}{m}\frac{\mathbf{X}}{\mathbf{Y}}$ называется статистикой Снедекора, а ее плотность $\frac{m}{n}w\left(\frac{m}{n}x\right)$ — *плотностью Снедекора* или *F*-плотностью. Величина $\mathbf{Z}=\log\frac{1}{2}\mathbf{F}$ есть *Z*-статистика Фишера, а ее плотность в свою очередь *Z*-плотность Фишера. Обе статистики являются, конечно, вариантами записи одной и той же величины.

б) *Если* $\mathbf{X}$ *имеет плотность* u, *а* $\mathbf{Y}^2$ — *плотность* $f_{1/2,\,1/2n}$, то *величина* $\mathbf{X}/\mathbf{Y}$ *имеет плотность*

$$w(t)=\frac{C_n}{(1+t^2)^{\frac{1}{2}(n+1)}},\quad \text{где } C_n=\frac{1}{\sqrt{\pi}}\,\frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\Gamma\left(\frac{1}{2}n\right)}. \tag{3.5}$$

Действительно, $\mathbf{Y}$ имеет плотность $2xf_{1/2,\,1/2n}(x^2)$, и выражение (3.2) принимает вид

$$2\cdot\frac{1}{\sqrt{2\pi}\,2^{\frac{1}{2}n}\,\Gamma\left(\frac{1}{2}n\right)}\int_0^\infty e^{-\frac{1}{2}(1+t^2)y^2}y^n\,dy. \tag{3.6}$$

Подстановка $s=\frac{1}{2}(1+t^2)y^2$ сводит (3.6) к (3.5).

В статистике выражение (3.5) называется *плотностью Стьюдента*. Это плотность распределения величины $\mathbf{X}/\sqrt{\mathbf{Y}_1^2+\ldots+\mathbf{Y}_n^2}$, где $\mathbf{X}, \mathbf{Y}_1, \ldots, \mathbf{Y}_n$ — независимые одинаково нормально распределенные величины с нулевым математическим ожиданием. (Дисперсия не играет роли ввиду однородности.) ▶

§ 4. Некоторые распространенные плотности

а) *Двусторонняя показательная* плотность определяется выражением $\frac{1}{2}\alpha e^{-\alpha|x|}$, где α — масштабный параметр. Ей соответствует нулевое математическое ожидание и дисперсия $2\alpha^{-2}$. Эта плотность представляет собой *свертку показательной плотности* $\alpha e^{-\alpha x}$ $(x>0)$ *со своим зеркальным отображением* $\alpha e^{\alpha x}$ $(x<0)$. Иными словами, двусторонняя показательная плотность является плотностью величины $\mathbf{X}_1-\mathbf{X}_2$, где $\mathbf{X}_1$ и $\mathbf{X}_2$ независимы и обладают одной и той же показательной плотностью $\alpha e^{-\alpha x}$ $(x>0)$. Во французской литературе это распределение обычно называют

«вторым законом распределения Лапласа» (первым считается нормальное распределение).

б) *Равномерная (или прямоугольная)* плотность ρ_a и *треугольная* плотность τ_a, сосредоточенные на $-a, a$, определяются соответственно выражениями

$$\rho_a(x)=\frac{1}{2a}, \quad \tau_a(x)=\frac{1}{a}\left(1-\frac{|x|}{a}\right), \quad |x|<a. \tag{4.1}$$

Легко видеть, что $\rho_a * \rho_a = \tau_{2a}$, или словами: сумма двух равномерно распределенных на $-a, a$ величин имеет треугольную плотность распределения на $-2a, 2a$. [Повторные свертки $\rho_a * \ldots * \rho_a$ были описаны в гл. I, (9.7).]

в) *Бета-плотности на* $\overline{0, 1}$ определяются как

$$\beta_{\mu,\nu}(x)=\frac{\Gamma(\mu+\nu)}{\Gamma(\mu)\,\Gamma(\nu)}(1-x)^{\mu-1}x^{\nu-1}, \quad 0<x<1, \tag{4.2}$$

где $\mu>0$ и $\nu>0$ — свободные параметры. То, что (4.2) действительно определяет плотность вероятностей, следует из (2.5). Из той же самой формулы видно, что $\beta_{\mu,\nu}$ имеет математическое ожидание $\nu/(\mu+\nu)$ и дисперсию $\mu\nu/[(\mu+\nu)^2(\mu+\nu+1)]$. Если $\mu<1$, $\nu<1$, то график $\beta_{\mu,\nu}$ является U-образным, уходящим в бесконечность при x, стремящемся к 0 или 1. При $\mu>1$, $\nu>1$ график колоколообразен.

Простой вариант бета-плотности определяется формулой

$$\frac{1}{(1+t)^2}\beta_{\mu,\nu}\left(\frac{1}{1+t}\right)=\frac{\Gamma(\mu+\nu)}{\Gamma(\mu)\,\Gamma(\nu)}\frac{t^{\mu-1}}{(1+t)^{\mu+\nu}}, \quad 0<t<\infty. \tag{4.3}$$

Если случайная величина $\mathbf{X}$ имеет плотность (4.2), то величина $\mathbf{Y}=\mathbf{X}^{-1}-1$ имеет плотность (4.3).

В системе плотностей Пирсона плотности (4.2) и (4.3) именуются плотностями типа I и VI соответственно. Плотность Снедекора (3.3) есть не что иное, как частный случай плотности (4.3). Плотности (4.3) иногда называются плотностями Парето (в честь экономиста *Парето*). Считалось (несколько наивно с современной статистической точки зрения), что распределения дохода должны иметь хвост с плотностью $\sim Ax^{-a}$ при $x\to\infty$, а (4.3) удовлетворяет этому требованию.

г) *Так называемая «арксинус-плотность»*

$$\frac{1}{\pi\sqrt{x(1-x)}}, \quad 0<x<1 \tag{4.4}$$

фактически то же самое, что и бета-плотность $\beta_{1/2,1/2}$, однако эта плотность заслуживает специального упоминания вследствие ее частого появления в теории флуктуаций. (Мы ввели ее в **1**, гл. III, 5 в связи с неожиданным поведением времени пребывания.) Вводящее в заблуждение название, к сожалению, обще-

употребительно; фактически функция распределения равна $2\pi^{-1}\arcsin\sqrt{x}$. (Бета-плотности с $\mu+\nu=1$ иногда называются «обобщенными арксинус-плотностями».)

д) *Плотность Коши* с центром в начале координат определяется как

$$\gamma_t(x)=\frac{1}{\pi}\cdot\frac{t}{t^2+x^2}, \qquad -\infty<x<\infty, \tag{4.5}$$

где $t>0$ — масштабный параметр. Соответствующая функция распределения равна $\pi^{-1}\operatorname{arctg}(x/t)$. График γ_t напоминает график нормальной плотности, но приближается к оси x столь медленно, что *математическое ожидание не существует.*

Важность плотностей Коши обусловлена формулой свертки для них

$$\gamma_s * \gamma_t = \gamma_{s+t}. \tag{4.6}$$

Этим утверждается, что *семейство плотностей Коши* (4.5) *замкнуто относительно свертки.* Формула (4.6) может быть доказана элементарным (но скучным) способом с помощью обычного разложения подинтегральной функции на элементарные дроби. Более простое доказательство опирается на анализ Фурье.

Формула свертки (4.6) имеет весьма интересное следствие, а именно: для независимых случайных величин $\mathbf{X}_1, \ldots, \mathbf{X}_n$, имеющих одинаковую плотность распределения (4.5), *среднее значение* $(\mathbf{X}_1+\ldots+\mathbf{X}_n)/n$ *распределено с той же самой плотностью, что и* $\mathbf{X}_j$.

Пример. Рассмотрим лабораторный опыт, заключающийся в том, что вертикальное зеркало проецирует горизонтальный световой луч на стену. Зеркало может свободно вращаться относительно вертикальной оси, проходящей через A. Мы предполагаем, что направление отраженного луча выбирается «случайно», т. е. угол φ между этим лучом и перпендикуляром AO к стене распределен равномерно между $-\frac{1}{2}\pi$ и $\frac{1}{2}\pi$. Световой луч пересекает стену в точке, которая лежит от точки O на расстоянии

$$\mathbf{X}=t\cdot\operatorname{tg}\varphi$$

(t равно расстоянию AO центра зеркала A от стены). Теперь очевидно, что случайная величина $\mathbf{X}$ имеет плотность (4.5) [1].

[1]) Простая переформулировка этого эксперимента приводит к физической интерпретации формулы свертки (4.6). Наше рассуждение показывает, что если единичный источник света находится в начале отсчета, то γ_t представляет распределение интенсивности света вдоль линии $y=t$ на x, y-плоскости. Тогда (4.6) выражает *принцип Гюйгенса,* согласно которому интенсивность света вдоль направления $y=s+t$ такая же, как если бы источник был распределен вдоль линии $y=t$ с плотностью γ_t. (Я обязан этим замечанием Дж. В. Уолшу.)

Если эксперимент повторен n раз, то среднее $(\mathbf{X}_1+\ldots+\mathbf{X}_n)/n$ имеет ту же самую плотность распределения и таким образом средние не накапливаются вокруг нуля, как можно было бы предположить по аналогии с законом больших чисел. ▶

Плотность распределения Коши обладает любопытным свойством: если $\mathbf{X}$ имеет плотность γ_t, то $2\mathbf{X}$ имеет плотность $\gamma_{2t}=\gamma_t*\gamma_t$. Таким образом, величина $2\mathbf{X}=\mathbf{X}+\mathbf{X}$ *является суммой двух зависимых величин, однако ее плотность задается формулой свертки.* Более того, если $\mathbf{U}$ и $\mathbf{V}$ — две независимые величины с одинаковой плотностью γ_t и $\mathbf{X}=a\mathbf{U}+b\mathbf{V}$, $\mathbf{Y}=c\mathbf{U}+d\mathbf{V}$, то величина $\mathbf{X}+\mathbf{Y}$ имеет плотность $\gamma_{(a+b+c+d)t}$, что представляет собой свертку плотностей $\gamma_{(a+b)t}$ величины $\mathbf{X}$ и $\gamma_{(c+d)t}$ величины $\mathbf{Y}$; однако $\mathbf{X}$ и $\mathbf{Y}$ не являются независимыми. (Схожий пример см. в задаче 1 в гл. III, 9).

[Плотность Коши соответствует частному случаю $n=1$ семейства (3.5) t-плотностей Стьюдента. Иными словами, если $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины с нормальной плотностью $\mathfrak{n}$, то величина $\mathbf{X}/|\mathbf{Y}|$ имеет плотность распределения Коши (4.5) с $t=1$. Некоторые другие родственные плотности см. в задачах 5, 6.]

Свойство свертки гамма-плотностей, выраженное формулой (2.3), выглядит точным подобием свойства (4.6), однако существует важное различие, состоящее в том, что для гамма-плотностей параметр α является существенным, в то время как формула (4.6) содержит только масштабный параметр. Тип распределений Коши является устойчивым *типом.* Эта устойчивость относительно операции свертки свойственна плотностям Коши и нормальным плотностям; различие заключается в том, что масштабные параметры определяются согласно правилам $\sigma^2=\sigma_1^2+\sigma_2^2$ и $\alpha=\alpha_1+\alpha_2$ соответственно. Имеются *другие устойчивые плотности* с подобными свойствами, и после систематизации терминологии мы будем называть нормальные плотности и плотности Коши «симметричными, устойчивыми с показателями 2 и 1» (см. гл. VI, 1).

е) *Одностороннее устойчивое распределение с показателем* $1/2$. Если $\mathfrak{N}$ — нормальное распределение, заданное формулой (1.6), то равенство

$$F_\alpha(x)=2\left[1-\mathfrak{N}\left(\frac{\alpha}{\sqrt{x}}\right)\right], \qquad x>0 \tag{4.7}$$

определяет функцию распределения с плотностью

$$f_\alpha(x)=\frac{\alpha}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{x^3}}e^{-\frac{1}{2}\alpha^2/x}, \qquad x>0. \tag{4.8}$$

Очевидно, что математическое ожидание не существует. Это распределение было найдено в **1**, гл. III, (8, в) и снова в **1**, гл. X, 1 как предел распределения времен возвращения. Отсюда легко вывести *правило композиции*

$$f_\alpha * f_\beta = f_\gamma, \quad \text{где} \quad \sqrt{\gamma} = \sqrt{\alpha} + \sqrt{\beta}. \tag{4.9}$$

(Возможна проверка элементарным, но несколько громоздким способом интегрирования. Доказательство, основанное на анализе Фурье, проще.) Если $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — независимые случайные величины с распределением (4.7), то из равенства (4.9) следует, что величина $(\mathbf{X}_1 + \ldots + \mathbf{X}_n)n^{-2}$ имеет то же самое распределение, и поэтому средние $|\mathbf{X}_1 + \ldots + \mathbf{X}_n|n^{-1}$ вероятно должны быть порядка n; вместо того чтобы сходиться, они безгранично возрастают (см. задачи 7 и 8).

ж) *Распределения вида* $e^{-x^{-\alpha}}$ $(x>0,\ \alpha>0)$ появляются в связи с порядковыми статистиками (см. задачу 8). Вместе с разновидностью $1-e^{-x^\alpha}$ они появляются (несколько таинственно) под именем распределений *Вейбула* в статистической теории надежности.

з) *Логистическая* функция *распределения*

$$F(t) = \frac{1}{1 + e^{-\alpha t - \beta}}, \qquad \alpha > 0 \tag{4.10}$$

может служить предостережением. Существует невероятно большая литература, где делаются попытки доказать трансцендентный «закон логистического развития». Предполагалось возможным представить практически все процессы развития (измеренные в соответствующих единицах) функцией типа (4.10) с t, изображающим время. Весьма длинные таблицы с хи-квадрат критериями подтверждали это положение для человеческих популяций, бактериальных колоний, развития железных дорог и т. д. Было обнаружено, что как высота, так и вес растений и животных подчиняются логистическому закону, хотя из теоретических соображений ясно, что эти две величины не могут подчиняться одному и тому же распределению. Лабораторные эксперименты на бактериях показали, что даже систематические нарушения не могут привести к другим результатам. Теория популяций была основана на логистических экстраполяциях (даже если они оказывались очевидным образом ненадежными). Единственное затруднение «логистической» теории заключается в том, что не только логистическое распределение, но также нормальное, Коши и другие распределения могут быть подогнаны под *тот же самый материал с тем же или лучшим согласием* [1]). В этой конкуренции логистическое распределение не играет никакой выдающейся роли: самые противоречивые теоретические модели могут быть подтверждены на том же наблюдательном материале.

Теории этого рода недолговечны, так как они не открывают новые пути, а новые подтверждения одних и тех же старых вещей очень скоро становятся надоедливыми. Однако наивное рассуждение само по себе не было за-

[1]) Feller W., On the logistic law of growth and its empirical verifications in biology, *Acta Biotheoretica*, **5** (1940), 51—66.

менено здравым смыслом, и поэтому может быть полезно иметь очевидное доказательство того, как могут вводить в заблуждение взятые сами по себе критерии согласия.

§ 5. Рандомизация и смеси

Пусть F — функция распределения, зависящая от параметра θ, и u — некоторая плотность вероятности. Тогда

$$W(x) = \int_{-\infty}^{+\infty} F(x, \theta)\, u(\theta)\, d\theta \tag{5.1}$$

есть монотонная функция x, возрастающая от 0 до 1, и, следовательно, функция распределения. Если F имеет непрерывную плотность f, то и W имеет плотность w, равную

$$w(x) = \int_{-\infty}^{+\infty} f(x, \theta)\, u(\theta)\, d\theta. \tag{5.2}$$

Вместо интегрирования относительно плотности u мы можем суммировать по отношению к дискретному распределению вероятностей: если $\theta_1, \theta_2, \ldots$ выбраны произвольно и если $p_k \geqslant 0$, $\Sigma p_k = 1$, то равенство

$$w(x) = \sum_k f(x, \theta_k)\, p_k \tag{5.3}$$

определяет новую плотность вероятностей. Процесс может быть описан вероятностно как *рандомизация*; параметр θ рассматривается как случайная величина, и новое распределение вероятностей определяется в x, θ-плоскости, которая служит выборочным пространством. Плотности вида (5.3) называются *смесями*, и это выражение здесь используется вообще для распределений, полученных рандомизацией (даже когда интегрирование проводится относительно общего распределения для θ).

Мы не предполагаем на этом этапе развивать общую теорию. Наша цель скорее в том, чтобы проиллюстрировать на нескольких примерах возможности этого метода и его вероятностное содержание. Примеры служат также подготовительным материалом к понятию условных вероятностей. Следующий параграф содержит примеры дискретных распределений, полученных посредством рандомизации непрерывного параметра. Наконец, § 7 иллюстрирует построение непрерывных процессов на основе случайных блужданий; попутно мы получим распределения, встречающиеся во многих приложениях и требующие при другом подходе затруднительных вычислений.

Примеры. а) *Отношения.* Если $\mathbf{X}$ — случайная величина с плотностью f, то при фиксированном $y > 0$ величина $\mathbf{X}/y$ имеет

плотность $f(xy)y$. Рассматривая параметр y как случайную величину с плотностью g, мы получаем новую плотность

$$w(x)=\int_{-\infty}^{+\infty} f(xy)\,yg(y)\,dy. \tag{5.4}$$

[Это то же самое, что и формула (3.2), которая была основой для обсуждений в § 3. Если $f=f_{1,\mu}$ и $g=f_{1,\nu}$, то результирующая плотность w дается формулой (4.3).]

На вероятностном языке рандомизация знаменателя y в $\mathbf{X}/y$ означает рассмотрение случайной величины $\mathbf{X}/\mathbf{Y}$, и мы можем просто повторить вывод формулы (3.2) для плотности величины $\mathbf{X}/\mathbf{Y}$. В этом частном случае терминология есть дело вкуса.

б) *Суммы случайного числа слагаемых.* Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ взаимно независимые случайные величины с одинаковой плотностью f. Сумма $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ имеет плотность f^{n*}, а именно n-кратную свертку f с собой (см. гл. I, 2). Число слагаемых n является параметром, который мы теперь рандомизируем с помощью распределения вероятностей $\mathbf{P}\{\mathbf{N}=n\}=p_n$. Плотность результирующей суммы $\mathbf{S}_\mathbf{N}$ со случайным числом слагаемых $\mathbf{N}$ равна

$$w=\sum_1^\infty p_n f^{n*}. \tag{5.5}$$

Возьмем, например, в качестве $\{p_n\}$ геометрическое распределение $p_n=qp^{n-1}$, а в качестве f — показательную плотность. Тогда $f^{n*}=g_{n-1}$ дается формулой (2.2) и

$$w(t)=qae^{-\alpha x}\sum_{n=1}^\infty p^{n-1}\frac{(\alpha x)^{n-1}}{(n-1)!}=qae^{-\alpha qx}. \tag{5.6}$$

Таким образом, w — снова показательная плотность.

в) *Применение к теории очередей.* Рассмотрим один обслуживающий прибор с показательным распределением времени обслуживания (плотность $f(t)=\mu e^{-\mu t}$) и предположим, что поступающая нагрузка — это нагрузка пуассоновского типа, т. е. что интервалы времени между вызовами независимы и имеют плотность распределения $\lambda e^{-\lambda t}$, $\lambda<\mu$. Данная модель описана в 1, гл. XVII, (7, в). Поступающие вызовы становятся в (возможно, пустую) «очередь» и обслуживаются в порядке поступления без перерывов. Было показано, что система стремится к стационарному состоянию, в котором вероятность наличия около прибора n абонентов (ожидающих или обслуживаемых) стремится к qp^n, где $p=\lambda/\mu$ [см. 1, XVII, (7.10) с $a=1$]. Здесь

$n=0, 1, \ldots$. Рассмотрим теперь вызов, прибывающий в момент времени t. Общее время $\mathbf{T}$, которое он проводит в системе обслуживания, равно сумме времени обслуживания его самого и времен обслуживания n абонентов, ставших в очередь до него; плотность $\mathbf{T}$ равна поэтому $f^{(n+1)*}=f_{\mu,\, n+1}$. Если предположить, что для очереди имеет место стационарное распределение, то последний пример показывает, что *полное время, проведенное вызовом в системе обслуживания, имеет плотность* $(\mu-\lambda)e^{-(\mu-\lambda)t}$ и математическое ожидание $1/(\mu-\lambda)$. (См. задачу 10.)

г) *Очереди на автобус.* Предполагается, что автобус появляется каждый час (в час, в два часа ... и т. д.), но может запаздывать. Мы рассмотрим последовательные запаздывания $\mathbf{X}_k$ как независимые случайные величины с одним и тем же распределением F и плотностью f. Для простоты мы предполагаем, что $0 \leqslant \mathbf{X}_k \leqslant 1$. Обозначим через $\mathbf{T}_x$ время ожидания для человека, прибывшего в момент $x<1$ после полудня. Вероятность, что автобус, назначенный на полдень, уже ушел, равна $F(x)$. Ясно видно, что

$$\mathbf{P}\{\mathbf{T}_x \leqslant t\} = \begin{cases} F(t+x)-F(x) & \text{при } 0<t<1-x, \\ 1-F(x)+F(x)F(t+x-1) & \text{при } 1-x<t<2-x, \end{cases} \tag{5.7}$$

и, конечно, $\mathbf{P}\{\mathbf{T}_x \leqslant t\}=1$ для всех бо́льших t. Соответствующая плотность равна

$$\begin{matrix} f(t+x) & \text{при} & 0<t<1-x, \\ F(x)f(t+x-1) & \text{при} & 1-x<t<2-x. \end{matrix} \tag{5.8}$$

Здесь момент x прибытия является свободным параметром, и его естественно рандомизировать. Например, для человека, прибывающего «наугад», момент прибытия представляет собой случайную величину, распределенную равномерно на $\overline{0, 1}$. Среднее время ожидания равно в данном случае $1/2+\sigma^2$, где σ^2 есть дисперсия времени запаздывания. Другими словами, *среднее время ожидания является наименьшим тогда, когда автобусы идут точно по расписанию*, и возрастает вместе с дисперсией запаздывания. (См. задачи 11,12.) ▶

§ 6. Дискретные распределения

Этот параграф посвящен беглому обзору некоторых результатов применения рандомизации к биномиальному и пуассоновскому распределениям.

Число $\mathbf{S}_n$ успехов в испытаниях Бернулли подчиняется распределению, зависящему от вероятности успеха p. Рассматривая p как случайную величину с плотностью u, мы придем к новому распределению

$$\mathbf{P}\{\mathbf{S}_n = k\} = \binom{n}{k} \int_0^1 p^k (1-p)^{n-k} u(p)\, dp, \quad k = 0, \ldots, n. \tag{6.1}$$

Примеры. а) Если $u(p) = 1$, то интегрированием по частям можно показать, что распределение, заданное формулой (6.1), не зависит от k и сводится к дискретному *равномерному распределению вероятностей* $\mathbf{P}\{\mathbf{S}_n = k\} = (n+1)^{-1}$. Более разъясняющей является аргументация Байеса. Рассмотрим $n+1$ независимых величин $\mathbf{X}_0, \ldots, \mathbf{X}_n$, распределенных равномерно между 0 и 1. Интеграл в выражении (6.1) (при $u = 1$) равен вероятности того, что ровно k величин из $\mathbf{X}_1, \ldots, \mathbf{X}_n$ будут меньше $\mathbf{X}_0$, или, другими словами, что при перечислении точек $\mathbf{X}_0, \ldots, \mathbf{X}_n$ в порядке возрастания величина $\mathbf{X}_0$ появляется на $(k+1)$-м месте. Но ввиду симметрии все положения равновероятны, и поэтому интеграл равен $(n+1)^{-1}$. ▶

На языке азартных игр (6.1) соответствует ситуации, когда несимметричная монета выбирается посредством случайного механизма, после чего испытания производятся с этой монетой (неизвестной структуры). Игроку испытания не кажутся независимыми; действительно, если наблюдается длинная последовательность гербов, то правдоподобно, что для нашей монеты p близко к 1, и поэтому безопасно делать ставку на дальнейшее появление гербов. Два формальных примера иллюстрируют оценивание и предсказание в задачах такого рода.

б) Дано, что n испытаний закончились k успехами (гипотеза H). Какова вероятность того, что $p < \alpha$? По определению условных вероятностей

$$\mathbf{P}\{A \mid H\} = \frac{\mathbf{P}\{AH\}}{\mathbf{P}\{H\}} = \frac{\int_0^\alpha p^k (1-p)^{n-k} u(p)\, dp}{\int_0^1 p^k (1-p)^{n-k} u(p)\, dp}. \tag{6.2}$$

Этот способ оценок использовался Байесом [при $u(p) = 1$]. В рамках нашей модели (т. е. если мы действительно занимаемся смешанной популяцией монет с *известной* плотностью u) не возникнет возражений против этого способа. Опасность в том, что этот способ постоянно использовался без оснований примени-

тельно к суждениям о «вероятностях причин», когда о рандомизации не было и речи; эта точка зрения широко обсуждалась в примере (2, д) в 1, гл. V в связи с вычислением пресловутой *вероятности того, что завтра солнце взойдет.*

в) Вариант предыдущей задачи можно сформулировать следующим образом. Дано, что n испытаний закончились k успехами; какова вероятность того, что следующие m испытаний закончатся j успехами? То же самое рассуждение приводит к ответу

$$\frac{\binom{m}{j}\int_0^1 p^{j+k}(1-p)^{m+n-j-k}\,u(p)\,dp}{\int_0^1 p^k(1-p)^{n-k}\,u(p)\,dp}. \tag{6.3}$$

▷

Обращаясь к *пуассоновскому распределению*, мы будем интерпретировать его как распределение числа «прибытий» за интервал времени длительностью t. Математическое ожидание числа прибытий равно αt. Мы проиллюстрируем две существенно различные процедуры рандомизации.

г) *Рандомизированное время.* Если длительность интервала времени является случайной величиной с плотностью u, то вероятность p_k ровно k прибытий равна

$$p_k = \int_0^\infty e^{-\alpha t}\frac{(\alpha t)^k}{k!}\,u(t)\,dt. \tag{6.4}$$

Например, если интервал времени распределен показательно, то вероятность $k=0, 1, \ldots$ новых прибытий равна

$$p_k = \int_0^\infty e^{-(\alpha+\beta)t}\frac{(\alpha t)^k}{k!}\beta\,dt = \frac{\beta}{\alpha+\beta}\cdot\left(\frac{\alpha}{\alpha+\beta}\right)^k, \tag{6.5}$$

что представляет собой геометрическое распределение.

д) *Расслоение.* Допустим, что имеется несколько независимых источников для случайных прибытий, каждый источник имеет пуассоновский выход, но параметры различны. Например, *аварии* в какой-либо установке в продолжение фиксированного промежутка времени t могут быть представлены пуассоновскими величинами, но параметр будет меняться от установки к установке. Аналогично *телефонные вызовы* от отдельных абонентов могут быть пуассоновскими с математическим ожиданием числа вызовов, меняющимся от абонента к або-

ненту. В таких процессах параметр α появляется как случайная величина с плотностью u, а вероятность ровно n вызовов за время t равна

$$P_n(t)=\int_0^\infty e^{-\alpha t}\frac{(\alpha t)^n}{n!}u(\alpha)\,d\alpha. \tag{6.6}$$

В частном случае гамма-плотности $u=f_{\beta,\nu+1}$ мы получим

$$P_n(t)=\binom{n+\nu}{n}\left(\frac{\beta}{\beta+t}\right)^{\nu+1}\left(\frac{t}{\beta+t}\right)^n, \tag{6.7}$$

что является предельной формой распределения Пойа, данного в **1**, гл. V, (8.3) и **1**, гл. XVII, (11.2) (если $\beta=a^{-1}$, $\nu=a^{-1}-1$). ▶

Замечание о ложном заражении. Курьезная и поучительная история приписывается распределению (6.7) и его двойственной природе.

Урновая модель Пойа и процесс Пойа, которые приводят к (6.7), являются моделями истинного заражения, где каждый случай эффективно увеличивает вероятность будущих случаев. Эта модель пользовалась большой популярностью. Для множества явлений эмпирически проверялось согласие с (6.7). При этом хороший результат воспринимался как *признак истинного заражения.*

По случайному совпадению то же самое распределение (6.7) было получено еще ранее (1920) М. Гринвудом и Дж. Юлом с тем замыслом, что хорошее согласие с ним *опровергает наличие заражения.* Их вывод приблизительно соответствует нашей модели расслоения, для которой начальным является предположение о том, что в основе лежит пуассоновский процесс, т. е. что полностью отсутствует последействие. Таким образом, мы имеем курьезный факт, что хорошее согласие наблюдений с одним и тем же распределением может быть истолковано двумя различными способами, диаметрально противоположными как по своей природе, так и по своим практическим последствиям. Это послужит предостережением против слишком поспешной интерпретации статистических данных.

Объяснение этому лежит в явлении *ложного* заражения, описанного в **1**, гл. V, (2, г) и выше в связи с (6.1). В данной ситуации, получив за интервал времени длиной s m прибытий, можно оценить вероятность n прибытий за следующий интервал времени t по формуле, аналогичной (6.3). Результат будет зависеть от m, но эта зависимость обусловлена способом выбора, а не природой данных; информация, относящаяся к прошлому, дает нам возможность делать лучшие предсказания относительно будущего поведения нашей выборки, но это не следует смешивать с будущим всей популяции.

§ 7. Бесселевы функции и случайные блуждания

Удивительно многие явные решения задач теории диффузии, теории очередей и других приложений содержат бесселевы функции. Обычно далеко не очевидно, что решения представляют собой распределения вероятностей, а аналитическая теория, требующаяся для получения их преобразований Лапласа и других

характеристик, довольно сложна. К счастью, рассматриваемые распределения (и многие другие) могут быть получены простой процедурой рандомизации. На этом пути многие соотношения теряют свой случайный характер, и можно избежать весьма трудного анализа.

Под *бесселевой функцией порядка* $\rho \geqslant -1$ мы будем понимать функцию I_ρ:

$$I_\rho(x) = \sum_{k=0}^{\infty} \frac{1}{k!\,\Gamma(k+\rho+1)} \left(\frac{x}{2}\right)^{2k+\rho}, \tag{7.1}$$

определенную для всех вещественных x [1]).

Мы приступаем к описанию трех процедур, ведущих к трем различным типам распределений, содержащих бесселевы функции.

а) Рандомизированные гамма-плотности

Пусть мы имеем при фиксированном $\rho > -1$ гамма-плотность $f_{1,\,\rho+k+1}$, заданную формулой (2.2). Рассматривая параметр k как целочисленную случайную величину, подчиненную пуассоновскому распределению, мы получаем в соответствии с (5.3) новую плотность

$$w_\rho(x) = e^{-t} \sum_{k=0}^{\infty} \frac{t^k}{k!} f_{1,\,\rho+k+1}(x) = e^{-t-x} \sum_{k=0}^{\infty} \frac{t^k x^{\rho+k}}{k!\,\Gamma(\rho+k+1)}. \tag{7.2}$$

Сравнив выражения (7.1) и (7.2), мы можем записать

$$w_\rho(x) = e^{-t-x} \sqrt{\left(\frac{x}{t}\right)^\rho}\, I_\rho(2\sqrt{tx}), \qquad x > 0. \tag{7.3}$$

Если $\rho > -1$, то w_ρ является *плотностью вероятностей*, сосредоточенной на $0, \infty$ (при $\rho = -1$ правая часть не интегрируема по x). Заметим, что t — не масштабный параметр, так что эти плотности принадлежат различным типам.

Между прочим, из этого построения и формулы свертки (2.3) для гамма-плотностей следует, что

$$w_\rho * f_{1,\,\nu} = w_{\rho+\nu}. \tag{7.4}$$

[1]) Согласно обычному словоупотреблению, I_ρ — это «модифицированная» бесселева функция или бесселева функция «с мнимым аргументом». «Обыкновенная» бесселева функция, всегда обозначаемая J_ρ, определяется добавлением множителя $(-1)^k$ в правую часть (7.1). Наше использование названия бесселевой функции должно рассматриваться как сокращение, а не как нововведение.

б) Рандомизированные случайные блуждания

При анализе случайных блужданий обычно считают, что последовательные переходы совершаются в моменты времени $1, 2, \ldots$. Должно быть, однако, ясно, что соглашение просто придает гармоничность описанию и что модель полностью независима от времени. Настоящие стохастические процессы с непрерывным временем получаются из обычных случайных блужданий при предположении, что *интервалы времени между последовательными скачками являются независимыми случайными величинами с одной и той же плотностью* e^{-t}. Другими словами, моменты скачков регулируются пуассоновским процессом, а скачки сами по себе являются случайными величинами, принимающими значения $+1$ и -1 с вероятностями p и q, независимо друг от друга и от пуассоновского процесса.

Каждому распределению, связанному со случайным блужданием, соответствует некоторое распределение для процесса с непрерывным временем, которое формально получается рандомизацией числа скачков. Чтобы увидеть процедуру в деталях, рассмотрим состояние процесса в заданный момент t. В исходном случайном блуждании n-й шаг приводит к состоянию $r \geqslant 0$ тогда и только тогда, когда среди первых n скачков $\frac{1}{2}(n+r)$ положительны и $\frac{1}{2}(n-r)$ отрицательны. Это возможно, если только $n-r=2\nu$ четно. В этом случае вероятность состояния r точно после n-го скачка равна

$$\binom{n}{\frac{1}{2}(n+r)} p^{\frac{1}{2}(n+r)} q^{\frac{1}{2}(n-r)} = \binom{r+2\nu}{r+\nu} p^{r+\nu} q^{\nu}. \tag{7.5}$$

Для нашего пуассоновского процесса вероятность того, что вплоть до момента t произойдет ровно n скачков, равна $e^{-t}t^n/n!$, и поэтому для нашего процесса с непрерывным временем вероятность состояния $r \geqslant 0$ в момент t равна

$$e^{-t} \sum_{\nu=0}^{\infty} \frac{t^{r+2\nu}}{(r+2\nu)!} \binom{r+2\nu}{r+\nu} p^{r+\nu} q^{\nu} = \sqrt{\left(\frac{p}{q}\right)^r}\, e^{-t} I_r(2\sqrt{pq}\, t), \tag{7.6}$$

и мы приходим к двум заключениям.

(i) Если мы определим $I_{-r} = I_r$ при $r = 1, 2, 3, \ldots$, тогда *при фиксированных* $t > 0$, p, q выражение

$$a_r(t) = \sqrt{\left(\frac{p}{q}\right)^r}\, e^{-t} I_r(2\sqrt{pq}\, t), \qquad r = 0, \pm 1, \pm 2, \ldots \tag{7.7}$$

представляет распределение вероятностей (т. е. $a_r \geqslant 0$, $\Sigma a_r = 1$).

(ii) В нашем случайном блуждании с непрерывным временем $a_r(t)$ *равно вероятности состояния* r *в момент* t.

Две известные формулы для бесселевых функций являются непосредственными следствиями этого результата. Во-первых, если изменить обозначения, полагая $2\sqrt{pqt}=x$ и $p/q=u^2$, то тождество $\Sigma a_r(t)=1$ превращается в

$$e^{\frac{1}{2}x(u+u^{-1})}=\sum_{-\infty}^{+\infty}u^r I_r(x). \tag{7.8}$$

Это так называемая производящая функция для *бесселевых функций* (*формула Шлемильха*, которая иногда служит в качестве определения для I_r).

Во-вторых, из природы нашего процесса ясно, что вероятности $a_r(t)$ должны удовлетворять уравнению Колмогорова — Чепмена

$$a_r(t+\tau)=\sum_{k=-\infty}^{\infty}a_k(t)\,a_{r-k}(\tau), \tag{7.9}$$

которое отражает тот факт, что в момент t частица должна быть в некотором состоянии k и что переход из k в r эквивалентен переходу из 0 в $r-k$. Мы возвратимся к этому соотношению в гл. XVII, 3. [Его легко проверить непосредственно, используя представление (7.6) и аналогичную формулу для вероятностей в случайном блуждании.] Уравнение Колмогорова — Чепмена (7.9) эквивалентно формуле

$$I_r(t+\tau)=\sum_{k=-\infty}^{\infty}I_k(t)\,I_{r-k}(\tau), \tag{7.10}$$

которая известна как *тождество К. Неймана.*

в) Первые прохождения

Для простоты сосредоточим наше внимание на симметричных случайных блужданиях $p=q=1/2$. Согласно **1**, гл. III, (4.11), вероятность того, что *первое прохождение* через точку $r>0$ произойдет во время скачка с номером $2n-r$, равна

$$\frac{r}{2n-r}\binom{2n-r}{n}2^{-2n+r}, \qquad n\geqslant r. \tag{7.11}$$

Случайное блуждание возвратно. Поэтому первое прохождение происходит с вероятностью единица, т. е. при фиксированном r величины (7.11) дают в сумме единицу. В нашем процессе момент k-го скачка имеет гамма-плотность $f_{1,k}$, заданную формулой (2.2). Отсюда следует, что момент первого прохождения через $r>0$ распределен с плотностью

$$\sum_n \frac{r}{2n-r}\binom{2n-r}{n}2^{-2n+r}f_{1,\,2n-r}(t)=$$

$$=e^{-t}\sum\frac{t^{2n-r-1}}{(2n-r-1)!}\left(\frac{r}{2n-r}\right)\cdot\frac{(2n-r)!}{n!\,(n-r)!}\,2^{-2n+r}=e^{-t}\,\frac{r}{t}\,I_r(t). \tag{7.12}$$

Таким образом, (i) *при фиксированном* $r=1, 2, \ldots$ *выражение*

$$v_r(t)=e^{-t}\frac{r}{t}\,I_r(t) \tag{7.13}$$

определяет плотность вероятности, сосредоточенную на $\overline{0, \infty}$.

(ii) *Момент первого прохождения через* $r>0$ *имеет плотность* v_r. (См. задачу 14.)

Этот результат позволяет сделать другие интересные заключения. Первое прохождение через $r+\rho$ в момент t предполагает предварительное первое прохождение через r в некоторый момент $s<t$. Вследствие независимости скачков в интервалах времени $\overline{0, s}$ и $\overline{s, t}$ и свойства отсутствия последействия у показательного времени ожидания мы должны иметь

$$v_r * v_\rho = v_{r+\rho}. \tag{7.14}$$

[Проверка этого соотношения по формуле (7.12) будет легкой, если использовать соответствующее свойство свертки для вероятностей (7.11).]

Фактически утверждение (i) и соотношение (7.14) справедливы для всех положительных значений параметров r и ρ [1]).

§ 8. Распределения на окружности

Полуоткрытый интервал $\overline{0, 1}$ можно рассматривать как изображение окружности единичной длины, но предпочтительнее обернуть всю прямую вокруг окружности. Окружность тогда приобретает ориентацию, а длина дуги изменяется от $-\infty$ до ∞, но x, $x\pm1$, $x\pm2$, ... интерпретируются как одна и та же точка. Сложение происходит по модулю 1 (так же как сложение углов по модулю 2π). *Плотность распределения вероятностей на окружности есть периодическая функция* $\varphi\geqslant0$, *такая, что*

$$\int_0^1 \varphi(x)\,dx = 1. \tag{8.1}$$

Примеры. а) *Задача Бюффона об игле* (1777 г.). Традиционная формулировка такова: плоскость разбивается на параллельные полосы единичной ширины. Произвольным образом бросается на плоскость игла единичной длины. Какова вероятность, что игла ляжет на плоскости так, что заденет две полосы? Чтобы поставить задачу формально, рассмотрим сначала *центр* иглы. Его положение определяется двумя координатами, но y мы пренебрегаем, а x приводим по модулю 1. Таким образом, «центр иглы» превращается в случайную величину $\mathbf{X}$, равномерно

[1]) Feller W., Infinitely divisible distributions and Bessel functions associated with random walks. *J. Soc. Indust. Appl Math.* (1966)

распределенную на окружности. *Направление* иглы может быть описано углом (измеренным по часовой стрелке) между иглой и x-осью. Поворот на угол π возвращает иглу в то же самое положение. Так как мы имеем дело с окружностью единичной длины, обозначим рассматриваемый угол через $\mathbf{Z}\pi$. В задаче Бюффона подразумевается, что $\mathbf{X}$ и $\mathbf{Z}$ — независимые равномерно распределенные на окружности единичной длины величины [1]).

Если мы выбираем значения $\mathbf{X}$ между 0 и 1, а значения $\mathbf{Z}$ между $-1/2$ и $1/2$, то игла пересекает границу тогда и только тогда, когда $\frac{1}{2}\cos \mathbf{Z}\pi > \mathbf{X}$ или $\frac{1}{2}\cos \mathbf{Z}\pi > 1 - \mathbf{X}$. По соображениям симметрии два события равновероятны, и поэтому искомая вероятность равна

$$\int_{-1/2}^{1/2} \cos z\pi \cdot dz = \frac{2}{\pi}. \tag{8.2}$$ ▶

Случайная величина $\mathbf{X}$ на прямой может быть преобразована по модулю 1 в величину $\mathbf{X}^0$ на окружности. Случайными величинами такого рода являются *ошибки округления* при численных подсчетах. Если $\mathbf{X}$ имеет плотность f, то плотность величины $\mathbf{X}^0$ равна [2])

$$\varphi(x) = \sum_{-\infty}^{+\infty} f(x+n). \tag{8.3}$$

Любая плотность на прямой, таким образом, индуцирует плотность вероятностей на окружности. [В гл. XIX, 5 мы увидим, что та же самая функция φ допускает совершенно иное представление в виде рядов Фурье. Частный случай нормальных плотностей см. в примере (5, д) гл. XIX.]

б) *Задача Пуанкаре о рулетке.* Рассмотрим число поворотов колеса рулетки как случайную величину $\mathbf{X}$ с плотностью f, сосредоточенной на положительной полуоси. Наблюденный результат, а именно точка $\mathbf{X}^0$, против которой колесо останавливается, является величиной $\mathbf{X}$, приведенной по модулю 1. Ее плотность дается формулой (8.3).

---

[1]) Выборочное пространство пары $(\mathbf{X}, \mathbf{Z})$ есть тор.

[2]) Читателям, которых беспокоит вопрос сходимости, следует рассматривать только плотности f, сосредоточенные на конечном интервале. Если f монотонна при достаточно больших x и $-x$, то, очевидно, имеет место равномерная сходимость. Если не накладывать на f никаких ограничений, то ряд может расходиться в некоторых точках, но φ всегда представляет собой плотность, так как частные суммы в (8.3) образуют монотонную последовательность функций, интегралы которых стремятся к 1 (см. гл. IV, 2).

Естественно ожидать, что «при обычных обстоятельствах» плотность $\mathbf{X}^0$ должна быть приблизительно равномерной. В 1912 г. Пуанкаре придал этой расплывчатой мысли форму предельной теоремы. Мы не будем повторять здесь его анализ, так как подобный результат легко следует из (8.3). Подразумевается, конечно, что данная плотность по существу «расплывается» на длинный интервал, так что ее *максимум* m *мал*. Предположим для простоты, что f возрастает вплоть до точки a, где она достигает своего максимума $m=f(a)$, и что f убывает при $x>a$. Хотя $f(s)=0$ при $s<0$, мы пишем

$$\varphi(x)-1=\sum_n f(x+n)-\int_{-\infty}^{+\infty} f(s)\,ds. \tag{8.4}$$

При фиксированном x обозначим x_k единственную точку вида $x+n$, такую, что $a+k\leqslant x_k<a+k+1$. Тогда формула (8.4) может быть переписана в виде

$$\varphi(x)-1=\sum_{k=-\infty}^{\infty}\int_{x_k}^{x_{k+1}}[f(x_k)-f(s)]\,ds. \tag{8.5}$$

При $k<-1$ подинтегральная функция $\leqslant 0$, и поэтому

$$\varphi(x)-1\leqslant f(a)+\sum_{k=0}^{\infty}[f(x_k)-f(x_{k+1})]\leqslant 2f(a)=2m.$$

Аналогичные аргументы доказывают, что $\varphi(x)-1\geqslant -m$. Таким образом, $|\varphi(x)-1|<2m$, и, следовательно, φ действительно приближенно постоянна.

Условия монотонности были наложены на плотность только ради простоты изложения и могут быть ослаблены многими способами. [Хорошие достаточные условия могут быть получены применением формулы суммирования Пуассона, гл. XIX, (5.2).]

в) *Распределение первых значащих цифр.* Известный ученый, прикладной математик, имел громадный успех в заключаемых пари на то, что число, выбранное наугад в Farmer's Almanac или Census Report или в подобном компендиуме, будет иметь первую значащую цифру меньше 5. Наивно предполагается, что все девять цифр равновероятны, в этом случае вероятность того, что цифра не превосходит 4, будет равна $^4/_9$. Практически [1]) она близка к 0,7.

[1]) Эмпирические данные см. в статье Benford F., The law of anomalous numbers, *Proc. Amer. Philos. Soc.*, 78 (1938), 551—572.

Рассмотрим дискретное распределение вероятностей, приписывающее цифре k вероятность $p_k = \mathrm{Log}\,(k+1) - \mathrm{Log}\,k$ (где Log обозначает логарифм по основанию 10 и $k=1, \ldots, 9$). Так как $p_1 \approx 0{,}3$, то наше распределение заметно отличается от равномерного распределения с весами $1/9 = 0{,}111\ldots$. Мы теперь покажем [следуя Р. С. Пинкхэму], что распределение $\{p_k\}$ правдоподобно как эмпирическое распределение первой значащей цифры чисел, выбранных наугад из большой совокупности физических или наблюденных данных. Действительно, такое число может быть рассмотрено как случайная величина $\mathbf{Y}>0$ с некоторым неизвестным распределением. Первая значащая цифра $\mathbf{Y}$ равна 1 тогда и только тогда, когда $10^k \leqslant \mathbf{Y} < 2 \cdot 10^k$, т. е. когда $k \leqslant \mathrm{Log}\,\mathbf{Y} < \mathrm{Log}\,2 + k$. Обозначим через $\mathbf{X}^0$ величину Log $\mathbf{Y}$, приведенную по модулю 1. Если размах $\mathbf{Y}$ очень велик, то правдоподобно, что распределение величины $\mathbf{X}^0$ близко к равномерному распределению. Но 1 является первой значащей цифрой $\mathbf{Y}$ тогда и только тогда, когда $\mathbf{X}^0$ лежит между 0 и Log 2, и вероятность этого события тогда близка к Log 2. Подобное рассуждение применимо и к другим цифрам. ▶

Формула свертки (1.5) и умозаключения, приводящие к ней, остаются справедливыми в случае, когда сложение производится по модулю 1. Таким образом, *свертка двух плотностей на окружности длины* 1 *есть плотность, определяемая формулой*

$$w(x) = \int_0^1 f_1(x-y)\, f_2(y)\, dy. \tag{8.6}$$

Если $\mathbf{X}_1$ и $\mathbf{X}_2$ — независимые величины с плотностями f_1 и f_2, то $\mathbf{X}_1 + \mathbf{X}_2$ имеет плотность w. Так как эти плотности — периодические функции, *свертка равномерной плотности с любой другой плотностью является равномерной.* (См. задачу 15.)

§ 9. Задачи

1. Покажите, что нормальное приближение для биномиального распределения, выведенное в I, гл. VII, позволяет обосновать формулу свертки (1.7) для нормальных плотностей.

2. Используя подстановку $x = {}^1/_2 y^2$, докажите, что

$$\Gamma\left(\frac{1}{2}\right) = \pi^{1/2}.$$

3. *Формула удвоения Лежандра.* Из формулы (2.5) при $\mu = \nu$ вывести, что

$$\Gamma(2\nu) = \frac{1}{\sqrt{\pi}}\, 2^{2\nu-1}\Gamma(\nu)\,\Gamma\left(\nu + \frac{1}{2}\right).$$

Указание. Используйте подстановку $4(y-y^2)=s$ в интервале $0<y<1/2$.

4. Пусть $g(x)=\frac{1}{2}e^{-|x|}$. Найдите свертки $g*g$ и $g*g*g$.

5. Пусть $\mathbf{X}$ и $\mathbf{Y}$ — одинаково распределенные независимые случайные величины с плотностью Коши $\gamma_1(x)$, определенной по (4.5). Докажите, что произведение $\mathbf{XY}$ имеет плотность $2\pi^{-2}\frac{\log|x|}{x^2-1}$.

Указание. Для вычисления не требуется ничего, кроме соотношения

$$\frac{a-1}{(1+s)(a+s)}=\frac{1}{1+s}-\frac{1}{a+s}.$$

6. Пусть

$$f(x)=\frac{2}{\pi}\cdot\frac{1}{e^x+e^{-x}}.$$

Докажите тогда, что

$$f*f(x)=\frac{4}{\pi^2}\frac{x}{e^x-e^{-x}}$$

двумя способами: а) рассматривая величины $\log\mathbf{X}$ и $\log\mathbf{Y}$ в обстановке предыдущей задачи; б) непосредственно при помощи подстановки $e^{2y}=t$ и разложения на элементарные дроби. (См. задачу 7 в гл. XV, 9.)

7. Если $\mathbf{X}$ имеет нормальную плотность n, то очевидно, что $\mathbf{X}^{-2}$ имеет устойчивую плотность (4.8). Используя это, покажите, что, если $\mathbf{X}$ и $\mathbf{Y}$ — независимые нормально распределенные величины с нулевым математическим ожиданием и дисперсиями σ_1^2 и σ_2^2, то величина $\mathbf{Z}=\mathbf{XY}/\sqrt{\mathbf{X}^2+\mathbf{Y}^2}$ *нормально распределена с дисперсией* σ_3^2, такой, что $1/\sigma_3^2=1/\sigma_1^2+1/\sigma_2^2$ (Л. Шепп).

8. Пусть величины $\mathbf{X}_1$, ..., $\mathbf{X}_n$ независимы и $\mathbf{X}_{(n)}$ — наибольшая среди них. Покажите, что, если $\mathbf{X}_j$ имеют

а) плотность Коши (4.5), то

$$\mathbf{P}\{n^{-1}\mathbf{X}_{(n)}\leqslant x\}\to e^{-t/(\pi x)},\qquad x>0;$$

б) устойчивую плотность (4.8), то

$$\mathbf{P}\{n^{-2}\mathbf{X}_{(n)}\leqslant x\}\to e^{-a\sqrt{2/(\pi x)}},\qquad x>0.$$

9. Пусть величины $\mathbf{X}$ и $\mathbf{Y}$ независимы с плотностями f и g, сосредоточенными на $\overline{0,\infty}$. Если $\mathbf{E}(\mathbf{X})<\infty$, то отношение $\mathbf{X}/\mathbf{Y}$ имеет конечное математическое ожидание тогда и только тогда, когда

$$\int_0^1\frac{1}{y}g(y)\,dy<\infty.$$

10. Пусть в примере (5, в) t произвольно. Обозначим через $t+\mathbf{T}$ момент *первой* разгрузки после t. Найдите плотность распределения времени ожидания $\mathbf{T}$ (имейте в виду возможность отсутствия вызовов в момент t).

11. В условиях примера (5, г) покажите, что

$$\mathrm{E}(\mathbf{T}_x) = F(x)(\mu + 1 - x) + \int_0^{1-x} tf(t+x)\,dt,$$

где μ есть математическое ожидание F. Используя это, проверьте утверждение относительно $\mathbf{E}(\mathbf{T})$, когда x распределено равномерно.

12. В примере (5, г) найдите распределение времени ожидания, когда $f(t)=1$ для $0<t<1$.

13. Пусть $\mathbf{X}$ и $\mathbf{Y}$ независимы и подчинены распределению Пуассона с одним и тем же параметром $\mathbf{P}\{\mathbf{X}=n\}=e^{-t}t^n/n!$. Покажите, что

$$\mathbf{P}\{\mathbf{X}-\mathbf{Y}=r\} = e^{-2t}I_{|r|}(2t). \qquad r=0,\ \pm 1,\ \pm 2,\ \ldots.$$

14. Результаты § 7 (в) остаются справедливыми для несимметричных случайных блужданий при условии, что вероятность первого прохождения через $r>0$ равна единице, т. е. при условии $p \geqslant q$. Покажите, что единственное изменение в (7.11) состоит тогда в том, что 2^{-2n+r} заменяется на $p^n q^{n-r}$, а вывод таков: при $p \geqslant q$ и $r=1, 2, \ldots$

$$\sqrt{\left(\frac{p}{q}\right)^r}\, e^{-t}\,\frac{r}{t}\, I_r(2\sqrt{pq}t)$$

определяет плотность вероятностей, сосредоточенную на $\overline{0, \infty}$.

15. Пусть $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины, а $\mathbf{X}^0$ и $\mathbf{Y}^0$ те же величины, приведенные по модулю 1. Покажите, что величина $\mathbf{X}+\mathbf{Y}$ приводится по модулю 1 в $\mathbf{X}^0+\mathbf{Y}^0$. Непосредственным вычислением проверьте соответствующую формулу для сверток.

ГЛАВА III

МНОГОМЕРНЫЕ ПЛОТНОСТИ НОРМАЛЬНЫЕ ПЛОТНОСТИ И ПРОЦЕССЫ

По понятным причинам многомерные распределения встречаются реже, чем одномерные распределения, и материал этой главы почти не будет играть роли в последующих главах. С другой стороны, глава содержит важный материал, например известную характеризацию нормального распределения и методы, применяемые в теории случайных процессов. Истинная природа этих методов становится более понятной, если отделить их от усложненных задач, с которыми они иногда связываются.

§ 1. Плотности

В ходе дальнейшего изложения будет очевидным, что число измерений несущественно И лишь ради типографских удобств мы будем иметь дело с декартовой плоскостью $\mathscr{R}^2$. Мы связываем с плоскостью фиксированную систему координат с координатными величинами $\mathbf{X}_1$, $\mathbf{X}_2$ (более удобное однобуквенное обозначение будет введено в § 5).

Неотрицательная интегрируемая функция f, заданная в $\mathscr{R}^2$ и такая, что ее интеграл равен единице, называется *плотностью вероятности* или просто плотностью. (Все плотности, встречающиеся в этой главе, — кусочно-непрерывны, и поэтому их интегрируемость не требует разъяснений.) Плотность f приписывает области Ω вероятность

$$\mathbf{P}(\Omega) = \int\int_{\Omega} f(x_1, x_2)\, dx_1\, dx_2, \tag{1.1}$$

при условии, конечно, что область Ω достаточно регулярна для того, чтобы интеграл существовал. Все такие вероятности однозначно определяются через вероятности, соответствующие прямоугольникам, параллельным осям, т. е.

$$\mathbf{P}\{a_1 < \mathbf{X}_1 \leqslant b_1,\ a_2 < \mathbf{X}_2 \leqslant b_2\} = \int_{a_1}^{b_1}\int_{a_2}^{b_2} f(x_1, x_2)\, dx_1\, dx_2 \tag{1.2}$$

при всех комбинациях $a_i<b_i$. Полагая $a_1=a_2=-\infty$, мы получим *функцию распределения* F, а именно

$$F(x_1, x_2)=\mathbf{P}\{\mathbf{X}_1\leqslant x_1,\ \mathbf{X}_2\leqslant x_2\}. \tag{1.3}$$

Очевидно, что $F(b_1, x_2)-F(a_1, x_2)$ представляет собой вероятность, соответствующую полубесконечной полосе шириной b_1-a_1, и так как прямоугольник в (1.2) выражается разностью двух таких полос, то вероятность (1.2) равна

$$F(b_1, b_2)-F(a_1, b_2)-F(b_1, a_2)+F(a_1, a_2)$$

(так называемая смешанная разность). Отсюда следует, что по функции распределения F однозначно определяются все вероятности (1.1). Несмотря на формальную аналогию с прямой линией, понятие функции распределения на плоскости гораздо менее полезно, поэтому лучше сосредоточиться на задании вероятностей (1.1) в терминах плотности. Это задание отличается от совместного распределения вероятностей двух дискретных случайных величин (**1**, гл. IX, 1) в двух отношениях. Во-первых, суммирование заменено интегрированием, и, во-вторых, вероятности теперь приписываются лишь «достаточно регулярным областям», тогда как в дискретных выборочных пространствах всем множествам соответствовали вероятности. Так как в настоящей главе рассматриваются простые примеры, в которых это различие едва уловимо, понятия и термины дискретной теории переносятся на новый случай в понятной форме. Так же как и в предыдущих главах, мы используем поэтому вероятностный язык без всякой попытки привлечения общей теории (которая будет изложена в гл. V).

Из (1.3) очевидно, что[1])

$$\mathbf{P}\{\mathbf{X}_1\leqslant x_1\}=F(x_1, \infty). \tag{1.4}$$

Таким образом, $F_1(x)=F(x, \infty)$ определяет функцию распределения величины $\mathbf{X}_1$, а ее *плотность* f_1 равна

$$f_1(x)=\int_{-\infty}^{+\infty} f(x, y)\,dy. \tag{1.5}$$

Если желательно подчеркнуть связь между $\mathbf{X}_1$ и парой $(\mathbf{X}_1, \mathbf{X}_2)$, мы говорим об F_1 как о частном (или *маргинальном*) распределении[2]) и об f_1 как о частной (маргинальной) плотности.

---

[1]) Здесь и в дальнейшем $U(\infty)=\lim U(x)$ при $x\to\infty$, и использование символа $U(\infty)$ подразумевает существование предела.

[2]) Другой принятый термин: *проекция* на ось.

Математическое ожидание μ_1 и *дисперсия* σ_1^2 величины $\mathbf{X}_1$, если они существуют, даются следующими формулами:

$$\mu_1 = \mathbf{E}(\mathbf{X}_1) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1 f(x_1, x_2)\, dx_1\, dx_2 \tag{1.6}$$

и

$$\sigma_1^2 = \mathrm{Var}(\mathbf{X}_1) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_1 - \mu_1)^2 f(x_1, x_2)\, dx_1\, dx_2. \tag{1.7}$$

По симметрии эти определения применимы также и к $\mathbf{X}_2$. Наконец, *ковариация* $\mathbf{X}_1$ и $\mathbf{X}_2$ равна

$$\mathrm{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_1 - \mu_1)(x_2 - \mu_2) f(x_1, x_2)\, dx_1\, dx_2. \tag{1.8}$$

Нормированные величины $\mathbf{X}_i\sigma_i^{-1}$ безразмерны, и их ковариация, а именно $\rho = \mathrm{Cov}(\mathbf{X}_1, \mathbf{X}_2)\sigma_1^{-1}\sigma_2^{-1}$, есть *коэффициент корреляции* $\mathbf{X}_1$ и $\mathbf{X}_2$ (см. 1, гл. IX, 8).

Случайная величина $\mathbf{U}$ — это функция координатных величин $\mathbf{X}_1$ и $\mathbf{X}_2$; здесь мы снова рассматриваем только такие функции, что вероятности $\mathbf{P}\{\mathbf{U} \leqslant t\}$ могут быть выражены интегралами вида (1.1). Таким образом, каждая случайная величина будет иметь единственную функцию распределения, каждая пара будет иметь совместное распределение и т. д.

Во многих ситуациях выгодно *переменить координатные величины*, т. е. приписать величинам $\mathbf{Y}_1$, $\mathbf{Y}_2$ роль, ранее исполнявшуюся $\mathbf{X}_1$, $\mathbf{X}_2$. В простейшем случае величины $\mathbf{Y}_j$ определяются линейным преобразованием

$$\mathbf{X}_1 = a_{11}\mathbf{Y}_1 + a_{12}\mathbf{Y}_2, \quad \mathbf{X}_2 = a_{21}\mathbf{Y}_1 + a_{22}\mathbf{Y}_2 \tag{1.9}$$

с детерминантом $\Delta = a_{11}a_{22} - a_{12}a_{21} > 0$. Обычно преобразование вида (1.9) можно описать или как отображение одной плоскости на другую, или как изменение координат в той же самой плоскости. Производя замену переменных (1.9) в интеграле (1.1), мы получаем

$$\mathbf{P}\{\Omega\} = \int_{\Omega_*}\int f(a_{11}y_1 + a_{12}y_2,\ a_{21}y_1 + a_{22}y_2)\cdot\Delta\, dy_1\, dy_2, \tag{1.10}$$

где область Ω_* содержит все точки (y_1, y_2), образ которых (x_1, x_2) принадлежит Ω. Так как события $(\mathbf{X}_1, \mathbf{X}_2) \in \Omega$ и $(\mathbf{Y}_1, \mathbf{Y}_2) \in \Omega_*$ идентичны, ясно, что *совместная плотность* $(\mathbf{Y}_1\ \mathbf{Y}_2)$ *равна*

$$g(y_1, y_2) = f(a_{11}y_1 + a_{12}y_2,\ a_{21}y_1 + a_{22}y_2)\cdot\Delta. \tag{1.11}$$

Все это в равной степени справедливо и при более высоком числе измерений.

Подобные аргументы применимы к более общим преобразованиям с той лишь разницей, что детерминант Δ заменяется якобианом. Мы будем использовать явно только *переход к полярным координатам*

$$\mathbf{X}_1 = \mathbf{R}\cos\mathbf{\Theta}, \quad \mathbf{X}_2 = \mathbf{R}\sin\mathbf{\Theta}, \tag{1.12}$$

где пара величин $(\mathbf{R}, \mathbf{\Theta})$ ограничена условиями $\mathbf{R} \geqslant 0$, $-\pi < \mathbf{\Theta} \leqslant \pi$. Плотность пары $(\mathbf{R}, \mathbf{\Theta})$ равна

$$g(r, \theta) = f(r\cos\theta, r\sin\theta)\, r. \tag{1.13}$$

В пространстве трех измерений используются географические долгота φ и широта θ ($-\pi < \varphi \leqslant \pi$ и $-{}^1/_2\pi \leqslant \theta \leqslant {}^1/_2\pi$). Тогда координатные величины выражаются в полярной системе следующим образом:

$$\mathbf{X}_1 = \mathbf{R}\cos\mathbf{\Phi}\cos\mathbf{\Theta}, \quad \mathbf{X}_2 = \mathbf{R}\sin\mathbf{\Phi}\cos\mathbf{\Theta}, \quad \mathbf{X}_3 = \mathbf{R}\sin\mathbf{\Theta}. \tag{1.14}$$

Их совместная плотность записывается как

$$g(r, \varphi, \theta) = f(r\cos\varphi\cos\theta,\ r\sin\varphi\cos\theta,\ r\sin\theta)\, r^2\cos\theta. \tag{1.15}$$

[При преобразовании (1.14) «плоскости» $\mathbf{\Theta} = -{}^1/_2\pi$ и $\mathbf{\Theta} = {}^1/_2\pi$ соответствуют полуосям оси x_3, но эта вырожденность не играет роли, так как эти полуоси имеют нулевую вероятность. Аналогичное замечание можно сделать и в отношении начала координат для полярной системы на плоскости.]

Примеры. а) *Независимые случайные величины.* В предыдущих главах мы рассматривали *независимые* случайные величины $\mathbf{X}_1$ и $\mathbf{X}_2$ с плотностями f_1 и f_2. Это равносильно определению двумерной плотности равенством $f(x_1, x_2) = f_1(x_1) f_2(x_2)$, где f_i — частные плотности.

б) *«Случайный выбор».* Пусть Γ — ограниченная область; предположим для простоты, что Γ — выпуклая область. Обозначим площадь Γ через γ. Пусть теперь функция f равна γ^{-1} внутри Γ и 0 вне Γ. Тогда f является плотностью, и вероятность, соответствующая любой области $\Omega \subset \Gamma$, равна отношению площадей Ω и Γ. По очевидной аналогии с одномерным случаем мы будем говорить, что пара $(\mathbf{X}_1, \mathbf{X}_2)$ *распределена равномерно в* Γ. Частная плотность $\mathbf{X}_1$ в точке с абсциссой x_1 равна ширине области Γ в этой точке в очевидном смысле этого слова. (См. задачу 1.)

в) *Равномерное распределение на сфере.* Единичная сфера Σ в пространстве трех измерений может быть описана при помощи

географических долготы φ и широты θ равенствами

$$x_1 = \cos\varphi\cos\theta, \quad x_2 = \sin\varphi\cos\theta, \quad x_3 = \sin\theta. \tag{1.16}$$

Каждой паре (φ, θ), такой, что $-\pi < \varphi \leqslant \pi$, $-\frac{1}{2}\pi < \theta < \frac{1}{2}\pi$, соответствует здесь ровно одна точка на сфере, и каждую точку Σ, исключая полюса, можно получить таким путем. Исключительность полюсов не должна нас беспокоить, так как им соответствует нулевая вероятность. Область Ω на сфере определяется своим образом на (φ, θ)-плоскости, и, следовательно, площадь Ω равна интегралу от $\cos\theta\, d\varphi\, d\theta$ по этому образу [см. (1.15)]. Для описания мысленного эксперимента, заключающегося в «случайном выборе точки на Σ», мы должны положить $4\pi \mathbf{P}(\Omega) =$ площадь Ω. Это эквивалентно определению в (φ, θ)-плоскости плотности

$$g(\varphi, \theta) = \begin{cases} \frac{1}{4\pi}\cos\theta & \text{при} \quad -\pi < \varphi \leqslant \pi, \quad |\theta| < \frac{1}{2}\pi, \\ 0 & \text{при других } \varphi \text{ и } \theta. \end{cases} \tag{1.17}$$

При таком определении координатные величины независимы и долгота распределена равномерно на $-\pi, \pi$.

Идея перехода от сферы Σ к (φ, θ)-плоскости хорошо знакома нам по географическим картам и полезна для теории вероятностей. Заметим, однако, что координатные величины в значительной степени произвольны и их математические ожидания и дисперсии ничего не значат для первоначального мысленного эксперимента.

г) *Двумерная нормальная плотность.* Многомерные нормальные плотности будут введены более систематично в § 6. Оправдание преждевременному появлению двумерного случая в том, что тем самым обеспечивается более легкий подход к нему в дальнейшем. По очевидной аналогии с нормальной плотностью и из II, (1.6) можно записать плотность вида $ce^{-q(x_1, x_2)}$, где $q(x_1, x_2) = a_1x_1^2 + 2bx_1x_2 + a_2x_2^2$. Нетрудно видеть, что функция e^{-q} будет интегрируемой тогда и только тогда, когда $a_1a_2 - b^2 > 0$ и $a_1, a_2 > 0$. В этом случае дисперсии и ковариация легко вычисляются. Для целей теории вероятностей предпочтительнее выразить коэффициенты a_i и b в терминах дисперсий и *определить двумерную нормальную плотность, центрированную в начале координат как*

$$\varphi(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2} - 2\rho\frac{x_1x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right], \tag{1.18}$$

где $\sigma_1>0$, $\sigma_2>0$ и $-1<\rho<1$. Интегрирование по x_2 легко выполняется с помощью подстановки $t=\frac{x_2}{\sigma_2}-\rho\frac{x_1}{\sigma_1}$ (дополнение до полного квадрата). Видно, что φ действительно представляет плотность в $\mathscr{R}^2$. Кроме того, становится очевидным то, что *частные* распределения $\mathbf{X}_1$ и $\mathbf{X}_2$ снова нормальны[1]) и что $\mathbf{E}(\mathbf{X}_i)=0$, $\mathrm{Var}(\mathbf{X}_i)=\sigma_i^2$, $\mathrm{Cov}(\mathbf{X}_1, \mathbf{X}_2)=\rho\sigma_1\sigma_2$. Иными словами, ρ есть коэффициент *корреляции* $\mathbf{X}_1$ и $\mathbf{X}_2$. Заменяя x_i на x_i-c_i в (1.18), мы приходим к нормальной плотности, центрированной в точке (c_1, c_2).

Важно то, что *линейное преобразование* (1.9) *переводит нормальное распределение в другое нормальное распределение*. Это очевидно из определения и формулы (1.11). [Продолжение см. в примере (2, а).]

д) *Симметричное распределение Коши в* $\mathscr{R}^2$. Положим

$$u(x_1, x_2)=\frac{1}{2\pi}\cdot\frac{1}{\sqrt{(1+x_1^2+x_2^2)^3}}. \tag{1.19}$$

Чтобы удостовериться, что это есть плотность, заметим[2]), что

$$\int_{-\infty}^{+\infty} u(x_1, y)\,dy=\frac{1}{2\pi}\cdot\frac{1}{1+x_1^2}\cdot\frac{y}{\sqrt{1+x_1^2+y^2}}\Bigg|_{-\infty}^{+\infty}=\frac{1}{\pi}\cdot\frac{1}{1+x_1^2}. \tag{1.20}$$

Отсюда следует, что u является плотностью и что частная плотность $\mathbf{X}_1$ не что иное, как *плотность Коши* γ_1 из гл. II, (4.5). Очевидно, что величина $\mathbf{X}_1$ не имеет математического ожидания.

При переходе к полярным координатам [как в (1.12)] видим, что плотность $\mathbf{R}$ не зависит от θ, и поэтому величины $\mathbf{R}$ и $\boldsymbol{\Theta}$ стохастически независимы. Следуя терминологии гл. I, 10, мы можем тогда сказать, что имеющая симметричное распределение Коши пара $(\mathbf{X}_1, \mathbf{X}_2)$ *изображает вектор со случайно выбранным направлением с длиной* $\mathbf{R}$, *плотность которой равна* $r\sqrt{(1+r^2)^{-3}}$, откуда $\mathbf{P}\{\mathbf{R}\leqslant r\}=1-\sqrt{(1+r^2)^{-1}}$. [Продолжение в примере (2, б).]

---

[1]) Вопреки широко распространенному мнению, *существуют не являющиеся нормальными двумерные* плотности с нормальными частными плотностями (два типа описаны в задачах 2, 3; еще два в задачах 4 и 6 из гл. V, 11). Для того чтобы иметь дело с нормальными плотностями, статистики иногда вводят новые координатные величины $\mathbf{Y}_1=g_1(\mathbf{X}_1)$, $\mathbf{Y}_2=g_2(\mathbf{X}_2)$, которые распределены нормально. При этом, увы, совместное распределение $(\mathbf{Y}_1, \mathbf{Y}_2)$ не делается нормальным.

[2]) Подстановка $y=\sqrt{1+x_1^2}\,\mathrm{tg}\,t$ облегчает вычисление.

е) *Симметричное распределение Коши в* $\mathscr{R}^3$. Положим

$$v(x_1, x_2, x_3) = \frac{1}{\pi^2} \cdot \frac{1}{(1+x_1^2+x_2^2+x_3^2)^2}. \tag{1.21}$$

Легко видеть[1]), что частная плотность $(\mathbf{X}_1, \mathbf{X}_2)$ является симметричной плотностью Коши u из (1.19). Частная плотность $\mathbf{X}_1$ есть поэтому плотность Коши γ_1. (Продолжение в задаче 5.)

Хотя это и не используется явно в последующем, но мы должны отметить, что *свертки* можно определить так же, как и в одномерном случае. Рассмотрим две пары $(\mathbf{X}_1, \mathbf{X}_2)$ и $(\mathbf{Y}_1, \mathbf{Y}_2)$ с совместными плотностями f и g соответственно. Утверждение о том, что *две пары независимы*, означает, что мы берем четырехмерное пространство с четырьмя координатными величинами $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Y}_1$, $\mathbf{Y}_2$ в качестве выборочного пространства и определяем в нем плотность произведением $f(x_1, x_2)g(y_1, y_2)$. Так же как в $\mathscr{R}^1$, тогда легко видеть, что совместная плотность v суммы $(\mathbf{X}_1+\mathbf{Y}_1, \mathbf{X}_2+\mathbf{Y}_2)$ задается формулой свертки

$$v(z_1, z_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(z_1-x_1, z_2-x_2)\, g(x_1, x_2)\, dx_1\, dx_2, \tag{1.22}$$

которая является очевидным аналогом формулы гл. I, (2.12). (См. задачи 14—16.)

§ 2. Условные распределения

Предположим, что пара $(\mathbf{X}_1, \mathbf{X}_2)$ имеет непрерывную плотность f и что частная плотность f_1 величины $\mathbf{X}_1$ строго положительна. Рассмотрим условную вероятность события $\mathbf{X}_2 \leqslant \eta$ при условии, что $\xi < \mathbf{X}_1 \leqslant \xi + h$, а именно

$$\mathbf{P}\{\mathbf{X}_2 \leqslant \eta \mid \xi < \mathbf{X}_1 \leqslant \xi + h\} = \frac{\int_{\xi}^{\xi+h} dx \int_{-\infty}^{\eta} f(x, y)\, dy}{\int_{\xi}^{\xi+h} f_1(x)\, dx}. \tag{2.1}$$

Деля числитель и знаменатель на h, устанавливаем, что при $h \to 0$ правая часть стремится к

$$U_\xi(\eta) = \frac{1}{f_1(\xi)} \int_{-\infty}^{\eta} f(\xi, y)\, dy. \tag{2.2}$$

---

[1]) Используя подстановку $z = \sqrt{1+x_1^2+x_2^2}\,\operatorname{tg} t$.

При фиксированном ξ это функция распределения по η с плотностью

$$u_\xi(\eta) = \frac{1}{f_1(\xi)} f(\xi, \eta). \tag{2.3}$$

Мы называем u_ξ *условной плотностью величины* $\mathbf{X}_2$ *при условии, что* $\mathbf{X}_1 = \xi$. *Условное математическое ожидание* $\mathbf{X}_2$ *при условии, что* $\mathbf{X}_1 = \xi$, *определяется как*

$$\mathbf{E}(\mathbf{X}_2 | \mathbf{X}_1 = \xi) = \frac{1}{f_1(\xi)} \int_{-\infty}^{+\infty} y f(\xi, y)\, dy \tag{2.4}$$

в предположении, что интеграл сходится абсолютно. Если рассматривать ξ как переменную величину, то правая часть есть функция от ξ. В частности, отождествляя ξ с координатной величиной $\mathbf{X}_1$, получим случайную величину, называемую *регрессией* $\mathbf{X}_2$ *на* $\mathbf{X}_1$ и обозначаемую $\mathbf{E}(\mathbf{X}_2 | \mathbf{X}_1)$. Присутствие $\mathbf{X}_2$ не должно затемнять тот факт, что эта случайная величина является функцией одной переменной $\mathbf{X}_1$ [ее значения даются формулой (2.4)].

До сих пор мы предполагали, что $f_1(\xi) > 0$ для всех ξ. Выражение (2.4) не имеет смысла при $f_1(\xi) = 0$, но множество таких точек имеет вероятность нуль, и мы условимся интерпретировать (2.4) как нуль во всех точках, где f_1 обращается в нуль. Тогда $\mathbf{E}(\mathbf{X}_2 | \mathbf{X}_1)$ определено всякий раз, когда плотность непрерывна. (В гл. V, 9—10 условные вероятности будут введены для произвольных распределений.)

Излишне говорить, что регрессия $\mathbf{E}(\mathbf{X}_1 | \mathbf{X}_2)$ величины $\mathbf{X}_1$ на $\mathbf{X}_2$ определяется подобным же образом. Кроме того, *условная дисперсия* $\mathrm{Var}(\mathbf{X}_2 | \mathbf{X}_1)$ определяется по очевидной аналогии с (2.4).

Эти определения переносятся на случаи большего числа измерений, за исключением того, что плотность в $\mathscr{E}^3$ порождает три двумерные и три одномерные условные плотности [см. пример (в)].

Примеры. а) *Нормальная плотность.* Для плотности (1.18), очевидно, имеем

$$u_\xi(y) = \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_2^2}} \exp\left[-\frac{\left(y - \rho\frac{\sigma_2}{\sigma_1}\xi\right)}{2(1-\rho^2)\sigma_2^2} \right], \tag{2.5}$$

что представляет собой нормальную плотность с математическим ожиданием $\rho\frac{\sigma_2}{\sigma_1}\xi$ и дисперсией $(1-\rho^2)\sigma_2^2$. Таким образом,

$$\mathbf{E}(\mathbf{X}_2 | \mathbf{X}_1) = \rho\frac{\sigma_2}{\sigma_1}\mathbf{X}_1, \quad \mathrm{Var}(\mathbf{X}_2 | \mathbf{X}_1) = (1-\rho^2)\sigma_2^2. \tag{2.6}$$

Одним из приятных свойств нормального распределения является то, что регрессии — *линейные* функции.

Возможно, самым ранним применением этих соотношений мы обязаны Гальтону, и одним из его примеров можно проиллюстрировать их *эмпирическое значение*. Предположим, что $\mathbf{X}_1$ и $\mathbf{X}_2$ характеризуют рост (измеренный по отклонению в дюймах от среднего) отцов и сыновей в некоторой популяции. Рост случайно выбранного сына есть тогда нормальная случайная величина с математическим ожиданием 0 и дисперсией σ_2^2. Однако в подпопуляции сыновей, отцы которых имеют фиксированный рост ξ, рост сыновей представляется нормальной величиной с математическим ожиданием $\rho\frac{\sigma_2}{\sigma_1}\xi$ и дисперсией $\sigma_2^2(1-\rho^2) < \sigma_2^2$. Таким образом, регрессия $\mathbf{X}_2$ на $\mathbf{X}_1$ показывает, как много статистической информации относительно $\mathbf{X}_2$ содержится в наблюдении над $\mathbf{X}_1$.

б) *Распределение Коши* в $\mathscr{R}^2$. Для двумерной плотности (1.19) частная плотность $\mathbf{X}_1$ задается формулой (1.20), и поэтому условная плотность $\mathbf{X}_2$ при данном $\mathbf{X}_1$ равна

$$u_\xi(y) = \frac{1}{2}\cdot\frac{1+\xi^2}{\sqrt{(1+\xi^2+y^2)^3}}. \tag{2.7}$$

Заметим, что u_ξ отличается от плотности $u_1(y)$ только масштабным множителем $\sqrt{1+\xi^2}$, и поэтому все плотности u_ξ принадлежат одному и тому же типу. Условных математических ожиданий в этом примере не существует.

в) *Распределение Коши в* $\mathscr{R}^3$. Нетрудно видеть, что [см. пример (1, е)] условная плотность $\mathbf{X}_3$ при данных $\mathbf{X}_1, \mathbf{X}_2$ равна

$$v_{\xi_1,\xi_2}(z) = \frac{2}{\pi}\cdot\frac{\sqrt{(1+\xi_1^2+\xi_2^2)^3}}{(1+\xi_1^2+\xi_2^2+z^2)^2}, \tag{2.8}$$

а двумерная условная плотность $(\mathbf{X}_2, \mathbf{X}_3)$ при данном $\mathbf{X}_1=\xi$ равна

$$v_\xi(y, z) = \frac{1}{\pi}\cdot\frac{1+\xi^2}{(1+\xi^2+y^2+z^2)^2}. \tag{2.9}$$

В терминах условных плотностей (2.3) функция распределения величины $\mathbf{X}_2$ принимает вид

$$\mathbf{P}\{\mathbf{X}_2 < y\} = \int_{-\infty}^{y}\int_{-\infty}^{+\infty} u_\xi(\eta)\cdot f_1(\xi)\,d\xi\,d\eta. \tag{2.10}$$

Иными словами, распределение $\mathbf{X}_2$ получается посредством *рандомизации* параметра ξ в условных плотностях u_ξ, и

поэтому каждое [1]) *распределение может быть представлено в виде смеси.* Несмотря на эту теоретическую универсальность, имеется большое различие в положении «ударения». В некоторых ситуациях, таких, как в примере (а), *начальным* является двумерное распределение для $(\mathbf{X}_1, \mathbf{X}_2)$, а затем выводятся условные распределения, тогда как при истинной рандомизации условные вероятности u_x являются исходным понятием и плотность $f(x, y)$ фактически *определяется* как $u_x(y)f_{\xi}(x)$. (Эта процедура определения вероятностей в терминах условных вероятностей объяснялась элементарным путем в 1, гл. V, 2.)

§ 3. Возвращение к показательному и равномерному распределениям

Задача этого параграфа в том, чтобы дать примеры-иллюстрации к предыдущим параграфам и в то же время дополнить теорию первой главы.

Примеры. а) *Характеристическое свойство показательного распределения.* Пусть $\mathbf{X}_1$ *и* $\mathbf{X}_2$ — две *независимые* случайные величины с плотностями f_1 и f_2. Обозначим плотность их суммы $\mathbf{S}=\mathbf{X}_1+\mathbf{X}_2$ через g. Пары $(\mathbf{X}_1, \mathbf{S})$ и $(\mathbf{X}_1, \mathbf{X}_2)$ связаны линейным преобразованием $\mathbf{X}_1=\mathbf{X}_1$, $\mathbf{X}_2=\mathbf{S}-\mathbf{X}_1$ с детерминантом 1, и *совместная плотность* пары $(\mathbf{X}_1, \mathbf{S})$ по формуле (1.11) равна $f_1(x)f_2(s-x)$. Интегрируя по всем x, мы получаем частную плотность g суммы $\mathbf{S}$. *Условная плотность* u_s *величины* $\mathbf{X}_1$ *при условии, что* $\mathbf{S}=s$, удовлетворяет соотношению

$$u_s(x)=\frac{f_1(x)f_2(s-x)}{g(s)}. \tag{3.1}$$

Для частного случая показательных плотностей $f_1(x)=f_2(x)=\alpha e^{-\alpha x}$ $(x>0)$ мы получаем $u_s(x)=s^{-1}$ при $0<x<s$. Иными словами, *при условии, что* $\mathbf{X}_1+\mathbf{X}_2=s$, *случайная величина* $\mathbf{X}_1$ *распределена равномерно* на интервале $0, s$. Образно говоря, знание того, что $\mathbf{S}=s$, не дает нам никакой информации о возможном положении случайной точки $\mathbf{X}_1$ внутри интервала $0, s$. Этот результат согласуется с представлением о полной случайности, свойственной показательному распределению. (Более сильный вариант содержится в примере (в). См. также задачу 11.)

б) *Случайные разбиения интервала.* Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — n точек, выбранных независимо и наудачу в (одномерном) ин-

[1]) Мы рассматривали до сих пор непрерывные плотности. Общий случай будет разобран в гл. V, 9. Понятие рандомизации обсуждалось в гл. II, 5.

тервале $\overline{0,1}$. Как и прежде, мы обозначим через $\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)}$, ..., $\mathbf{X}_{(n)}$ случайные точки $\mathbf{X}_1$, ..., $\mathbf{X}_n$, расположенные в возрастающем порядке. Эти точки делят интервал $\overline{0,1}$ на $n+1$ подинтервалов, которые мы обозначим $I_1, I_2, \ldots, I_{n+1}$, пронумеровав их слева направо, так что $\mathbf{X}_{(j)}$ является правым концом интервала I_j. В первую очередь мы вычислим совместную плотность $(\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n)})$.

Выборочное пространство, соответствующее $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, представляет n-мерный гиперкуб Γ: $0<x_k<1$, а вероятности равны (n-мерному) объему. Естественное выборочное пространство, где $\mathbf{X}_{(k)}$ — координатные величины, есть подмножество $\Omega\subset\Gamma$, состоящее из всех точек, таких, что $0<x_1\leqslant\ldots\leqslant x_n<1$. Объем Ω равен $1/n!$. Очевидно, что гиперкуб Γ содержит $n!$ конгруэнтных «копий» множества Ω, и в каждой из них упорядоченная строка $(\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n)})$ совпадает с фиксированной перестановкой величин $\mathbf{X}_1, \ldots, \mathbf{X}_n$. (Внутри Γ, в частности, $\mathbf{X}_{(k)}=\mathbf{X}_k$.) Вероятность того, что $\mathbf{X}_j=\mathbf{X}_k$ для некоторой пары $j\neq k$ равна нулю, и только это событие вызывает перекрытия между различными «копиями». Отсюда следует, что для любого подмножества $A\subset\Omega$ вероятность того, что $(\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n)})$ принадлежит A, равна вероятности того, что $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ принадлежит одной из $n!$ «копий» A, а эта вероятность в свою очередь равна умноженному на $n!$ объему A. Таким образом, вероятность $\mathbf{P}\{(\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n)})\in A\}$ равна отношению объемов A и Ω, а это означает, что *строка* $(\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n)})$ *распределена равномерно на множестве* Ω таких точек, что

$$0<x_1\leqslant x_2\leqslant\ldots<1.$$

Совместная плотность величин этой строки равна $1/n!$ внутри Ω и нулю вне Ω.

Из совместной плотности $(\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(n)})$, предполагая x_k фиксированным и интегрируя по всем оставшимся переменным, можно вычислить плотность $\mathbf{X}_{(k)}$. Легко видеть, что этот результат согласуется с формулой для плотности, вычисленной другими методами в гл. I, (7.2).

Этот пример был рассмотрен детально в качестве упражнения в обращении с многомерными плотностями.

в) *Распределение длин.* В обстановке предыдущего примера обозначим длину k-го интервала I_k через $\mathbf{U}_k$. Тогда

$$\mathbf{U}_1=\mathbf{X}_{(1)},\quad \mathbf{U}_k=\mathbf{X}_{(k)}-\mathbf{X}_{(k-1)}\qquad k=2,\ 3,\ \ldots,\ n. \tag{3.2}$$

Это не что иное, как линейное преобразование типа (1.9) с детерминантом 1. Множество Ω точек $0<x_1\leqslant\ldots\leqslant x_n<1$ отображается в множестве Ω^* таких точек, что $u_j\geqslant 0$, $u_1+\ldots+u_n<1$,

и, следовательно, строка $(\mathbf{U}_1, \ldots, \mathbf{U}_n)$ *распределена равномерно в этой области.* Этот результат сильнее ранее установленного факта, что величины $\mathbf{U}_k$ одинаково распределены [пример гл. I, (7, б)].

г) *Еще раз о случайности показательного распределения.* Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_{n+1}$ независимы и одинаково распределены с плотностью $\alpha e^{-\alpha x}$ при $x>0$. Положим $\mathbf{S}_j=\mathbf{X}_1+\ldots+\mathbf{X}_j$. Тогда строка $(\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{n+1})$ получается из $(\mathbf{X}_1, \ldots, \mathbf{X}_{n+1})$ линейным преобразованием типа (1.9) с детерминантом 1. Обозначим через Ω «октант», состоящий из точек $x_j>0$ $(j=1, \ldots, n+1)$. Плотность $(\mathbf{X}_1, \ldots, \mathbf{X}_{n+1})$ сосредоточена на Ω и равна

$$\alpha^{n+1}e^{-\alpha(x_1+\ldots+x_{n+1})},$$

если $x_j>0$. Величины $\mathbf{S}_1, \ldots, \mathbf{S}_{n+1}$ отображают Ω на область Ω^*, определяемую неравенствами $0<s_1\leqslant s_2\leqslant\ldots\leqslant s_{n+1}<\infty$, и [см. (1.11)] внутри Ω^* плотность $(\mathbf{S}_1, \ldots, \mathbf{S}_{n+1})$ равна $\alpha^{n+1}e^{-\alpha s_{n+1}}$. Частная плотность $\mathbf{S}_{n+1}$ есть не что иное, как гамма-плотность $\alpha^{n+1}s^n e^{-\alpha s}/n!$, и, следовательно, *условная плотность строки* $(\mathbf{S}_1, \ldots, \mathbf{S}_n)$ *при условии, что* $\mathbf{S}_{n+1}=s$, равна $n!\,s^{-n}$ *при* $0<s_1<\ldots<s_n<s$ (и нулю в других случаях). Иными словами, при условии, что $\mathbf{S}_{n+1}=s$ величины $(\mathbf{S}_1, \ldots, \mathbf{S}_n)$ равномерно распределены в области их возможных значений. Сравнивая это с примером (б), мы можем сказать, что *при условии* $\mathbf{S}_{n+1}=s$ *величины* $(\mathbf{S}_1, \ldots, \mathbf{S}_n)$ *изображают n точек, выбранных независимо и наудачу в интервале* $\overline{0, s}$, занумерованных в их естественном порядке слева направо.

д) *Другое распределение, связанное с показательным.* Имея в виду последующее неожиданное применение, мы дадим еще один пример преобразования. Пусть снова $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — независимые величины, имеющие одинаковое показательное распределение и $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$. Определим величины $\mathbf{U}_1, \ldots, \mathbf{U}_n$ следующим образом:

$$\mathbf{U}_k=\frac{\mathbf{X}_k}{\mathbf{S}_n} \quad \text{при} \quad k=1, \ldots, n-1, \quad \mathbf{U}_n=\mathbf{S}_n, \tag{3.3}$$

или, что равносильно,

$$\begin{aligned} &\mathbf{X}_k=\mathbf{U}_k\mathbf{U}_n \quad \text{при} \quad k=1, \ldots, n-1, \\ &\mathbf{X}_n=\mathbf{U}_n(1-\mathbf{U}_1-\ldots-\mathbf{U}_{n-1}). \end{aligned} \tag{3.4}$$

Якобиан преобразования (3.4) равен $\mathbf{U}_n^{n-1}$. Совместная плотность $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ сосредоточена в области Ω, определяемой неравенствами $x_k>0$, и внутри нее эта плотность равна $\alpha^n e^{-\alpha(x_1+\ldots+x_n)}$. Отсюда следует, что совместная плотность

$(\mathbf{U}_1, \ldots, \mathbf{U}_n)$ равна $\alpha^n u_n^{n-1} e^{-\alpha u_n}$ в области Ω^*, определенной неравенствами

$$u_1 + \ldots + u_{n-1} < 1, \quad u_k > 0, \quad k = 1, \ldots, n,$$

и равна нулю вне Ω^*. Интегрирование относительно u_n показывает, что совместная плотность $(\mathbf{U}_1, \ldots, \mathbf{U}_{n-1})$ равна $(n-1)!$ в Ω^* и нулю в других случаях. Сравнивая с примером (в), мы видим, что $(\mathbf{U}_1, \ldots, \mathbf{U}_{n-1})$ *имеет такое же распределение, как если бы величина* $\mathbf{U}_k$ *была длиной* k*-го интервала при случайном разбиении интервала* $\overline{0, 1}$ *набором* $n-1$ *точек.*

е) *Критерий значимости в анализе периодограмм и теорема о покрытии.* Практически любая непрерывная функция времени t может быть приближена тригонометрическим полиномом. Если эта функция представляет собой выборочную функцию стохастического процесса, то коэффициенты становятся случайными величинами и аппроксимирующий полином может быть записан в виде

$$\sum_{\nu=1}^{n} (\mathbf{X}_\nu \cos \omega_\nu t + \mathbf{Y}_\nu \sin \omega_\nu t) \equiv \sum_{\nu=1}^{n} \mathbf{R}_\nu \cos (\omega_\nu t - \mathbf{\Phi}_\nu), \qquad (3.5)$$

где $\mathbf{R}_\nu^2 = \mathbf{X}_\nu^2 + \mathbf{Y}_\nu^2$ и $\operatorname{tg} \mathbf{\Phi}_\nu = \mathbf{Y}_\nu / \mathbf{X}_\nu$. Обратно, разумные предположения относительно случайных величин $\mathbf{X}_\nu$, $\mathbf{Y}_\nu$ приводят к стохастическому процессу с выборочной функцией, задаваемой (3.5). В свое время было модным вводить модели такого рода и открывать «скрытую периодичность» для солнечных пятен, цен на пшеницу, поэтического творчества и т. д. Подобные скрытые периодичности отыскивали с такой же легкостью, как ведьм в средневековое время. Но даже сильная вера должна быть подкреплена статистическим критерием. Грубо говоря, метод заключается в следующем. Тригонометрический полином вида (3.5) с как-то выбранными частотами $\omega_1, \ldots, \omega_n$ сравнивается с некоторыми результатами наблюдений. Допустим, что при этом обнаруживается особенно большая амплитуда $\mathbf{R}_\nu$. Желательно доказать, что это не может быть случайным и, следовательно, что ω_ν — истинный период. Для проверки этого предположения выясним, сколь правдоподобно совместимо большое наблюденное значение $\mathbf{R}_\nu$ с гипотезой, что все n компонент играют одинаковую роль. В соответствии с этим предполагается, что коэффициенты $\mathbf{X}_1, \ldots, \mathbf{Y}_n$ взаимно независимы и имеют одинаковое нормальное распределение с нулевым математическим ожиданием и дисперсией σ^2. В этом случае (см. гл. II, 3) величины $\mathbf{R}_\nu^2$ взаимно независимы и имеют одно и то же показательное распределение с математическим ожиданием $2\sigma^2$. Если наблюденное

значение $\mathbf{R}_\nu^2$ «значимо» отклонилось от ожидаемого, то мы приходим к заключению, что гипотеза равенства весов была несостоятельна и $\mathbf{R}_\nu$ изображает «скрытую периодичность».

Ошибочность этих соображений была раскрыта Р. А. Фишером (1929 г.). Он отметил, что максимальный из результатов n независимых наблюдений не подчиняется тому же самому распределению вероятностей, которое имеет каждый из них в отдельности. Ошибка в статистической трактовке худшего случая так, как будто он выбирался случайно, до сих пор распространена в медицинской статистике, но причиной обсуждения этого вопроса здесь является неожиданная и занятная связь критерия значимости Фишера с теоремами о покрытии.

Так как важны только отношения компонент, то мы нормируем коэффициенты, полагая

$$\mathbf{V}_j = \frac{\mathbf{R}_j^2}{\mathbf{R}_1^2 + \ldots + \mathbf{R}_n^2}, \qquad j = 1, \ldots, n. \tag{3.6}$$

Поскольку $\mathbf{R}_j^2$ имеют одно и то же показательное распределение, мы можем использовать предыдущий пример с $\mathbf{X}_j = \mathbf{R}_j^2$. Тогда $\mathbf{V}_1 = \mathbf{U}_1, \ldots, \mathbf{V}_{n-1} = \mathbf{U}_{n-1}$, но $\mathbf{V}_n = 1 - \mathbf{U}_1 - \ldots - \mathbf{U}_{n-1}$. Соответственно строка $(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ *распределена так, как если бы* $\mathbf{V}_j$ *были длинами* n *интервалов, на которые разбивается интервал* $\overline{0, 1}$ *при случайном расположении в нем* $n - 1$ *точек. Вероятность того, что все* $\mathbf{V}_j$ *будут меньше* a, *задается поэтому формулой гл. I,* (9.9) *теоремы о покрытии.* Этот результат иллюстрирует факт неожиданных отношений между на вид не связанными задачами [1]). ▶

§ 4*. Характеризация нормального распределения

Рассмотрим невырожденное линейное преобразование координатных величин

$$\mathbf{Y}_1 = a_{11}\mathbf{X}_1 + a_{12}\mathbf{X}_2, \quad \mathbf{Y}_2 = a_{21}\mathbf{X}_1 + a_{22}\mathbf{X}_2 \tag{4.1}$$

и допустим (не ограничивая общности), что детерминант $\Delta = 1$. Если $\mathbf{X}_1$ и $\mathbf{X}_2$ независимые нормально распределенные величины

[1]) Фишер получил распределение максимальной из величин $\mathbf{V}_j$ в 1929 г. без знания теоремы о покрытии, а в 1940 г. объяснил связь с теоремой о покрытии, после того как Стивенс доказал последнюю. [См. Fisher, Contributions to Mathematical Statistics, John Wiley, N. Y. (1950), статьи 16 и 37.] Другой вывод, использующий анализ Фурье, см. Grenander U., Rosenblatt M. (1957).

*) Этот параграф затрагивает специальную тему и не используется в дальнейшем.

с дисперсиями σ_1^2 и σ_2^2, то распределение пары $(\mathbf{Y}_1, \mathbf{Y}_2)$ нормально с ковариацией $a_{11}a_{21}\sigma_1^2 + a_{12}a_{22}\sigma_2^2$ [см. пример (1, г)]. В этом случае существуют нетривиальные наборы коэффициентов a_{jk}, такие, что $\mathbf{Y}_1$ и $\mathbf{Y}_2$ независимы. Следующая теорема показывает, что это свойство одномерного нормального распределения *не разделяется никаким другим распределением*. Мы докажем это только для распределений с непрерывными плотностями, в случае которых наше утверждение сводится к лемме относительно функционального уравнения (4.3). Более общий случай сводится (с использованием характеристических функций) к *тому же самому* уравнению. Поэтому наше доказательство фактически устанавливает теорему в ее наибольшей общности (см. гл. XV, 8). Элементарное рассмотрение плотностей лучше раскрывает основную суть теоремы.

Теорема. *Допустим, что* $\mathbf{X}_1$ *и* $\mathbf{X}_2$ *независимы и что величины* $\mathbf{Y}_1$ *и* $\mathbf{Y}_2$ *из* (4.1) *также не зависят одна от другой. Тогда все четыре величины являются нормальными, исключая случай, когда преобразование тривиально, т. е. или* $(\mathbf{Y}_1, \mathbf{Y}_2) = (a\mathbf{X}_1, b\mathbf{X}_2)$, *или* $(\mathbf{Y}_1, \mathbf{Y}_2) = (a\mathbf{X}_2, b\mathbf{X}_1)$.

Наиболее интересный чистый случай (4.1) доставляется вращениями, т. е. преобразованиями вида

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{X}_1 \cos\omega + \mathbf{X}_2 \sin\omega, \\ \mathbf{Y}_2 &= -\mathbf{X}_1 \sin\omega + \mathbf{X}_2 \cos\omega, \end{aligned} \tag{4.2}$$

где ω не кратно $^1/_2\pi$. Применяя теорему к подобным вращениям, мы получаем

Следствие. *Если* $\mathbf{X}_1$ *и* $\mathbf{X}_2$ *— независимые случайные величины и существует вращение* (4.2), *такое, что* $\mathbf{Y}_1$ *и* $\mathbf{Y}_2$ *также независимы, то* $\mathbf{X}_1$ *и* $\mathbf{X}_2$ *имеют нормальные распределения с одинаковой дисперсией. В этом случае* $\mathbf{Y}_1$ *и* $\mathbf{Y}_2$ *независимы при каждом* ω.

Пример. *Максвелловское распределение скоростей.* При изучении распределений скоростей молекул в $\mathscr{R}^3$ Максвелл предполагал, что в *любой* декартовой системе координат три компоненты скорости являются независимыми случайными величинами с нулевым математическим ожиданием. Примененное к вращениям, оставляющим одну ось фиксированной, наше следствие непосредственно показывает, что три компоненты нормально распределены с одинаковой дисперсией. Как мы видели в гл. II, 3, это влечет распределение Максвелла для скоростей. ▶

Эта теорема имеет длинную историю, восходящую к исследованиям Максвелла. Чисто вероятностное изучение было предпринято М. Кацем (1940 г.) и С. Бернштейном (1941 г.), который доказал наше следствие в предположении.

конечности дисперсий. Многие авторы вносили улучшения и исследовали варианты, иногда при помощи более сильных методов. Кульминации это развитие достигло в результате, доказанном В. П. Скитовичем [1]).

Перейдем к доказательству для случая непрерывных плотностей. Обозначим плотности $\mathbf{X}_j$ через w_j, а плотности $\mathbf{Y}_j$ через f_j. В предположениях теоремы имеем [см. (1.11)]

$$f_1(a_{11}x_1 + a_{12}x_2)\, f_2(a_{21}x_1 + a_{22}x_2) = w_1(x_1)\, w_2(x_2). \tag{4.3}$$

Отсюда мы выведем, что

$$f_1(x) = \pm e^{\varphi_1(x)}, \tag{4.4}$$

где φ_1 — полином не выше второй степени. Отсюда следует наша теорема, так как лишь нормальные плотности имеют такой вид.

Для распределений с непрерывными плотностями теорема содержится поэтому в следующей лемме.

Лемма. *Пусть функции f_1 и f_2 непрерывны и отличны от постоянных. Допустим, что выполняется функциональное уравнение вида* (4.3) *и*

$$\Delta = a_{11}a_{22} - a_{12}a_{21} \neq 0.$$

Если никакой из коэффициентов a_{ij} не равен нулю, то (4.4) *выполняется, причем φ_1 является квадратным полиномом.*

Если $a_{11}a_{12} \neq 0$, но $a_{21}a_{22} = 0$, то (4.4) *выполняется, причем φ_1 является линейной функцией.*

(Излишне говорить, что те же самые выводы применимы к f_2.)

Правдоподобное рассуждение. Предположим, что функции в (4.3) являются строго положительными и дважды дифференцируемыми. Записывая сокращенно

$$\varphi_k = \log f_k, \quad y_k = a_{k1}x_1 + a_{k2}x_2, \tag{4.5}$$

получаем в таком случае

$$\varphi_1(y_1) + \varphi_2(y_2) = \log w_1(x_1) + \log w_2(x_2) \tag{4.6}$$

и отсюда посредством последовательного дифференцирования по x_1 и x_2

$$a_{11}a_{12}\varphi_1''(y_1) + a_{21}a_{22}\varphi_2''(y_2) = 0. \tag{4.7}$$

Меняя x_1 и x_2 так, чтобы сохранить y_2 постоянным, мы видим, что $\varphi_1''(y_1) = \text{const}$, и поэтому на самом деле φ_1 представляет собой полином первой или второй степени. Следующее ниже доказательство основано на этих аргументах, за исключением того, что производные заменены разностями.

---

[1]) Изв. АН СССР, т. 18 (1954), стр. 185—200. Теорема. Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ взаимно независимы, $\mathbf{Y}_1 = \Sigma a_i \mathbf{X}_i$ и $\mathbf{Y}_2 = \Sigma b_i \mathbf{X}_i$, где ни один коэффициент не равен нулю. Если $\mathbf{Y}_1$ и $\mathbf{Y}_2$ независимы, то величины $\mathbf{X}_i$ распределены нормально.

Доказательство леммы. Предположим вначале, что *ни один из коэффициентов a_{jk} не равен нулю.* Мы начинаем с доказательства того, что функции f_k *не имеют нулей.* Предположим противное. Тогда существует открытая область Ω, внутри которой обе части равенства (4.3) не имеют нулей и на границе которой они обращаются в нуль. Но правая часть (4.3) равна нулю в точке (x_1, x_2) тогда и только тогда, когда либо $w_1(x_1)=0$, либо $w_2(x_2)=0$. Поэтому граница состоит из отрезков прямых, параллельных осям. Аналогичные рассуждения, примененные к левой части, показывают, что граница состоит из отрезков прямых, на которых либо $y_1=\text{const}$, либо $y_2=\text{const}$. Это противоречие показывает, что граница не существует.

Мы теперь можем предположить, что f_k строго положительны, и исходить из (4.6). Определим разностный оператор Δ (зависящий от двух произвольных констант) как

$$\Delta u(x_1x_2) = u(x_1+h_1,\ x_2+h_2) - u(x_1+h_1,\ x_2-h_2) - \\ - u(x_1-h_1,\ x_2+h_2) + u(x_1-h_1,\ x_2-h_2). \quad (4.8)$$

Если u зависит только от одной переменной x_j, мы имеем $\Delta u=0$, и из (4.6) поэтому получаем $\Delta\varphi_1+\Delta\varphi_2=0$ [вместо (4.7)]. Теперь легко проверить, что $\Delta\varphi_k$ есть функция исключительно y_k, и, следовательно, соотношение $\Delta\varphi_1+\Delta\varphi_2=0$ влечет, что $\Delta\varphi_1$ есть постоянная, зависящая, конечно, от h_1 и h_2. При подходящем выборе h_1 и h_2 мы заключаем, что при произвольном t выражение

$$\varphi_1(y_1+t)+\varphi_1(y_1-t)-2\varphi_1(y_1)=\lambda_1(t) \quad (4.9)$$

не зависит от y_1.

Это равенство заменяет результат $\varphi_1''(y_1)=\text{const}$, полученный правдоподобными рассуждениями. Мы покажем, что *каждое непрерывное решение уравнения* (4.9) *приводит к квадратному полиному.*

Функция

$$\Phi(x)=\varphi_1(x)+\alpha+\beta x+\gamma x^2 \quad (4.10)$$

удовлетворяет тождеству

$$\Phi(x+t)+\Phi(x-t)-2\Phi(x)=\lambda(t), \quad (4.11)$$

где $\lambda(t)=\lambda_1(t)+2\gamma t^2$. Мы выбираем α, β, γ так, что

$$\Phi(0)=\Phi(-1)=\Phi(1)=0,$$

и докажем, что каждая непрерывная функция Φ, удовлетворяющая этим условиям, тождественно равна нулю. Предполагая, что Φ имеет положительный максимум, мы выводим из (4.11),

что $\lambda(t) \leqslant 0$ в некоторой окрестности начала координат и $\lambda(t) < 0$ при некотором t. Отсюда следует, что Φ не может иметь минимальное значение, и поскольку Φ равна нулю в трех точках $-1, 0, 1$, то она должна быть равна нулю тождественно в интервале $\overline{-1, 1}$. Из (4.11) тогда ясно, что $\Phi(x) = 0$ при всех x и, следовательно, $\varphi_1(x) = -\alpha - \beta x - \gamma x^2$. Этим завершается доказательство в случае, когда ни один из коэффициентов a_{jk} не равен нулю.

Случай, когда какой-либо коэффициент равен нулю, тривиален. Предположим, например, что $a_{21} = 0$, но $a_{11} \neq 0$ и $a_{12} \neq 0$. Тогда $f_1(a_{11}x_1 + a_{12}x_2)$ является произведением функции от x_1 и функции от x_2, и потому $f_1(s+t) = f_1(s)f_1(t)$. Это функциональное уравнение появлялось неоднократно, и мы знаем, что оно влечет $f_1(s) = e^{\lambda s}$. ▶

§ 5. Матричные обозначения. Матрица ковариаций

Обозначения, использованные в § 1, громоздки и становятся еще более громоздкими в случае большего числа измерений. Изящества и экономии мысли можно достичь при использовании матричной системы обозначений.

Чтобы облегчить ссылки, мы кратко изложим немногие факты из теории матриц и матричных обозначений, нужные в дальнейшем. Основное правило таково: сначала строки, затем столбцы. Таким образом, $(\alpha \times \beta)$-матрица A имеет α строк и β столбцов; ее элементы обозначаются a_{jk}, при этом первый индекс указывает строку. Если B есть $(\beta \times \gamma)$-матрица с элементами b_{jk}, то произведение AB представляет собой $(\alpha \times \gamma)$-матрицу с элементами $a_{j1}b_{1k} + a_{j2}b_{2k} + \ldots + a_{j\beta}b_{\beta k}$. Если число столбцов матрицы A не совпадает с числом строк матрицы B, то произведение не определено. Выполняется ассоциативный закон, $(AB)C = A(BC)$, хотя, вообще говоря, $AB \neq BA$. *Транспонированная* матрица A^T есть $(\beta \times \alpha)$-матрица с элементами $a^T_{jk} = a_{kj}$. Очевидно, что $(AB)^T = B^T A^T$.

$(1 \times \alpha)$-матрица с единственной строкой называется *вектор-строкой*; а матрица с единственным столбцом — *вектор-столбцом* [1]). Вектор-строка $r = (r_1, \ldots, r_\alpha)$ легко изображается при печати, и вектор-столбец лучше определять результатом его транспонирования $c^T = (c_1, \ldots, c_\alpha)$. Заметим, что cr *есть* $(\alpha \times \alpha)$-*матрица* (типа «таблицы умножения»), *в то время как* rc *есть* (1×1)-*матрица* или скаляр. В случае $\alpha = 2$

$$cr = \begin{pmatrix} c_1 r_1 & c_1 r_2 \\ c_2 r_1 & c_2 r_2 \end{pmatrix}, \qquad rc = (r_1 c_1 + r_2 c_2).$$

Нулевой вектор имеет нулевые компоненты.

Под *единичной матрицей* подразумевается квадратная матрица с единицами по главной диагонали и с нулями на всех остальных местах. Если I —

[1]) На самом деле это — плохое обращение с языком. В конкретном случае x_1 может означать количество фунтов, а x_2 — число коров; тогда (x_1, x_2) не является «вектором» в строгом смысле.

единичная матрица с r строками и r столбцами и A — любая $(r\times r)$-матрица, то очевидно, что $IA=AI=A$. Матрицей, *обратной* для A, называется матрица A^{-1}, такая, что $AA^{-1}=A^{-1}A=I$. [Только квадратные матрицы могут иметь обратные. Обратная матрица единственна: так, если B — любая обратная матрица для A, то $AB=I$ и по ассоциативному закону $A^{-1}=(A^{-1}A)B=B$.] Квадратная матрица, не имеющая обратной, называется *вырожденной*. Имеет место следующий элементарный критерий. Матрица A является вырожденной тогда и только тогда, когда существует ненулевая вектор-строка x, такая, что $xA=0$. Если A — вырожденная матрица, то такова же и транспонированная матрица A^T.

Квадратная матрица A *симметрична*, если $a_{jk}=a_{kj}$, т. е. если $A^T=A$. *Квадратичная форма, соответствующая симметричной* $(r\times r)$*-матрице* A, определяется как

$$xAx^T=\sum_{j,\,k=1}^{r}a_{jk}x_jx_k,$$

где $x_1, \ldots, x_r$ являются переменными. Матрица называется *положительно определенной*, если $xAx^T>0$ для всех ненулевых векторов x. Из последнего критерия следует, что положительно определенная матрица не вырождена. (Об ортогональных матрицах или вращениях см. в § 6а.)

Отныне мы будем обозначать точки n-мерного пространства $\mathscr{R}^n$ одной буквой, интерпретируя ее как *вектор-строку*. Таким образом, $x=(x_1, \ldots, x_r)$ и $f(x)=f(x_1, \ldots, x_r)$ и т. д. Неравенства должны интерпретироваться покоординатно: $x<y$ тогда и только тогда, когда $x_k<y_k$ для $k=1, \ldots, r$; аналогично и для других неравенств. На плоскости $\mathscr{R}^2$ соотношение $x<y$ может быть прочитано как «x лежит юго-западнее y». Новая особенность этого обозначения в том, что две точки не обязаны состоять в одном из двух соотношений: либо $x\leqslant y$, либо $y<x$, т. е. в многомерном случае неравенство $x<y$ определяет только частичное упорядочение.

Мы вводим обозначение $\mathbf{X}=(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ для вектора-строки из координатных величин и будем использовать это обозначение для случайных величин вообще (в основном для нормально распределенных величин).

Если величины $\mathbf{X}_1, \ldots, \mathbf{X}_r$ имеют математические ожидания $\mathbf{E}(\mathbf{X}_j)$, то мы обозначаем через $\mathbf{E}(\mathbf{X})$ *вектор-строку с компонентами* $\mathbf{E}(\mathbf{X}_j)$. Вектор $\mathbf{X}-\mathbf{E}(\mathbf{X})$ имеет нулевое математическое ожидание. Вообще, если $\mathbf{M}$ — матрица, элементы которой $\mathbf{M}_{jk}$ являются случайными величинами, мы обозначаем через $\mathbf{E}(\mathbf{M})$ матрицу элементов $\mathbf{E}(\mathbf{M}_{jk})$ в предположении, что она существует.

Определение. *Если* $\mathbf{E}(\mathbf{X})=0$, *то ковариационная матрица* $\mathrm{Var}\,(\mathbf{X})$ *для* $\mathbf{X}$ *есть симметричная* $(r\times r)$*-матрица с элементами* $\mathbf{E}(\mathbf{X}_j\mathbf{X}_k)$ (*при условии, что все они существуют*). *Иными словами,*

$$\mathrm{Var}\,(\mathbf{X})=\mathbf{E}(\mathbf{X}^T\mathbf{X}). \tag{5.1}$$

В случае произвольного $\mathbf{X}$ *мы определим* $\operatorname{Var}\mathbf{X}$, *как и* $\operatorname{Var}(\mathbf{X}-\mathbf{E}(\mathbf{X}))$.

Использование векторов-строк приводит к необходимости записывать *линейное преобразование* $\mathscr{R}^r$ в $\mathscr{R}^m$ в виде

$$\mathbf{Y}=\mathbf{X}A, \tag{5.2}$$

т. е.

$$y_k=\sum_{j=1}^{r} a_{jk}x_j, \qquad k=1, \ldots, m, \tag{5.3}$$

где A — $(r\times m)$-матрица. Очевидно, что $\mathbf{E}(\mathbf{Y})=\mathbf{E}(\mathbf{X})A$ всякий раз, когда $\mathbf{E}(\mathbf{X})$ существует. Для того чтобы найти дисперсии, мы положим, не ограничивая общности, $\mathbf{E}(\mathbf{X})=0$. Тогда $\mathbf{E}(\mathbf{Y})=0$ и

$$\mathbf{E}(\mathbf{Y}^T\mathbf{Y})=\mathbf{E}(A^T\mathbf{X}^T\mathbf{X}A)=A^T\mathbf{E}(\mathbf{X}^T\mathbf{X})A. \tag{5.4}$$

Мы получаем, таким образом, важный результат

$$\operatorname{Var}(\mathbf{Y})=A^T\operatorname{Var}(\mathbf{X})A. \tag{5.5}$$

Особый интерес представляет частный случай $m=1$, когда

$$\mathbf{Y}=a_1\mathbf{X}_1+\ldots+a_r\mathbf{X}_r \tag{5.6}$$

есть обыкновенная случайная величина. Здесь $\operatorname{Var}(\mathbf{Y})$ — (скалярная) квадратичная форма

$$\operatorname{Var}(\mathbf{Y})=\sum_{j,\,k=1}^{r}\mathbf{E}(\mathbf{X}_j\mathbf{X}_k)a_ja_k. \tag{5.7}$$

Линейная форма (5.6) равна нулю с вероятностью единица, если $\operatorname{Var}(\mathbf{Y})=0$, и в этом случае любая область за пределами гиперплоскости $\sum a_kx_k=0$ имеет вероятность нуль. Распределение вероятностей сосредоточено тогда на $(r-1)$-мерном многообразии и *вырождено*, если его рассматривать в r измерениях. Мы доказали, что *ковариационная матрица любого невырожденного распределения вероятностей положительно определена.* Обратно, каждая такая матрица может служить ковариационной матрицей нормальной плотности (см. теорему 3 следующего параграфа).

§ 6. Нормальные плотности и распределения

Всюду в этом параграфе Q обозначает симметричную $(r\times r)$-матрицу, а $q(x)$ — соответствующую ей квадратичную форму

$$q(x)=\sum_{j,\,k=1}^{r}q_{jk}x_jx_k=xQx^T, \tag{6.1}$$

где $x=(x_1, \ldots, x_r)$ есть вектор-*строка*. Плотности в $\mathscr{R}^r$, определенные как показательные с квадратичной формой в экспоненте, являются естественным обобщением нормальной плотности на прямой, и мы начнем поэтому с того, что введем следующее

Определение. *Плотность* φ *в пространстве* r *измерений называется нормальной*[1]) (*с центром в начале координат*), *если она представляется в виде*

$$\varphi(x)=\gamma^{-1}\cdot e^{-\frac{1}{2}q(x)}, \tag{6.2}$$

где γ *— некоторая постоянная. Нормальная плотность с центром в точке* $a=(a_1, a_2, \ldots, a_r)$ *определяется как* $\varphi(x-a)$.

Специальный случай двух измерений обсуждался в примерах (1, г) и (2, а).

Рассмотрим результат невырожденного линейного преобразования $\mathbf{X}=(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ в $\mathbf{Y}=(\mathbf{Y}_1, \ldots, \mathbf{Y}_r)$. Для этого запишем преобразование в форме $\mathbf{Y}=\mathbf{X}A$, где A — $(r\times r)$-матрица с обратной матрицей $B=A^{-1}$. Как было показано в § 1, плотность $\mathbf{Y}$ получается из (6.2) при помощи подстановки $x=yB$ и умножения на детерминант матрицы B. Теперь

$$q(x)=xQx^T=yBQB^Ty^T. \tag{6.3}$$

Отсюда следует, что плотность $\mathbf{Y}$ снова имеет вид (6.2) и нами доказана таким образом следующая важная

Лемма. *Если* $\mathbf{X}$ *имеет плотность* (6.2) *и матрица* A *невырождена, то* $\mathbf{Y}=\mathbf{X}A$ *имеет нормальную плотность, порождаемую матрицей* $BQB^T=A^{-1}Q(A^{-1})^T$.

Рассмотрим, в частности, подстановку

$$y_1=x_1, \ \ldots, \ y_{r-1}=x_{r-1}, \quad y_r=q_{1r}x_1+\ \ldots\ +q_{rr}x_r. \tag{6.4}$$

Чтобы показать, что она невырождена, мы должны проверить, что $q_{rr}\neq 0$. Однако если бы мы имели $q_{rr}=0$, то при фиксированных значениях $x_1, \ldots, x_{r-1}$ плотность (6.2) должна была бы принять форму $\gamma^{-1}e^{-\alpha x_r+b}$ и интеграл относительно x_r должен был бы расходиться. Таким образом, $q_{rr}\neq 0$, и стандартный метод дополнения до квадрата дает нам, что $q(x)$ сводится к сумме y_r^2/q_{rr} и квадратичной форме q' от оставшихся $r-1$ величин

$$q(x)=\frac{1}{q_{rr}}y_r^2+q'(y). \tag{6.5}$$

Это порождает соответствующее разложение на множители плотности $\mathbf{Y}$, а отсюда видно, что частные плотности $\mathbf{Y}_r$ и $(\mathbf{Y}_1, \ldots$

[1]) «*Вырожденные*» нормальные распределения будут введены в конце этого параграфа.

$\ldots, \mathbf{Y}_{r-1}) = (\mathbf{X}_1, \ldots, \mathbf{X}_{r-1})$ снова нормальны. Продолжая по индукции, получаем следующую теорему.

Теорема 1. *Все частные плотности нормальной плотности снова нормальны.*

Линейное преобразование, приводящее к (6.5), дает нормальный вектор $(\mathbf{X}_1, \ldots, \mathbf{X}_{r-1}, \mathbf{Y}_r)$, такой, что $\mathbf{Y}_r$ не зависит от $(\mathbf{X}_1, \ldots, \mathbf{X}_{r-1})$. Применяя ту же самую редукцию к $(\mathbf{X}_1, \ldots, \mathbf{X}_{r-1})$, мы получим вектор

$$(\mathbf{X}_1, \ldots, \mathbf{X}_{r-2}, \mathbf{Z}_{r-1}, \mathbf{Y}_r),$$

такой, что $\mathbf{Y}_r$ и $\mathbf{Z}_{r-1}$ не зависят друг от друга и от $(\mathbf{X}_1, \ldots, \mathbf{X}_{r-2})$. Продолжая подобным же образом и вспоминая, что последовательное осуществление линейных преобразований эквивалентно однократному линейному преобразованию, мы заключаем, что существует невырожденное линейное преобразование $\mathbf{X} = \mathbf{Y}A^{-1}$, такое, что $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_r)$ имеет взаимно независимые нормальные компоненты $\mathbf{Y}_k$. Ввиду симметрии $\mathbf{E}(\mathbf{Y}_k) = 0$. Полагая $\mathrm{Var}(\mathbf{Y}_k) = \sigma_k^2$, мы видим, что $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_r)$ имеет нормальную плотность вида (6.2), порожденную *диагональной матрицей* D с элементами σ_k^{-2}. [Нормирующая константа γ определяется равенством $\gamma^2 = (2\pi)^r \det D^{-1}$.] Из леммы следует, что $Q = ADA^T$. С другой стороны, из (5.5) следует, что ковариационная матрица $\mathrm{Var}(\mathbf{X}) = \mathbf{E}(\mathbf{X}\mathbf{X}^T)$ существует и что $D^{-1} = A^T \mathrm{Var}(\mathbf{X}) A$. Таким образом, $\mathrm{Var}(\mathbf{X}) = Q^{-1}$, и после перемножения детерминантов мы имеем следующую теорему.

Теорема 2. *Если* $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_r)$ *имеет нормальную плотность* (6.2), *то* $Q^{-1} = \mathrm{Var}(\mathbf{X})$ *и*

$$\gamma^2 = (2\pi)^r \det Q^{-1}. \tag{6.6}$$

Следствие. *Если* $(\mathbf{X}_1, \mathbf{X}_2)$ *имеет нормальное распределение, то* $\mathbf{X}_1$ *и* $\mathbf{X}_2$ *независимы в том и только том случае, когда* $\mathrm{Cov}(\mathbf{X}_1, \mathbf{X}_2) = 0$, *т. е. когда* $\mathbf{X}_1$ *и* $\mathbf{X}_2$ *некоррелированы.*

Вообще, если $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ имеет нормальную плотность, то $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ и $(\mathbf{X}_{n+1}, \ldots, \mathbf{X}_r)$ независимы тогда и только тогда, когда $\mathrm{Cov}(\mathbf{X}_j, \mathbf{X}_k) = 0$, при $j \leqslant n$ $k > n$.

Предостережение. Для справедливости следствия существенно то, что совместная плотность пары $(\mathbf{X}_1, \mathbf{X}_2)$ нормальна. Следствие *неприменимо, если известно только, что частные плотности* $\mathbf{X}_1$ *и* $\mathbf{X}_2$ *являются нормальными.* В последнем случае плотность $(\mathbf{X}_1, \mathbf{X}_2)$ не обязана быть нормальной и в действительности даже не обязана существовать. Этот факт часто понимается неправильно (см. задачи 2,3).

Теорема 3. *Матрица M представляет собой ковариационную матрицу нормальной плотности тогда и только тогда, когда она положительно определена.*

Поскольку плотность порождается матрицей $Q=M^{-1}$, можно сформулировать эквивалентное утверждение: *матрица Q порождает нормальную плотность* (6.2) *тогда и только тогда, когда она положительно определена.*

Доказательство. Мы видели в конце § 5, что всякая ковариационная матрица плотности положительно определена. Обратное тривиально в случае $r=1$. К бо́льшим r мы перейдем по индукции. Предположим, что матрица Q положительно определенная. При $x_1=0, \ldots, x_{r-1}=0$ мы получаем $q(x)=q_{rr}x_r^2$ и, следовательно, $q_{rr}>0$. При этом предположении, как мы видели, q можно привести к виду (6.5). Выбирая x_r так, что $y_r=0$, мы видим, что положительная определенность Q влечет $q'(x)>0$ при всех наборах $x_1, \ldots, x_{r-1}$. Поэтому по предположению индукции q' соответствует нормальной плотности в пространстве $r-1$ измерений. Из (6.5) теперь очевидным образом следует, что q соответствует нормальной плотности в пространстве r измерений, и этим завершается доказательство. ▶

Пример. *Выборочные среднее значение и дисперсия.* В статистике случайные величины

$$\hat{\mathbf{X}}=\frac{1}{r}(\mathbf{X}_1+\ \ldots\ +\mathbf{X}_r),\quad \hat{\boldsymbol{\sigma}}^2=\frac{1}{r}\sum_{k=1}^{r}(\mathbf{X}_k-\hat{\mathbf{X}})^2 \tag{6.7}$$

называются выборочным средним значением и выборочной дисперсией для $\mathbf{X}=(\mathbf{X}_1, \ldots, \mathbf{X}_r)$. Весьма любопытен тот факт, что *если величины $\mathbf{X}_1, \ldots, \mathbf{X}_r$ независимы и нормальны с $\mathbf{E}(\mathbf{X}_k)=0$, $\mathbf{E}(\mathbf{X}_k^2)=\sigma^2$, то случайные величины $\hat{\mathbf{X}}$ и $\hat{\boldsymbol{\sigma}}^2$ являются независимыми*[1]). Доказательство демонстрирует применимость предыдущих результатов. Мы положим $\mathbf{Y}_k=\mathbf{X}_k-\hat{\mathbf{X}}$ при $k=1, \ldots, r-1$, но $\mathbf{Y}_r=\hat{\mathbf{X}}$. Преобразование $\mathbf{X}$ в $\mathbf{Y}=(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ линейно и невырождено. Поэтому $\mathbf{Y}$ имеет нормальную плотность. Далее $\mathbf{E}(\mathbf{Y}_k\mathbf{Y}_r)=0$ при $k=1, \ldots, r-1$, и поэтому $\mathbf{Y}_r$ не зависит от $(\mathbf{Y}_1, \ldots, \mathbf{Y}_{r-1})$. Однако

$$r\hat{\boldsymbol{\sigma}}^2=\mathbf{Y}_1^2+\ \ldots\ +\mathbf{Y}_{r-1}^2+(\mathbf{Y}_1+\ \ldots\ +\mathbf{Y}_{r-1})^2 \tag{6.8}$$

зависит только от $\mathbf{Y}_1, \ldots, \mathbf{Y}_{r-1}$, и, таким образом, $\hat{\boldsymbol{\sigma}}^2$ действительно не зависит от $\mathbf{Y}_r=\hat{\mathbf{X}}$. ▶

[1]) То, что этот факт характеризует нормальное распределение в $\mathcal{R}^1$, было показано Гири и Лукачем.

Мы закончим эту общую теорию интерпретацией (6.5) в терминах *условных плотностей*, которая приводит к общим формулировкам теории регрессии, разобранной для двумерного случая в примере (2, а).

Чтобы найти условную плотность $\mathbf{X}_r$ для данных $\mathbf{X}_1=x_1, \ldots, \mathbf{X}_{r-1}=x_{r-1}$, мы должны разделить плотность $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ на частную плотность для $\mathbf{X}_1, \ldots, \mathbf{X}_{r-1}$. Ввиду (6.5) в результате получим одномерную нормальную плотность с математическим ожиданием

$$-(q_{1,r}x_1+\ldots+q_{r-1,r}x_{r-1})/q_{rr}=a_1x_1+\ldots+a_{r-1}x_{r-1} \qquad (6.9)$$

и дисперсией q_{rr}^{-1}. (Здесь мы полагаем для сокращения $a_k=-q_{kr}/q_{rr}$.) Соответственно

$$\mathbf{E}(\mathbf{X}_r \mid \mathbf{X}_1, \ldots, \mathbf{X}_{r-1})=a_1\mathbf{X}_1+\ldots+a_{r-1}\mathbf{X}_{r-1}. \qquad (6.10)$$

Мы видели также, что

$$\mathbf{T}=\mathbf{X}_r-a_1\mathbf{X}_1-\ldots-a_{r-1}\mathbf{X}_{r-1} \qquad (6.11)$$

не зависит от $(\mathbf{X}_1, \ldots, \mathbf{X}_{r-1})$, и это свойство однозначно характеризует коэффициенты $a_k=-q_{kr}/q_{rr}$. Нами получена таким образом

Теорема 4. *Если* $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ *имеет нормальную плотность, то условная плотность* $\mathbf{X}_r$ *при данных* $(\mathbf{X}_1, \ldots, \mathbf{X}_{r-1})$ *снова нормальна. Кроме того, условное математическое ожидание* (6.10) *есть единственная линейная функция от* $\mathbf{X}_1, \ldots, \mathbf{X}_{r-1}$, *делающая величину* $\mathbf{T}$ *не зависящей от* $(\mathbf{X}_1, \ldots, \mathbf{X}_{r-1})$. *Условная дисперсия равна* $\mathrm{Var}(\mathbf{T})=q_{rr}^{-1}$.

Нормальные распределения общего вида

Из леммы следует, что если $\mathbf{X}=(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ имеет нормальную плотность, то каждая ненулевая линейная комбинация $\mathbf{Y}_1=a_1\mathbf{X}_1+\ldots+a_r\mathbf{X}_r$ также имеет нормальную плотность. То же самое верно для каждой пары $(\mathbf{Y}_1, \mathbf{Y}_2)$ при условии, что не выполняется линейное соотношение типа $c_1\mathbf{Y}_1+c_2\mathbf{Y}_2=0$. В этом исключительном случае распределение вероятностей пары $(\mathbf{Y}_1, \mathbf{Y}_2)$ сосредоточено на прямой с уравнением $c_1y_1+c_2y_2=0$ и, следовательно, оно вырождено, если его *рассматривать как двумерное распределение*. Для многих целей желательно сохранить термин нормального распределения также и для вырожденных распределений, сосредоточенных на многообразиях низшей размерности, скажем на какой-либо прямой. Простейшее общее определение выглядит следующим образом: *распределение* $\mathbf{Y}=(\mathbf{Y}_1, \ldots, \mathbf{Y}_\rho)$ *является нормальным, если существует вектор*

$\mathbf{X}=(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ *с нормальной* r*-мерной плотностью, такой, что* $\mathbf{Y}=a+\mathbf{X}A$, где A — (постоянная) $(r\times\rho)$-матрица и $a=(a_1, \ldots, a_\rho)$. Если $\rho>r$, то распределение $\mathbf{Y}$ вырождено в ρ измерениях. При $\rho\leqslant r$ это распределение невырождено тогда и только тогда, когда ρ форм, определяющих $\mathbf{Y}_k$, линейно независимы.

6а. Дополнение. Вращения

Построения в доказательствах теорем 1 и 2 показывают, что *существует* $(r\times r)$*-матрица* A*, такая, что* $\mathbf{Y}=\mathbf{X}A$ *имеет взаимно независимые компоненты* $\mathbf{Y}_k$*, причем* $\mathbf{Y}_k$ *зависят только от* $\mathbf{X}_1, \ldots, \mathbf{X}_k$. (Матрица A имеет ниже диагонали только нули.) Должно быть понятно, что подобные линейные преобразования иногда с вероятностной точки зрения лишены смысла. Так, если $\mathbf{X}_1$ означает размер популяции, а $\mathbf{X}_2$ — температуру, то имеется по существу только одна естественная система координатных величин. С другой стороны, если выборочное пространство представляет собой «подлинную» евклидову плоскость (а не комбинаторное произведение двух осей), то координатная система может быть выбрана многими эквивалентными способами. В таких ситуациях желательно выбирать систему координат так, чтобы получить для заданной нормальной плотности наиболее простую форму. В этой связи мы не будем более касаться общих линейных преобразований, а рассмотрим лишь *вращения координатных осей.*

Следующая теорема является непосредственным следствием хорошо известной задачи о приведении квадратичных форм (или симметричных матриц) к нормальному виду. Теорема не будет использоваться в этой книге и указывается только для ссылок. Доказательство легче, чем могло бы появиться в общей теории матриц (где в то же время доказывается более сильное утверждение). Мы приводим здесь необходимые определения и отдельные этапы доказательства в надежде, что некоторые читатели найдут это интересным и завершат доказательство. Это будет неплохим упражнением в матричном анализе. Обозначения и центрирование из § 6 сохраняются. В частности, $M=(m_{jk})$ есть ковариационная матрица в данной координатной системе.

Теорема. *Пусть дана нормальная плотность в* $\mathscr{R}^r$*. Тогда оси координат могут быть повернуты таким образом, что новые координатные величины будут взаимно независимыми нормальными величинами.*

Примем следующие определения. Вектор-строка x называется единичным вектором, если $xx^T=\Sigma x_k^2=1$. Угол ω между

двумя единичными векторами x и $\bar{x}$ определяется равенством $\cos\omega = x\bar{x}^T = \Sigma x_k \bar{x}_k$. Матрица A порядка $r \times r$ *ортогональна*, если преобразование $y = xA$ переводит единичные векторы в единичные. Детерминант ортогональной матрицы равен ± 1. *Вращениями* называются преобразования, порождаемые ортогональными матрицами с детерминантом 1.

Нам необходимы следующие факты, касающиеся вращений.

1) Матрица A ортогональна тогда и только тогда, когда $AA^T = A^TA = I$.

2) Вращение сохраняет углы между единичными векторами.

3) Существует ортогональная матрица A, последней строкой которой служит произвольно заданный единичный вектор.

Для *доказательства* теоремы необходимо последовательно проверить следующие факты.

1) Существует единичный вектор ξ, такой, что $\Sigma m_{jk}\xi_j\xi_k \geqslant \geqslant \Sigma m_{jk}x_jx_k$ для всех единичных векторов x.

2) Если $\xi = (0, 0, \ldots, 0, 1)$ является таким максимизирующим вектором, то $m_{rk} = 0$ при $k = 1, \ldots, r-1$.

3) Существует вращение, переводящее максимизирующий вектор ξ в $(0, \ldots, 0, 1)$. В повернутой системе координат последняя координатная величина не зависит от первых $r-1$ координатных величин. Поэтому задача сводится от r к $r-1$ измерениям, и мы продолжаем в том же духе, используя вращения, которые оставляют последнюю ось фиксированной. Тот факт, что $\mathbf{Y} = \mathbf{X}A$ имеет независимые компоненты, означает, что матрица A^TMA диагональна.

§ 7 *). Стационарные нормальные процессы

Цель этого параграфа состоит отчасти в том, чтобы дать примеры нормальных распределений, отчасти в том, чтобы установить некоторые соотношения, имеющие важное применение в теории дискретных стохастических процессов и временных рядов. Они носят *аналитический характер* и легко отделимы от более глубокого стохастического анализа. Фактически мы будем иметь дело только с конечномерными нормальными плотностями, или, что равносильно, с их ковариационными матрицами. Ссылки на случайные величины существенны для вероятностной интуиции и как подготовка к применениям, но на данном этапе мы касаемся только их совместных распределений; случайные величины сами используются единственно как удобный способ

*) Не используется в последующем. В частности, § 8 может быть прочитан независимо (см. также гл. XIX, 8).

описания всех частных плотностей посредством указания соответствующих совокупностей $(\mathbf{X}_{\alpha_1}, \ldots, \mathbf{X}_{\alpha_k})$. Кроме того, ссылка на бесконечную последовательность $\{\mathbf{X}_k\}$ влечет лишь то, что число элементов в $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ может быть взято сколь угодно большим.

Мы будем фактически рассматривать бесконечную в обе стороны последовательность $\{\ldots, \mathbf{X}_{-2}, \mathbf{X}_{-1}, \ldots\}$. Под этим мы просто подразумеваем, что для каждой конечной совокупности $(\mathbf{X}_{n_1}, \ldots, \mathbf{X}_{n_r})$ задана *нормальная* плотность с очевидными условиями согласованности. Последовательность называется *стационарной*, если эти распределения инвариантны относительно сдвигов во времени, т. е. если все строки вида $(\mathbf{X}_{n_1+\nu}, \ldots, \mathbf{X}_{n_r+\nu})$ с фиксированными $(n_1, \ldots, n_r)$ имеют одно и то же распределение, не зависящее от ν. При $r=1$ отсюда следует, что математические ожидания и дисперсии постоянны и, следовательно, можно предположить, не ограничивая общности, что $\mathbf{E}(\mathbf{X}_n)=0$. Совместные распределения полностью определяются через ковариации $\rho_{jk}=\mathbf{E}(\mathbf{X}_j\mathbf{X}_k)$, а свойство стационарности требует, чтобы ρ_{jk} зависели только от разности $|k-j|$. Поэтому мы положим $\rho_{jr\,j+n}=r_n$. Таким образом,

$$r_n=\mathbf{E}(\mathbf{X}_k\mathbf{X}_{k+n})=\mathbf{E}(\mathbf{X}_{k-n}, \mathbf{X}_k), \tag{7.1}$$

откуда $r_n=r_{-n}$. Мы будем иметь дело только с последовательностями чисел r_n, которые могут служить ковариациями стохастического процесса.

Всюду в этом параграфе $\{\mathbf{Z}_n\}$ изображает бесконечную в обе стороны последовательность *взаимно независимых нормально распределенных случайных величин*, нормированных так, что

$$\mathbf{E}(\mathbf{Z}_n)=0, \quad \mathbf{E}(\mathbf{Z}_n^2)=1. \tag{7.2}$$

Будут описаны три метода построения стационарных последовательностей в терминах данной последовательности $\{\mathbf{Z}_n\}$. Эти методы постоянно используются в анализе временных рядов и могут служить упражнением в стандартных рассуждениях.

Примеры. а) *Обобщенные процессы скользящего среднего.* При произвольных константах $b_0, b_1, \ldots, b_N$ положим

$$\mathbf{X}_n=b_0\mathbf{Z}_n+b_1\mathbf{Z}_{n-1}+\ldots+b_N\mathbf{Z}_{n-N}. \tag{7.3}$$

В частном случае равных коэффициентов $b_k=1/(N+1)$ случайная величина $\mathbf{X}_n$ является средним арифметическим типа, применяемого в анализе временных рядов к «сглаживанию данных» (т. е. к устранению локальных иррегулярностей). В общем случае (7.3) представляет собой линейный оператор, переводящий

стационарную последовательность $\{\mathbf{Z}_n\}$ в новую стационарную последовательность $\{\mathbf{X}_n\}$. Такие операции модно называть «фильтрами». Последовательность $\{\mathbf{X}_n\}$ имеет ковариации

$$r_k = r_{-k} = \mathbf{E}(\mathbf{X}_n \mathbf{X}_{n+k}) = \sum_{\nu} b_\nu b_{\nu+k} \quad (k \geqslant 0), \tag{7.4}$$

где сумма справа содержит ряды, имеющие лишь конечное число членов.

Так как $2|b_\nu b_{\nu+k}| \leqslant b_\nu^2 + b_{\nu+k}^2$, то выражение (7.4) имеет смысл также и для бесконечных последовательностей, для которых $\Sigma b_\nu^2 < \infty$. Легко видеть, что предел последовательности ковариационных матриц есть снова ковариационная матрица. Полагая $N \to \infty$, мы заключаем, что *для любой последовательности* $b_0, b_1, b_2, \ldots$, *такой, что* $\Sigma b_n^2 < \infty$, *числа* r_k, *определенные в* (7.4), *могут служить ковариациями стационарного процесса* $\{\mathbf{X}_n\}$. Формально мы получаем для нового процесса

$$\mathbf{X}_n = \sum_{k=0}^{\infty} b_k \mathbf{Z}_{n-k}. \tag{7.5}$$

Можно без затруднений показать, что любой стационарный процесс с ковариациями (7.4) представим в такой форме, но выражение (7.5) включает бесконечно много членов, и мы не можем сделать это в настоящий момент (см. гл. XIX, 8).

б) *Процесс авторегрессии.* С тех пор как родился анализ временных рядов, предлагались многие теоретические модели для объяснения эмпирических явлений, например, таких, как поведение экономических временных рядов, солнечных пятен и наблюденные (или воображаемые) периодичности. Наиболее распространенная модель предполагает, что величины $\mathbf{X}_n$ исследуемого процесса связаны с нашей последовательностью $\mathbf{Z}_n$ независимых нормальных величин посредством *уравнения авторегрессии* вида

$$a_0 \mathbf{X}_n + a_1 \mathbf{X}_{n-1} + \ldots + a_N \mathbf{X}_{n-N} = \mathbf{Z}_n. \tag{7.6}$$

Эта модель основывается на эмпирическом предположении, что значение величины $\mathbf{X}_n$ в момент n (цена, содержание или интенсивность) зависит от ее прошлого развития и наложенного на него «случайного возмущения» $\mathbf{Z}_n$, которое не связано с прошлым. Как часто бывает, предположение о *линейной* зависимости упрощает (или делает возможным) теоретический анализ. Более общие модели получаются, если положить $N \to \infty$ или выбрать в качестве $\mathbf{Z}_n$ величины другого стационарного процесса.

Если $a_0 \neq 0$, то можно произвольным образом выбрать $(\mathbf{X}_0, \ldots, \mathbf{X}_{N-1})$ и затем последовательно вычислить $\mathbf{X}_N, \mathbf{X}_{N+1}, \ldots$ и $\mathbf{X}_{-1}, \mathbf{X}_{-2}, \ldots$ В этом смысле (7.6) определяет некоторый процесс, но нас интересует, существует ли *стационарное* решение.

Для ответа на этот вопрос мы перепишем (7.6) в форме, не включающей элементов, непосредственно предшествующих $\mathbf{X}_n$. Рассмотрим (7.6), заменяя n последовательно на $n-1, n-2, \ldots, n-\nu$. Умножим эти равенства на $b_1, b_2, \ldots, b_\nu$ соответственно и прибавим к (7.6). Величины $\mathbf{X}_{n-1}, \ldots, \mathbf{X}_{n-\nu}$ не появятся в новом уравнении в том и только в том случае, если b_j таковы, что

$$a_0 b_1 + a_1 b_0 = 0, \ \ldots, \quad a_0 b_\nu + a_1 b_{\nu-1} + \ \ldots \ + a_\nu b_0 = 0, \tag{7.7}$$

где $b_0 = 1$. В конце концов приходим к выражению вида

$$a_0 \mathbf{X}_n = b_0 \mathbf{Z}_n + b_1 \mathbf{Z}_{n-1} + \ \ldots \ + b_\nu \mathbf{Z}_{n-\nu} + \mathbf{Y}_{n,\nu}, \tag{7.8}$$

где $\mathbf{Y}_{n,\nu}$ представляет собой линейную комбинацию $\mathbf{X}_{n-\nu-1}, \ldots, \mathbf{X}_{n-N-\nu}$ (с коэффициентами, которые нас не интересуют). В (7.8) мы представим величину $\mathbf{X}_n$ как сумму случайных возмущений в моменты $n, n-1, \ldots, n-\nu$ и величины $\mathbf{Y}_{n,\nu}$, изображающей влияние «прошлого» до момента $n-\nu$. При $\nu \to \infty$ это время становится «бесконечно далеким прошлым», и в большинстве ситуаций оно не будет иметь влияния. При переходе к пределу мы (по крайней мере временно) предположим именно этот случай, т. е. мы разыскиваем процесс, удовлетворяющий предельному соотношению вида

$$a_0 \mathbf{X}_n = \sum_{k=0}^{\infty} b_k \mathbf{Z}_{n-k}. \tag{7.9}$$

(Грубо говоря, мы предполагаем, что остаточные величины $\mathbf{Y}_{n,\nu}$ стремятся к нулю. Другие возможные пределы будут изучены в следующем примере.)

Процессы вида (7.9) рассматривались в примере (а), и мы видели, что всякий раз, когда $\Sigma b_k^2 < \infty$, существует стационарное решение (если ряд расходится, то даже выражение для ковариаций теряет смысл). Для решения уравнений (7.7) относительно b_k мы используем формальные производящие функции

$$A(s) = \Sigma a_k s^k, \quad B(s) = \Sigma b_k s^k. \tag{7.10}$$

Уравнения (7.7) выполняются тогда и только тогда, когда $A(s)B(s) = a_0 b_0$. Так как A полином, то B — рациональная функция. Поэтому мы можем использовать теорию разложения на простые дроби, развитую в **1**, гл. XI, 4. Если полином $A(s)$

имеет различные корни $s_1, \ldots, s_N$, мы получаем

$$B(s) = \frac{A_1}{s_1 - s} + \ldots + \frac{A_N}{s_N - s} \tag{7.11}$$

и, следовательно,

$$b_n = A_1 s_1^{-n-1} + \ldots + A_N s_N^{-n-1}. \tag{7.12}$$

Очевидно, что $\Sigma b_n^2 < \infty$ *тогда и только тогда, когда все корни удовлетворяют неравенству* $|s_j| > 1$. Легко проверить, что это остается справедливым также в случае кратных корней. Мы таким образом сейчас показали, что *стационарное решение авторегрессионной модели* (7.6) *существует всегда, когда все корни полинома* $A(s)$ *лежат вне единичного круга.* Ковариации нашего процесса задаются формулой (7.4), и «бесконечно далекое прошлое» не играет в нем никакой роли. (Другие решения см. в следующем примере.)

в) *Вырожденные процессы.* Обратимся к стационарным последовательностям $\{\mathbf{Y}_n\}$, удовлетворяющим *«стохастическому разностному уравнению»*

$$a_0 \mathbf{Y}_n + a_1 \mathbf{Y}_{n-1} + \ldots + a_N \mathbf{Y}_{n-N} = 0. \tag{7.13}$$

Типичными являются частные случаи

$$\mathbf{Y}_n = \lambda (\mathbf{Z}_1 \cos n\omega + \mathbf{Z}_{-1} \sin n\omega), \tag{7.14}$$

$$\mathbf{Y}_n = \alpha_1 \mathbf{Z}_1 + (-1)^n \alpha_2 \mathbf{Z}_{-1}, \tag{7.15}$$

где коэффициенты и величина ω постоянны, а $\mathbf{Z}_1$ и $\mathbf{Z}_{-1}$ являются независимыми нормальными случайными величинами, нормированными условиями (7.2). Эти процессы удовлетворяют уравнению (7.13), первый с $a_0 = a_2 = 1$ и $a_1 = -2\cos\omega$, второй с $a_0 = -a_2 = 1$ и $a_1 = 0$. Эти процессы вырождены в том смысле, что весь процесс полностью определяется двумя наблюдениями, скажем $\mathbf{Y}_{k-1}$ и $\mathbf{Y}_k$. Эти два наблюдения можно взять так далеко в прошлом, как нам угодно, и в этом смысле процесс полностью определяется своим «бесконечно далеким прошлым». Аналогичные рассуждения применимы к любому процессу, удовлетворяющему разностному уравнению вида (7.13), и, следовательно, эти процессы составляют противоположность примеру (б), где бесконечно далекое прошлое совсем не имеет влияния. ▶

Теория разностного уравнения (7.13) дает также *наиболее общее решение задачи в примере* (б). Действительно, мы нашли *одно* решение (7.6), и разность двух решений удовлетворяет (7.13). Отсюда следует, что мы можем добавить к решению в примере (б) произвольную стационарную последовательность,

удовлетворяющую (7.13), причем последняя будет описывать вклад бесконечно далекого прошлого.

Сделаем одно замечание, касающееся природы разностного уравнения (7.13). Если $\{\mathbf{Y}_k\}$ удовлетворяет этому уравнению, то существует линейная зависимость между $N+1$ величинами $\mathbf{Y}_0, \mathbf{Y}_1, \ldots, \mathbf{Y}_N$ и, следовательно, их распределение должно быть вырождено (т. е. оно не может иметь плотность в $N+1$ измерениях). Без ограничения общности мы можем предположить, что распределение $(\mathbf{Y}_0, \ldots, \mathbf{Y}_{N-1})$ невырождено, в противном случае некоторая линейная комбинация этих величин была бы равна нулю и вследствие предположенной стационарности это привело бы к тому, что процесс $\{\mathbf{Y}_k\}$ удовлетворял бы разностному уравнению порядка $N-1$ или ниже. Иными словами, (7.13) должно представлять собой разностное уравнение *наинизшего* порядка, которому удовлетворяет $\{\mathbf{Y}_n\}$ Тогда строка $(\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$ имеет плотность.

Теперь можно показать, что теория стационарных решений разностного уравнения (7.13) тесно связана с «характеристическим уравнением»

$$\xi^N + a_1\xi^{N-1} + \ldots + a_N = 0. \tag{7.16}$$

Каждому квадратному множителю полинома, стоящего слева здесь соответствует стохастическое разностное уравнение второго порядка и вследствие этого процесс вида (7.14) или (7.15). В соответствии с разложением на множители характеристического полинома мы представим, таким образом, общее решение уравнения (7.13) как сумму компонент вида (7.14) и (7.15).

Как и раньше, мы предполагаем $\mathbf{E}(\mathbf{Y}_n)=0$. Вся теория строится на следующей лемме.

Лемма 1. *Стационарная последовательность с* $\mathbf{E}(\mathbf{Y}_n\mathbf{Y}_{n+k}) = r_k$ *удовлетворяет стохастическому разностному уравнению* (7.13) *в том и только том случае, когда*

$$r_n + a_1 r_{n-1} + \ldots + a_N r_{n-N} = 0. \tag{7.17}$$

Доказательство. Умножив (7.13) на $\mathbf{Y}_0$ и взяв математические ожидания, придем к (7.17). Возводя в квадрат левую часть равенства (7.13) и вновь вычисляя математические ожидания, получаем в результате $\Sigma a_j(\Sigma a_k r_{k-j})$, и поэтому из (7.17) следует, что левая часть в (7.13) имеет нулевую дисперсию. Это доказывает лемму. ▷

Мы приступаем к выводу канонической формы (7.22) для r_n. Эти величины, конечно, вещественны, но, так как в формулу входят корни характеристического уравнения (7.16), мы должны прибегнуть к временному использованию комплексных чисел.

Лемма 2. *Если* $\{\mathbf{Y}_n\}$ *удовлетворяет уравнению* (7.13), *но не удовлетворяет никакому разностному уравнению низшего порядка, то характеристическое уравнение* (7.16) *обладает* N *различными корнями* $\xi_1, \ldots, \xi_N$, *по модулю равными единице. В этом случае*

$$r_n = c_1\xi_1^n + \ldots + c_N\xi_N^n, \tag{7.18}$$

где $c_j > 0$ *при* $j = 1, \ldots, N$.

Доказательство. Предположим вначале, что характеристическое уравнение (7.16) имеет N *различных* корней $\xi_1, \ldots, \xi_N$. Определим r_n формулой (7.18) с произвольными постоянными $c_1, \ldots, c_N$. Очевидно тогда, что эти r_n удовлетворяют разностному уравнению (7.17) и что свободные параметры $c_1, \ldots, c_N$ могут быть подобраны так, чтобы привести к предписанным значениям $r_1, \ldots, r_N$. Отсюда следует, что (в случае различных корней) *каждое* решение $\{r_n\}$ разностного уравнения (7.17) имеет вид (7.18). Теперь ковариации r_n стационарной последовательности остаются ограниченными при $n \to \infty$ и $n \to -\infty$, что возможно только если или $|\xi_j| = 1$, или $c_j = 0$. Те же самые рассуждения применимы в случае, когда $\xi_1 = \xi_2$ есть двойной корень, за исключением того, что первый член в правой части (7.18) заменяется на $(c_1 n + c_2)\xi_1^n$. Из ограниченности r_n снова с необходимостью следует, что $c_1 = 0$. Мы видим, таким образом, что даже в случае кратных корней r_n должны иметь вид

$$r_n = c_1\xi_1^n + \ldots + c_\rho\xi_\rho^n, \tag{7.19}$$

где $\rho \leqslant N$ и $\xi_1, \ldots, \xi_\rho$ — различные корни характеристического уравнения (7.16), причем все они по модулю равны единице. Нам остается доказать, что фактически число членов, присутствующих в (7.19), не может быть меньше N. Для этой цели мы отметим, что $\xi_1, \ldots, \xi_\rho$ удовлетворяют уравнению степени ρ с (возможно комплексными) коэффициентами α_k. Отсюда следует, что r_n удовлетворяет разностному уравнению

$$r_n + \alpha_1 r_{n-1} + \ldots + \alpha_\rho r_{n-\rho} = 0, \tag{7.20}$$

являющемуся аналогом (7.17). Теперь ковариации r_n являются вещественными числами, и, следовательно, они удовлетворяют также разностному уравнению, полученному отделением вещественной части уравнения (7.20). Но по предположению r_n не удовлетворяют никакому вещественному разностному уравнению порядка, меньшего N, и поэтому должно быть $\rho = N$ и $c_j \neq 0$ при $j = 1, \ldots, N$. Этим завершается доказательство.

Для определенности мы сформулируем окончательный результат для случая, когда N — *нечетное* целое число.

Теорема. *Предположим, что стационарная последовательность* $\{\mathbf{Y}_n\}$ *удовлетворяет разностному уравнению* (7.13) *с* $N=2\nu+1$, *но не удовлетворяет никакому разностному уравнению низшего порядка. Тогда характеристическое уравнение* (7.16) *обладает* ν *парами комплексных корней* $\xi_j=\cos\omega_j \pm i\sin\omega_j$ (*где* ω_j *вещественны*) *и одним вещественным корнем* $\omega_0=\pm 1$. *Последовательность* $\{\mathbf{Y}_n\}$ *имеет вид*

$$\mathbf{Y}_n=\lambda_0\mathbf{Z}_0\cdot\omega_0^n+\sum_{j=1}^{\nu}\lambda_j[\mathbf{Z}_j\cos n\omega_j+\mathbf{Z}_{-j}\sin n\omega_j], \qquad (7.21)$$

где $\mathbf{Z}_j$ — *взаимно независимые нормальные величины, нормированные условиями* (7.2), *а* λ_j — *постоянные. Для этой последовательности*

$$r_n=\lambda_0^2\omega_0^n+\sum_{j=1}^{\nu}\lambda_j^2\cos n\omega_j. \qquad (7.22)$$

Обратно, выберем произвольно вещественные $\lambda_\nu\neq 0$ *и* $\omega_0=\pm 1$. *Пусть* $\omega_1,\ldots,\omega_\nu$ — *различные вещественные числа, такие, что* $0<\omega_j<\pi$. *Тогда* (7.21) *определяет стационарный процесс с ковариациями* (7.22), *удовлетворяющий разностному уравнению порядка* $2\nu+1$ *(но не удовлетворяющий никакому разностному уравнению низшего порядка).*

Доказательство. Так как r_n вещественны, то из леммы 2 следует, что r_n необходимо имеют вид (7.22). Простейшие вычисления показывают, что ковариации последовательности $\{\mathbf{Y}_n\}$, определенной в (7.21), удовлетворяют (7.22) и, следовательно, (7.17). Поэтому по лемме 1 $\{\mathbf{Y}_n\}$ удовлетворяет заданному разностному уравнению (7.13). По построению r_n они не удовлетворяют никакому разностному уравнению порядка меньше чем $N=2\nu+1$. Следовательно, $\{\mathbf{Y}_n\}$ не может удовлетворять никакому разностному уравнению порядка меньше N. Таким образом, $\{\mathbf{Y}_n\}$ удовлетворяет сформулированным условиям, но мы должны еще показать, что *каждая* последовательность $\{\mathbf{Y}_n\}$, удовлетворяющая данному разностному уравнению, имеет вид (7.21). Для этой цели рассмотрим (7.21) при $n=0,\ldots,2\nu$ как линейное преобразование $(\mathbf{Z}_{-\nu},\ldots,\mathbf{Z}_\nu)$ в $(\mathbf{Y}_0,\ldots,\mathbf{Y}_{2\nu})$. Легко видеть, что это преобразование невырождено, и поэтому ковариационные матрицы для $(\mathbf{Z}_{-\nu},\ldots,\mathbf{Z}_\nu)$ и $(\mathbf{Y}_0,\ldots,\mathbf{Y}_{2\nu})$ определяют одна другую единственным образом. Мы предположили, что ковариации величин $\mathbf{Z}_j$ задаются единичной матрицей, и доказали, что ковариации величин $\mathbf{Y}_n$ задаются формулой (7.22). Ввиду невырожденного характера преобразования в равной степени

справедливо, что ковариации (7.22) для $\{\mathbf{Y}_n\}$ определяют ковариации (7.2) для $\{\mathbf{Z}_n\}$

Это доказывает прямую часть теоремы. Обратная часть сводится к ней рассмотрением характеристического уравнения с $2\nu+1$ корнями ω_0 и $\cos\omega_j \pm i\sin\omega_j$.

§ 8. Марковские нормальные плотности

Мы перейдем к обсуждению одного класса нормальных плотностей, встречающихся в марковских процессах. Не ограничивая общности, *мы рассмотрим только плотности, центрированные в начале координат*. Тогда $\mathbf{E}(\mathbf{X}_k)=0$ и мы используем обычные сокращения

$$\mathbf{E}(\mathbf{X}_k^2)=\sigma_k^2, \quad \mathbf{E}(\mathbf{X}_j\mathbf{X}_k)=\sigma_j\sigma_k\rho_{jk}. \tag{8.1}$$

Величины ρ_{jk} суть *коэффициенты корреляции*, и $\rho_{kk}=1$.

Определение. *Нормальная r-мерная плотность* $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ *является марковской, если при* $k\leqslant r$ *условная плотность* $\mathbf{X}_k$ *при данных* $\mathbf{X}_1, \ldots, \mathbf{X}_{k-1}$ *тождественна с условной плотностью* $\mathbf{X}_k$ *при данном* $\mathbf{X}_{k-1}$.

Грубо говоря, если мы знаем $\mathbf{X}_{k-1}$ («настоящее»), то дополнительное знание «прошлого» $\mathbf{X}_1, \ldots, \mathbf{X}_{k-2}$ не привносит никакой полезной информации относительно «будущего», т. е. относительно любой из величин $\mathbf{X}_j$ с $j\geqslant k$.

Теорема 1. *Для того чтобы конечная последовательность* $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ *была марковской*[1]), *каждое из следующих двух условий является необходимым и достаточным*:

(i) *для* $k\leqslant r$

$$\mathbf{E}(\mathbf{X}_k \mid \mathbf{X}_1, \ldots, \mathbf{X}_{k-1}) = \mathbf{E}(\mathbf{X}_k \mid \mathbf{X}_{k-1}); \tag{8.2}$$

(ii) *для* $j\leqslant \nu < k \leqslant r$

$$\rho_{jk}=\rho_{j\nu}\rho_{\nu k}. \tag{8.3}$$

Доказательство. Ввиду того что совпадение плотностей влечет равенство математических ожиданий, необходимость (8.2) тривиальна. Положим

$$\mathbf{T}=\mathbf{X}_k-\frac{\sigma_k}{\sigma_{k-1}}\rho_{k-1,k}\mathbf{X}_{k-1}. \tag{8.4}$$

Тогда $\mathbf{E}(\mathbf{T}\mathbf{X}_{k-1})=0$ и $\mathbf{T}$ — единственная величина вида $\mathbf{X}_k-\lambda\mathbf{X}_{k-1}$, некоррелированная с $\mathbf{X}_{k-1}$. По теореме 6.4 мы имеем $\mathbf{T}=\mathbf{X}_k$ —

[1]) Как обычно в подобных случаях, мы применяем термин «марковский» и к последовательности $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ и к ее плотности.

$-\mathbf{E}(\mathbf{X}_k \mid \mathbf{X}_{k-1})$. Следовательно, (8.2) выполняется тогда и только тогда, когда величина $\mathbf{T}$ некоррелирована с $\mathbf{X}_1, \ldots, \mathbf{X}_{k-1}$, т. е. тогда и только тогда, когда

$$\rho_{jk} = \rho_{j,\,k-1} \cdot \rho_{k-1,\,k}, \quad j \leqslant k. \tag{8.5}$$

В этом случае выполняется не только (8.2), но также и соотношение

$$\mathrm{Var}(\mathbf{X}_k \mid \mathbf{X}_1 \ldots \mathbf{X}_{k-1}) = \mathrm{Var}(\mathbf{X}_k \mid \mathbf{X}_{k-1}) = \mathrm{Var}(\mathbf{T}). \tag{8.6}$$

Следовательно, соответствующие условные плотности совпадают. Таким образом, (8.2) выполняется при всех $k \leqslant r$ тогда и только тогда, когда справедливо (8.5), и в этом случае последовательность $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ — марковская. Поскольку (8.5) есть частный случай (8.3), то последнее условие является достаточным. Оно также и необходимо, так как повторное применение (8.5) показывает, что при $j<v<k \leqslant r$

$$\frac{\rho_{jk}}{\rho_{vk}} = \frac{\rho_{j,\,k-1}}{\rho_{v,\,k-1}} = \frac{\rho_{j,\,k-2}}{\rho_{v,\,k-2}} = \ldots = \frac{\rho_{jv}}{\rho_{vv}} = \rho_{jv}. \tag{8.7}$$

Поэтому (8.5) влечет (8.3). ▶

Следствие. *Если последовательность* $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ *марковская, то каждая подпоследовательность* $(\mathbf{X}_{\alpha_1}, \ldots, \mathbf{X}_{\alpha_v})$ *с* $\alpha_1<\alpha_2<\ldots<\alpha_v \leqslant r$ *является марковской.*

Это очевидно, так как (8.3) автоматически распространяется на все подпоследовательности. ▶

Примеры. а) *Независимые приращения.* Говорят, что подпоследовательность (конечная или бесконечная) $\{\mathbf{X}_k\}$ нормальных случайных величин с $\mathbf{E}(\mathbf{X}_k)=0$ образует процесс с независимыми приращениями, если при $j<k$ приращение $\mathbf{X}_k - \mathbf{X}_j$ не зависит от $(\mathbf{X}_1, \ldots, \mathbf{X}_j)$. Отсюда следует, в частности, что $\mathbf{E}(\mathbf{X}_j(\mathbf{X}_k - \mathbf{X}_j))=0$ или

$$\rho_{jk} = \frac{\sigma_j}{\sigma_k}, \qquad j < k. \tag{8.8}$$

Сравнивая это с (8.3), видим, что нормальный процесс с независимыми приращениями автоматически оказывается марковским. Его структура чрезвычайно проста: $\mathbf{X}_k$ есть сумма k взаимно независимых нормальных величин

$$\mathbf{X}_1, \ \mathbf{X}_2 - \mathbf{X}_1, \ \ldots, \ \mathbf{X}_k - \mathbf{X}_{k-1}.$$

б) *Модели авторегрессии.* Рассмотрим нормальную марковскую последовательность $\mathbf{X}_1, \mathbf{X}_2, \ldots$ с $\mathbf{E}(\mathbf{X}_k)=0$. Существует единственная константа a_k, такая, что $\mathbf{X}_k - a_k\mathbf{X}_{k-1}$ не зависит

от $\mathbf{X}_{k-1}$ и, следовательно, от $\mathbf{X}_1, \ldots, \mathbf{X}_{k-1}$. Положим

$$\lambda_k^2 = \operatorname{Var}(\mathbf{X}_k - a_k\mathbf{X}_{k-1})$$

и последовательно

$$\begin{aligned} \mathbf{X}_1 &= \lambda_1\mathbf{Z}_1, \\ \mathbf{X}_k &= a_k\mathbf{X}_{k-1} + \lambda_k\mathbf{Z}_k \qquad k = 2, 3, \ldots . \end{aligned} \tag{8.9}$$

Легко видеть, что определенные таким образом величины $\mathbf{Z}_k$ независимы и

$$\mathbf{E}(\mathbf{Z}_k) = 0, \quad \mathbf{E}(\mathbf{Z}_k^2) = 1. \tag{8.10}$$

Верно и обратное. Если $\mathbf{Z}_k$ нормальны и удовлетворяют (8.10), то (8.9) определяет последовательность $\{\mathbf{X}_n\}$ и сама по себе структура (8.9) показывает, что последовательность $\{\mathbf{X}_n\}$ — марковская. В качестве упражнения мы проверим это утверждение вычислением. Умножим (8.9) на $\mathbf{X}_j$ и возьмем математические ожидания. Так как $\mathbf{Z}_k$ не зависит от $\mathbf{X}_1, \ldots, \mathbf{X}_{k-1}$, мы получаем при $j<k$

$$a_k = \frac{\sigma_k}{\sigma_{k-1}} \frac{\rho_{jk}}{\rho_{j,\,k-1}}. \tag{8.11}$$

Теперь (8.5) есть простое следствие (8.11), а мы знаем, что выполнение (8.5) влечет марковский характер $\mathbf{X}_k$. Таким образом, $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ *является марковской тогда и только тогда, когда выполняются соотношения вида* (8.9), где $\mathbf{Z}_j$ — независимые нормальные величины, удовлетворяющие (8.10). [Это есть частный случай примера (7, б).] ▶

До сих пор мы рассматривали только конечные последовательности $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$, но число r не играет никакой роли, и мы можем также говорить о бесконечных последовательностях $\{\mathbf{X}_n\}$. Это *не* затрагивает пространств бесконечных последовательностей или какую-либо новую теорию, это просто указание на то, что распределение для $(\mathbf{X}_1, \ldots, \mathbf{X}_r)$ определяется при всех r. Аналогично мы говорим о *марковском семействе* $\{\mathbf{X}(t)\}$, когда любая *конечная* совокупность $\mathbf{X}_1 = \mathbf{X}(t_1), \ldots, \mathbf{X}_r = \mathbf{X}(t_r)$ является марковской. Ее описание зависит от функций

$$\mathbf{E}(\mathbf{X}^2(t)) = \sigma^2(t), \quad \mathbf{E}(\mathbf{X}(s)\mathbf{X}(t)) = \sigma(s)\sigma(t)\rho(s, t). \tag{8.12}$$

В силу критерия (8.3) очевидно, что *семейство является марковским тогда и только тогда, когда при* $s<t<\tau$

$$\rho(s, t)\rho(t, \tau) = \rho(s, \tau). \tag{8.13}$$

Вопреки порождаемым терминологией образам, мы фактически будем иметь дело только с семействами конечномерных нор-

мальных распределений с ковариациями, которые удовлетворяют (8.13).

Как обстоятельно объяснялось в начале § 7, последовательность $\{\mathbf{X}_n\}$ называют *стационарной*, если для каждой фиксированной строки $\alpha_1, \ldots, \alpha_n$ распределение $(\mathbf{X}_{\alpha_1+\nu}, \ldots, \mathbf{X}_{\alpha_n+\nu})$ не зависит от ν. Конечный отрезок такой последовательности может быть продолжен в обе стороны, и, следовательно, естественно рассматривать только бесконечные в обе стороны последовательности $\{\ldots, \mathbf{X}_{-2}, \mathbf{X}_{-1}, \mathbf{X}_0, \mathbf{X}_1 \ldots\}$. Эти понятия тривиальным образом переносятся на семейства $\{\mathbf{X}(t)\}$.

Для стационарной *последовательности* $\{\mathbf{X}_n\}$ дисперсия σ_n^2 не зависит от n и в марковском случае (8.3) влечет равенство $\rho_{jk} = \rho_{12}^{|k-j|}$. Таким образом, *для стационарной марковской последовательности*

$$\mathbf{E}(\mathbf{X}_j\mathbf{X}_k) = \sigma^2\rho^{|k-j|}, \tag{8.14}$$

где σ^2 и ρ — постоянные, $|\rho|<1$. Обратно, последовательность с нормальными распределениями, удовлетворяющая (8.14), является марковской и стационарной.

В случае стационарного семейства $\{\mathbf{X}(t)\}$ корреляция $\rho(s, t)$ зависит только от разности $|t-s|$ и (8.13) приобретает вид

$$\rho(t)\rho(\tau) = \rho(t+\tau) \quad \text{для} \quad t, \ \tau>0.$$

Очевидно, из равенства $\rho(\tau)=0$ должно следовать $\rho(t)=0$ при всех $t>\tau$ и также $\rho(\frac{1}{2}\tau)=0$. Поэтому ρ не может иметь нулей, за исключением случая $\rho(t)=0$ при всех $t>0$. Следовательно, $\rho(t)=e^{-\lambda t}$ (здесь используется результат из **1**, гл. XVII, 6). Соответственно *для стационарного марковского семейства*

$$\mathbf{E}(\mathbf{X}(s)\mathbf{X}(s+t)) = \sigma^2 e^{-\lambda t}, \qquad t>0, \tag{8.15}$$

за исключением случая, когда $\mathbf{X}(s)$ и $\mathbf{X}(t)$ некоррелированы при всех $s\neq t$.

Пример. в) *Стационарные последовательности* могут быть построены по схеме последнего примера. Вследствие (8.11) мы должны иметь

$$\mathbf{X}_k = \rho\mathbf{X}_{k-1} + \sigma\sqrt{1-\rho^2}\,\mathbf{Z}_k. \tag{8.16}$$

При каждом k можно выразить $\mathbf{X}_k$ как линейную комбинацию $\mathbf{Z}_k, \mathbf{Z}_{k-1}, \ldots, \mathbf{Z}_{k-\nu}$ и $\mathbf{X}_{k-\nu+1}$. Формальный переход к пределу приводит к представлению

$$\mathbf{X}_k = \sigma(1-\rho^2)\sum_{j=0}^{\infty}\rho^j\mathbf{Z}_{k-j}, \tag{8.17}$$

т. е. к представлению $\{\mathbf{X}_k\}$ через бесконечную в обе стороны последовательность независимых нормальных величин $\mathbf{Z}_j$,

нормированных условием (8.10). Так как $|\rho|<1$, то правдоподобно, что ряд сходится, однако формула по существу связана с пространством бесконечных последовательностей. [См. замечания, касающиеся формулы (7.5), частным случаем которой является (8.17).] ▶

Полезно, быть может, обсудить связь теоремы 1 с непосредственным описанием марковских последовательностей в терминах плотностей. Обозначим через g_i плотность $\mathbf{X}_i$ и через $g_{ik}(x, y)$ — значение в точке y условной плотности $\mathbf{X}_k$ при условии $\mathbf{X}_i=x$. (В теории стохастических процессов g_{ik} называется переходной плотностью от $\mathbf{X}_i$ к $\mathbf{X}_k$.) Из теории нормальных марковских последовательностей мы знаем, что g_i представляет собой нормальную плотность с нулевым математическим ожиданием и дисперсией σ_i^2, в то время как

$$g_{ik}(x, y)=\frac{1}{\sigma_k\sqrt{1-\rho_{ik}^2}}\,n\left(\frac{y-\sigma_i^{-1}\rho_{ik}\sigma_k x}{\sigma_k\sqrt{1-\rho_{ik}^2}}\right) \tag{8.18}$$

при фиксированном x есть нормальная плотность с математическим ожиданием $\sigma_i^{-1}\rho_{ik}\sigma_k x$ и дисперсией $\sigma_k^2(1-\rho_{ik}^2)$. Как бы то ни было, мы начнем все заново, не используя теорему 1.

По самому определению совместная плотность $(\mathbf{X}_i, \mathbf{X}_j)$ задается как $g_i(x)g_{ij}(x, y)$. Совместная плотность $(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k)$ равна произведению предыдущей плотности на условную плотность $\mathbf{X}_k$ при данных $\mathbf{X}_j$ и $\mathbf{X}_i$, но ввиду марковского характера индекс i может быть опущен, если $i<j<k$, и плотность $(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k)$ становится равной

$$g_i(x)\,g_{ij}(x, y)\,g_{ik}(y, z). \tag{8.19}$$

В марковском случае плотность каждой строки $(\mathbf{X}_{\alpha_1}, \ldots, \mathbf{X}_{\alpha_n})$ задается произведением вида (8.19), однако плотности g_{ik} не могут быть выбраны произвольно. В самом деле, интегрирование (8.19) относительно y дает частную плотность для $(\mathbf{X}_i, \mathbf{X}_k)$. Поэтому мы получаем условие согласованности

$$g_{ik}(x, z)=\int_{-\infty}^{+\infty} g_{ij}(x, y)\,g_{ik}(y, z)\,dy \tag{8.20}$$

при всех $i<j<k$. Это есть специальный случай *уравнения Колмогорова — Чепмена* для марковских процессов [1]). Говоря очень

[1]) Другие специальные случаи встречались в 1, гл. XV, (10.3) и гл. XVII, (9.1). Заметим, что формула (8.19) представляет собой аналог определения [т. I, гл. XV, (1.1)] вероятностей для марковских цепей, за исключением того, что здесь суммирование заменено интегрированием и что были рассмотрены только стационарные переходные вероятности.

приблизительно, это уравнение означает, что переход из x в момент i в z в момент k происходит через все промежуточные состояния y, причем переход от y к z не зависит от прошлого. Очевидно, что при любой системе переходных вероятностей g_{ik}, удовлетворяющих уравнению Колмогорова—Чепмена, схема умножения (8.19) приводит к согласованной системе плотностей для $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_r)$, и эта последовательность является марковской. Таким образом, мы пришли к следующему аналитическому дополнению теоремы 1.

Теорема 2. *Плотности семейства $\{g_{ik}\}$ тогда и только тогда могут служить переходными плотностями в нормальном марковском процессе, когда они удовлетворяют уравнению Колмогорова — Чепмена и когда $g_{ik}(x, y)$ при каждом фиксированном x является нормальной плотностью относительно y.*

Обе теоремы содержат необходимые и достаточные условия, и поэтому они в некотором смысле эквивалентны. Тем не менее они имеют различную природу. Вторая из них на самом деле не ограничивается нормальными процессами; примененная к семействам $\{\mathbf{X}(t)\}$, она приводит к дифференциальным и интегральным уравнениям для переходных вероятностей и на этом пути способствует введению новых классов марковских процессов. С другой стороны, из теоремы 2 трудно угадать, что g_{ik} необходимо имеет вид (8.18) — результат, содержащийся в более специальной теореме 1.

Для последующих ссылок и сравнений мы приводим здесь два наиболее важных марковских семейства $\{\mathbf{X}(t)\}$.

Пример. г) *Броуновское движение или процесс Винера — Башелье.* Процесс определяется тем условием, что $\mathbf{X}(0)=0$ и что при $t>s$ величина $\mathbf{X}(t) - \mathbf{X}(s)$ не зависит от $\mathbf{X}(s)$ и имеет дисперсию, зависящую только от $t-s$. Иными словами, процесс имеет независимые приращения [пример (а)] и стационарные переходные вероятности [но он не стационарный, поскольку $\mathbf{X}(0)=0$]. Очевидно, что $\mathbf{E}(X^2(t))=\sigma^2 t$ и $\mathbf{E}(\mathbf{X}(s)\mathbf{X}(t))=\sigma^2 s$ при $s<t$. При $\tau>t$ переходные плотности, характеризующие переход из (t, x) в (τ, y), нормальны с математическим ожиданием x и дисперсией $\sigma^2(\tau-t)$. Они зависят только от $(y-x)/(\tau-t)$, и уравнения Колмогорова — Чепмена сводятся к свертке.

д) *Процесс Орнстейна — Уленбека.* Под этим понимается наиболее общий нормальный *стационарный марковский процесс* с нулевыми математическими ожиданиями. Его ковариации определяются формулой (8.15). Иными словами, при $\tau>t$ переходные плотности, характеризующие переход из (t, x) в (τ, y), нормальны с математическим ожиданием $e^{-\lambda(\tau-t)}x$ и дисперсией

$\sigma^2(1-e^{-2\lambda(\tau-t)})$. При $\tau\to\infty$ математическое ожидание стремится к нулю, а дисперсия — к σ^2. Этот процесс был рассмотрен Орнстейном и Уленбеком с совершенно другой точки зрения. Его связь с диффузией будет обсуждена в гл. X, 4. ▶

§ 9. Задачи

1. Рассмотрим Ω — область на плоскости (площадью $1/2$), ограниченную четырехугольником с вершинами $(0,0)$, $(1,1)$, $(0,1/2)$, $(1/2,1)$ и треугольником с вершинами $(1/2,0)$, $(1,0)$, $(1,1/2)$. (Единичный квадрат представляется объединением Ω и области, симметричной Ω относительно биссектрисы.) Пусть пара $(\mathbf{X},\mathbf{Y})$ распределена равномерно на Ω. Докажите, что частные распределения являются равномерными и что $\mathbf{X}+\mathbf{Y}$ имеет такую же плотность, как если бы $\mathbf{X}$ и $\mathbf{Y}$ были независимыми [1]).

Указание. Рисунок делает вычисления излишними.

2. *Плотности с частными нормальными плотностями.* Пусть u — *нечетная* непрерывная функция на прямой, равная нулю вне интервала $-1,1$. Если $|u|<(2\pi e)^{-1/2}$, то выражение

$$n(x)\,n(y)+u(x)\,u(y)$$

изображает двумерную плотность, которая *не является нормальной, но частные плотности которой нормальны.* (Э. Нелсон.)

3. *Второй пример.* Пусть φ_1 и φ_2 — две двумерные нормальные плотности с единичными дисперсиями, но с различными коэффициентами корреляции. Смесь $1/2(\varphi_1+\varphi_2)$ *не является нормальной плотностью, но две ее частные плотности совпадают с* n.

Замечание. В последующем все случайные величины рассматриваются в $\mathscr{R}^1$. *Векторные величины обозначаются парами* $(\mathbf{X}_1,\mathbf{X}_2)$ *и т. д.*

4. Пусть $\mathbf{X}_1,\ \ldots,\ \mathbf{X}_n$ — независимые случайные величины с одинаковой плотностью f и функцией распределения F. Если $\mathbf{X}$ и $\mathbf{Y}$ являются соответственно наименьшей и наибольшей из них, то совместная плотность пары $(\mathbf{X},\mathbf{Y})$ равна

$$n(n-1)f(x)f(y)[F(y)-F(x)]^{n-2}.$$

5. Покажите, что распределение Коши в $\mathscr{R}^3$ [определенное в (1.21)] соответствует случайному вектору, длина которого распределена с плотностью $v(r)=4\pi^{-1}r^2(1+r^2)^{-2}$ при $r>0$.

Указание. Используйте полярные координаты и либо (1.15), либо же общее соотношение гл. I, (10.4) для проекций.

6. Пусть $a>0$ и $f(x,y)=[(1+ax)(1+ay)-a]e^{-x-y-axy}$ при $x>0$, $y>0$ и $f(x,y)=0$ в других случаях.

а) Докажите, что f есть плотность пары $(\mathbf{X},\mathbf{Y})$. Найдите частные плотности и функцию распределения.

б) Найдите условную плотность $u_x(y)$ и $\mathbf{E}(\mathbf{Y}|\mathbf{X})$, $\mathrm{Var}(\mathbf{Y}|\mathbf{X})$. (Э. Гумбел.)

7. Пусть f — плотность, сосредоточенная на $\overline{0,\infty}$. Положим $u(x,y)=f(x+y)/(x+y)$ при $x>0$, $y>0$ и $u(x,y)=0$ в других случаях. Докажите, что u есть плотность в $\mathscr{R}^2$ и найдите ее ковариационную матрицу.

---

[1]) Иными словами, распределение суммы может быть задано сверткой, даже если величины зависимы. Этот интуитивно понятный пример предложен Г. Роббинсом. Другой парадокс такого же типа см. в гл. II, (4, д).

8. Пусть $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ взаимно независимы и равномерно распределены на $\overline{0,1}$. Пусть $\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)}$, $\mathbf{X}_{(3)}$ — соответствующие порядковые статистики. Найдите плотность пары

$$\left(\frac{\mathbf{X}_{(1)}}{\mathbf{X}_{(2)}}, \frac{\mathbf{X}_{(2)}}{\mathbf{X}_{(3)}}\right)$$

и покажите, что эти два отношения независимы. Обобщите на n измерений.

9. Пусть $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ независимы и одинаково показательно распределены. Найдите плотность $(\mathbf{X}_2-\mathbf{X}_1, \mathbf{X}_3-\mathbf{X}_1)$.

10. Частица массы m расщепляется на два осколка с массами $\mathbf{X}m$ и $(1-\mathbf{X})m$. Здесь $\mathbf{X}$ — случайная величина с плотностью $f(x)>0$ при $0<x<1$ и $f(x)=0$ в других случаях; по соображениям симметрии $f(x)=f(1-x)$. Предположим, что эти два осколка расщепляются независимо таким же образом. Найдите (а) совместную плотность наименьшего и наибольшего из четырех осколков; (б) частную плотность наименьшего из четырех осколков.

11. Пусть $\mathbf{X}_1$, $\mathbf{X}_2$, ... независимы и имеют одну и ту же нормальную плотность n. Обозначим $\mathbf{S}_k=\mathbf{X}_1+\ldots+\mathbf{X}_k$. При $m<n$ найдите совместную плотность $(\mathbf{S}_m, \mathbf{S}_n)$ и условную плотность для $\mathbf{S}_m$ при условии, что $\mathbf{S}_n=t$.

12. В предыдущей задаче найдите условную плотность $\mathbf{X}_1^2+\ldots+\mathbf{X}_m^2$ при данном значении $\mathbf{X}_1^2+\ldots+\mathbf{X}_n^2$.

13. Пусть $(\mathbf{X}, \mathbf{Y})$ имеет двумерную нормальную плотность, центрированную в начале координат, с $\mathbf{E}(\mathbf{X}^2)=\mathbf{E}(\mathbf{Y}^2)=1$ и $\mathbf{E}(\mathbf{XY})=\rho$. При переходе к полярным координатам $(\mathbf{X}, \mathbf{Y})$ превращается в $(\mathbf{R}, \mathbf{\Phi})$, где $\mathbf{R}^2=\mathbf{X}^2+\mathbf{Y}^2$. Докажите, что $\mathbf{\Phi}$ имеет плотность, равную

$$\frac{\sqrt{1-\rho^2}}{2\pi(1-2\rho\sin\varphi\cos\varphi)} \qquad 0<\varphi<2\pi,$$

и распределено равномерно тогда и только тогда, когда $\rho=0$. Выведите, что

$$\mathbf{P}\{\mathbf{XY}>0\}=\frac{1}{2}+\pi^{-1}\arcsin\rho \quad \text{и} \quad \mathbf{P}\{\mathbf{XY}<0\}=\pi^{-1}\arccos\rho.$$

14. Пусть f — равномерная плотность в треугольнике с вершинами (0,0), (0, 1), (1, 0) и g — равномерная плотность в симметричном треугольнике в третьем квадранте. Найдите $f*f$ и $f*g$.

Предостережение. Требуется утомительное раздельное рассмотрение отдельных интервалов.

15. Пусть f — равномерное распределение в единичном круге. Найдите $f*f$ в полярных координатах.

16. Пусть u и v — плотности в $\mathscr{R}^2$ вида

$$u(x, y)=f(\sqrt{x^2+y^2}), \qquad v(x, y)=g(\sqrt{x^2+y^2}).$$

Найдите $u*v$ в полярных координатах.

17. Найдите общий нормальный стационарный процесс, удовлетворяющий соотношениям

а) $\mathbf{X}_{n+2}+\mathbf{X}_n=0$;

б) $\mathbf{X}_{n+2}-\mathbf{X}_n=0$;

в) $\mathbf{X}_{n-3}-\mathbf{X}_{n+2}+\mathbf{X}_{n+1}-\mathbf{X}_n=0$.

18. *Сервостохастический процесс.* (Г. Д. Милс) Сервомеханизм подвергается случайным толчкам, но в любое время могут быть введены поправки. Таким образом, ошибка $\mathbf{Y}_n$ в момент n имеет (в подходящих единицах) вид $\mathbf{Y}_{n+1}=\mathbf{Y}_n+\mathbf{C}_n+\mathbf{X}_{n+1}$, где $\mathbf{C}_n$ — поправка, а $\mathbf{X}_n$ — независимые величины с $\mathbf{E}(\mathbf{X}_n)=0$, $\mathbf{E}(\mathbf{X}_n^2)=1$. Величины $\mathbf{C}_n$ в принципе являются произвольными функциями прошлых наблюдений, т. е. $\mathbf{Y}_k$ и $\mathbf{X}_k$ при $k \leqslant n$. Желательно выбрать их так, чтобы минимизировать дисперсию $\mathrm{Var}(\mathbf{Y}_n)$ (которая представляет собой меру того, сколь *хорошо* работает механизм).

а) Рассмотрите ковариационную функцию $\{\mathbf{Y}_n\}$ и покажите, что $\mathrm{Var}(\mathbf{Y}_n)>1$.

б) В предположении, что $\mathrm{Var}(\mathbf{C}_n)\to\alpha^2$, $\mathrm{Var}(\mathbf{Y}_n)\to\sigma^2$ (тенденция к стационарности), покажите, что $\sigma>{}^1/_2(\alpha+\alpha^{-1})$.

в) Рассмотрите, в частности, *линейную схему* $\mathbf{C}_n=a-p(\mathbf{Y}_n-b)$, $0<p\leqslant 1$. Найдите ковариационную функцию и представление вида (7.8) для $\mathbf{Y}_n$.

19. *Продолжение.* Если существует *запаздывание во времени* в поступлении информации или корректировке, то модель остается по существу той же самой, за исключением того, что $\mathbf{C}_n$ следует заменить на $\mathbf{C}_{n+N}$. Обсудите эту ситуацию.

ГЛАВА IV

ВЕРОЯТНОСТНЫЕ МЕРЫ И ПРОСТРАНСТВА

Как уже говорилось во введении, технический аппарат теории меры используется в этой книге лишь в небольшой степени, и для понимания большей ее части настоящая глава не нужна [1]). Тем не менее желательно дать краткий обзор важнейших понятий, образующих теоретический фундамент этой книги, а также привести основные теоремы для удобства ссылок. Идеи и факты, имеющиеся в виду, не трудны, однако доказательства в теории меры включают в себя неприятные технические детали. Для начинающего и неспециалиста подступ затрудняется многогранностью теории меры, а также разнообразием ее применений. Существуют превосходно написанные вводные курсы, однако материал в них в силу необходимости дается в слишком общей форме и, кроме того, в них излагаются различные аспекты теории меры, не являющиеся существенными для настоящей книги. Даваемый ниже обзор содержит, главным образом, лишь то, что нужно для дальнейшего; многие доказательства и технические детали опускаются [2]). (Следует отметить, что простота теории обманчива: значительно более трудные задачи теории меры возникают в связи со случайными процессами, зависящими от непрерывного временного параметра. Изложение условных математических ожиданий откладывается до гл. V. 10; теорема Радона — Никодима содержится в гл. V. 3.)

---

[1]) Это относится к читателям, знакомым с основами теории меры, а также к тем читателям, которые интересуются главным образом результатами и фактами. Для удобства последних определение интеграла повторяется в гл. V. 1. За пределами этого материала они могут полагаться на свою интуицию, поскольку по существу теория меры дает обоснование простым формальным действиям.

[2]) Бэровские функции и интегрирование по Лебегу — Стильтьесу превосходно изложены в книге Маклейна и Т. А. Боттса, Real analysis, D. Van Nostrand, Princeton, 1959. Результаты общей теории меры широко используются в изложении Халмоша, Халмош П. Р., Теория меры, ИЛ, М., 1953, Бурбаки Н., Elements de mathematiques [Livre VI, chapitres 3—5], Herman, Paris, 1952, 1956. Общая теория меры, написанная с целью применения в теории вероятностей, содержится в книгах Дуба, Лоэва, Невё и Аншкина и Тортра.

Формулы, относящиеся к евклидовым пространствам $\mathscr{R}^r$, не зависят от числа измерений и x в них нужно понимать как сокращенную запись для $(x_1, \ldots, x_r)$.

§ 1. Бэровские функции

Нам нужно указать класс множеств, для которых определены вероятности, а также класс функций, являющихся случайными величинами. Эти две задачи связаны между собой и, более того, их изложение объединяется применением современных обозначений. Введем сначала эти обозначения и напомним определение сходимости в терминах монотонных пределов.

Индикатором[1]) *множества* A называется функция, равная единице во всех точках A и равная нулю во всех точках множества A', дополнительного к A. Индикатор A обозначается $\mathbf{1}_A$, так что $\mathbf{1}_A(x)=1$ при $x\in A$ и $\mathbf{1}_A(x)=0$ в противном случае. Каждому множеству отвечает его индикатор, и каждая функция, принимающая лишь значения 0 и 1, является индикатором некоторого множества. Если f — произвольная функция, то произведение $\mathbf{1}_A f$ есть функция, равная f на A и нулю в остальных точках.

Рассмотрим теперь пересечение $C=A\cap B$ двух множеств. Его индикатор $\mathbf{1}_C$ равен нулю, когда либо $\mathbf{1}_A$, либо $\mathbf{1}_B$ обращается в нуль, так что $\mathbf{1}_C=\inf(\mathbf{1}_A, \mathbf{1}_B)$, т. е. значение $\mathbf{1}_C$ равно наименьшему из значений $\mathbf{1}_A$, $\mathbf{1}_B$. Этот параллелизм отражается в обозначении $f\cap g$, которое используется вместо $\inf(f, g)$ для функции, равной в каждой точке x наименьшему из значений $f(x)$ и $g(x)$. Аналогично $f\cup g=\sup(f, g)$ обозначает наибольшее из двух значений[2]). Операторы $\cap$ и $\cup$ применимы к любому множеству функций. Обычно пользуются сокращенными обозначениями

$$f_1\cap\ldots\cap f_n=\bigcap_{k=1}^{n} f_k,\quad f_1\cup\ldots\cup f_n=\bigcup_{k=1}^{n} f_k. \tag{1.1}$$

По определению в каждой точке x эти функции равны соответственно минимуму и максимуму среди n значений $f_1(x)$, ..., $f_n(x)$. Если f_k есть индикатор множества A_k, то функ-

[1]) Этот термин был введен Лоэвом. Более старый термин «характеристическая функция» неудачен в теории вероятностей, поскольку его принято относить к другому объекту.

[2]) Многие авторы предпочитают символы $\wedge$ и $\vee$ для функций, оставляя символы $\cap$ и $\cup$ лишь для множеств. В нашем изложении двойные обозначения не дают никаких преимуществ.

ции (1.1) являются индикаторами соответственно пересечения $A_1 \cap \ldots \cap A_n$ и объединения $A_1 \cup \ldots \cup A_n$.

Рассмотрим теперь *бесконечную* последовательность $\{f_n\}$. Функции, определяемые формулой (1.1), монотонны по n, так что пределы $\bigcap_{k=1}^{\infty} f_k$ и $\bigcup_{k=1}^{\infty} f_k$ вполне определены, хотя, возможно, и бесконечны. Для фиксированного j функция

$$w_j = \bigcap_{k=j}^{\infty} f_k \tag{1.2}$$

есть предел монотонной последовательности функций $f_j \cap \ldots \cap f_{j+n}$; и сама последовательность $\{w_j\}$ тоже монотонна, именно $w_n = w_1 \cap \ldots \cap w_n$. В наших обозначениях $w_n \to \bigcup_{k=1}^{\infty} w_k$. По определению $w_n(x)$ есть нижняя грань (инфимум) числовой последовательности $f_n(x), f_{n+1}(x), \ldots$. Следовательно, предел последовательности w_n совпадает с нижним пределом $\liminf f_n$, так что

$$\liminf f_n = \bigcup_{j=1}^{\infty} \bigcap_{k=j}^{\infty} f_k. \tag{1.3}$$

При таком подходе нижний предел lim inf получается двумя последовательными переходами к пределу в монотонных последовательностях. Аналогичное справедливо и для верхнего предела lim sup, равенство (1.3) сохранится, если в левой его части lim inf заменить на lim sup, а в правой части знаки $\cap$ и $\cup$ поменять местами.

Все сказанное переносится на множества. Так, например, по определению $A = \lim A_n$ тогда и только тогда, когда $\mathbf{1}_A = \lim \mathbf{1}_{A_n}$. Это означает, что последовательность множеств $\{A_n\}$ сходится к множеству A тогда и только тогда, когда каждая точка A принадлежит *всем* A, за исключением не более чем конечного числа, и каждая точка дополнения A' принадлежит не более чем конечному числу множеств A_n'.

Пример. а) *Множество* $\{A_n$ б. ч.$\}$ [1]). В качестве важного с вероятностной точки зрения примера предельных операций над множествами рассмотрим событие A, состоящее в «реализации бесконечного числа событий из данной последовательности событий $A_1, A_2, \ldots$». [Частные случаи были рассмотрены в **1**, гл. VIII (леммы Бореля — Кантелли) и в **1**, гл. XIII (рекуррентные события).] На более формальном языке это означает, что точка x принадлежит множеству A тогда и только тогда,

[1]) В подлиннике «A_n i. o.» (infinitely often). — *Прим. перев.*

когда она принадлежит бесконечному числу множеств последовательности $\{A_k\}$. Поскольку индикаторы могут принимать лишь два значения 0 и 1, то данное определение равносильно следующему: $\mathbf{1}_A = \limsup \mathbf{1}_{A_n}$. В обычных обозначениях это записывается как $A = \limsup A_n$, однако обозначение $\{A_n$ б. ч.$\}$ (читается A_n «бесконечно часто») более приятно и наглядно. Оно было введено Чжуном и теперь общепринято в вероятностной литературе. ▶

Наша следующая задача — ограничить класс функций[1]) на $\mathscr{R}^r$, с которыми мы предполагаем иметь дело. Понятие произвольной функции слишком общо, чтобы быть полезным для наших целей. Более подходящим является модернизированный вариант эйлерового определения функции. Если считать, что непрерывные функции заданы, то единственный эффективный способ конструирования новых функций состоит в предельных переходах. Оказывается, все наши потребности будут удовлетворены, если мы будем знать, как обращаться с функциями, являющимися пределами последовательностей непрерывных функций или пределами последовательностей, в которых каждая $\{f_n\}$ является таким пределом и т. д. Иными словами, нам интересен класс $\mathfrak{B}$ функций, обладающий следующими свойствами: 1) каждая непрерывная функция принадлежит $\mathfrak{B}$ и 2) если функции $f_1, f_2, \ldots$ принадлежат $\mathfrak{B}$, и предел $f(x) = \lim f_n(x)$ существует при всех x, то f тоже принадлежит $\mathfrak{B}$. Такой класс называется *замкнутым* относительно поточечных пределов. Ясно, что такие классы существуют, например класс *всех* функций. Пересечение всех таких классов является замкнутым семейством и образует, очевидно, наименьший такой класс. Интуитивно ясно, что разумно ограничить наши рассмотрения этим наименьшим классом.

Наименьший замкнутый класс функций, содержащий все непрерывные функции, называется классом Бэра и будет обозначаться $\mathfrak{B}$. Функции из $\mathfrak{B}$ называются бэровскими функциями[2]).

---

[1]) В принципе мы интересуемся лишь конечнозначными функциями, однако иногда удобно допустить также в качестве возможных значений $\pm\infty$. Например, простая теорема, утверждающая сходимость всякой монотонной последовательности, неверна для конечнозначных функций, а без нее многие формулировки становятся тяжеловесными. По этой причине мы следуем обычному соглашению, по которому все функции принимают значения из расширенной вещественной прямой, т. е. их значения суть числа или $\pm\infty$. На практике значения $\pm\infty$ не будут играть роли. Чтобы сумма и произведение двух функций были определены, вводят следующие соглашения: $\infty + \infty = \infty$, $\infty - \infty = 0$, $\infty \cdot \infty = \infty$, $0 \cdot \infty = 0$ и т. д.

[2]) Это определение основано лишь на понятии непрерывности и не зависит от других свойств евклидовых пространств. Его, следовательно, можно перенести на любое топологическое пространство.

Понятие бэровской функции будет использоваться не только по отношению к функциям, заданным на всем пространстве, но и по отношению к функциям, определенным лишь на некотором подмножестве (например, $\sqrt{x}$ и $\log x$ в случае пространства $\mathscr{E}^1$).

Из данного определения ясно, что сумма и произведение двух бэровских функций тоже являются бэровскими функциями. Более того, если w — непрерывная функция от r переменных и $f_1, \ldots, f_r$ — бэровские функции, то функция $w(f_1, \ldots, f_r)$ тоже бэровская. Заменяя w на w_n и переходя к пределу, можно показать, что вообще *всякая бэровская функция от бэровских функций есть бэровская функция.* Фиксация значений одной или нескольких переменных приводит опять к бэровской функции и т. д. Короче говоря, никакая из обычных операций над бэровскими функциями не выводит за пределы класса $\mathfrak{B}$, и, следовательно, класс $\mathfrak{B}$ представляет собой естественный объект для изучения. Оказывается, что никакого упрощения не достигается, если ограничиться меньшими классами.

§ 2. Функции интервалов и интегралы в $\mathscr{R}^r$

Интервалом мы будем называть множество точек, удовлетворяющих двойному неравенству одного из следующих четырех типов (попутно вводятся очевидные обозначения):

$\overline{a, b}$: $a < x < b$ $\quad$ $\overline{a, b}|$: $a < x \leqslant b$

$|\overline{a, b}|$: $a \leqslant x \leqslant b$ $\quad$ $|\overline{a, b}$: $a \leqslant x < b$

В одномерном случае выписанные неравенства исчерпывают все возможные интервалы, включая вырожденный интервал нулевой длины. В двумерном случае эти неравенства понимаются в покоординатном смысле, и интервалы представляют собой (возможно, вырожденные) прямоугольники со сторонами, параллельными осям координат. При числе измерений, большем или равном двум, возможны и другие типы интервалов, однако мы их исключаем из рассмотрения. Допускается предельный случай, когда одна или более координат a или b равны $\pm\infty$. В частности, все пространство есть интервал $\overline{-\infty, \infty}$.

Функция точек f сопоставляет значения $f(x)$ отдельным точкам. *Функция множеств* F сопоставляет значения отдельным множествам или областям пространства. Объем в $\mathscr{R}^3$, площадь в $\mathscr{R}^2$ или длина в $\mathscr{R}^1$ являются типичными примерами функций множеств. Помимо этих есть много других функций множеств, и среди них вероятности представляют собой наиболее интересный

для нас частный случай. Мы будем интересоваться лишь функциями множеств, обладающими следующим свойством: если множество A разбито на две части A_1 и A_2, то $F\{A\}=F\{A_1\}+F\{A_2\}$. Такие функции называются аддитивными [1]).

Как мы видели, вероятности $F\{I\}$ часто определены для всех интервалов r-мерного пространства $\mathscr{R}^r$, и их желательно доопределить на более общих множествах. Та же самая задача возникает в элементарном анализе, где площадь первоначально определена только для прямоугольников, и желательно определить ее для более общей области A. Простейший подход состоит в том, что сначала определяется интеграл от функции двух переменных и затем «площадь A» приравнивается к интегралу от индикатора $\mathbf{1}_A$ (т. е. к функции, равной 1 на A и обращающейся в нуль вне A). Аналогично мы определим интеграл

$$\mathbf{E}(u)=\int_{\mathscr{R}^r} u(x)F\{dx\} \tag{2.1}$$

от функции точек u по отношению к функции интервалов F. Вероятность A определяется затем как $\mathbf{E}(\mathbf{1}_A)$. При построении интеграла (2.1) неважно, как интерпретируется функция F, и мы по существу дадим общее понятие интеграла Лебега — Стильтьеса. Приступим теперь к аккуратному выполнению намеченной программы.

Пусть F — функция, сопоставляющая каждому интервалу I конечное значение $F\{I\}$. Такая функция называется (конечно) *аддитивной*, если для каждого разбиения интервала на конечное число непересекающихся интервалов $I_1, \ldots, I_n$

$$F\{I\}=F\{I_1\}+\ldots+F\{I_n\}. \tag{2.2}$$

Примеры. а) *Распределения в* $\mathscr{R}^1$. В первом томе мы рассматривали дискретные вероятностные распределения, сопоставляющие вероятности $p_1, p_2, \ldots$ точкам $a_1, a_2, \ldots$. Здесь $F\{I\}$ есть сумма весов p_n всех точек a_n, содержащихся в I, и $\mathbf{E}(u)=\Sigma u(a_n)p_n$.

Если G — какая-либо непрерывная монотонная функция, возрастающая от 0 на $-\infty$ до 1 на ∞, то можно положить $F\{\overline{a,\ b}\}=G(b)-G(a)$.

---

[1]) Вот несколько примеров аддитивных функций из практики: масса и количество тепла в области, цена земли, площадь под пшеницей и число жителей в географическом районе, ежегодная добыча угля, расстояние, которое пролетел пассажир, или число израсходованных киловатт-часов в течение некоторого периода, число телефонных вызовов и т. д.

б) *Случайные векторы.* Вектор единичной длины выходит из начала координат в случайном направлении. Вероятность того, что его конечная точка лежит в двумерном интервале I, пропорциональна длине дуги, по которой I пересекается с единичной окружностью. Определяемое здесь непрерывное распределение вероятностей не имеет плотности. Оно *сингулярно* в том смысле, что вся вероятность сосредоточена на окружности. Можно думать, что такие распределения искусственны и что в рассматриваемом случае скорее окружность, а не плоскость должна быть естественным выборочным пространством. Это возражение неосновательно, поскольку сумма двух таких независимых случайных векторов может принять любое значение между 0 и 2, и имеет положительную плотность в круге радиуса 2 (см. V, пример (4, д)). Следовательно, в некоторых задачах, связанных с единичными случайными векторами, плоскость является естественным выборочным пространством. Во всяком случае нашей целью было только показать на простом примере, что может происходить в более сложных ситуациях.

в) В заключение приведем пример, иллюстрирующий одно обстоятельство, которое будет *исключено* в дальнейшем. В пространстве $\mathscr{R}^1$ положим $F\{I\}=0$ для любого интервала $I=\overline{a, b}$ с $b<\infty$ и $F\{I\}=1$, когда $I=\overline{a, \infty}$. Так определенная функция интервалов, будучи аддитивной, является патологической в том смысле, что она не удовлетворяет естественному требованию непрерывности, согласно которому $F\{\overline{a, b}\}$ должно стремиться к $F\{\overline{a, \infty}\}$ при $b \to \infty$. ▶

Последний пример показывает, что желательно усилить требование (2.2) конечной аддитивности. Мы скажем, что *функция интервалов F счетно- или σ-аддитивна, если для каждого разбиения интервала I на счетное число интервалов $I_1, I_2, \ldots$*

$$F\{I\} = \sum F\{I_k\}. \tag{2.3}$$

[Счетное число обозначает либо конечное число, либо бесконечное счетное число. Термины «вполне аддитивная» и «счетно-аддитивная» означают одно и то же. Условие (2.3) явно нарушено в последнем примере.]

В дальнейшем мы будем рассматривать только счетно-аддитивные функции множеств. Это ограничение, с одной стороны, позволяет успешно развить теорию, а с другой — может быть оправдано a priori на эвристической и прагматической основе. В самом деле, если $A_n = I_1 \cup \ldots \cup I_n$ есть объединение первых n интервалов, то $A_n \to I$. Можно считать, что «при n достаточно большом A_n практически неотличимо от I». Если $F\{A_n\}$ может

быть найдено из экспериментов, то $F\{A_n\}$ должно быть «практически неотличимо» от $F\{I\}$, т. е. $F\{A_n\}$ должно стремиться к $F\{I\}$. Счетная аддитивность (2.3) и есть точное выражение этого требования.

Поскольку нас интересуют в первую очередь вероятности, мы ограничимся рассмотрением лишь неотрицательных функций интервалов F, нормированных условием $F\{-\infty, \infty\}=1$. Требование нормированности не накладывает никаких серьезных ограничений, когда $F\{-\infty, \infty\}<\infty$, однако исключает такие функции интервалов, как длина в $\mathscr{R}^1$ или площадь в $\mathscr{R}^2$. Чтобы использовать излагаемую ниже теорию в подобных случаях, достаточно разбить прямую или плоскость на единичные интервалы и рассматривать их по отдельности. Этот прием так очевиден и настолько хорошо известен, что не требует никаких дальнейших пояснений.

Функция на $\mathscr{R}^r$ называется *ступенчатой функцией*, если она принимает лишь конечное число значений, причем каждое значение на некотором интервале. Пусть ступенчатая функция u принимает значения $a_1, \ldots, a_n$ на интервалах $I_1, \ldots, I_n$ (с вероятностями $F\{I_1\}, \ldots, F\{I_n\}$) соответственно. По аналогии с определением математического ожидания дискретных случайных величин положим

$$\mathbf{E}(u) = a_1 F_1\{I_1\} + \ldots + a_n F\{I_n\}. \tag{2.4}$$

[Конечно, разбиение пространства на интервалы, на которых функция u постоянна, не единственно, однако, как и в дискретном случае, легко видеть, что определение (2.4) не зависит от разбиения.] Математическое ожидание $\mathbf{E}(u)$ обладает следующими свойствами:

а) *Аддитивностью* относительно линейных комбинаций:

$$\mathbf{E}(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 \mathbf{E}(u_1) + \alpha_2 \mathbf{E}(u_2). \tag{2.5}$$

б) *Положительностью*:

$$\text{если } u \geqslant 0, \text{ то } \mathbf{E}(u) \geqslant 0. \tag{2.6}$$

в) *Нормированностью*: для функции, равной во всех точках 1,

$$\mathbf{E}(1) = 1. \tag{2.7}$$

Последние два свойства эквивалентны теореме о среднем значении: если $\alpha \leqslant u \leqslant \beta$, то $\alpha \leqslant \mathbf{E}(u) \leqslant \beta$, так что функция

$\mathbf{E}(u)$ представляет собой один из видов *среднего значения*[1]).

Теперь задача состоит в том, чтобы продолжить определение $\mathbf{E}(u)$ на более широкий класс функций с сохранением свойств (а) — (в). В классическом интегрировании по Риману используется тот факт, что для каждой непрерывной функции u на $\overline{0, 1}$ существует последовательность ступенчатых функций u_n, для которой $u_n \to u$ равномерно на $\overline{0, 1}$. По определению полагают $\mathbf{E}(u) = \lim \mathbf{E}(u_n)$. Оказывается, что требовать равномерность сходимости необязательно, и то же самое определение $\mathbf{E}(u)$ можно использовать, когда сходимость $u_n \to u$ лишь поточечная. Таким способом $\mathbf{E}(u)$ можно продолжить на *все ограниченные бэровские функции*, и это продолжение *единственно*. При распространении $\mathbf{E}(u)$ на неограниченные функции неизбежно появляются расходящиеся интегралы, однако, по крайней мере для *положительных* бэровских функций можно определить $\mathbf{E}(u)$ либо как число, либо как символ ∞ (указывающий расходимость). Никаких неприятностей в связи с этим не возникает, поскольку в лебеговской теории интегрирования рассматривается лишь абсолютная интегрируемость. Грубо говоря, исходя из определения (2.4) математического ожидания для простых функций, можно доопределить $\mathbf{E}(u)$ для произвольных бэровских функций, используя очевидные приближения и переходы к пределу. Так определенное число $\mathbf{E}(u)$ есть интеграл Лебега — Стильтьеса от функции u по отношению к F. (Когда функция F фиксирована, более предпочтителен термин «математическое ожидание» — никаких двусмысленностей при этом не возникает.) Ниже приводится без доказательства[2]) основной факт

[1]) Когда функция F задает вероятности, среднее $\mathbf{E}(u)$ можно интерпретировать как ожидаемый выигрыш игрока, возможные выигрыши которого суть $a_1, a_2, \ldots$. Чтобы интуитивно осознать смысл среднего в других ситуациях, рассмотрим три примера, в которых $u(x)$ является соответственно температурой во время x, числом телефонных разговоров во время x, расстоянием материальной точки от начала координат, а F представляет собой соответственно длительность временно́го интервала, цену временно́го интервала (цену разговора) и механическую массу. В каждом случае интегрирование производится лишь по конечному интервалу и $\mathbf{E}(u)$ есть соответственно средняя температура, накопленный доход и статический момент. Эти примеры показывают, что определенное нами интегрирование по отношению к произвольной функции множеств проще и нагляднее, чем интегрирование по Риману, в котором независимая переменная играет несколько ролей и интерпретация как «площадь под кривой» ничуть не помогает начинающему. Следует иметь в виду, что понятие среднего (математического ожидания) встречается не только в теории вероятностей.

[2]) Метод доказательства намечен в § 5. Как обычно u^+ и u^- обозначают положительную и отрицательную части u, т. е. $u^+ = u \cup 0$ и $-u^- = u \cap 0$. Очевидно, $u = u^+ - u^-$.

лебеговской теории; его природа и значение будут проанализированы в следующих параграфах. [Конструктивное определение $\mathbf{E}(u)$ дается в § 4.]

Основная теорема. *Пусть F — счетно-аддитивная функция интервалов в $\mathscr{R}^2$ и $F\{-\infty, \infty\}=1$. Существует единственный интеграл Лебега — Стильтьеса $\mathbf{E}(u)$ на классе бэровских функций, обладающий следующими свойствами:*

Если $u\geqslant 0$, то $\mathbf{E}(u)$ равно либо неотрицательному числу, либо ∞. В случае произвольного u $\mathbf{E}(u)$ существует тогда и только тогда, когда либо $\mathbf{E}(u^+)$, либо $\mathbf{E}(u^-)$ конечно; при этом $\mathbf{E}(u)=\mathbf{E}(u^+)-\mathbf{E}(u^-)$. Функция u называется интегрируемой, если $\mathbf{E}(u)$ конечно. Далее

I) *если u — ступенчатая функция, то $\mathbf{E}(u)$ дается формулой (2.4);*

II) *условия (2.5)—(2.7) выполняются для всех интегрируемых функций;*

III) *(принцип монотонной сходимости), если $u_1\leqslant u_2\leqslant \ldots \ldots\to u$ и u_n интегрируемы, то $\mathbf{E}(u_n)\to\mathbf{E}(u)$.*

Сделав замену переменных $v_n=u_{n+1}-u_n$, принцип монотонной сходимости можно переформулировать в терминах рядов. Если v_n интегрируемы и $v_n\geqslant 0$, то

$$\sum \mathbf{E}(v_n)=\mathbf{E}\left(\sum v_n\right), \tag{2.8}$$

причем обе стороны равенства конечны или бесконечны одновременно. Отсюда, в частности, следует, что если $v\geqslant u\geqslant 0$ и $\mathbf{E}(u)=\infty$, то также $\mathbf{E}(v)=\infty$.

Что будет, если в (III) условие монотонности отбросить? Ответ на этот вопрос содержится в следующей важной лемме, имеющей широкую область применений.

Лемма Фату. *Если $u_n\geqslant 0$ и u_n интегрируемы, то*

$$\mathbf{E}(\liminf u_n)\leqslant \liminf \mathbf{E}(u_n). \tag{2.9}$$

В частности, если $u_n\to u$, то $\liminf \mathbf{E}(u_n)\geqslant \mathbf{E}(u)$.

Доказательство. Положим $v_n=u_n\cap u_{n+1}\cap\ldots$. Тогда $v_n\leqslant u_n$ и, следовательно,

$$\mathbf{E}(v_n)\leqslant \mathbf{E}(u_n).$$

Но (как мы видели в § 1) v_n монотонно стремится к $\liminf u_n$, так что $\mathbf{E}(v_n)$ стремится к левой части в (2.9) и лемма доказана. [Отметим, что обе стороны неравенства (2.9) могут быть равны ∞.] ▶

Приводимый ниже пример (г) показывает, что условие положительности не может быть отброшено, однако его можно за-

менить на формально более слабое условие, состоящее в том, что существует интегрируемая функция U, для которой $u_n \geqslant U$ (достаточно заменить u_n на $u_n - U$). Заменяя u_n на $-u_n$, устанавливаем следующий результат: если $u_n < U$ и $\mathbf{E}(U) < \infty$, то

$$\limsup \mathbf{E}(u_n) \leqslant \mathbf{E}(\limsup u_n). \tag{2.10}$$

Для сходящихся последовательностей левая часть неравенства (2.9) совпадает с правой частью неравенства (2.10), и мы получаем следующий важный

Принцип мажорированной сходимости. *Пусть u_n интегрируемы и $u_n \to u$ поточечно. Если существует интегрируемая функция U, такая, что $|u_n| \leqslant U$ при всех n, то и u интегрируема и $\mathbf{E}(u_n) \to \mathbf{E}(u)$.*

Эта теорема относится к единственному месту в лебеговской теории интегрирования, где наивные формальные действия могут привести к неверному результату. Необходимость условия $|u_n| \leqslant U$ иллюстрируется нижеследующим примером.

Пример. г) Рассмотрим гамма-плотности $u_n(x) = n^2 x e^{-nx}$ $(x > 0)$, отличающиеся друг от друга лишь масштабным параметром. Здесь $1 = \mathbf{E}(u_n) \to 1$, в то время как $u_n(x) \to 0$ при всех $x > 0$. Заменяя u_n на $-u_n$, видим, что неравенство Фату (2.9) может не выполняться для неположительных функций. ▶

Отметим без доказательства одно правило обычного интегрального исчисления, применимое и в более общей ситуации.

Теорема Фубини *о повторных интегралах. Если $u \geqslant 0$ — бэровская функция и F, G — вероятностные распределения, то с очевидной интерпретацией в случае расходящихся интегралов имеет место равенство*

$$\int_{-\infty}^{\infty} F\{dx\} \int_{-\infty}^{\infty} u(x, y) G\{dy\} = \int_{-\infty}^{\infty} G\{dy\} \int_{-\infty}^{\infty} u(x, y) F\{dx\}. \tag{2.11}$$

Здесь x и y можно понимать как точки в $\mathscr{R}^m$ и $\mathscr{R}^n$, и теорема содержит в себе утверждение о том, что оба внутренних интеграла суть бэровские функции. (Эта теорема применима к произведению произвольных пространств, и более общий ее вариант приводится в § 6.) ▶

Теорема об аппроксимации в среднем. *Для всякой интегрируемой функции u и любого $\varepsilon > 0$ можно найти ступенчатую функцию v, такую, что $\mathbf{E}(|u - v|) < \varepsilon$.*

Замечание об обозначениях. Обозначение $\mathbf{E}(u)$ подчеркивает зависимость от u и хорошо применимо в тех случаях, когда функция интервалов F фиксирована. Если F меняется или нужно подчеркнуть зависимость от F, то предпочтительнее интегральное обозначение (2.1). То же самое относится к интегралам по подмножеству A, поскольку интеграл от u по подмножеству A есть (по определению) интеграл от произведения $\mathbf{1}_A u$ по всему пространству

$$\int_A u(x)F\{dx\} = \mathbf{E}(\mathbf{1}_A u)$$

(при этом предполагается, конечно, что индикатор $\mathbf{1}_A$ есть бэровская функция). Обе части выписанного равенства означают в точности одно и то же, причем выражение, стоящее в левой части, подчеркивает зависимость от F. Когда $A=\overline{a, b}$ есть интервал, иногда предпочитают обозначение $\int_b^a$, однако чтобы сделать его точным, необходимо указать, включаются ли в интервал его конечные точки. Это можно сделать, используя запись $a+$ или $a-$.

В соответствии с намеченной в начале этого параграфа программой вероятность множества A полагается теперь, по определению равной $\mathbf{E}(\mathbf{1}_A)$ для всех множеств, для которых $\mathbf{1}_A$ есть бэровская функция. Для множеств, не обладающих этим свойством, вероятности не определяются. Теперь мы обсудим следствия из данного определения в более общей ситуации произвольного выборочного пространства. ▶

§ 3. Вероятностные меры и пространства

Основные свойства введенных нами в пространстве $\mathscr{R}^r$ вероятностей настолько удовлетворительны, что их можно использовать для аксиоматического построения вероятностей в произвольном выборочном пространстве. Конечно, никакого аналога интервалов в произвольном выборочном пространстве может не существовать, и действительное задание вероятностей меняется от случая к случаю (две типичные ситуации описаны в § 5 и 6). В общем случае необходимо обратить рассуждения § 2 и определять математические ожидания в терминах заданной вероятностной меры (этот подход описан в § 5). Понятие вероятностной меры (или распределения) является при этом первичным понятием.

Множество A в $\mathscr{R}^r$, индикатор которого $\mathbf{1}_A$ есть бэровская функция, называется *борелевским множеством*. Для борелевских множеств мы положим

$$\mathbf{P}\{A\}=\mathbf{E}(\mathbf{1}_A). \tag{3.1}$$

Поскольку интервалы являются борелевскими множествами[1]), это определение согласуется с требованием равенства $\mathbf{P}\{I\}=F\{I\}$ для всех интервалов I. Далее из (2.8) вытекает, что для любой счетной совокупности непересекающихся борелевских множеств A_n

$$\mathbf{P}\{\cup A_n\}=\Sigma\mathbf{P}\{A_n\}. \tag{3.2}$$

Свойство (3.2) есть свойство полной аддитивности. Рассуждения предыдущего параграфа показывают, что (3.1) является *единственно* возможным способом наделить борелевские множества вероятностями, если требовать выполнения условия (3.2) и равенств $\mathbf{P}\{I\}=F\{I\}$ для интервалов.

Мы наделили вероятностями лишь борелевские множества, однако класс борелевских множеств настолько обширен, что никакие стандартные операции не выводят за его пределы. Именно дополнение A' борелевского множества есть борелевское множество и то же самое относится к объединениям и пересечениям счетных совокупностей борелевских множеств. Отсюда и из § 1 вытекает, что верхний и нижний пределы любой последовательности $\{A_n\}$ борелевских множеств тоже являются борелевскими множествами. Семейства, обладающие этими свойствами, чрезвычайно широко используются, и мы формализуем понятие такого семейства, дав определение, применимое не только в $\mathscr{R}^n$, но и в *любом выборочном пространстве*.

Определение 1. *σ-алгеброй*[2]) *называется семейство* $\mathfrak{A}$ *подмножеств данного множества* $\mathfrak{S}$, *обладающее следующими*

---

[1]) Чтобы убедиться, что *открытый* интервал I есть борелевское множество, возьмем непрерывную функцию v, строго положительную на I и равную нулю вне I. Тогда $\overset{n}{V}\, v\to\mathbf{1}_I$. Замкнутый или полузамкнутый интервал I есть предел убывающей последовательности открытых интервалов I_n и, следовательно, $\mathbf{1}_I=\lim\mathbf{1}_{I_n}$, так что $\mathbf{1}_I$ — тоже бэровская функция. (Из того, что всякая непрерывная функция есть предел ступенчатых функций, вытекает, что класс бэровских функций можно охарактеризовать как наименьший класс, замкнутый относительно переходов к пределу и содержащий все ступенчатые функции.)

[2]) Если в этом определении слово «счетного» заменить на слово «конечного», то получим определение алгебры. σ-алгебру иногда называют борелевской алгеброй, однако этот термин лучше сохранить для наименьшей σ-алгебры, содержащей все открытые множества топологического пространства $\mathfrak{S}$. При $\mathfrak{S}=\mathscr{R}^r$ эта σ-алгебра совпадает с определенным нами семейством борелевских множеств.

свойствами: I) *для любого множества* A *из* $\mathfrak{A}$ *его дополнение* $A'=\mathfrak{S}-A$ *тоже принадлежит* $\mathfrak{A}$, II) *объединение* $\bigcup A_n$ *и пересечение* $\bigcap A_n$ *любого счетного семейства* $\{A_n\}$ *множеств из* $\mathfrak{A}$ *тоже принадлежат* $\mathfrak{A}$.

Короче говоря, σ-алгебра есть система множеств, замкнутая относительно взятия дополнений и образования счетных объединений и пересечений. Все пространство, будучи равным объединению A и A', всегда принадлежит σ-алгебре.

Примеры. *Наибольшей* σ-алгеброй в пространстве $\mathfrak{S}$ является семейство *всех* его подмножеств. Эта σ-алгебра хороша в случае дискретного выборочного пространства, однако в общем случае она слишком обширна, чтобы ее можно было использовать. Другим крайним случаем является тривиальная σ-алгебра, состоящая лишь из всего пространства и из пустого множества. В качестве нетривиального примера рассмотрим совокупность всех подмножеств действительной прямой, обладающих следующим свойством: если $x\in A$, то все точки $x\pm 1$, $x\pm 2$, ... тоже принадлежат A (периодические множества). Очевидно, семейство таких множеств образует σ-алгебру. ▶

Мы рассмотрели частный случай евклидова пространства $\mathscr{R}^1$ и показали, как можно определить вероятности на σ-алгебре борелевских множеств. Этот подход служит моделью для следующей более общей вероятностной конструкции.

Определение 2. *Вероятностной мерой* $\mathbf{P}$ *на* σ-*алгебре* $\mathfrak{A}$ *в пространстве* $\mathfrak{S}$ *называется функция, наделяющая каждое множество* $A\in\mathfrak{A}$ *числом* $\mathbf{P}\{A\}\geqslant 0$ *так, что* $\mathbf{P}\{\mathfrak{S}\}=1$ *и для любого счетного семейства взаимно непересекающихся множеств* A_n *из* $\mathfrak{A}$ *выполняется равенство* (3.2).

Вероятностным пространством[1]) *называется тройка* $(\mathfrak{S},\mathfrak{A},\mathbf{P})$, *состоящая из выборочного пространства* $\mathfrak{S}$, σ-*алгебры* $\mathfrak{A}$ *его подмножеств и вероятностной меры* $\mathbf{P}$ *на* $\mathfrak{A}$.

Естественно, введенное определение лишь закладывает основу общей конструкции, и в отдельных случаях нужно выби-

[1]) Условие $\mathbf{P}\{\mathfrak{S}\}=1$ вводится лишь для нормировки, и никаких существенных изменений не произойдет, если его заменить на $\mathbf{P}\{\mathfrak{S}\}<\infty$. В этом случае мы получаем *пространство с конечной мерой*. В теории вероятностей в связи с различными обстоятельствами появляются ситуации, когда $\mathbf{P}\{\mathfrak{S}\}<1$. В этом случае $\mathbf{P}$ называют несобственной (или) *дефектной вероятностной мерой*. Даже условие $\mathbf{P}\{\mathfrak{S}\}<\infty$ может быть ослаблено: можно потребовать лишь, чтобы $\mathfrak{S}$ было объединением счетного числа подмножеств $\mathfrak{S}_n$, для которых $\mathbf{P}\{\mathfrak{S}_n\}<\infty$. (Типичными примерами являются длина и площадь.) Такие меры называются *σ-конечными*.

рать подходящую σ-алгебру и строить на ней вероятностную меру. Эта процедура будет проиллюстрирована на примере «последовательности независимых случайных величин» (§ 6). Произвольное «вероятностное пространство» не всегда представляет из себя интересный объект, однако данное определение включает в себя все, что требуется для формального развития теории, следуя схеме первого тома, и было бы бесплодным обсуждать заранее типы выборочных пространств, которые могут встретиться в той или иной вероятностной ситуации.

Аппроксимация борелевских множеств интервалами. Вернемся к важному частному случаю евклидова пространства $\mathscr{R}^r$ и σ-алгебры борелевских множеств Исходя из вполне аддитивной функции интервалов, вероятности определяются по формуле $\mathbf{P}\{A\}=\mathbf{E}(\mathbf{1}_A)$ В силу теоремы об аппроксимации в среднем (§ 2) существует ступенчатая функция v, для которой $\mathbf{E}(|\mathbf{1}_A - v|)<\varepsilon/2$. Рассмотрим функцию w, такую, что $w(x)=1$ при $v(x)>1/2$ и $w(x)=0$ при $v(x)\leqslant 1/2$ Легко проверить, что $|\mathbf{1}_A(x)-w(x)|\leqslant 2|\mathbf{1}_A(x)-v(x)|$ при всех x, и, следовательно, $|\mathbf{P}\{A\}-\mathbf{E}(w)|<\varepsilon$. Но w есть ступенчатая функция, принимающая лишь значения 0 и 1, и поэтому $w=\mathbf{1}_B$ есть индикатор. Тем самым доказано, что *для каждого борелевского множества A можно найти множество B, образованное конечной совокупностью прямоугольников, для которого*

$$|\mathbf{P}\{A\} - \mathbf{P}\{B\}|<\varepsilon, \tag{3.3}$$

где $\varepsilon>0$ произвольно Значение этой теоремы станет более ясным, если обратиться к частному случаю площади плоской фигуры (рассматривая, например, внутренность треугольника или эллипса). Теорема эта свидетельствует о том, что борелевские множества представляют собой естественный объект, не слишком удаленный от интервалов.

§ 4. Случайные величины. Математические ожидания

(Теперь мы рассматриваем случай произвольного вероятностного пространства, введенного в предыдущем параграфе. Задача состоит в том, чтобы определить случайные величины по аналогии с бэровскими функциями в $\mathscr{R}^r$ и далее определить их математические ожидания в терминах соответствующей вероятностной меры.)

Случайные величины — обычный объект теории вероятностей. В дискретных выборочных пространствах вероятности определены на σ-алгебре всех подмножеств, и, следовательно, произвольная действительная функция на таком выборочном пространстве будет «случайной величиной». Вообще минимальное требование, предъявляемое к случайной величине,— это существование у нее функции распределения. Иными словами, если u — функция на вероятностном пространстве, то мы требуем, чтобы множество точек x, для которых $u(x)\leqslant t$, принадлежало бы нашей σ-алгебре $\mathfrak{A}$. Это дает возможность написать

$$F(t)=\mathbf{P}\{u\leqslant t\}. \tag{4.1}$$

Довольно неожиданно оказывается, что этого «невинного» требования достаточно для характеризации столь широкого класса функций, что никогда не возникает необходимости выходить за его пределы. Введем

Определение 1[1]). *Случайной величиной называется функция u на вероятностном пространстве, такая, что при любом действительном t множество точек x, для которых $u(x)\leqslant t$, принадлежит соответствующей σ-алгебре $\mathfrak{A}$. Функция F в* (4.1) *называется функцией распределения случайной величины u.*

Некоторые свойства случайных величин почти очевидны. Например, множество точек, на котором $u(x)<t$, есть объединение множеств, на которых $u(x)\leqslant t-n^{-1}$ при $n=1, 2, \ldots$, и, следовательно, принадлежит $\mathfrak{A}$. Множество, где $a<u(x)\leqslant b$, есть разность двух множеств из $\mathfrak{A}$ и поэтому также принадлежит $\mathfrak{A}$ и т. д. Покажем теперь, что вообще, грубо говоря, всякое множество, которое можно определить с помощью случайной величины, принадлежит $\mathfrak{A}$ и что все функции от u являются случайными величинами. Иными словами, никакие операции не выводят за пределы первоначальных рамок. Чтобы показать это, приведем одно необходимое и достаточное условие для того, чтобы функция была случайной величиной.

Определение 2. *Функция называется простой, если она принимает лишь счетное число значений $a_1, a_2, \ldots$, причем каждое значение — на множестве из σ-алгебры $\mathfrak{A}$ вероятностного пространства.*

В силу самого определения каждая простая функция является случайной величиной. С другой стороны, пусть заданы случайная величина u и число $\varepsilon>0$. Обозначим A_n множество всех точек x, для которых $(n-1)\varepsilon<u(x)\leqslant n\varepsilon$ (где $n=0, \pm1, \pm2, \ldots$). Эти множества, очевидно, не пересекаются, и мы

[1]) Пусть задана *произвольная* σ-алгебра $\mathfrak{A}$ множеств в произвольном пространстве. Функция f называется $\mathfrak{A}$ *-измеримой*, если при любом t множество точек x, для которых $f(x)\leqslant t$, принадлежит $\mathfrak{A}$. (Термин неудачен, поскольку на $\mathfrak{A}$ может быть не определено никакой меры.) Определение 1 означает, что функция на вероятностном пространстве является случайной величиной, если она измерима по отношению к соответствующей σ-алгебре. Следующее за определением 1 обсуждение без всяких изменений применимо к классу $\mathfrak{A}$-измеримых функций для любой σ-алгебры $\mathfrak{A}$ (этот факт используется при рассмотрении условных математических ожиданий в гл. V. 10). Приведенная в тексте ограниченная формулировка выбрана ради простоты, а также чтобы не уклоняться от основной цели.

можем определить две простые функции $\underline{\sigma}$ и $\overline{\sigma}$, положив $\underline{\sigma}(x)=(n-1)\varepsilon$ и $\overline{\sigma}=n\varepsilon$ при $x\in A_n$. Тогда

$$\underline{\sigma}(x)<u(x)\leqslant\overline{\sigma}(x) \text{ и } \overline{\sigma}(x)-\underline{\sigma}(x)=\varepsilon. \tag{4.2}$$

Таким образом, каждую случайную величину можно аппроксимировать простой функцией с ошибкой, не превосходящей ε. Выбирая $\varepsilon=1, 1/2, 1/3, \ldots$, получим две последовательности простых функций σ_n, для которых $\sigma_n \to u$ равномерно. Обратно, рассмотрим неубывающую последовательность простых функций σ_n и положим $\lim \sigma_n(x)=v(x)$. Пусть $v(x)<\infty$ при всех x. Обозначим S_n и S множества точек, для которых $\sigma_n(x)>t$ и $v(x)>t$ соответственно. Тогда $S_n \to S$ и, следовательно, S принадлежит $\mathfrak{A}$, так что v есть случайная величина. Тем самым нами доказана

Лемма. Функция и является случайной величиной тогда и только тогда, когда она есть поточечный предел монотонной последовательности простых функций. Для каждой случайной величины существуют две простые функции, удовлетворяющие (4.2) *с любым заданным* $\varepsilon>0$.

Последнее рассуждение показывает, что вообще, если монотонная последовательность $\{v_n\}$ случайных величин сходится к конечному пределу v, то этот предел тоже есть случайная величина. То же самое верно относительно конечных пределов произвольных последовательностей $\{v_n\}$ случайных величин, поскольку они могут быть представлены как пределы монотонных последовательностей (§ 1). Далее если $W(x_1, \ldots, x_n)$ — непрерывная функция действительных переменных $x_1, \ldots, x_n$ и $\sigma_1, \ldots, \sigma_n$ — простые функции в выборочном пространстве $\mathfrak{S}$, то $W(\sigma_1, \ldots, \sigma_n)$ тоже есть простая функция. Простой переход к пределу показывает, что *всякая непрерывная функция от случайных величин есть случайная величина.* Таким образом, класс случайных величин замкнут относительно непрерывных операций и относительно переходов к пределу. В частности, суммы и произведения случайных величин снова будут случайными величинами. Это утверждение, уже достаточное для наших целей, может быть следующим образом усилено. Рассмотрим функцию W от n действительных переменных, обладающую тем свойством, что $W(u_1, \ldots, u_n)$ есть случайная величина для любого набора случайных величин $u_1, \ldots, u_n$ на $\mathfrak{S}$. Класс таких функций W замкнут относительно переходов к пределу и содержит все непрерывные функции на $\mathscr{R}^n$. Он содержит, таким образом, все бэровские функции на $\mathscr{R}^n$, и нами доказана следующая

Теорема. *Класс случайных величин замкнут относительно переходов к пределу. Всякая бэровская функция от конечного числа случайных величин тоже есть случайная величина.*

Пример. *Пусть выборочное пространство есть $\mathscr{R}^r$, σ-алгебра $\mathfrak{A}$ совпадает с классом борелевских множеств и пусть* $\mathbf{P}\{A\}=\mathbf{E}(\mathbf{1}_A)$. (Предполагается, что математическое ожидание определено как в § 2.) В этом случае *классы случайных величин, бэровских функций и измеримых по Борелю функций совпадают.* ▶

Математические ожидания

В частной ситуации, рассмотренной в только что приведенном примере, сначала вводились математические ожидания, а вероятности определялись затем как математические ожидания индикаторов. Этот подход удобен для введения различных понятий и методов, однако он неприменим, когда исходным является произвольное вероятностное пространство. К счастью, прямое определение математических ожиданий в терминах вероятностей весьма несложно и единственно. Для простой функции σ, принимающей значения $a_1, a_2, \ldots$ на множества $A_1, A_2, \ldots$, в любом случае мы должны иметь

$$\mathbf{E}(\sigma)=\Sigma a_n \mathbf{P}\{A_n\}, \tag{4.3}$$

если, конечно, ряд сходится абсолютно. [В противном случае $\mathbf{E}(\sigma)$ не определяется.] Каждая случайная величина u есть равномерный предел последовательности простых функций σ_n. Это означает, что $|\sigma_n(x)-\sigma_m(x)|<\varepsilon$ при всех достаточно больших n и m, и поэтому либо $\mathbf{E}(\sigma_n)$ существует для всех достаточно больших n, либо при всех достаточно больших n $\mathbf{E}(\sigma_n)$ не существует. В первом случае $|\mathbf{E}(\sigma_n)-\mathbf{E}(\sigma_m)|<\varepsilon$ и, следовательно, $\{\mathbf{E}(\sigma_n)\}$ есть фундаментальная последовательность с конечным пределом. Поскольку две аппроксимирующие последовательности можно объединить в одну, то этот предел не зависит от аппроксимирующей последовательности и можно положить

$$\mathbf{E}(u)=\lim \mathbf{E}(\sigma_n). \tag{4.4}$$

Иными словами, математическое ожидание *u либо не существует, либо есть общий предел для* $\mathbf{E}(\sigma_n)$ *по всем последовательностям простых функций, сходящихся равномерно к u.*

Определение $\mathbf{E}(u)$ с вычислительной точки зрения может несколько прояснить ситуацию. Пусть $\varepsilon>0$ задано. Обозначим, по-прежнему, через A_n, $n=0, \pm 1, \pm 2, \ldots$ множество точек x, для которых $(n-1)\varepsilon<u(x)\leqslant n\varepsilon$. Рассмотрим опять простые функции $\underline{\sigma}$ и $\bar{\sigma}$, принимающие на A_n значения $(n-1)\varepsilon$ и $n\varepsilon$ со-

ответственно. Тогда имеет место (4.2) и, следовательно, либо оба математических ожидания

$$\mathbf{E}(\underline{\sigma}_n) = \varepsilon \sum (n-1)\,\mathbf{P}\{A_n\}, \quad \mathbf{E}(\overline{\sigma}_n) = \varepsilon \sum n\mathbf{P}\{A_n\} \tag{4.5}$$

конечны, либо оба не существуют. Во втором из этих случаев $\mathbf{E}(u)$ не определено, в то время как в первом случае

$$\mathbf{E}(\underline{\sigma}_n) \leqslant \mathbf{E}(u) \leqslant \mathbf{E}(\overline{\sigma}_n). \tag{4.6}$$

Тем самым мы получаем $\mathbf{E}(u)$ с ошибкой, не превосходящей ε.

Случайная величина u отображает выборочное пространство $\mathfrak{S}$ в действительную прямую, и ее функция распределения (4.1) порождает вероятностную меру на действительной прямой. В соответствии с этим случайную величину u можно рассматривать либо как функцию на выборочном пространстве $\mathfrak{S}$, либо как координатную переменную на прямой, наделенную вероятностным распределением F. Далее если F есть распределение u и I_n — интервал $(n-1)\varepsilon < t \leqslant n\varepsilon$, то

$$F\{I_n\} = \mathbf{P}\{(n-1)\varepsilon < u \leqslant n\varepsilon\} = \mathbf{P}\{A_n\}. \tag{4.7}$$

Простые функции $\underline{\sigma}_n$ и $\overline{\sigma}_n$ отображаются в простые функции на прямой, и, очевидно, математическое ожидание координатной переменной (определенное в § 2) удовлетворяет неравенствам (4.6). Таким образом,

$$\mathbf{E}(u) = \int_{-\infty}^{\infty} tF\{dt\}. \tag{4.8}$$

Иными словами, *математическое ожидание случайной величины можно (корректно) определить либо непосредственно в исходном вероятностном пространстве, либо в терминах ее функции распределения.* (В **1**, гл. IX имеется аналогичное замечание для случая дискретных выборочных пространств.)

§ 5. Теорема о продолжении

При построении вероятностных пространств вероятности обычно определены a priori на не очень богатом классе множеств и требуется подходящим образом расширить их область определения. Например, при рассмотрении в **1** бесконечных последовательностей испытаний и рекуррентных событий были заданы вероятности всех событий, зависящих от конечного числа испытаний, и эту область определения вероятностей требовалось расширить так, чтобы включить такие события, как разорение, возвращение и полное вырождение. Аналогично при построении в § 2 мер в $\mathscr{R}^r$ сначала наделялись вероятностями $F\{I\}$

интервалы, а затем область определения продолжалась до класса всех борелевских множеств. Такое продолжение возможно благодаря теореме, имеющей значительно более широкую область применения, и на которой основываются многие конструкции вероятностных пространств. Схема рассуждения состоит в следующем.

Аддитивность F позволяет однозначно определить

$$F\{A\}=\Sigma F\{I_k\} \tag{5.1}$$

для любого множества A, являющегося объединением *конечного числа непересекающихся интервалов* I_k. Эти множества образуют *алгебру* $\mathfrak{A}_0$ (т. е. объединения и пересечения конечного числа множеств из $\mathfrak{A}_0$, а также дополнения множеств из $\mathfrak{A}_0$ тоже принадлежат $\mathfrak{A}_0$). Начиная с этого пункта, природа пространства $\mathscr{R}^r$ не играет никакой роли и можно рассматривать произвольную алгебру $\mathfrak{A}_0$ множеств в произвольном пространстве Ξ. Всегда существует наименьшая алгебра $\mathfrak{A}$ множеств, содержащая $\mathfrak{A}_0$ и замкнутая относительно *счетных* объединений и пересечений. Иными словами, существует наименьшая σ-алгебра $\mathfrak{A}^r$ содержащая $\mathfrak{A}_0$ (см. определение 3.1). При построении мер в $\mathscr{R}^r$ σ-алгебра $\mathfrak{A}$ совпадала с σ-алгеброй всех борелевских множеств. Расширение области определения вероятностей $\mathfrak{A}_0$ до $\mathfrak{A}$ основывается на следующей общей теореме.

Теорема о продолжении. *Пусть $\mathfrak{A}_0$ — произвольная алгебра множеств в некотором пространстве Ξ. Пусть F — определенная на $\mathfrak{A}_0$ функция множеств, такая, что $F\{A\}\geqslant 0$ при всех $A\in\mathfrak{A}_0$, $F\{\Xi\}=1$, и, кроме того, соотношение* (5.1) *выполняется для любого разбиения A на счетное число непересекающихся множеств $I_k\in\mathfrak{A}_0$.*

Тогда существует единственное продолжение F до счетно-аддитивной функции множеств (т. е. до вероятностной меры) на наименьшей σ-алгебре $\mathfrak{A}$, содержащей $\mathfrak{A}_0$.

В следующем параграфе будет приведен типичный пример применения этой теоремы. Теперь мы обратимся к более общему и гибкому варианту теоремы о продолжении, ближе примыкающему к рассмотрениям § 2 и 3. Мы исходили из математического ожидания (2.4) для ступенчатых функций (т. е. функций, принимающих лишь конечное число значений, причем каждое из значений — на некотором интервале). Область определения этого математического ожидания была затем продолжена с довольно бедного класса ступенчатых функций на более широкий класс, включающий все ограниченные бэровские функции. Указанное продолжение непосредственно приводит к интегралу Ле-

бега — Стильтьеса, и мера множества A получается как математическое ожидание его индикатора $\mathbf{1}_A$. Наметим соответствующую абстрактную схему.

Рассмотрим вместо алгебры множеств $\mathfrak{A}_0$ класс функций $\mathfrak{B}_0$, замкнутый относительно линейных комбинаций и операций $\cap$ и $\cup$. Иными словами, предполагается, что если функции u_1 и u_2 принадлежат $\mathfrak{B}_0$, то функции

$$\alpha_1 u_1 + \alpha_2 u_2, \quad u_1 \cap u_2, \quad u_1 \cup u_2 \tag{5.2}$$

тоже принадлежат $\mathfrak{B}_0$[1]). Отсюда, в частности, следует, что всякая функция u из $\mathfrak{B}_0$ может быть записана как разность двух неотрицательных функций, именно $u = u^+ - u^-$, где $u^+ = u \cup 0$ и $u^- = u \cap 0$. Под *линейным функционалом* на $\mathfrak{B}_0$ понимается сопоставление значений $\mathbf{E}(u)$ всем функциям из $\mathfrak{B}_0$ с выполнением условия

$$\mathbf{E}(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 \mathbf{E}(u_1) + \alpha_2 \mathbf{E}(u_2). \tag{5.3}$$

Функционал называется *положительным*, если $\mathbf{E}(u) \geqslant 0$ при $u \geqslant 0$. *Нормой* $\mathbf{E}$ называется верхняя грань значений $\mathbf{E}(|u|)$ по всем функциям $u \in \mathfrak{B}_0$, удовлетворяющим условию $|u| \leqslant 1$. Если функция, равная всюду 1, принадлежит $\mathfrak{B}_0$, то норма $\mathbf{E}$ равна $\mathbf{E}(1)$. Наконец, скажем, что $\mathbf{E}$ счетно-аддитивен на $\mathfrak{B}_0$, если

$$\mathbf{E}\left(\sum_1^\infty u_k\right) = \sum_1^\infty \mathbf{E}(u_k) \tag{5.4}$$

всякий раз, когда $\sum u_k$ принадлежит $\mathfrak{B}_0$[2]). Условие счетной аддитивности *эквивалентно* следующему: если $\{v_n\}$ — последовательность функций из $\mathfrak{B}_0$, сходящаяся *монотонно* к нулю, то[3])

$$\mathbf{E}(v_n) \to 0. \tag{5.5}$$

Для каждого заданного класса функций $\mathfrak{B}_0$ существует *наименьший класс $\mathfrak{B}$, содержащий $\mathfrak{B}_0$ и замкнутый относительно переходов к пределу в смысле поточечной сходимости.* [Класс $\mathfrak{B}$ автоматически замкнут относительно операций (5.2).] Проведем теперь соответствующую формулировку теоремы о продолже-

---

[1]) Наши предположения равносильны требованию, чтобы класс $\mathfrak{B}_0$ был линейной структурой.

[2]) Здесь следует добавить требование $u_k \geqslant 0$, иначе последующее утверждение неверно. — *Прим. перев.*

[3]) Для доказательства эквивалентности (5.4) и (5.5) достаточно рассмотреть случай $u_k \geqslant 0$, $v_k \geqslant 0$. Тогда (5.4) следует из (5.5), если положить $v_n = u_{n+1} + u_{n+2} + \ldots$, и (5.5) вытекает из (5.4), если положить $u_k = v_k - v_{k+1}$ (при этом $\Sigma u_k = v_1$).

нии[1]). *Каждый положительный счетно-аддитивный линейный функционал нормы* 1 *на* $\mathfrak{B}_0$ *может быть единственным образом продолжен до положительного счетно-аддитивного линейного функционала нормы* 1 *на всех ограниченных (и многих неограниченных) функциях из* $\mathfrak{B}$.

В качестве примера применения этой теоремы докажем следующий важный результат.

Теорема Ф. Рисса о представлении[2]). *Пусть* $\mathbf{E}$ — *положительный линейный функционал нормы* 1 *на классе всех непрерывных функций на* $\mathscr{R}^r$, *обращающихся в нуль на бесконечности*[3]). *Существует мера* $\mathbf{P}$ *на* σ-*алгебре борелевских множеств, для которой* $\mathbf{P}\{\mathscr{R}^r\}=1$ *и такая, что* $\mathbf{E}(u)$ *совпадает с интегралом от* u *по* $\mathbf{P}$.

Иными словами, определенные нами интегралы дают представление наиболее общих положительных линейных функционалов.

Доказательство. В основе доказательства лежит тот факт, что если последовательность обращающихся в нуль на бесконечности непрерывных функций монотонно сходится к нулю, то сходимость эта автоматически *равномерна*. Пусть $v_n \geqslant 0$, и обозначим $\|v_n\| = \max v_n(x)$. Тогда $\mathbf{E}(v_n) \leqslant \|v_n\|$, и поэтому для $\mathbf{E}$ выполнено условие счетной аддитивности (5.5). В силу теоремы о продолжении $\mathbf{E}$ можно продолжить на все ограниченные бэровские функции. Положив $\mathbf{P}\{A\}=\mathbf{E}(\mathbf{1}_A)$, полу-

---

[1]) Основная идея доказательства (восходящая к Лебегу) проста и остроумна. Нетрудно видеть, что если две последовательности $\{u_n\}$ и $\{u'_n\}$ функций из $\mathfrak{B}_0$ сходятся *монотонно* к одному и тому же пределу u, то последовательности $\mathbf{E}(u_n)$ и $\mathbf{E}(u'_n)$ тоже сходятся к одному и тому же пределу. Для таких монотонных пределов u можно, следовательно, определить $\mathbf{E}(u) = \lim \mathbf{E}(u_n)$. Рассмотрим теперь класс $\mathfrak{B}_1$ функций u, для которых при любом $\varepsilon>0$ существуют две функции $\underline{u}$ и $\bar{u}$ либо принадлежащие $\mathfrak{B}_0$, либо являющиеся монотонными пределами последовательностей из $\mathfrak{B}_0$, и такие, что $\underline{u}<u<\bar{u}$, $\mathbf{E}(\bar{u}) - \mathbf{E}(\underline{u})<\varepsilon$. Класс $\mathfrak{B}_1$ замкнут относительно переходов к пределам, и для функций из $\mathfrak{B}_1$ значение $\mathbf{E}(u)$ определяется очевидным образом, поскольку должно быть $\mathbf{E}(\underline{u}) \leqslant \mathbf{E}(u) \leqslant \mathbf{E}(\bar{u})$.

Замечательной особенностью этого рассуждения является следующее обстоятельство: обычно класс $\mathfrak{B}_1$ больше класса $\mathfrak{B}$ и простое доказательство оказывается возможным благодаря тому, что доказывается больше, чем требуется. (Относительно сравнения классов $\mathfrak{B}$ и $\mathfrak{B}_1$ см. § 7.)

[2]) Эта теорема справедлива для произвольных локально компактных пространств.

[3]) u обращается в нуль на бесконечности, если для любого $\varepsilon>0$ существует шар (компактное множество), вне которого $|u(x)|<\varepsilon$.

чаем меру на σ-алгебре борелевских множеств. Как мы видели, по заданной мере $\mathbf{P}\{A\}$ интеграл Лебега — Стильтьеса однозначно определяется с помощью двойного неравенства (4.6). Следовательно, интеграл Лебега — Стильтьеса по $\mathbf{P}$ от обращающейся в нуль на бесконечности непрерывной функции u совпадает со значением на u заданного функционала $E(u)$. ▶

§ 6. Произведения пространств. Последовательности независимых случайных величин

Понятие прямого произведения пространств (1, гл. V, 4) является весьма важным в теории вероятностей и используется всякий раз, когда речь идет о повторных испытаниях. Представление точки плоскости $\mathscr{R}^2$ посредством двух координат означает, что $\mathscr{R}^2$ рассматривается как прямое произведение двух своих осей. Обозначим две координатные переменные через $\mathbf{X}$ и $\mathbf{Y}$. Рассматриваемые как функции на плоскости они представляют собой бэровские функции, и если на σ-алгебре борелевских множеств в $\mathscr{R}^2$ определена вероятностная мера $\mathbf{P}$, то существуют две функции распределения $\mathbf{P}\{\mathbf{X} \leqslant x\}$ и $\mathbf{P}\{\mathbf{Y} \leqslant y\}$. Эти две функции распределения индуцируют вероятностные меры на обеих осях, называемые частными (маргинальными) распределениями (или проекциями). Здесь у нас в качестве первичного понятия появляется плоскость, однако часто более естественной является противоположная ситуация. Например, когда говорят о двух независимых случайных величинах с заданными распределениями, два частных распределения являются первичным понятием и вероятности на плоскости получаются из них с помощью «правила умножения». Общий случай ничуть не сложнее случая плоскости.

Рассмотрим теперь два произвольных вероятностных пространства. Иными словами, нам заданы два выборочных пространства $\mathfrak{E}^{(1)}$ и $\mathfrak{E}^{(2)}$, две σ-алгебры $\mathfrak{A}^{(1)}$ и $\mathfrak{A}^{(2)}$ подмножеств $\mathfrak{E}^{(1)}$ и $\mathfrak{E}^{(2)}$ соответственно и определенные на $\mathfrak{A}^{(1)}$ и $\mathfrak{A}^{(2)}$ вероятностные меры $\mathbf{P}^{(1)}$ и $\mathbf{P}^{(2)}$. Прямое произведение $(\mathfrak{E}^{(1)}, \mathfrak{E}^{(2)})$ есть множество всех упорядоченных пар $(x^{(1)}, x^{(2)})$, где $x^{(i)}$ — точка пространства $\mathfrak{E}^{(i)}$. Рассмотрим в этом произведении множества, являющиеся «прямоугольниками», т. е. прямые произведения $(A^{(1)}, A^{(2)})$ множеств $A^{(1)} \in \mathfrak{A}^{(i)}$. Множества такого вида мы наделяем вероятностями, согласно правилу умножения:

$$\mathbf{P}\{A^{(1)}, A^{(2)}\} = \mathbf{P}^{(1)}\{A^{(1)}\}\mathbf{P}^{(2)}\{A^{(2)}\}. \tag{6.1}$$

Далее множества, представляющие собой конечные объединения непересекающихся прямоугольников, образуют алгебру $\mathfrak{A}_0$, на

которой с помощью (6.1) единственным образом определяется счетно-аддитивная [1]) функция. Применяя теперь теорему о продолжении, можем утверждать, что *на наименьшей содержащей все прямоугольники σ-алгебре существует единственная вероятностная мера* **P**, *такая, что вероятности прямоугольников задаются правилом о произведении* (6.1). *Эта наименьшая, содержащая все прямоугольники σ-алгебра обозначается* $\mathfrak{A}^{(1)} \times \mathfrak{A}^{(2)}$, *а так определенная мера на ней называется произведением мер.*

Конечно, на произведении пространств можно определить и другие вероятности, например, определить их в терминах условных вероятностей. Рассматриваемая σ-алгебра множеств $\mathfrak{A}$ всегда будет не меньше $\mathfrak{A}^{(1)} \times \mathfrak{A}^{(2)}$, и в то же время за пределы $\mathfrak{A}^{(1)} \times \mathfrak{A}^{(2)}$ приходится выходить весьма редко. Ниже в рассуждении о случайных величинах предполагается, что $\mathfrak{A} = \mathfrak{A}^{(1)} \times \mathfrak{A}^{(2)}$.

Понятие случайной величины (измеримой функции) связано с соответствующей σ-алгеброй, и при рассмотрении произведений пространств мы должны различать случайные величины на произведении и на пространствах $\mathfrak{S}^{(1)}$ и $\mathfrak{S}^{(2)}$. Весьма приятным является тот факт, что отношение между этими тремя классами случайных величин весьма простое. Пусть u и v — случайные величины на $\mathfrak{S}^{(1)}$ и $\mathfrak{S}^{(2)}$ и рассмотрим в произведении пространств функцию w, которая в точке $(x^{(1)}, x^{(2)})$ принимает значение

$$w(x^{(1)}, x^{(2)}) = u(x^{(1)})\, v(x^{(2)}). \tag{6.2}$$

Покажем, что *класс случайных величин на произведении пространств* $(\mathfrak{S}^{(1)}, \mathfrak{S}^{(2)})$ *совпадает с минимальным классом конечнозначных функций, замкнутым относительно поточечных перехо-*

---

[1]) *Идея* доказательства состоит в следующем. Пусть $(A^{(1)}, A^{(2)})$ есть объединение непересекающихся прямоугольников $(A_p^{(1)}, A_p^{(2)})$. Рассмотрим всевозможные пересечения $A_j^{(i)} \cap A_k^{(i)}$ (где $i = 1, 2$). Среди этих пересечений можно выбрать последовательность $B_1^{(i)}, B_2^{(i)}, \ldots$ непересекающихся множеств, такую, что каждое множество $A_p^{(i)}$ есть объединение некоторых $B_k^{(i)}$. Но тогда $(A^{(1)}, A^{(2)})$ есть объединение непересекающихся прямоугольников $(B_j^{(1)}, B_k^{(2)})$ и

$$\sum_p \mathbf{P}^{(1)}\{A_p^{(1)}\}\, \mathbf{P}^{(2)}\{A_p^{(2)}\} = \sum_{j,k} \mathbf{P}^{(1)}\{B_j^{(1)}\}\, \mathbf{P}^{(2)}\{B_k^{(2)}\}.$$

Последняя двойная сумма равна произведению

$$\sum_j \mathbf{P}^{(1)}\{B_j^{(1)}\} = \mathbf{P}\{A^{(1)}\} \text{ и } \sum_k \mathbf{P}^{2}\{B_k^{(2)}\} = \mathbf{P}\{A^{(2)}\}.$$

дов к пределу и содержащим все линейные комбинации функций вида (6.2).

Ясно, во-первых, что каждый множитель в правой части (6.2) является случайной величиной и в том случае, когда он рассматривается как функция на произведении пространств. Следовательно, w есть случайная величина, и поэтому класс случайных величин на $(\mathfrak{S}^{(1)}, \mathfrak{S}^{(2)})$ не меньше минимального класса, о котором говорилось выше. С другой стороны, случайные величины образуют наименьший класс, замкнутый относительно переходов к пределу и содержащий все линейные комбинации индикаторов прямоугольников. Эти индикаторы имеют вид (6.2), и, следовательно, класс случайных величин не больше минимального класса, о котором говорилось выше.

Частный случай произведения двух пространств $\mathscr{R}^m$ и $\mathscr{R}^n$ с вероятностными мерами F и G неявно встречался в связи с теоремой Фубини (2.11) о повторных интегралах. Теперь мы можем утверждать справедливость более общей теоремы, относящейся не только к $\mathscr{R}^r$.

Теорема Фубини для произведений мер. *Интеграл от любой неотрицательной бэровской функции и по произведению мер равен повторным интегралам в формуле* (2.12). (При этом имеется в виду, что интегралы могут расходиться. Теорема очевидна для простых функций, а в общем случае получается с помощью повторной аппроксимации.) Обобщение на случай произведения трех или более пространств вполне очевидно и не требует специальных пояснений.

Обратимся теперь к проблеме *бесконечной последовательности случайных величин*, с которой мы уже встречались в 1 в связи с бесконечными последовательностями испытаний Бернулли, случайными блужданиями, рекуррентными событиями и т. д., а также в гл. III в связи с нормальными случайными процессами. Проблемы не возникает, когда имеется бесконечное множество случайных величин, определенных на заданном вероятностном пространстве. Например, действительная прямая с нормальным распределением есть вероятностное пространство и $\{\sin nx\}$ есть бесконечная последовательность определенных на нем случайных величин. Нас интересует только такая ситуация, когда вероятности должны быть определены в терминах заданных случайных величин. Более точно проблема состоит в следующем.

Пусть $\mathscr{R}^\infty$ — пространство, точками которого являются бесконечные последовательности действительных чисел $(x_1, x_2, \ldots)$ (т. е. $\mathscr{R}^\infty$ есть счетное прямое произведение действительных

прямых). Обозначим n-ю координатную переменную через $\mathbf{X}_n$ ($\mathbf{X}_n$ есть функция на $\mathscr{R}^\infty$, принимающая в точке $x=(x_1, x_2, \ldots)$ значение x_n). Предположим, что нам заданы вероятностные распределения для $\mathbf{X}_1$, $(\mathbf{X}_1, \mathbf{X}_2)$, $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$, ... и требуется определить соответствующие вероятности в $\mathscr{R}^\infty$. Конечно, заданные вероятности должны быть взаимно согласованы в том смысле, что распределения $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ являются частными распределениями для $(\mathbf{X}_1, \ldots, \mathbf{X}_{n+1})$ и т. д.

Формализуем теперь интуитивное понятие «события, определяемого исходом конечного множества испытаний». Скажем, что *множество A в $\mathscr{R}^\infty$ зависит лишь от первых r координат, если существует борелевское множество A_r в $\mathscr{R}^r$, такое, что $x=(x_1, x_2, \ldots)$ принадлежит A тогда и только тогда, когда $(x_1, \ldots, x_r)$ принадлежит A_r.* В теории вероятностей часто встречается ситуация, когда вероятности таких множеств заданы, и задача состоит в расширении этой области определения. Приведем без доказательства принадлежащую А. П. Колмогорову фундаментальную теорему (доказанную им в несколько большей общности), впервые появившуюся в его ставшем теперь уже классическим аксиоматическом обосновании теории вероятностей (1933). Эта работа предвосхитила и стимулировала развитие современной теории меры.

Теорема 1. *Согласованную систему вероятностных распределений величин $\mathbf{X}_1$, $(\mathbf{X}_1, \mathbf{X}_2)$, $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$, ... можно единственным способом продолжить до вероятностной меры на наименьшей σ-алгебре $\mathfrak{A}$ множеств в $\mathscr{R}^\infty$, содержащей все множества, зависящие лишь от конечного числа координат*[1]).

Весьма важным является то обстоятельство, что все вероятности определяются последовательными переходами к пределам, исходя из конечномерных множеств. *Каждое множество A из $\mathfrak{A}$ можно аппроксимировать конечномерными множествами* в следующем смысле. Для любого $\varepsilon>0$ при некотором n существует зависящее лишь от первых n координат множество A_n, такое, что

$$\mathbf{P}\{A - A \cap A_n\}<\varepsilon, \quad \mathbf{P}\{A_n - A \cap A_n\}<\varepsilon. \tag{6.3}$$

Иными словами, множество тех точек, которые принадлежат либо A, либо A_n, но не принадлежат обоим множествам A и A_n, имеет вероятность, меньшую 2ε. Отсюда следует, что можно так выбрать множества A_n, чтобы

$$\mathbf{P}\{A_n\} \to \mathbf{P}\{A\}. \tag{6.4}$$

[1]) Эта теорема справедлива и в более общем случае произведения локально компактных пространств. Например, она справедлива, когда переменные $\mathbf{X}_n$ являются векторными переменными (т. е. значения их — точки в $\mathscr{R}^r$).

Теорема 1 позволяет говорить о *бесконечной последовательности взаимно независимых случайных величин с произвольными заданными распределениями.* Такие последовательности уже встречались в первом томе, однако нам приходилось определять интересующие нас вероятности с помощью специальных переходов к пределу, в то время как теорема 1 является общим результатом, охватывающим все случаи. Это обстоятельство хорошо иллюстрируется следующими двумя важными теоремами, первая из которых принадлежит А. Колмогорову (1933), а вторая — Е. Хьюитту и Л. Дж. Сэвиджу (1955). Обе они являются типичными вероятностными предложениями и играют центральную роль во многих рассуждениях.

Теорема 2. *(Закон нуля или единицы для остаточных событий.) Предположим, что случайные величины* $\mathbf{X}_k$ *взаимно независимы и что при каждом* n *событие* A *не зависит от*[1]) $\mathbf{X}_1, \ldots, \mathbf{X}_n$. *Тогда или* $\mathbf{P}\{A\}=0$, *или* $\mathbf{P}\{A\}=1$.

Доказательство. Вообще говоря, случайные величины $\mathbf{X}$[2]) могут быть определены на произвольном вероятностном пространстве, однако они отображают это пространство в произведение прямых $\mathscr{R}^\infty$, в котором они играют роль координатных переменных. Следовательно, не уменьшая общности, можно ограничиться рассмотрением ситуации, описанной в настоящем параграфе. В обозначениях, использованных в (6.3), множества A и A_n независимы, и поэтому из (6.3) вытекает, что $\mathbf{P}\{A\}=\mathbf{P}^2\{A\}$. ▶

Пример. а) Ряд $\Sigma\mathbf{X}_n$ сходится с вероятностью единица или нуль. Точно так же множество точек, в которых $\limsup \mathbf{X}_n=\infty$, имеет вероятность либо единица, либо нуль. ▶

Теорема 3. *(Закон нуля или единицы для симметричных событий.) Предположим, что случайные величины* $\mathbf{X}_k$ *взаимно независимы и одинаково распределены. Если множество* A *инвариантно относительно конечных перестановок координат*[3]), *то либо* $\mathbf{P}\{A\}=0$, либо $\mathbf{P}\{A\}=1$.

---

[1]) Более точно A не зависит от любого события, определяемого в терминах $\mathbf{X}_1, \ldots, \mathbf{X}_n$. Иными словами, индикатор A есть случайная величина, независимая от $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

[2]) Предполагается, конечно, что A определяется последовательностью $\mathbf{X}_1, \mathbf{X}_2, \ldots$. — *Прим. перев.*

[3]) Более точно предполагается, что если a-$(a_1, a_2, \ldots)$ — точка из A и n_1, n_2 — два произвольных натуральных числа, то A содержит также точку, у которой n_1-я координата есть a_{n_2} n_2-я координата равна a_{n_1}, а остальные координаты — те же, что и у a. Это условие автоматически распространяется на перестановки, включающие k координат.

Доказательство. Как и в доказательстве теоремы 2, мы рассматриваем $\mathbf{X}_k$ как координатные величины и под множествами A_n понимаем то же, что и в (6.3). Пусть B_n — множество, полученное из A_n обращением порядка первых $2n$ координат, в то время как остальные координаты остаются без изменения. В силу сделанных предположений соотношения (6.3) остаются справедливыми при замене A_n на B_n. Отсюда вытекает, что множество точек, принадлежащих либо A, либо $A_n \cap B_n$, но не принадлежащих обоим множествам A и $A_n \cap B_n$, имеет вероятность, меньшую 4ε. Следовательно,

$$\mathbf{P}\{A_n \cap B_n\} \to \mathbf{P}\{A\}. \tag{6.5}$$

Далее A_n зависит лишь от первых n координат, и поэтому B_n зависит лишь от координат с номерами $n+1, \ldots, 2n$. Таким образом, A_n и B_n независимы, что вместе с (6.5) приводит к равенству $\mathbf{P}\{A\} = \mathbf{P}^2\{A\}$. ▶

Пример. б) Положим $\mathbf{S}_n = \mathbf{X}_1 + \ldots + X_n$, и пусть $A = \{\mathbf{S}_n \in I \text{ б. ч.}\}$, где I — *произвольный интервал на прямой*. Событие A инвариантно относительно конечных перестановок. [По поводу обозначений см. пример (1, а).] ▶

§ 7. Нулевые множества. Пополнение

Множеством вероятности нуль обычно можно пренебрегать, и две случайные величины, различающиеся лишь на таком нулевом множестве, «практически совпадают». Выражаясь более формально, их называют *эквивалентными*. Это означает, что если изменить определение случайной величины на нулевом множестве, то все вероятностные соотношения сохраняются без изменения и, следовательно, случайную величину можно считать неопределенной на нулевом множестве. Типичным примером является время осуществления первого рекуррентного события: с вероятностью единица оно есть некоторое число, а с вероятностью нуль — неопределено (или считается равным ∞). Таким образом, мы чаще имеем дело с классами эквивалентных случайных величин, а не с самими случайными величинами. Тем не менее обычно проще выбрать подходящие представители классов и оперировать с ними, а не с классами.

С нулевыми множествами связано единственное место в теории вероятностей, которое идет вразрез с нашей интуицией. Ситуация одна и та же во всех вероятностных пространствах и достаточно описать ее для случая вероятностной прямой. В рамках нашей конструкции вероятности определены лишь на борелевских множествах, но каждое борелевское множество, вообще говоря, содержит много неборелевских подмножеств. Сле-

довательно, нулевое множество может содержать подмножества, на которых вероятности не определены, в то время как естественно ожидать, что каждое подмножество нулевого множества должно быть нулевым множеством. Это несоответствие не является серьезным и легко может быть устранено. В самом деле, введем такой *постулат*: *если* $A \subset B$ *и* $\mathbf{P}\{B\}=0$, *то* $\mathbf{P}\{A\}=0$. При этом мы должны расширить σ-алгебру $\mathfrak{A}$ борелевских множеств до (по меньшей мере) наименьшей σ-алгебры $\mathfrak{A}_1$, содержащей все множества из $\mathfrak{A}$ и все подмножества нулевых множеств. Можно дать прямое описание σ-алгебры $\mathfrak{A}_1$. Именно множество A принадлежит $\mathfrak{A}_1$ тогда и только тогда, когда оно отличается от некоторого борелевского множества A^0 лишь на нулевое множество [1]). Область определения вероятности можно продолжить с $\mathfrak{A}$ на $\mathfrak{A}_1$, просто положив $\mathbf{P}\{A\}=\mathbf{P}\{A^0\}$. Без труда показывается, что это продолжение единственно и дает вполне аддитивную меру на $\mathfrak{A}_1$. Таким образом, мы построили вероятностное пространство, которое удовлетворяет введенному постулату и в котором сохранились вероятности борелевских множеств.

Описанная конструкция называется *лебеговым пополнением* (заданного вероятностного пространства). В действительности упоминавшаяся в § 5 лебеговская конструкция приводит к пополненной σ-алгебре $\mathfrak{A}_1$, а не к борелевской σ-алгебре $\mathfrak{A}$. Использование пополнения естественно в задачах, связанных с единственным основным вероятностным распределением. По этой причине длину интервалов в $\mathscr{R}^1$ обычно продолжают до лебеговой меры, которая определена не только на борелевских множествах. Однако, когда рассматриваются семейства распределений (например, бесконечные последовательности испытаний Бернулли с различными вероятностями p), пополнение обычно приводит к неприятностям. Именно $\mathfrak{A}_1$ зависит от рассматриваемого распределения, и поэтому случайная величина по отношению к $\mathfrak{A}_1$ может не быть таковой, если перейти к другому распределению.

Пример. Пусть a_1, a_2, — последовательность точек в $\mathscr{R}^1$, имеющих вероятности p_1, p_2, ... соответственно, причем $\Sigma p_k=1$. Дополнение к $\{a_j\}$ имеет вероятность нуль, и поэтому $\mathfrak{A}_1$ содержит все подмножества $\mathscr{R}^1$. Каждая ограниченная функция u является при этом случайной величиной с математическим ожиданием $\Sigma p_k u(a_k)$, однако, когда рассматриваемое распределение не дискретно, оперировать с «произвольными функциями» становится опасным.

[1]) Более точно требуется, чтобы оба множества $A - A \cap A^0$ и $A^0 - A \cap A^0$ содержались в некотором нулевом множестве.

ГЛАВА V

ВЕРОЯТНОСТНЫЕ РАСПРЕДЕЛЕНИЯ в $\mathscr{R}^r$

В этой главе изучаются вероятностные распределения в r-мерном пространстве $\mathscr{R}^r$. Идейно понятие вероятностного распределения основывается на изложенной в предыдущей главе теории интегрирования, однако по существу никаких сложных размышлений для понимания настоящей главы не требуется, поскольку вводимые понятия и формулы интуитивно близки к понятиям и формулам, уже известным из первого тома и первых трех глав.

Новая черта рассматриваемой теории состоит в том, что (в противоположность случаю дискретного выборочного пространства) не *каждое* множество имеет вероятность и не *каждая* функция является случайной величиной. К счастью, эта теоретическая сложность не очень заметна на практике, поскольку можно исходить из интервалов и непрерывных функций и ограничиться соответственно рассмотрением множеств и функций, которые могут быть получены из них посредством элементарных операций и (возможно, бесконечно многих) предельных переходов. Этим выделяются классы борелевских множеств и бэровских функций. Читатели, которых интересуют больше факты, нежели логические связи, могут не беспокоиться по поводу точных определений (приводимых в гл. IV). Им разумнее полагаться на собственную интуицию и считать, что все рассматриваемые множества и функции являются «хорошими». Приводимые теоремы настолько просты[1]), что для их понимания достаточно знакомства с элементарным анализом. Изложение является строгим, если условиться, что *слова «множество» и*

[1]) Следует понимать, что этой простоты нельзя достичь ни в какой теории, использующей лишь непрерывные функции или какой-либо другой класс «хороших» функций. Например, в гл. II, (8.3) мы определили плотность φ бесконечным рядом. Скучно и бессмысленно устанавливать условия, при которых φ является хорошей функцией, однако в простых случаях формула очевидна, и она всегда имеет смысл, если ограничиваться бэровскими функциями, что и можно считать точным смыслом неопределенного выражения «формула всегда справедлива». Между прочим, будут указаны несколько случаев, где ограничение бэровскими функциями нетривиально [примером такого рода является теория выпуклых функций см. (8, б)].

«функция» сокращенно означают «борелевское множество» и «бэровская функция» соответственно.

При первоначальном чтении можно ограничиться § 1—4 и 9. Параграфы 5—8 содержат результаты и неравенства, собранные для удобства ссылок. В последних параграфах, посвященных теории условных распределений и математических ожиданий, изложение ведется более полно, чем это необходимо для настоящего тома. В дальнейшем результаты этой теории будут использоваться лишь от случая к случаю при рассмотрении мартингалов в гл. VI, 11 и гл. VII, 7.

1. Распределения и математические ожидания

Даже совсем безобидное использование термина «случайная величина» может содержать косвенную ссылку на весьма непростое вероятностное пространство или на сложный умозрительный эксперимент. Например, теоретическая модель может включать в себя положения и скорости 10^{28} частиц, в то время как нам интересны лишь температура и энергия. Эти две случайные величины отображают исходное выборочное пространство в плоскость $\mathscr{R}^2$, порождая в ней вероятностное распределение. По существу мы имеем дело с двумерной задачей, и первоначальное выборочное пространство утрачивает ясные очертания и превращается в фон. Конечномерные евклидовы пространства $\mathscr{R}^r$ представляют собой, следовательно, наиболее важные выборочные пространства, и мы переходим к систематическому изучению соответствующих вероятностных распределений.

Рассмотрим сначала случай прямой $\mathscr{R}^1$. Интервалы, определяемые неравенствами $a<x<b$ и $a\leqslant x\leqslant b$, обозначим $\overline{a,\ b}$ и $\overline{|\!-\!|}$ $a,\ b$ соответственно. (Мы не исключаем предельный случай замкнутого интервала, сводящегося к одной точке.) Полуоткрытые интервалы обозначаются $\overline{-|}$ $\overline{|-}$ $a,\ b$ и $a,\ b$. В одномерном случае все случайные величины являются функциями координатной переменной $\mathbf{X}$ (иными словами, функции, которая в точке x принимает значение x). Все вероятности можно, следовательно, выразить в терминах функции распределения

$$F(x)=\mathbf{P}\{\mathbf{X}\leqslant x\},\quad -\infty<x<\infty. \tag{1.1}$$

В частности, интервал $I=\overline{|\ |}$ $a,\ b$ имеет вероятность

$$\mathbf{P}\{I\}=F(b)-F(a).$$

Гибкое стандартное обозначение $\mathbf{P}\{\ \}$ становится неудобным, когда рассматриваются различные распределения. Применение новой буквы было бы неэкономичным, а использование

обозначения $\mathbf{P}_F\{\ \}$ для указания зависимости от F слишком неуклюже. Намного проще *использовать одну и ту же букву F как для функции точки* (1.1), *так и для функции интервала*, и в дальнейшем вместо $\mathbf{P}\{I\}$ мы будем писать $F\{I\}$. Иными словами, использование фигурных скобок $\{\ \}$ будет означать, что аргумент в $F\{A\}$ есть интервал или множество и что F представляет собой функцию интервалов (или меру). Когда используются круглые скобки, аргумент в $F(a)$ есть точка. Соотношение между функцией точки $F(\ )$ и функцией интервала $F\{\ \}$ выражается равенствами

$$F(x)=F\{-\infty,\ x\},\quad F\{a,\ b\}=F(b)-F(a). \tag{1.2}$$

В действительности понятие функции точки $F(x)$ излишне и служит лишь для удобства обозначения и записи. Основным и первичным является наделение вероятностями *интервалов*. Функция точки $F(\ )$ называется *функцией распределения* функции интервалов $F\{\ \}$. Символы $F(\ )$ и $F\{\ \}$ представляют собой различное выражение одного и того же, и никакой двусмысленности не появляется, когда говорят о «вероятностном распределении F». Следует привыкнуть думать в терминах функций *интервалов* и мер и использовать функцию распределения только как обозначение [1]).

Определение. *Определенная на прямой функция точки F есть функция распределения, если она*

i) *неубывающая, т. е. $F(a)\leqslant F(b)$ при $a<b$,*

ii) *непрерывна справа* [2]), *т. е. $F(a)=F(a+)$,*

iii) *удовлетворяет условиям $F(-\infty)=0$, $F(\infty)<\infty$.*

Функция распределения F, для которой $F(\infty)=1$, называется вероятностной функцией распределения. Функция распределения называется дефектной (или несобственной), если $F(\infty)<1$.

Покажем, что каждая функция распределения задает (соответствующие ей) вероятности всех множеств на прямой. Первым шагом является наделение вероятностями интервалов. Поскольку F монотонна, предел слева $F(a-)$ существует в каждой точ-

[1]) Для вводных курсов представляется разумной педантичная тщательность в использовании обозначений, однако автор надеется, что читатели *не будут* требовать такого рода «состоятельности» и *найдут в себе смелость* писать без различия $F\{I\}$ и $F(x)$. Это не повлечет никакой двусмысленности. В лучшей математической литературе стало привычным, к счастью, использовать один и тот же символ (в частности, 1 и $=$) на одной и той же странице в нескольких смыслах.

[2]) Как обычно, $f(a+)$ обозначает предел, если он существует, значений $f(x)$, когда $x\to a$ при условии $x>a$. Под $f(\infty)$ понимается предел $f(x)$ при $x\to\infty$. Аналогично определяется $f(a-)$ и $f(-\infty)$. Эти обозначения распространяются и на многомерный случай.

ке a. Определим функцию интервалов $F\{I\}$, положив

$$F\{\overline{|a,\ b|}\}=F(b)-F(a),\quad F\{\overline{(a,\ b)}\}=F(b-)-F(a),$$
$$F\{\overline{(a,\ b|}\}=F(b)-F(a),\quad F\{\overline{|a,\ b)}\}=F(b-)-F(a-). \tag{1.3}$$

Для интервала $\overline{|a,\ a|}$, сводящегося к единственной точке a, имеем

$$F\{\overline{|a,\ a|}\}=F(a)-F(a-),$$

что представляет собой скачок F в точке a. (Вскоре мы увидим, что F непрерывна «почти всюду».)

Чтобы показать, что определяемые формулами (1.3) значения вероятностей для интервалов удовлетворяют требованиям теории вероятностей, докажем следующую простую лемму (читатели могут принять ее как интуитивно очевидную).

Лемма 1. (*Счетная аддитивность.*) *Если интервал I представляет собой объединение счетного числа непересекающихся интервалов $I_1, \ldots, I_2$, то*

$$F\{I\}=\sum F\{I_k\}. \tag{1.4}$$

Доказательство. В частном случае, когда $I=\overline{|a,\ b|}$ и $I_1=\overline{|a,\ a_1|}$, $I_2=\overline{(a_1,\ a_2|}$, $\ldots$, $I_n=\overline{(a_{n-1},\ b|}$, утверждение леммы тривиально. Произвольное *конечное* разбиение $I=\overline{|a,\ b|}$ получается из этого разбиения перераспределением конечных точек a_k подинтервалов, так что правило сложения (1.4) для конечных разбиений выполняется.

При рассмотрении случая бесконечного множества интервалов I_k, не ограничивая общности, можно допустить, что интервал I замкнут. В силу непрерывности справа функции распределения F для любого заданного $\varepsilon>0$ можно найти содержащий I_k открытый интервал $I_k^{\#}$, такой, что $0\leqslant F\{I_k^{\#}\}-F\{I_k\}\leqslant\varepsilon 2^{-k}$. Поскольку существует конечная совокупность интервалов $I_{k_1}^{\#}, \ldots, I_{k_n}^{\#}$, покрывающая I, то

$$F\{I\}\leqslant F\{I_{k_1}^{\#}\}+\ldots+F\{I_{k_n}^{\#}\}\leqslant F\{I_1\}+\ldots+F\{I_n\}+\varepsilon, \tag{1.5}$$

откуда

$$F\{I\}\leqslant\sum F\{I_k\}. \tag{1.6}$$

Обратное неравенство тоже справедливо, так как для каждого n существует *конечное* разбиение I, содержащее в качестве подинтервалов интервалы $I_1, \ldots, I_n$. Тем самым лемма доказана. ▶

Как уже говорилось в гл. IV, 2, теперь можно определить

$$F\{A\}=\Sigma F\{A_k\} \tag{1.7}$$

для любого множества A, представляющего собой конечное или счетное объединение непересекающихся интервалов A_k. Интуитивно можно ожидать, что любое множество допускает аппроксимацию такими объединениями интервалов, и теория меры это подтверждает [1]). Используя естественные приближения и переходы к пределу, можно продолжить определение F на все множества так, чтобы свойство счетной аддитивности (1.7) сохранялось. Это продолжение единственно, и получающаяся в результате функция множеств называется вероятностным распределением или мерой.

Замечание о терминологии. Термин «распределение» свободно используется в литературе в различных смыслах, и поэтому здесь уместно точно оговорить, что понимается под распределением в настоящей книге.

Вероятностное распределение, или *вероятностная мера*, есть функция, ставящая в соответствие множествам числа $F\{A\}\geqslant 0$ и удовлетворяющая условию счетной аддитивности (1.7) и условию нормировки $F\{-\infty, \infty\}=1$. Отбрасывая условие нормировки, получаем определение более общей меры (распределения массы). Весьма важным примером такой меры является мера Лебега (или обычная длина).

Иногда приходится иметь дело с мерами, относительно которых масса всей прямой $p=F\{-\infty, \infty\}$ меньше 1 (достаточно вспомнить рассматривавшуюся в первом томе теорию рекуррентных событий). Меру с таким свойством мы называем вероятностной *несобственной или дефектной* мерой с *дефектом*, *равным* $1-p$. Иногда с целью стилистической ясности или желая подчеркнуть, что рассматривается не мера вообще и не дефектная вероятностная мера, а именно вероятностная мера, мы будем называть ее *собственной* вероятностной мерой, хотя прилагательное собственная и излишне.

Аргументом меры $m\{A\}$ является множество; при письме буква, обозначающая это множество, заключается в фигурные скобки. Каждой ограниченной мере m соответствует *функция распределения* — функция точки, определяемая равенством $m(x)=m\{-\infty, x\}$. Функция распределения обозначается той же буквой, что и соответствующая ей мера, только аргумент ее заключается в круглые скобки. Двоякое использование одной

[1]) Следует иметь в виду принятое нами соглашение, согласно которому слова «множество» и «функция» сокращенно означают соответственно *«борелевское множество»* и *«бэровская функция»*.

и той же буквы не вызывает никаких недоразумений, и, более того, термин «распределение» может сокращенно обозначать как вероятностное распределение, так и его функцию распределения. ▶

В 1, гл. IX *случайная величина* была определена как вещественная функция на выборочном пространстве. Мы сохраняем здесь это определение. В случае когда выборочным пространством является прямая, каждая вещественная функция есть случайная величина. Координатная случайная величина $\mathbf{X}$ является основной, и все остальные случайные величины можно рассматривать как функции от нее. Функция распределения случайной величины u, по определению равная $\mathbf{P}\{u(\mathbf{X}) \leqslant x\}$, может быть выражена в терминах функции распределения координатной случайной величины $\mathbf{X}$. Например, функция распределения величины $\mathbf{X}^3$ есть $F(\sqrt[3]{x})$.

Функция u называется *простой*, если она принимает лишь счетное множество значений $a_1, a_2, \ldots$. Пусть u — простая функция и A_n — то множество, на котором u равна a_n. Скажем, что *математическое ожидание* функции u *существует и равно*

$$\mathbf{E}(u) = \sum a_k F\{A_k\}, \tag{1.8}$$

если ряд в правой части (1.8) *абсолютно сходится*. В противном случае u называется неинтегрируемой по отношению к F. Таким образом, u имеет математическое ожидание тогда и только тогда, когда существует $\mathbf{E}(|u|)$. Исходя из (1.8), можно следующим образом определить математическое ожидание произвольной ограниченной функции u. Возьмем $\varepsilon > 0$ и обозначим через A_n множество тех точек x, в которых $(n-1)\varepsilon < x \leqslant n\varepsilon$. При любом разумном определении $\mathbf{E}(u)$ должно быть

$$\sum (n-1)\varepsilon \cdot F\{A_n\} \leqslant \mathbf{E}(u) \leqslant \sum n\varepsilon F\{A_n\}. \tag{1.9}$$

(Суммы в (1.9) представляют собой математические ожидания двух аппроксимирующих простых функций $\underline{\sigma}$ и $\overline{\sigma}$, таких, что $\underline{\sigma} \leqslant u \leqslant \overline{\sigma}$ и $\overline{\sigma} - \underline{\sigma} = \varepsilon$.) Поскольку функция u, согласно предположению, ограничена, то суммы в (1.9) содержат лишь конечное число отличных от нуля членов. Разность этих сумм равна $\varepsilon \Sigma F\{A_n\} = \varepsilon$. При замене ε на $\frac{1}{2}\varepsilon$ левая сумма, очевидно, увеличивается, а правая — уменьшается. Отсюда легко вывести, что при $\varepsilon \to 0$ обе суммы в (1.9) стремятся к одному и тому же пределу, который и полагается равным $\mathbf{E}(u)$. Для неограниченной функции применимо то же самое рассуждение, если оба ряда в (1.9) *абсолютно* сходятся; в противном случае $\mathbf{E}(u)$ не определяется (u не интегрируема по отношению к F).

Определенное таким простым способом математическое ожидание представляет собой *интеграл Лебега — Стильтьеса* от функции u по отношению к распределению F. Когда желают подчеркнуть зависимость математического ожидания от F, предпочтительно использовать интегральное обозначение

$$\mathbf{E}(u) = \int_{-\infty}^{\infty} u(x) F\{dx\}. \tag{1.10}$$

За исключением редких случаев, мы будем иметь дело лишь с монотонными функциями u или с кусочно-непрерывными функциями, такими, что для них множества A_n сводятся к объединению конечного числа интервалов. При этом суммы в (1.9) представляют собой, с точностью до порядка суммирования, верхние и нижние суммы, используемые при элементарном определении обычных интегралов. Общий интеграл Лебега — Стильтьеса, обладая основными свойствами обычного интеграла, имеет дополнительное преимущество — при обращении с ним меньше приходится заботиться о законности формальных операций и переходов к пределу. Мы будем использовать математические ожидания лишь в очень простых ситуациях, и никакой общей теории не потребуется для возможности проследить за отдельными этапами рассуждений. Читатель, интересующийся теоретическим фундаментом и основными фактами, может обратиться к гл. IV.

Примеры. а) Пусть F — дискретное распределение, наделяющее точки $a_1, a_2, \ldots$ массами $p_1, p_2, \ldots$. Ясно, что в этом случае математическое ожидание $\mathbf{E}(u)$ функции u, если оно существует, равно $\sum u(a_k) p_k$, и условием существования $\mathbf{E}(u)$ является абсолютная сходимость ряда $\sum u(a_k) p_k$. Это согласуется с определением, данным в **1**, гл. IX.

б) Для распределений, задаваемых непрерывной плотностью, математическое ожидание u представляет собой интеграл

$$\mathbf{E}(u) = \int_{-\infty}^{\infty} u(x) f(x)\, dx, \tag{1.11}$$

если только этот интеграл сходится абсолютно. По поводу общего понятия плотности смотри § 3. ▶

Обобщение вышеизложенного на многомерные пространства можно описать в нескольких словах. Точка в $\mathscr{R}^2$ представляет собой пару действительных чисел $x = (x_1, x_2)$. Неравенства следует понимать в покоординатном смысле [1]); так $a < b$ означает,

[1]) Это соглашение было введено в гл. III, 5.

что $a_1<b_1$ и $a_2<b_2$ (или, иными словами, «a лежит юго-западнее b»). Это отношение порядка является лишь частичным упорядочением, т. е. две точки a и b не обязаны находиться в одном из двух отношений $a<b$, $a\geqslant b$. *Интервалами* мы называем множества, определяемые четырьмя возможными типами двойных неравенств $a<x<b$ и т. д. Интервалы представляют собой прямоугольники со сторонами, параллельными осям координат, и могут также вырождаться в отрезки или точки.

Единственно новым является то обстоятельство, что если $\overline{a,\ c}$ — двумерный интервал и $a<b<c$, то $a,\ c$ не есть объединение $\overline{a,\ b}$ и $\overline{b,\ c}$. Каждой функции интервала $F\{I\}$ можно поставить в соответствие отвечающую ей функцию распределения, которая, как и в одномерном случае, определяется равенством $F(x)=F\{-\infty,\ x\}$. Теперь, однако, в выражение $F\{\overline{a,\ b}\}$ в терминах функции распределения входят все четыре вершины интервала. Если $\alpha=(\alpha_1,\ \alpha_2)$ — вершина с координатами $\alpha_1=a_1$, $\alpha_2=b_2$, то разность $\overline{-\infty,\ b}-\overline{-\infty,\ \alpha}$ есть полубесконечная («вертикальная») полоса ширины b_1-a_1 и интервал $I=\overline{a,\ b}$ представляет собой разность двух таких полос. Таким образом, при $a\leqslant b$

$$F\{\overline{a,\ b}\}=F(b_1,\ b_2)-F(a_1,\ b_2)-F(b_1,\ a_2)+F(a_1,\ a_2). \quad (1.12)$$

Для функции распределения $F(x)=F(x_1,\ x_2)$ *«смешанные разности»* типа стоящей в правой части (1.12) неотрицательны. Отсюда следует монотонность $F(x_1,\ x_2)$ по x_1 и x_2, однако обратное неверно: монотонность F не обеспечивает положительности (1.12).

Пример. в) Пусть $F(x)=0$ при $x<0$ и $F(x)=1$ во всех остальных точках. Очевидно, F монотонна по каждой из координат, однако при $a<0$ и $b>0$ величина (1.12) для такой функции принимает отрицательное значение, равное -1. ▶

Ясно, что в многомерном случае функции распределения являются не очень подходящим аппаратом: если бы не соображение аналогии со случаем пространства $\mathscr{R}^1$, все рассмотрения, возможно, велись бы в терминах функций интервалов. Формально определение функции распределения в $\mathscr{R}^1$ распространяется на $\mathscr{R}^2$, если условие монотонности (1) заменить на условие неотрицательности смешанных разностей (1.12) при $a\leqslant b$. Так определенная функция распределения индуцирует функцию интервалов (соотношения подобны (1.3), в случае $\mathscr{R}^1$ нужно

лишь обычные разности заменить на смешанные). Лемма 1 сохраняется вместе с доказательством [1]).

Одно простое, но существенно важное свойство математических ожиданий принимается иногда как само собой разумеющееся. Любая функция $u(\mathbf{X}) = u(\mathbf{X}_1, \mathbf{X}_2)$ от координатных переменных является случайной величиной и, как таковая, имеет функцию распределения G. Математическое ожидание $\mathbf{E}(u(\mathbf{X}))$ можно определить двумя способами — как интеграл от $u(x_1, x_2)$ по отношению к заданной вероятностной мере на плоскости, а также в терминах функции распределения величины u:

$$\mathbf{E}(u) = \int_{-\infty}^{\infty} y\, G\{dy\}. \tag{1.13}$$

Оба определения эквивалентны в силу самого определения первого из этих интегралов с помощью аппроксимирующих сумм гл. IV, (4.6) [2]). Существенным здесь является то обстоятельство, что математическое ожидание случайной величины $\mathbf{Z}$ (если оно существует) имеет внутренний смысл и не зависит от того, рассматривается ли $\mathbf{Z}$ как функция на исходном вероятностном пространстве $\mathscr{S}$ или на пространстве, полученном при соответствующем отображении $\mathscr{S}$; в частности, $\mathbf{Z}$ сама отображает $\mathscr{S}$ на прямую и превращается на ней в координатную величину.

Начиная с этого места, не будет никакой разницы между случаями пространства $\mathscr{R}^1$ и пространства $\mathscr{R}^2$. В частности, определение математического ожидания не зависит от числа измерений.

Резюмируя, можно сказать, что *любая функция распределения индуцирует вероятностную меру на σ-алгебре борелевских множеств в $\mathscr{R}^r$ и тем самым определяет вероятностное пространство.* Иначе говоря, мы показали, что вероятностная конструкция дискретных выборочных пространств, а также пространств с распределениями, задаваемыми плотностями, распространяется без формальных изменений на общий случай. Тем самым оправдана вероятностная терминология, использовавшаяся в первых трех главах. Когда мы говорим об r случайных величинах $\mathbf{X}_1, \ldots, \mathbf{X}_r$, мы подразумеваем, что все они определены на одном и том же вероятностном пространстве, и поэтому суще-

---

[1]) В доказательстве используется тот факт, что при *конечном* разбиении одномерного интервала подинтервалы имеют естественный порядок слева направо. Столь же хорошее упорядочение имеет место при *шахматных разбиениях* двумерного интервала $\overline{a, b}$, т. е. разбиениях на mn подинтервалов, получаемых разбиениями отдельно обеих сторон $\overline{a, b}$ и проведением прямых, параллельных осям, через все точки подразбиений. Доказательство конечной аддитивности для таких шахматных разбиений не меняется, и каждому разбиению соответствует его шахматное подразбиение. Переход от конечных разбиений к счетным одинаков при любом числе измерений.

[2]) Специальный случай содержится в теореме 1, гл. IX, 2.1. См. также задачу 1.

ствует их совместное вероятностное распределение. При этом мы можем, если пожелаем, рассматривать $\mathbf{X}_k$ как координатные величины в выборочном пространстве $\mathscr{S}^r$.

Вряд ли необходимо объяснять способ распространения таких понятий, как *частное (маргинальное) распределение* (см. гл. III, 1, **1**, гл. IX, 1) или *независимость случайных величин*. Основные утверждения о независимых случайных величинах переносятся с дискретного на общий случай без изменений.

i) Независимость случайных величин $\mathbf{X}$ и $\mathbf{Y}$ с (одномерными) распределениями F и G означает, что функция совместного распределения $(\mathbf{X}, \mathbf{Y})$ задается произведениями $F(x_1)G(x_2)$. Это может относиться к двум случайным величинам, определенным на заданном вероятностном пространстве, но может также пониматься и как сокращенное выражение того, что мы вводим плоскость с $\mathbf{X}$ и $\mathbf{Y}$ в качестве координатных переменных и *определяем* вероятности по правилу произведения. Сказанное в равной степени применимо к парам, тройкам и вообще к любому числу случайных величин.

ii) Если совокупность m случайных величин $(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ не зависит от совокупности n случайных величин $(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$, то случайные величины $u(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ и $v(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ независимы, каковы бы ни были функции u и v.

iii) Для независимых $\mathbf{X}$ и $\mathbf{Y}$ справедливо равенство $\mathbf{E}(\mathbf{XY}) = \mathbf{E}(\mathbf{X})\mathbf{E}(\mathbf{Y})$, если только математические ожидания $\mathbf{X}$ и $\mathbf{Y}$ существуют (т. е. если соответствующие интегралы сходятся абсолютно).

В заключение этого параграфа приведем один простой результат[1]), который часто приходится использовать.

Лемма 2. *Вероятностная мера однозначно определяется значениями* $\mathbf{E}(u)$ *для всех непрерывных функций* u, *обращающихся в нуль вне некоторого конечного интервала.*

Доказательство. Пусть I — конечный открытый интервал и v — непрерывная функция, которая положительна на I и равна нулю вне I. Тогда $\sqrt[n]{v(x)} \to 1$ при всех $x \in I$ и, следова-

[1]) Этот результат содержится в гл. IV, 5 и повторяется здесь потому, что доказать его можно совсем просто. Для читателя, интересующегося общей теорией, заметим, что наши рассуждения в значительной степени основывались на использовании интервалов. В более общих ситуациях не существует таких отличительных множеств и часто предпочтительно *начинать* с математических ожиданий непрерывных функций и затем с их помощью определять меры. При этом используется рассуждение, обратное к содержащемуся в приводимом ниже доказательстве.

тельно, $\mathbf{E}\left(\sqrt[n]{v}\right) \to F\{I\}$. Таким образом, значения математических ожиданий непрерывных функций, обращающихся в нуль вне некоторого интервала, однозначно определяют значения $F\{I\}$ для всех открытых интервалов, которые в свою очередь однозначно определяют F. ▷

Замечание о независимости и корреляции. Теория статистической корреляции восходит к тому времени, когда формализация теории была еще невозможна и понятие стохастической независимости по необходимости носило мистический характер. Уже тогда понимали, что независимость двух ограниченных случайных величин с нулевыми математическими ожиданиями влечет равенство $\mathbf{E}(\mathbf{XY}) = 0$, однако сначала думали, что это равенство должно быть и достаточным для независимости $\mathbf{X}$ и $\mathbf{Y}$. После того как была обнаружена ложность этого заключения, долгое время искали условия, при которых обращение в нуль корреляции влечет стохастическую независимость. Как часто случается, история задачи и красота частных результатов затемнили тот факт, что современные методы позволяют дать чрезвычайно простое ее решение. Следующая теорема содержит различные результаты, доказанные ранее трудоемкими методами.

Теорема. *Случайные величины* $\mathbf{X}$ и $\mathbf{Y}$ *независимы тогда и только тогда, когда*

$$\mathbf{E}(u(\mathbf{X})v(\mathbf{Y})) = \mathbf{E}(u(\mathbf{X}))\,\mathbf{E}(v(\mathbf{Y})) \tag{1.14}$$

для всех непрерывных функций u и v, *обращающихся в нуль вне конечного интервала.*

Доказательство. Необходимость условия очевидна. Достаточность его будет доказана, если показать, что для всякой непрерывной функции $w(\mathbf{X}, \mathbf{Y})$, равной нулю вне некоторого конечного интервала, математическое ожидание $\mathbf{E}(w)$ совпадает с математическим ожиданием w относительно пары *независимых* случайных величин, распределенных как $\mathbf{X}$ и $\mathbf{Y}$. Но, согласно (1.14), это верно для функций вида $w(\mathbf{X}, \mathbf{Y}) = u(\mathbf{X})v(\mathbf{Y})$. Так как каждую непрерывную функцию w, обращающуюся в нуль вне некоторого конечного интервала, можно равномерно приблизить[1]) линейными комбинациями вида $\Sigma c_k u_k(\mathbf{X}) v_k(\mathbf{Y})$, то, переходя к пределу, убеждаемся в справедливости нашего утверждения. ▶

§ 2. Предварительные сведения

Настоящий параграф посвящен в основном известным или очевидным понятиям и фактам, связанным с функциями распределения в $\mathscr{R}^1$.

Как и в дискретном случае, *k-й момент* случайной величины $\mathbf{X}$ по определению полагается равным $\mathbf{E}(\mathbf{X}^k)$. При этом

[1]) См. задачу 10 в гл. VIII, 10.

предполагается, что интеграл

$$\mathbf{E}(\mathbf{X}^k) = \int_{-\infty}^{\infty} x^k F\{dx\} \tag{2.1}$$

существует, т. е. сходится абсолютно. Таким образом, $\mathbf{E}(\mathbf{X}^k)$ существует тогда и только тогда, когда $\mathbf{E}(|\mathbf{X}|^k) < \infty$. Величина $\mathbf{E}(|\mathbf{X}|^k)$ называется k-м *абсолютным моментом* X. Она определена и для нецелых $k > 0$. Так как при $0 < a < b$ имеет место неравенство $|x|^a \leqslant |x|^b + 1$, то из существования абсолютного момента порядка b вытекает существование абсолютных моментов всех порядков $a < b$.

Второй момент случайной величины $\mathbf{X} - m$, где m — математическое ожидание $\mathbf{X}$, называется *дисперсией* $\mathbf{X}$ и обозначается $\operatorname{Var}\mathbf{X}$:

$$\operatorname{Var}(\mathbf{X}) = \mathbf{E}((\mathbf{X} - m)^2) = \mathbf{E}(\mathbf{X}^2) - m^2. \tag{2.2}$$

В рассматриваемом общем случае дисперсия обладает теми же свойствами и играет ту же роль, что и в дискретном случае. В частности, если $\mathbf{X}$ *и* $\mathbf{Y}$ *независимы* и дисперсии $\operatorname{Var}\mathbf{X}$ и $\operatorname{Var}\mathbf{Y}$ существуют, то

$$\operatorname{Var}(\mathbf{X} + \mathbf{Y}) = \operatorname{Var}\mathbf{X} + \operatorname{Var}\mathbf{Y}. \tag{2.3}$$

[Две случайные величины, удовлетворяющие равенству (2.3), называются *некоррелированными*. В **1**, гл. IX, 5 было показано, что существуют зависимые некоррелированные случайные величины.]

Вспомним, что часто мы заменяли случайную величину на «нормированную величину» $\mathbf{X}^* = (\mathbf{X} - m)/\sigma$, где $m = \mathbf{E}(\mathbf{X})$ и $\sigma^2 = \operatorname{Var}(\mathbf{X})$. Выражаясь физическим языком, можно сказать, что $\mathbf{X}^*$ есть запись $\mathbf{X}$ в «безразмерных единицах». Вообще переход от $\mathbf{X}$ к $(\mathbf{X} - \beta)/\alpha$ при $\alpha > 0$ есть изменение начала отсчета и единицы измерения. Функция распределения новой случайной величины равна $F(\alpha x + \beta)$, и нередко мы имеем дело по существу со всем классом функций распределения этого вида, а не с отдельным представителем класса. Введем для удобства следующее

Определение 1. *Два распределения* F_1 *и* F_2 *в* $\mathscr{R}^1$ *называются отличающимися лишь параметрами расположения, если* $F_2(x) = F_1(\alpha x + \beta)$, $\alpha > 0$. *Про такие распределения говорят также, что они одного и того же типа. Параметр* α *называют масштабным множителем, а параметр* β*—центрирующей постоянной.*

Данное определение позволяет использовать выражения вида «F центрировано к нулевому математическому ожиданию» или «центрирование не меняет дисперсии».

Медианой распределения F называется число ξ, удовлетворяющее условиям $F(\xi) \geqslant \frac{1}{2}$ и $F(\xi-) \leqslant \frac{1}{2}$. Медиана может быть определена неоднозначно; если $F(x) = \frac{1}{2}$ при всех x из некоторого интервала $\overline{a, b}$, то каждое x из этого интервала есть медиана. Распределение можно центрировать так, чтобы его медиана стала равной нулю.

За исключением медианы, все рассмотренные понятия переносятся на случай любого конечного числа измерений, или, иными словами, на векторные случайные величины вида $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$. Соответствующие векторные обозначения были введены в гл. III, 5 и не требуют никаких изменений. Математическое ожидание $\mathbf{X}$ представляет собой в этом случае вектор, а дисперсия $\mathbf{X}$ — матрицу.

Первое, на что обращается внимание при рассмотрении графика функции распределения, это разрывы и интервалы постоянства. Нередко приходится оговаривать, что точка не принадлежит интервалу постоянства. Введем следующую удобную терминологию, относящуюся к любому конечному числу измерений.

Определение 2. *Точка x является атомом, если она имеет положительную массу*[1]). *Точка x, для которой $F\{I\} > 0$, каков бы ни был открытый интервал I, содержащий x, называется точкой роста.*

Распределение F сосредоточено на множестве A, если дополнение A' имеет вероятность $F\{A'\} = 0$.

Распределение F называется атомическим, если оно сосредоточено на множестве своих атомов.

Пример. Расположим рациональные числа из отрезка $\overline{0, 1}$ в последовательность $r_1, r_2, \ldots$ так, чтобы с ростом номера числа рос и его знаменатель. Определим распределение F, положив вероятность каждой точки r_k равной 2^{-k}. Распределение F атомическое, однако все точки замкнутого интервала $\overline{0, 1}$ являются точками роста F. ▶

В силу условия счетной аддитивности (1.7) сумма масс атомов не превосходит единицы, так что имеется не больше одного атома с массой $> \frac{1}{2}$, не больше двух атомов с массой $> \frac{1}{3}$ и т. д. Все атомы, следовательно, можно расположить в прос-

[1]) Это, конечно, есть краткое выражение следующего: «множество, состоящее из одной точки x, имеет положительную вероятность».

тую последовательность $a_1, a_2, \ldots$ так, чтобы соответствующие им массы не возрастали: $p_1 \geqslant p_2 \geqslant \ldots$. Иными словами, *число атомов не более чем счетно.*

Распределение, не имеющее атомов, называется *непрерывным.* Предположим, что распределение имеет атомы с массами $p_1, p_2, \ldots$, и обозначим $p = \sum p_k$. Положим

$$F_a(x) = \frac{1}{p} \sum_{a_k \leqslant x} p_k, \tag{2.4}$$

где суммирование распространяется на все атомы из интервала $(-\infty, x]$. Функция $F_a(x)$, представляющая собой, очевидно, функцию распределения, называется *атомической компонентой* F. Если $p=1$, то распределение F является атомическим. При $p<1$ положим $q=1-p$. Легко видеть, что $[F - pF_a]/q = F_c$ есть *непрерывная* функция распределения. Таким образом, имеем

$$F = pF_a + qF_c, \tag{2.5}$$

т. е. F является линейной комбинацией двух функций распределения F_a и F_c, из которых F_a — атомическая, а F_c — непрерывная. Если F есть атомическая функция распределения, то (2.5) по-прежнему имеет место с $p=1$, причем в качестве F_c можно взять произвольную непрерывную функцию распределения. При отсутствии атомов в (2.5) будет $p=0$. Нами доказана, таким образом,

Теорема Жордана о разложении. *Всякое вероятностное распределение является смесью атомического и непрерывного распределений вида* (2.5) *с* $p \geqslant 0$, $q \geqslant 0$, $p+q=1$.

Имеется один класс атомических распределений, из-за которых приходится иногда загромождать простые формулировки перечислением тривиальных исключений. Распределения этого класса лишь произвольным масштабным множителем отличаются от распределений, сосредоточенных в целочисленных точках, однако встречаются они весьма часто и поэтому заслуживают специального названия для удобства ссылок.

Определение 3. *Распределение F на $\mathscr{R}^1$ называется арифметическим* [1]*), если оно сосредоточено на множестве точек вида*

[1]) Пожалуй, более общепринятым является термин «решетчатое распределение», однако его употребляют в разных смыслах: некоторые авторы под решетчатым распределением понимают распределение, сосредоточенное на множестве a, $a \pm \lambda$, $a \pm 2\lambda$, ..., где a — произвольно. (Биномиальное распределение с атомами в точках ± 1 в нашей терминологии является арифметическим распределением с шагом 1, в то время как в упомянутой другой терминологии оно — решетчатое распределение с шагом 2.)

$0, \pm\lambda, \pm 2\lambda, \ldots$. *Наибольшее число* λ, *обладающее этим свойством, называется шагом* F.

§ 3. Плотности

Первые две главы были посвящены изучению вероятностных распределений в $\mathscr{R}^1$, для которых вероятность произвольного интервала (а, следовательно, и вообще произвольного множества) задается интегралом

$$F\{A\} = \int_A \varphi(x)\,dx. \tag{3.1}$$

Распределения, рассматривавшиеся в гл. III, имеют тот же вид, с той лишь разницей, что интегрирование происходит по лебеговой мере (площади или объему) в $\mathscr{R}^r$. Если плотность φ в (3.1) сосредоточена на интервале $\overline{0, 1}$, то (3.1) можно переписать в виде

$$F\{A\} = \int_A \varphi(x)\,U\{dx\}, \tag{3.2}$$

где U обозначает равномерное распределение на $\overline{0, 1}$. Последняя формула имеет смысл для произвольного вероятностного распределения U, и, когда $F\{\overline{-\infty, \infty}\} = 1$, она определяет новое вероятностное распределение F. В этом случае мы говорим, что φ *есть плотность* F *по отношению к* U.

В формуле (3.1) мера U бесконечна, в то время как в формуле (3.2) имеем $U\{\overline{-\infty, \infty}\} = 1$. Это различие несущественно, поскольку интеграл в (3.1) может быть разбит на интегралы вида (3.2), распространенные на конечные интервалы. Хотя мы будем использовать (3.2) лишь в случаях, когда U есть либо вероятностное распределение, либо мера Лебега, как в (3.1), дадим следующее общее определение.

Определение. *Распределение* F *называется абсолютно непрерывным по отношению к мере* U, *если для некоторой функции* φ *имеет место соотношение* (3.2). *При этом* φ *называется плотностью*[1]) F *по отношению к* U.

Наиболее важным частным случаем является, конечно, тот случай, когда имеет место (3.1), т. е. когда U есть мера Лебега. Функция φ в этом случае называется *«обычной» плотностью.*

[1]) Это определение является стандартным определением в теории меры, а функция φ называется производной Радона — Никодима распределения F по отношению к U.

Введем теперь следующую *сокращенную запись*:

$$F\{dx\}=\varphi(x)\ U\{dx\}. \tag{3.3}$$

Эта запись имеет лишь тот смысл, что для всех множеств выполняется равенство (3.2), и никакого специального смысла символу dx придавать не следует. Пользуясь этой сокращенной записью, (3.1) мы будем записывать в виде $F\{dx\}=\varphi(x)dx$, и если U имеет обычную плотность u, то (3.2) принимает вид $F\{dx\}=\varphi(x)u(x)dx$.

Примеры. а) Пусть U — вероятностное распределение на $\mathscr{R}^1$ со вторым моментом, равным m_2. Тогда

$$F\{dx\}=\frac{1}{m_2}x^2U\{dx\}$$

есть новое вероятностное распределение. В частности, если U — равномерное распределение на $\overline{0,\ 1}$, то $F(x)=x^3$ при $0<x<1$, а если U имеет плотность e^{-x} $(x>0)$, то F есть гамма-распределение с обычной плотностью $\frac{1}{2}x^2e^{-x}$.

б) Пусть U — атомическое распределение, наделяющее атомы $a_1,\ a_2,\ldots$ массами $p_1,\ p_2,\ldots$ (здесь $\sum p_k=1$). Распределение F имеет плотность φ по отношению к U тогда и только тогда, когда оно атомическое и все его атомы являются также атомами распределения U. Если F имеет в a_j массу q_j, то плотность φ в a_j равна $\varphi(a_k)=q_k/p_k$. Значение φ в точках, отличных от точек a_j, не играет никакой роли и лучше всего считать φ в этих точках неопределенной. ▶

Теоретически подинтегральная функция φ в (3.2) определена неоднозначно. В самом деле, если N — такое множество, что $U\{N\}=0$, то φ можно как угодно изменить на N и при этом (3.2) останется в силе. К этому, однако, и сводится вся неоднозначность в определении φ: *плотность определена однозначно с точностью до значений на нулевом множестве* [1]). Практически условиями непрерывности диктуется обычно однозначный выбор плотности, и поэтому, как правило, говорят о некоторой опреде-

[1]) В самом деле, пусть φ и φ_1 — две плотности F по отношению к U. Рассмотрим множество A всех точек x, для которых $\varphi(x)>\varphi_1(x)+\varepsilon$. Из равенств

$$F\{A\}=\int\limits_A\varphi(x)\,U\{dx\}=\int\limits_A\varphi_1(x)\,U\{dx\}$$

следует, что $U\{A\}=0$, и поскольку это верно для любого $\varepsilon>0$, то $\varphi(x)=\varphi_1(x)$ всюду, за исключением множества N, такого, что $U\{N\}=0$.

ленной плотности, хотя, конечно, более правильным было бы говорить о некоторой из плотностей.

Из (3.3), очевидно, следует, что для любой ограниченной функции v [1])

$$v(x)\,F\{dx\}=v(x)\varphi(x)\,U\{dx\}. \tag{3.4}$$

В частности, если все значения функции φ больше (или меньше) нуля на некоторое число, то, взяв $v=\varphi^{-1}$, получим *формулу, обращающую* (3.2):

$$U\{dx\}=\frac{1}{\varphi(x)}F\{dx\}. \tag{3.5}$$

Полезный критерий абсолютной непрерывности содержится в одной из основных теорем теории меры, которую мы приведем здесь без доказательства.

Теорема Радона — Никодима [2]). *Распределение F абсолютно непрерывно по отношению к мере U тогда и только тогда, когда*

$$U\{A\}=0 \text{ влечет } F\{A\}=0. \tag{3.6}$$

Условие (3.5) можно перефразировать следующим образом: U — нулевые множества являются также F-нулевыми множествами. Приведем одно важное следствие теоремы Радона — Никодима, которое, однако, нигде в дальнейшем непосредственно использоваться не будет.

Критерий. *Распределение F абсолютно непрерывно по отношению к мере U тогда и только тогда, когда каждому $\varepsilon>0$ соответствует $\delta>0$, такое, что для любой совокупности непересекающихся интервалов $I_1, \ldots, I_n$*

$$\sum_1^n U\{I_k\}<\delta \text{ влечет } \sum_1^n F\{I_k\}<\varepsilon. \tag{3.7}$$

---

[1]) Читателям, которых смущает справедливость соотношения (3.4), следует обратить внимание на его тривиальность в случае, когда F и U имеют непрерывные плотности. Следующее доказательство (3.4) в общем случае использует стандартное рассуждение, применимое и в более общих ситуациях. Формула (3.4) тривиальна, когда v есть простая функция, принимающая лишь конечное число различных значений. Для любой ограниченной функции v существуют две простые функции с конечным числом значений $\underline{v}$ и $\bar{v}$, такие, что $\underline{v}<v\leqslant\bar{v}$ и $\bar{v}-\underline{v}<\varepsilon$, и поэтому из справедливости формулы (3.4) для всех простых функций вытекает ее справедливость в общем случае.

[2]) Эту теорему нередко называют теоремой Лебега — Никодима. Соотношение (3.6) может быть принято в качестве определения абсолютной непрерывности, и в этом случае теорема будет утверждать существование плотности.

Весьма важным частным случаем выполнения (3.7) является тот случай, когда для всех интервалов

$$F\{I\} \leqslant aU\{I\}. \tag{3.8}$$

При этом (3.7) выполняется с $\delta = \varepsilon/a$ и, как легко видеть, плотность F по отношению к U не превосходит a.

3а [1]) Сингулярные распределения

Условие (3.6) теоремы Радона — Никодима наводит на рассмотрение случая, прямо противоположного случаю абсолютно непрерывных распределений.

Определение. *Вероятностное распределение F сингулярно по отношению к U, если оно сосредоточено на множестве N, таком, что $U\{N\} = 0$.*

Особую роль здесь играет мера Лебега $U\{dx\} = dx$, и когда говорят, что распределение «сингулярно», и не делают при этом дополнительных пояснений, имеют в виду, что оно сингулярно по отношению к мере Лебега. Каждое атомическое распределение сингулярно по отношению к dx, но имеются и непрерывные распределения в $\mathscr{R}^1$, сингулярные относительно dx: таковым является канторовское распределение, рассмотренное в примере (11, г) гл. I. Сингулярные распределения не поддаются аналитическому изучению, и их явное представление практически невозможно. Чтобы иметь возможность применять аналитические методы, приходится, следовательно, накладывать ограничения, которые обеспечивают либо абсолютную непрерывность, либо атомичность рассматриваемых распределений. Сингулярные распределения, однако, играют важную принципиальную роль, и многие статистические тесты основываются на их существовании. Это обстоятельство затемняется бытующим мнением, что «в практике» сингулярные распределения не встречаются.

Примеры. а) *Испытания Бернулли.* В примере (11, в) гл. I было показано, что выборочное пространство последовательностей $SS\cdots F\cdots$ можно отобразить на единичный интервал, пользуясь простым приемом замены символов S и F на 1 и 0 соответственно. Выборочным пространством становится тогда единичный интервал, и результат бесконечной последовательности испытаний представляется случайной величиной $\mathbf{Y} = \sum 2^{-k}\mathbf{X}_k$, где $\mathbf{X}_k$ — независимые случайные величины, принимающие

[1]) Хотя принципиально сингулярные распределения очень важны, они появляются в этой книге лишь между прочим.

значения 1 и 0 с вероятностями p и q. Обозначим распределение $\mathbf{Y}$ через F_p. Если испытания симметричны, то $F_p = F_{\frac{1}{2}}$ есть *равномерное* распределение, и модель принимает привлекательный простой вид. По существу эквивалентность симметричных испытаний Бернулли со «случайным выбором точки из $\overline{0,1}$» использовалась с самого возникновения теории вероятностей. Далее в силу закона больших чисел распределение F_p сосредоточено на множестве N_p, состоящем из тех точек, в двоичном разложении которых частота цифры 1 стремится к p. При $p \neq \alpha$ множество N_α имеет вероятность нуль и, следовательно, *распределения F_p сингулярны по отношению друг к другу*. В частности, при $p \neq \frac{1}{2}$ распределение F_p сингулярно по отношению к равномерному распределению dx. Точного представления для F_p практически не существует, и поэтому описанная конструкция редко используется при $p \neq \frac{1}{2}$. Два обстоятельства заслуживают особого внимания.

Во-первых, рассмотрим, что было бы, если бы частный случай $p = \frac{1}{3}$ представлял особый интерес или часто встречался в приложениях. Мы заменили бы тогда двоичное разложение чисел троичным и ввели бы новый масштаб так, чтобы $F_{\frac{1}{3}}$ представляло собой равномерное распределение. «В практике» мы по-прежнему имели бы дело лишь с абсолютно непрерывными распределениями, однако в основе этого лежит скорее выбор подхода, чем истинная природа вещей.

Во-вторых, симметричность или несимметричность монеты можно проверить статистически и практически определенного решения можно достигнуть после некоторого конечного числа испытаний. Это оказывается возможным лишь благодаря тому, что события, являющиеся правдоподобными при гипотезе $p = \frac{1}{2}$, весьма неправдоподобны при гипотезе $p = \frac{1}{3}$. Небольшое размышление показывает, что возможность принять решение после конечного числа испытаний основана на сингулярности F_p при $p \neq \frac{1}{2}$ по отношению к $F_{\frac{1}{2}}$. Таким образом, существование сингулярных распределений является существенным для статистической практики.

б) *Случайные направления.* В гл. I, 10 было введено понятие единичного вектора в $\mathscr{R}^2$ со случайным направлением. Распреде-

ление такого вектора сосредоточено на единичной окружности и, следовательно, является сингулярным по отношению к мере Лебега (площади) в плоскости. Можно было бы думать, что в этом случае в качестве выборочного пространства следует взять окружность, однако в некоторых практических задачах такой вариант выборочного пространства невозможен [см. пример (4, д)]. ▶

Возвращаясь к общей теории, предположим, что F есть вероятностное распределение, не являющееся абсолютно непрерывным по отношению к мере U. Множество N, для которого $U\{N\}=0$, будем называть для краткости нулевым множеством. Пусть p есть верхняя грань значений $F\{N\}$ по всем нулевым множествам N. Для каждого n существует нулевое множество N_n, такое, что $F\{N_n\} > p - \frac{1}{n}$. При этом $F\{A\} \leqslant \frac{1}{n}$ для всякого нулевого множества A, лежащего в дополнении N_n'. Для множества $N = \bigcup N_n$ имеем $U\{N\}=0$ и $F\{N\}=p$, так что никакое нулевое множество, лежащее в дополнении N', не может иметь положительной вероятности. Положим $q=1-p$ и определим две новые меры из равенств

$$p \cdot F_s\{A\} = F\{AN\},\quad q \cdot F_{ac}\{A\} = F\{AN'\}. \tag{3.9}$$

Если $q=0$, то $F=F_s$; в противном случае обе меры F_s и F_{ac} имеют полную массу, равную 1 и, следовательно, представляют собой вероятностные распределения. По самому построению распределение F_s сингулярно относительно U, в то время как $U\{B\}=0$ влечет $F_{ac}\{B\}=0$. Складывая обе формулы (3.9), мы получаем, таким образом, следующий результат.

Теорема Лебега о разложении. *Каждое вероятностное распределение F является смесью вида*

$$F = pF_s + qF_{ac}, \tag{3.10}$$

где $p \geqslant 0$, $q \geqslant 0$, $p+q=1$ *двух вероятностных распределений* F_s и F_{ac}, *из которых* F_s *сингулярно, а* F_{ac} *абсолютно непрерывно по отношению к заданной мере* U.

Применяя к F_s жорданово разложение (2.5), получаем, что F можно представить в виде смеси трех вероятностных распределений, из которых первое — атомическое, второе — абсолютно непрерывное по отношению к $U\{dx\}$, а третье непрерывно, но сингулярно.

§ 4. Свертки

Трудно переоценить важность операции свертки во многих областях математики. Нам придется иметь дело со сверткой с двух точек зрения: и как с операцией, применяемой к распре-

делениям, и как с операцией, применяемой к распределению и непрерывной функции.

Для определенности мы будем рассматривать распределения в $\mathscr{R}^1$, однако при использовании векторных обозначений параграфа 1 все формулы остаются в силе и для распределений в пространстве любой размерности. Определение *свертки на окружности* следует схеме, описанной в гл. II, 8, и не требует дополнительных пояснений. (Более общие свертки можно определить на произвольных группах.)

Пусть F — вероятностное распределение и φ — ограниченная функция точек. (В рассматриваемых нами приложениях φ будет либо непрерывной функцией, либо функцией распределения.) Определим новую функцию u, положив

$$u(x) = \int_{-\infty}^{\infty} \varphi(x-y)\, F\{dy\}. \tag{4.1}$$

Если F имеет плотность f (по отношению к dx), то

$$u(x) = \int_{-\infty}^{\infty} \varphi(x-y)\, f(y)\, dy. \tag{4.2}$$

Условие ограниченности φ является более жестким, чем это необходимо, однако при плохих подинтегральных функциях интегралы могут потерять смысл.

Определение 1. *Функция u в* (4.1) *называется сверткой функции φ с вероятностным распределением F. Для свертки будет использоваться обозначение $u = F * \varphi$. Если F имеет плотность f, то мы будем также писать $u = f * \varphi$.*

Отметим, что порядок компонент в свертке существен: символ $\varphi * F$, вообще говоря, смысла не имеет. С другой стороны, интеграл в формуле (4.2) имеет смысл для любых интегрируемых функций f и φ (даже если f может принимать значения разных знаков) и символ $*$ используется в этом обобщенном смысле. Ограниченность функции φ была предположена, конечно, лишь ради простоты и вовсе не является необходимой.

Примеры. а) Если F — равномерное распределение на $\overline{0,a}$, то

$$u(x) = \frac{1}{a} \int_{x-a}^{x} \varphi(s)\, ds. \tag{4.3}$$

Ясно, что в этом случае функция u непрерывна; если предположить еще, что функция φ непрерывна, то u будет обладать не-

прерывной производной и т. д. Вообще говоря, u ведет себя лучше, чем φ, так что свертка является *сглаживающим оператором.*

б) Формулы свертки для экспоненциальных и равномерных распределений [гл. I, (3.6) и гл. I, (9.1)] являются частными случаями общей формулы. Примеры сверток распределений в $\mathscr{R}^2$ имеются в задачах 14—16 гл. III, см. также гл. III, (1.22). ▶

Теорема 1. *Если функция φ ограничена и непрерывна, то свертка $u=F*\varphi$ тоже ограничена и непрерывна. Если φ — вероятностное распределение, то и тоже является вероятностным распределением.*

Доказательство. Если φ — ограниченная непрерывная функция, то $\varphi(x+h-y)\to\varphi(x-y)$ при $h\to 0$, и поэтому, согласно принципу мажорированной сходимости, $u(x+h)\to u(x)$. Если φ является функцией распределения, то сказанное выше остается верным при стремлении h к нулю справа. Таким образом, функция u в этом случае непрерывна справа и, поскольку она монотонно меняется от 0 до 1, она действительно является функцией распределения. ▶

Следующая теорема дает вероятностную интерпретацию свертки $F*\varphi$, когда φ есть функция распределения.

Теорема 2. *Пусть* $\mathbf{X}$ *и* $\mathbf{Y}$ *— независимые случайные величины с распределениями F и G. Тогда*

$$\mathbf{P}\{\mathbf{X}+\mathbf{Y}\leqslant t\}=\int_{-\infty}^{\infty}G(t-x)F\{dx\}. \tag{4.4}$$

Доказательство[1]). Возьмем $\varepsilon>0$ и обозначим через I_n интервал $n\varepsilon<x\leqslant(n+1)\varepsilon$, $n=0,\pm1,\ldots$. Появление при некотором n события $\{\mathbf{X}\in I_{n-1},\ \mathbf{Y}\leqslant t-n\varepsilon\}$ влечет появление события $\{\mathbf{X}+\mathbf{Y}\leqslant t\}$. Следовательно, поскольку события $\{\mathbf{X}\in I_{n-1},\ \mathbf{Y}\leqslant t-n\varepsilon\}$ взаимно исключают друг друга и случайные величины $\mathbf{X}$, $\mathbf{Y}$ независимы, имеем

$$\mathbf{P}\{\mathbf{X}+\mathbf{Y}\leqslant t\}\geqslant\sum G(t-n\varepsilon)F\{I_{n-1}\}. \tag{4.5}$$

Справа в этой формуле стоит интеграл от ступенчатой функции G_ε, которая на I_{n-1} равна $G(t-n\varepsilon)$. Поскольку $G_\varepsilon(y)\geqslant$

[1]) Формула (4.4) представляет собой частный случай теоремы Фубини [гл. IV, (2.11)], и доказательство ее приводится здесь лишь ради иллюстрации. Утверждение, обратное теореме 2, неверно: мы видели в гл. II, (4, д) и в примере 1 из гл. III, 9, что в исключительных случаях формула (4.4) может иметь место и для пары зависимых случайных величин $\mathbf{X}$ и $\mathbf{Y}$.

$\geqslant G(t-\varepsilon-y)$, то

$$\mathbf{P}\{\mathbf{X}+\mathbf{Y}\leqslant t\}\leqslant \int_{-\infty}^{\infty} G(t-\varepsilon-x)F\{dx\}. \tag{4.6}$$

Аналогичное рассуждение приводит к противоположному неравенству с ε, замененным на $-\varepsilon$. Устремляя ε к нулю, получаем (4.4). ▶

Пример. в) Пусть распределение F и G сосредоточены на множестве целых неотрицательных чисел 0, 1, 2, ..., причем соответствующие им массы точки k равны p_k и q_k. В этом случае интеграл в формуле (4.4) сводится к сумме $\sum G(t-k)p_k$. Функция, представляемая этой суммой, обращается в нуль при $t<0$ и постоянна на каждом интервале $n-1<t<n$. Скачок ее при $t=n$ равен

$$\sum_{k=0}^{n} q_{n-k}p_k = q_n p_0 + q_{n-1}p_1 + \cdots + q_0 p_n, \tag{4.7}$$

что согласуется с формулой свертки [1, гл. XI, (2.1)] для неотрицательных целочисленных случайных величин. ▶

Обе предыдущие теоремы показывают, что свертка двух функций распределения $F*G$ снова приводит к функции распределения U. Из коммутативности сложения $\mathbf{X}+\mathbf{Y}$ вытекает, что $F*G=G*F$. При совершенной системе обозначений возможно был бы введен новый символ для свертки функций распределения, однако это вряд ли было бы полезным [1]). Конечно, U следует рассматривать как функцию интервалов или меру:

---

[1]) Иными словами, символ $A*B$ используется, когда интегрирование происходит по мере A. Свертка $A*B$ является функцией точки или мерой в соответствии с тем, является ли B функцией точки [как в (4.1)] или мерой [как в (4.6)]. Звездочка $*$ используется для операции между двумя функциями, когда интегрирование происходит по лебеговой мере. Мы применяем этот тип свертки почти исключительно к вероятностным плотностям.

Можно дать более общее определение свертки двух функций, положив

$$f*g(x)=\int_{-\infty}^{\infty} f(x-y)\,g(y)\,m\{dy\},$$

где m — произвольная мера. Суммы вида (4.7) представляют собой частный случай этой формулы, а именно, когда m сосредоточена на множестве неотрицательных целых чисел и каждому такому числу сопоставлена единичная масса. В этом смысле наше теперешнее использование звездочки $*$ согласуется с ее использованием для сверток последовательностей в 1, гл. XI, 2.

для любого интервала $I=\overline{a, b}$ имеем, очевидно,

$$U\{I\} = \int_{-\infty}^{\infty} G\{I-y\}\, F\{dy\}, \tag{4.8}$$

где, как обычно, $I-y$ обозначает интервал $\overline{a-y, b-y}$. (Эта формула автоматически переносится на произвольные множества.) В силу коммутативности F и G в формуле (4.8) можно поменять местами.

Рассмотрим теперь три распределения F_1, F_2, F_3. Из ассоциативности сложения случайных величин следует, что $(F_1 * F_2) * F_3 = F_1 * (F_2 * F_3)$. Следовательно, при обозначении свертки трех распределений можно не пользоваться скобками и писать просто $F_1 * F_2 * F_3$. Резюмируем сказанное выше в следующих теоремах.

Теорема 3. *В классе распределений операция свертки коммутативна и ассоциативна.*

Теорема 4. *Если распределение G непрерывно (т. е. не имеет атомов), то свертка $U=F*G$ тоже непрерывна. Если G имеет обычную плотность φ, то U имеет обычную плотность u, задаваемую формулой* (4.1).

(В силу симметрии F и G можно поменять ролями.)

Доказательство. Первое утверждение содержится в теореме 1. Если φ есть плотность распределения G, то, интегрируя (4.1) по интервалу I, справа получим правую часть (4.8), так что u действительно является плотностью распределения U, определяемого формулой (4.8) [1]). ▶

Суммы $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ взаимно независимых случайных величин с общей функцией распределения F встречаются очень часто, и поэтому для их распределения удобно ввести специальное обозначение. Распределение суммы $\mathbf{S}_n$ есть *n-кратная свертка распределения F с собой* и обозначается F^{n*}. Имеем, таким образом

$$F^{1*} = F, \quad F^{(n+1)*} = F^{n*} * F. \tag{4.9}$$

Сумму без слагаемых условно считают равной нулю. В соответствии с этим мы определим F^{0*} как атомическое распределение, сосредоточенное в нуле. Тогда (4.9) будет выполняться и при $n=0$.

[1]) Из критерия (3.7) непосредственно вытекает, что из абсолютной непрерывности G по отношению к мере V следует и абсолютная непрерывность U по отношению к V. Более того, для U критерий выполняется с тем же самым δ.

Если F имеет плотность f, то F^{n*} имеет плотность $f*f*\ldots*f$ (n раз), которую мы *обозначим* f^{n*}. Эти обозначения согласуются с обозначениями, введенными в гл. I, 2.

Приведем в заключение две леммы[1]), которые будут использоваться главным образом в теории восстановления и теории случайных блужданий.

Лемма 1. *Если a и b — точки роста распределений F и G, то $a+b$ есть точка роста свертки $F*G$. Если a и b — атомы F и G, то $a+b$ есть атом $F*G$ и каждый атом $F*G$ может быть так получен.*

Доказательство. Для независимых $\mathbf{X}$ и $\mathbf{Y}$ имеем

$$\mathbf{P}\{|\mathbf{X}+\mathbf{Y}-a-b|<\varepsilon\}\geqslant \mathbf{P}\left\{|\mathbf{X}-a|<\tfrac{1}{2}\varepsilon\right\}\cdot\mathbf{P}\left\{|\mathbf{Y}-b|<\tfrac{1}{2}\varepsilon\right\}.$$

Если a и b суть точки роста, то правая часть этого неравенства положительна при любом $\varepsilon>0$, и поэтому $a+b$ тоже является точкой роста $F*G$.

Обозначим F_a и G_a атомические компоненты F и G в жордановом разложении (2.5). Очевидно, атомическая компонента композиции $F*G$ равна F_a*G_a, и поэтому все атомы $F*G$ имеют вид $a+b$, где a и b суть атомы F и G соответственно. ▷

Внутренняя простота следующей леммы немало теряет из-за той специальной роли, которую играют в ней, с одной стороны, арифметические распределения, а с другой, — распределения *положительных* случайных величин.

Лемма 2. *Пусть F — распределение в $\mathscr{R}^1$ и Σ — множество, образованное точками роста распределение F, F^{2*}, F^{3*}, ...* .

а) *Предположим, что F не сосредоточено на полуоси. Тогда если F не является арифметическим распределением, то Σ плотно в $-\infty$, ∞, а если F — арифметическое распределение с шагом λ, то $\Sigma=\{0, \pm\lambda, \pm 2\lambda, \ldots\}$.*

б) *Пусть F сосредоточено на $[0, \infty)$, но не сосредоточено в нуле. Если F не является арифметическим распределением, то Σ «асимптотически плотно на ∞» в том смысле, что для любого заданного $\varepsilon>0$ при всех достаточно больших x интервал $(x, x+\varepsilon)$ содержит точки из Σ. Если F — арифметическое распределение с шагом λ, то Σ содержит все точки вида $n\lambda$ при достаточно больших n.*

Доказательство. Пусть a и b — две точки множества Σ, такие, что $0<a<b$. Из леммы 1 вытекает, что множество Σ со-

[1]) Оставшуюся часть этого параграфа можно пропустить при первом чтении.

держит все точки вида $ma+nb$, где m, $n=0, 1, \ldots$. Точки ma, $(m-1)a+b$, ..., mb (всего их $m+1$) расположены на одинаковом расстоянии друг от друга, равном $b-a$. Ясно, что если $mb>(m+1)a$, то каждый подинтервал интервала $\overline{ma, (m+1)a}$ длины $\rho>b-a$ содержит точку множества Σ. Иными словами, если Σ содержит пару точек $a>0$, $b>0$, отстоящих друг от друга на расстояние, меньшее ρ, то при достаточно больших x каждый интервал x, $x+\rho$ содержит точку из Σ. В соответствии с этим либо Σ асимптотически плотно на ∞, либо точки Σ, принадлежащие $\overline{0, \infty}$, образуют простую последовательность $a_1<a_2<\ldots$, такую, что $a_{n+1}-a_n \to \delta>0$. Но если некоторая точка a принадлежит Σ, то все точки a_n+a принадлежат Σ, и поэтому a есть целое кратное δ (положительное или отрицательное). Следовательно, Σ содержится в множестве точек 0, $\pm\delta$, $\pm 2\delta$, ..., так что F является арифметическим распределением с шагом $\lambda=k\delta$, где k — некоторое целое число. Но тогда Σ автоматически содержится в множестве точек 0, $\pm\lambda$, $\pm 2\lambda$, ... и поэтому $\delta=\lambda$. Из сказанного вытекает теперь, что Σ содержит все точки $n\lambda$ при достаточно больших n. Тем самым пункт (б) теоремы доказан.

Если Σ содержит как положительные, так и отрицательные точки, то приведенное рассуждение применимо также к левой полуоси, и поэтому в случае неарифметического распределения F множество Σ асимптотически плотно и на $-\infty$ и на $+\infty$. Далее если a и b — точки из Σ, лежащие соответственно в ε-окрестностях точек $x+\frac{1}{2}t$ и $x-\frac{1}{2}t$, то $a+b$ лежит в 2ε окрестности t, так что каждая окрестность любой точки t содержит точки из Σ. В случае арифметического распределения F аналогичное рассуждение показывает, что все целые кратные λ принадлежат Σ и теорема доказана. ▶

Один частный случай этой теоремы представляет особый интерес. Каждое число $x>0$ можно представить единственным образом в виде $x=m+\xi$, где m — целое и $0\leqslant\xi<1$. Число ξ в этом представлении называется дробной долей x. Рассмотрим теперь распределение, сосредоточенное в двух точках -1 и $\alpha>0$. Соответствующее множество Σ, включая в себя все точки вида $n\alpha-m$ [1]), содержит дробные доли всех чисел α, $2\alpha, \ldots$. Если $\alpha=p/q$, где p и q — взаимно простые натуральные числа, то F — арифметическое распределение с шагом $1/q$. Мы приходим, таким образом, к следующему результату (в гл. VIII, 6 будет доказана теорема о равномерном распределении, уточняющая этот результат).

[1]) Множество Σ здесь в точности совпадает с множеством точек $n\alpha-m$. $m, n=0,1, \ldots$. — *Прим. перев.*

Следствие. *Для любого иррационального $\alpha>0$ дробные доли чисел α, 2α, 3α, ... образуют множество, плотное в $\overline{0,1}$.*
(Если $\alpha=p/q$, то множество дробных долей чисел α, 2α, 3α, ... состоит из чисел 0, $1/q$, ..., $(q-1)/q$.) ▶

Следующие примеры показывают, что *свертка двух сингулярных распределений может иметь непрерывную плотность.* Из этих примеров видно также, что можно эффективно вычислять свертки, не используя определяющую их формулу.

Примеры. *г) Равномерное распределение на $\overline{0,1}$ является сверткой двух сингулярных распределений канторовского типа.* В самом деле, пусть $\mathbf{X}_1$, $\mathbf{X}_2$, ... — взаимно независимые случайные величины, принимающие значения 0 и 1 с вероятностью $\frac{1}{2}$. В примере (11, в) гл. I мы видели, что случайная величина $\mathbf{X}=\Sigma 2^{-k}\mathbf{X}_k$ имеет равномерное распределение. Обозначим сумму членов ряда с четными и нечетными номерами через $\mathbf{U}$ и $\mathbf{V}$ соответственно. Очевидно, $\mathbf{U}$ и $\mathbf{V}$ независимы и $\mathbf{X}=\mathbf{U}+\mathbf{V}$. Следовательно, равномерное распределение есть композиция распределений величин $\mathbf{U}$ и $\mathbf{V}$. Ясно далее, что величины $\mathbf{U}$ и $2\mathbf{V}$ имеют одинаковые распределения, а величина $\mathbf{V}$ лишь обозначением отличается от величины $\frac{1}{3}\mathbf{Y}$ примера (11, г) гл. I. Иными словами, распределения величин $\mathbf{U}$ и $\mathbf{V}$ лишь масштабными множителями отличаются от рассмотренного в этом примере канторовского распределения.

д) *Случайное блуждание в $\mathcal{R}^2$.* Распределение единичного вектора со случайным направлением (см. гл. I, 10) сосредоточено на единичной окружности и, следовательно, сингулярно относительно лебеговой меры на плоскости. Тем не менее *длина* $\mathbf{L}$ суммы двух таких независимых векторов *имеет распределение с плотностью* $\frac{2}{\pi}\frac{1}{\sqrt{4-r^2}}$, сосредоточенное на $\overline{0,2}$. В самом деле, по закону косинусов $\mathbf{L}=\sqrt{2-2\cos\omega}=\left|2\sin\frac{1}{2}\omega\right|$, где $\pi-\omega$ — угол между двумя векторами. Так как ω имеет равномерное распределение на $\overline{0,2\pi}$, то

$$\mathbf{P}\{\mathbf{L}\leqslant r\}=\mathbf{P}\left\{\left|2\sin\frac{1}{2}\omega\right|\leqslant r\right\}=\frac{2}{\pi}\arcsin\frac{1}{2}r,\qquad 0<r<2, \tag{4.10}$$

что и требовалось доказать. ▶

§ 5. Симметризация

Пусть $\mathbf{X}$ — случайная величина с распределением F. Распределение случайной величины $-\mathbf{X}$ мы будем обозначать ^{-}F. В точках непрерывности имеем

$$^{-}F(x)=1-F(x), \tag{5.1}$$

и это соотношение однозначно определяет ^{-}F. Распределение F называется *симметричным*, если $^{-}F=F$. [В случае когда существует плотность f, симметричность означает, что $f(-x)=f(x)$.]

Пусть $\mathbf{X}_1$, $\mathbf{X}_2$ — независимые случайные величины с одним и тем же распределением F. Случайная величина $\mathbf{X}_1 - \mathbf{X}_2$ имеет симметричное распределение ${}^0F$, равное

$$ {}^0F = F * {}^-F. \tag{5.2} $$

Из свойства симметрии $F^0(x) = 1 - {}^0F(x-)$ непосредственно следует, что

$$ {}^0F(x) = \int_{-\infty}^{\infty} F(x+y)\, F\{dy\}. \tag{5.3} $$

О распределении ${}^0F$ говорят, что оно получено в результате *симметризации* F.

Примеры. а) Симметризация показательного распределения приводит к двустороннему показательному распределению [гл. II, 4(а)]; результатом симметризации равномерного на $\overline{0, 1}$ распределения является треугольное распределение τ_2 (см. гл. II, (4.1).

б) Симметричное распределение, имеющее атомы массы $\frac{1}{2}$ в точках ± 1, не является результатом симметризации никакого распределения.

в) Пусть F — атомическое распределение, наделяющее точки $0, 1, \ldots$ массами $p_0, p_1, \ldots$. Тогда симметризованное распределение ${}^0F$ тоже является атомическим, имеющим в точках $\pm n$ массу q_n, где

$$ q_n = \sum_{k=0}^{\infty} p_k p_{k+n}, \quad n \geqslant 0, \tag{5.4} $$

$q_{-n} = q_n$. В частности, для *пуассоновского* распределения F при $n \geqslant 0$ имеем

$$ q_n = e^{-2\alpha} \sum_{k=0}^{\infty} \frac{\alpha^{n+2k}}{k!\,(n+k)!} = e^{-2\alpha} I_n(2\alpha), \tag{5.5} $$

где I_n — *функция Бесселя*, определенная в гл. II, (7.1) (см. задачу 9). ▶

Применение симметризации позволяет значительно упростить многие рассуждения. В связи с этим весьма важным является то обстоятельство, что хвосты распределений F и ${}^0F$ имеют сравнимую величину. Более точно это выражено в нижеследующих неравенствах. Смысл этих неравенств представляется более ясным,

когда они выражены в терминах случайных величин, а не в терминах самих функций распределения.

Лемма 1. *Неравенства симметризации. Если* $\mathbf{X}_1, \mathbf{X}_2$ — *независимые и одинаково распределенные случайные величины, то при* $t>0$

$$\mathbf{P}\{|\mathbf{X}_1-\mathbf{X}_2|>t\}\leqslant 2\mathbf{P}\left\{|\mathbf{X}_1|>\frac{1}{2}t\right\}. \tag{5.6}$$

Если $a\geqslant 0$ *таково, что* $\mathbf{P}\{\mathbf{X}_i>a\}\leqslant 1-p$ *и* $\mathbf{P}\{\mathbf{X}_i<-a\}\leqslant 1-p$, *то*

$$\mathbf{P}\{|\mathbf{X}_1-\mathbf{X}_2|>t\}\geqslant p\mathbf{P}\{|\mathbf{X}_1|>t+a\}. \tag{5.7}$$

В частности, когда медиана $\mathbf{X}_j$ *равна нулю, имеем*

$$\mathbf{P}\{|\mathbf{X}_1-\mathbf{X}_2|>t\}\geqslant\frac{1}{2}\mathbf{P}\{|\mathbf{X}_1|>t\}. \tag{5.8}$$

Доказательство. Неравенство (5.6) является следствием того очевидного обстоятельства, что событие $|\mathbf{X}_1-\mathbf{X}_2|>t$ осуществляется лишь тогда, когда осуществляется хотя бы одно из событий $|\mathbf{X}_1|>\frac{1}{2}t$, $|\mathbf{X}_2|>\frac{1}{2}t$. Для доказательства (5.7) достаточно заметить, что каждое из непересекающихся событий $\mathbf{X}_1>t+a$, $\mathbf{X}_2\leqslant a$ и $\mathbf{X}_1<-t-a$, $\mathbf{X}_2\geqslant -a$ влечет событие $|\mathbf{X}_1-\mathbf{X}_2|\geqslant t$. ▶

Симметризация часто используется при оценке сумм независимых случайных величин. В этой связи особенно полезно следующее неравенство.

Лемма 2. *Если* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — *независимые случайные величины с симметрическими распределениями, то сумма* $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ *тоже имеет симметрическое распределение и*

$$\mathbf{P}\{|\mathbf{X}_1+\ldots+\mathbf{X}_n|>t\}\geqslant\frac{1}{2}\mathbf{P}\{\max|\mathbf{X}_j|>t\}. \tag{5.9}$$

Если $\mathbf{X}_j$ имеют одинаковое распределение F, то во всех точках непрерывности

$$\mathbf{P}\{\max|\mathbf{X}_j|\leqslant t\}=(F(t)-F(-t))^n\leqslant e^{-n[1-F(t)+F(-t)]} \tag{5.10}$$

и, следовательно, согласно (5.9), при всех $t>0$

$$\mathbf{P}\{|\mathbf{X}_1+\ldots+\mathbf{X}_n|\geqslant t\}\geqslant\frac{1}{2}(1-e^{-n[1-F(t)+F(-t)]}). \tag{5.11}$$

Доказательство. Пусть $\mathbf{M}$ — случайная величина, равная первой из величин среди случайных величин $\mathbf{X}_1, \ldots, \mathbf{X}_n$, имеющих наибольшее значение по абсолютной величине. Поло-

жим $\mathbf{T}=\mathbf{S}-\mathbf{M}$. Двумерная случайная величина $(\mathbf{M}, \mathbf{T})$ имеет симметричное распределение в том смысле, что все четыре случайные величины $(\pm\mathbf{M}, \pm\mathbf{T})$ имеют одинаковое распределение. Имеем, очевидно,

$$\mathbf{P}\{\mathbf{M}>t\}\leqslant\mathbf{P}\{\mathbf{M}>t,\ \mathbf{T}\geqslant 0\}+\mathbf{P}\{\mathbf{M}>t,\ \mathbf{T}\leqslant 0\}. \tag{5.12}$$

Слагаемые справа в (5.12) равны по величине, и поэтому

$$\mathbf{P}\{\mathbf{S}>t\}=\mathbf{P}\{\mathbf{M}+\mathbf{T}>t\}\geqslant\mathbf{P}\{\mathbf{M}>t,\ \mathbf{T}\geqslant 0\}\geqslant\frac{1}{2}\mathbf{P}\{\mathbf{M}>t\}. \tag{5.13}$$

что равносильно (5.9). ▸

§ 6. Интегрирование по частям. Существование моментов

Известную формулу интегрирования по частям можно применять также к интегралам по распределениям в $\mathscr{R}^1$. *Если функция u ограничена и имеет непрерывную производную u', то*

$$\int_a^{b+} u(x)\,F\{dx\}=u(b)\,F(b)-u(a)\,F(a)-\int_a^b u'(x)\,F(x)\,dx. \tag{6.1}$$

В самом деле, рассмотрим разность между двумя сторонами (пока еще не доказанного) равенства (6.1) как функцию от верхнего предела t и обозначим ее $\varphi(t)$. Производя элементарную выкладку, получаем, что при $h>0$

$$\varphi(t+h)-\varphi(t)=\int_t^{(t+h)+}[u(x)-u(t+h)]\,F\{dx\}-\int_t^{t+h}u'(x)\,[F(t)-F(x)]\,dx. \tag{6.2}$$

Используя (6.2), покажем, что $\varphi(t+h)-\varphi(t)=o(h)$ при $h\to 0$. Пусть $|u'|<m$ в окрестности точки t. Тогда $|u(t+h)-u(t)|\leqslant mh$ и, следовательно, первый интеграл в (6.2) есть $o(h)$. Второй интеграл в (6.2) тоже есть $o(h)$, поскольку $F(t)-F(x)\to 0$ при $h\to 0$. Аналогичное рассуждение применимо при $h<0$, и поэтому функция φ в точке t имеет производную, равную нулю. Так как t произвольно, то тем самым доказано, что разность между левой и правой частями (6.1) не зависит от верхнего предела и, стало быть, равна нулю.

В качестве приложения выведем одну часто используемую формулу [обобщающую 1, гл. XI, (1.8)].

Лемма 1. *Для любого* $\alpha>0$

$$\int_0^\infty x^\alpha F\{dx\} = \alpha \int_0^\infty x^{\alpha-1}[1-F(x)]\,dx, \tag{6.3}$$

причем из сходимости одного из этих интегралов следует сходимость другого.

Доказательство. Поскольку интегрирование в (6.3) происходит по бесконечному интервалу, мы не можем применять (6.1) непосредственно, однако после тривиальных преобразований при $b<\infty$ из (6.1) следует, что

$$\int_0^b x^\alpha F\{dx\} = -b^\alpha[1-F(b)] + \alpha \int_0^b x^{\alpha-1}[1-F(x)]\,dx. \tag{6.4}$$

Предположим сначала, что интеграл слева сходится при $b\to\infty$. Вклад в предельный интеграл по бесконечному промежутку от интегрирования по b, ∞ не меньше $b^\alpha[1-F(b)]$, так что первое слагаемое в правой части (6.4) стремится к нулю. Переходя теперь в (6.4) к пределу при $b\to\infty$, получаем (6.3). С другой стороны, в (6.4) интеграл слева не превосходит интеграла справа и, следовательно, из сходимости правого интеграла вытекает сходимость левого, что, как мы видели выше, приводит к (6.3). ▶

Для левой полуоси имеет место соотношение, аналогичное (6.3). Комбинируя обе формулы, получаем следующий результат.

Лемма 2. *Распределение F обладает абсолютным моментом порядка* $\alpha>0$ *тогда и только тогда, когда функция* $|x|^{\alpha-1}[1-F(x)+F(-x)]$ *интегрируема на* $-\infty, \infty$.

В качестве приложения покажем, что имеет место

Лемма 3. *Пусть* $\mathbf{X}, \mathbf{Y}$ *— независимые случайные величины, и* $\mathbf{S}=\mathbf{X}+\mathbf{Y}$. *Тогда* $\mathbf{E}(|\mathbf{S}|^\alpha)$ *существует тогда и только тогда, когда существуют* $\mathbf{E}(|\mathbf{X}|^\alpha)$ и $\mathbf{E}(|\mathbf{Y}|^\alpha)$.

Доказательство. Поскольку случайные величины $\mathbf{X}$ и $\mathbf{X}-c$ обладают абсолютными моментами одних и тех же порядков, то, не ограничивая общности, можно предположить, что медианы величин $\mathbf{X}$ и $\mathbf{Y}$ равны нулю. Но тогда $\mathbf{P}\{|\mathbf{S}|>t\}\geqslant \frac{1}{2}\mathbf{P}\{|\mathbf{X}|>t\}$, и поэтому в силу предыдущей леммы $\mathbf{E}(|\mathbf{S}|^\alpha)<\infty$ влечет $\mathbf{E}(|\mathbf{X}|^\alpha)<\infty$. Тем самым доказана необходимость. Достаточность следует из неравенства $|\mathbf{S}|^\alpha\leqslant 2^\alpha(|\mathbf{X}|^\alpha+|\mathbf{Y}|^\alpha)$, справедливость которого вытекает из того очевидного факта, что во всех точках $|\mathbf{S}|$ не превосходит наибольшего из чисел $2|\mathbf{X}|$ и $2|\mathbf{Y}|$. ▶

§ 7. Неравенство Чебышева

Неравенство Чебышева является одним из наиболее часто используемых инструментов теории вероятностей. Как само неравенство, так и его доказательство не отличаются от их дискретного варианта (1, гл. IX, 6) и повторяются здесь главным образом для удобства ссылок. Интересные приложения неравенства будут даны в гл. VII, 1.

Неравенство Чебышева. *Если существует* $\mathbf{E}(\mathbf{X}^2)$, *то*

$$\mathbf{P}\{|\mathbf{X}| \geqslant t\} \leqslant \frac{1}{t^2}\mathbf{E}(\mathbf{X}^2), \qquad t > 0. \tag{7.1}$$

В частности, если $\mathbf{E}(\mathbf{X}) = m$ *и* $\mathrm{Var}(\mathbf{X}) = \sigma^2$, *то*

$$\mathbf{P}\{|\mathbf{X} - m| \geqslant t\} \leqslant \frac{1}{t^2}\sigma^2. \tag{7.2}$$

Доказательство. Обозначим распределение величины $\mathbf{X}$ через F. Тогда

$$\mathbf{E}(\mathbf{X}^2) \geqslant \int_{|x| \geqslant t} x^2 F\{dx\} \geqslant t^2 \int_{|x| \geqslant t} F\{dx\}.$$

Поделив эти неравенства на t^2, получаем (7.1). ▶

Полезность неравенства Чебышева определяется не его точностью (она невелика), а простотой, и тем, что оно особенно удобно в применении к суммам случайных величин. Среди многих возможных обобщений неравенства Чебышева ни одно не обладает этими ценными его качествами. (Большинство из обобщений весьма просты и их лучше выводить по мере надобности. Например, полезная комбинация неравенства Чебышева и урезания распределений описана в гл. VII, 7.)

Приведем один довольно общий метод получения нетривиальных неравенств. Если $u \geqslant 0$ на всей прямой и $u(x) > a > 0$ при всех x из некоторого интервала I, то

$$F\{I\} \leqslant a^{-1}\mathbf{E}(u(\mathbf{X})). \tag{7.3}$$

С другой стороны, если $u \leqslant 0$ вне I и $u \leqslant 1$ на I, то мы получаем обратное неравенство $F\{I\} \geqslant \mathbf{E}(u(\mathbf{X}))$. Беря в качестве u многочлены, приходим к неравенствам, зависящим лишь от моментов F.

Примеры. а) Пусть $u(x) = (x+c)^2$, где $c > 0$. Тогда $u(x) \geqslant 0$ при всех x и $u(x) \geqslant (t+c)^2$ при $x \geqslant t > 0$. Следовательно,

$$\mathbf{P}\{\mathbf{X} > t\} \leqslant \frac{1}{(t+c)^2}\mathbf{E}((\mathbf{X}+c)^2). \tag{7.4}$$

При $\mathbf{E}(\mathbf{X}) = 0$ и $\mathbf{E}(\mathbf{X}^2) = \sigma^2$ правая часть в (7.4) достигает минимума, когда $c = \sigma^2/t$, и мы получаем следующее *общее неравенство*:

$$\mathbf{P}\{\mathbf{X} > t\} \leqslant \frac{\sigma^2}{\sigma^2 + t^2}, \qquad t > 0, \tag{7.5}$$

которое было обнаружено независимо многими авторами.

б) Пусть $\mathbf{X}$ — положительная случайная величина (т. е. $F(0) = 0$) и $\mathbf{E}(\mathbf{X}) = 1$, $\mathbf{E}(\mathbf{X}^2) = b$. Многочлен $u(x) = h^{-2}(x-a)(a+2h-x)$ положителен, лишь

когда $a<x<a+2h$ и $u(x)\leqslant 1$ при всех x. Если $0<a<1$, то, как легко видеть, $\mathbf{E}(u(X))\geqslant[2h(1-a)-b]h^{-2}$. Беря $h=b(1-a)^{-1}$ и используя замечание, предшествовавшее примерам, получаем

$$\mathbf{P}\{\mathbf{X}>a\}\geqslant(1-a)^2 b^{-1}. \tag{7.6}$$

в) Пусть $\mathbf{E}(\mathbf{X}^2)=1$ и $\mathbf{E}(\mathbf{X}^4)=M$. Применяя последнее неравенство к $\mathbf{X}^2$, находим, что при $0<t<1$

$$\mathbf{P}\{|\mathbf{X}|>t\}\geqslant(1-t^2)^2 M^{-1}. \tag{7.7}$$

По поводу обобщения, принадлежащего Колмогорову, см. гл. VII, 8. ▶

§ 8. Дальнейшие неравенства. Выпуклые функции

Собранные в этом параграфе неравенства широко используются в математике и ни в коей мере не являются специфическими для теории вероятностей. Наиболее общее из них — неравенство Шварца. Другие неравенства приведены по причине их применений в теории случайных процессов и математической статистике. (Настоящий параграф предназначается главным образом не для чтения, а для ссылок.)

Неравенство Шварца

В вероятностной интерпретации это неравенство означает, что для всяких двух случайных величин φ и ψ, определенных на одном и том же выборочном пространстве,

$$(\mathbf{E}(\varphi\psi))^2\leqslant\mathbf{E}(\varphi^2)\,\mathbf{E}(\psi^2), \tag{8.1}$$

если только входящие в неравенство математические ожидания существуют. Равенство в (8.1) осуществляется только тогда, когда некоторая линейная комбинация $a\varphi+b\psi$ величин φ и ψ равна нулю с вероятностью единица. Более общим образом, если F — произвольная мера на множестве A, то

$$\left(\int_A \varphi(x)\psi(x)F\{dx\}\right)^2\leqslant\int_A\varphi^2(x)F\{dx\}\int_A\psi^2(x)F\{dx\}, \tag{8.2}$$

каковы бы ни были функции φ и ψ, для которых интегралы справа существуют. Беря в качестве F атомическую меру, наделяющую единичными массами целочисленные точки, получаем *неравенство Шварца для сумм*

$$(\sum\varphi_i\psi_i)^2\leqslant\sum\varphi_i^2\sum\psi_i^2. \tag{8.3}$$

Принимая во внимание важность неравенства (8.1), мы дадим два его доказательства, намечающие различные обобщения. Аналогичные доказательства можно дать для (8.2) и (8.3).

Первое доказательство. Можно предположить, что $\mathbf{E}(\psi^2)>0$. Тогда математическое ожидание

$$\mathbf{E}(\varphi+t\psi)^2=\mathbf{E}(\varphi^2)+2t\mathbf{E}(\varphi\psi)+t^2\mathbf{E}(\psi^2) \tag{8.4}$$

является квадратным многочленом по t Этот многочлен, будучи неотрицательным, имеет либо два комплексных корня, либо двойной действительный корень λ. Из формулы для решения квадратного уравнения следует, что в первом случае (8.1) выполняется как строгое неравенство. Во втором случае (8.1) является равенством и $\mathbf{E}(\varphi+\lambda\psi)^2=0$, так что $\varphi+\lambda\psi=0$, за исключением множества вероятности нуль.

Второе доказательство. Поскольку при замене функций φ и ψ в (8.1) на их кратные $a\varphi$ и $b\psi$, где a, b — некоторые числа, получается равносильное неравенство, то достаточно доказать (8.1) в случае, когда $\mathbf{E}(\varphi^2)=\mathbf{E}(\psi^2)=1$. Для получения же (8.1) в этом случае достаточно взять математические ожидания от обеих частей неравенства $2|\varphi\psi|\leqslant\varphi^2+\psi^2$. ▶

Выпуклые функции. Неравенство Иенсена

Пусть u — функция, определенная на открытом интервале I, и $P=(\xi,\ u(\xi))$ — точка ее графика. Проходящая через P прямая L называется *опорной прямой функцией u в точке* ξ, если график u целиком лежит над или на L. (Это условие исключает вертикальные прямые.) На аналитическом языке это означает, что при всех x из I

$$u(x)\geqslant u(\xi)+\lambda(x-\xi), \tag{8.5}$$

где λ — тангенс угла наклона L. Функция u называется *выпуклой в интервале I, если в каждой точке x из I существует ее опорная прямая.* (Функция u называется вогнутой, если выпукла функция $-u$.)

Покажем сначала, что из этого определения вытекают различные свойства выпуклых функций, которые интуитивно ассоциируются нами с выпуклостью (исходя из примера выпуклых ломаных линий).

Пусть $\mathbf{X}$ — случайная величина с распределением F, сосредоточенным на I, и с математическим ожиданием $\mathbf{E}(\mathbf{X})$. Положив $\xi=\mathbf{E}(\mathbf{X})$ и взяв математическое ожидание от обеих частей неравенства (8.5), получим

$$\mathbf{E}(u)\geqslant u(\mathbf{E}(\mathbf{X})), \tag{8.6}$$

если только математическое ожидание слева существует. Неравенство (8.6) известно как *неравенство Иенсена.*

Наиболее важным является тот случай, когда F сосредоточено в двух точках x_1 и x_2 и имеет в них массы $1-t$ и t. При

этом (8.6) принимает вид

$$(1-t)u(x_1)+tu(x_2) \geqslant u((1-t)x_1+tx_2). \tag{8.7}$$

Это неравенство допускает простую геометрическую интерпретацию, которую мы сформулируем в виде следующей теоремы.

Теорема 1. *Функция u выпукла тогда и только тогда, когда все хорды ее графика лежат не ниже самого графика.*

Доказательство. i) *Необходимость.* Пусть u — выпуклая функция, и рассмотрим хорду под произвольным интервалом x_1, x_2. Как показано выше, при всех $0 \leqslant t \leqslant 1$ выполняется неравенство (8.7). Когда t меняется от 0 до 1 точка $(1-t)x_1+tx_2$ пробегает интервал x_1, x_2, а левая часть (8.7) представляет собой ординату хорды в точке $(1-t)x_1+tx_2$. Таким образом, (8.7) означает, что точки хорды лежат не ниже графика u.

ii) *Достаточность.* Пусть u обладает указанным в формулировке теоремы свойством, и рассмотрим треугольник, образованный тремя точками P_1, P_2, P_3 на графике u с абсциссами $x_1<x_2<x_3$. Точка P_2 лежит ниже хорды P_1P_3, и из трех сторон треугольника хорда P_1P_2 имеет наименьший наклон, а хорда P_2P_3 — наибольший. Следовательно, вне интервала x_2, x_3 график u лежит над прямой P_2P_3. Будем считать теперь, что x_3 — переменная точка и пусть $x_3 \to x_2+0$. Наклон хорды P_2P_3 при этом монотонно убывает, оставаясь ограниченным снизу наклоном хорды P_1P_2. Таким образом, прямые P_2P_3 стремятся к предельной прямой L, проходящей через P_2. Поскольку вне x_2x_3 график u лежит над прямой P_2P_3, то весь график u лежит не ниже L. Следовательно, L является опорной прямой для u в точке x_2 и, поскольку точка x_2 произвольна, выпуклость u тем самым доказана. ➢

Как предел хорд, прямая L является правой касательной. В рассмотренном предельном процессе абсцисса x_3 точки P_3 стремится к x_2, а сама точка P_3 — к некоторой точке на прямой L. Следовательно, $P_3 \to P_2$. Аналогичное рассуждение применимо при приближении к x_2 слева, и поэтому график u непрерывен и обладает правой и левой касательной в каждой точке. Далее эти касательные являются опорными прямыми, и их наклоны меняются монотонно. Поскольку монотонная функция имеет не более счетного множества точек разрыва, то нами доказана

Теорема 2. *Выпуклая функция имеет правую и левую производные в каждой точке. Эти производные являются монотон-*

но не убывающими функциями, совпадающими друг с другом всюду, за исключением, быть может, счетного множества точек.

Свойства выпуклых функций, указанные в теореме 2, являются также и достаточными для выпуклости. В частности, если u обладает второй производной, то она выпукла тогда и только тогда, когда $u''\geqslant 0$.

Обычно в качестве определения выпуклости функции берется (8.7). При $t=\frac{1}{2}$ из (8.7) получаем

$$u\left(\frac{x_1+x_2}{2}\right)\leqslant\frac{1}{2}u(x_1)+\frac{1}{2}u(x_2). \tag{8.8}$$

Геометрически (8.8) означает, что *середина* хорды лежит не ниже соответствующей точки графика u. Если функция u непрерывна, то отсюда вытекает, что график не может пересечь хорду и, следовательно, u выпукла. Можно показать, что вообще *любая бэровская функция*[1]), *обладающая свойством* (8.8), *выпукла.*

Неравенства для моментов

Покажем, что для любой случайной величины $\mathbf{X}$ *функция*

$$u(t)=\log \mathbf{E}(|\mathbf{X}|^t),\quad t\geqslant 0, \tag{8.9}$$

выпукла по t в каждом интервале, где она конечна. В самом деле, согласно неравенству Шварца (8.1),

$$\mathbf{E}^2(|\mathbf{X}|^t)\leqslant\mathbf{E}(|\mathbf{X}|^{t+h})\,\mathbf{E}(|\mathbf{X}|^{t-h}),\qquad 0\leqslant h\leqslant t. \tag{8.10}$$

если только математические ожидания существуют. Положив в (8.10) $x_1=t-h$, $x_2=t+h$ и прологарифмировав, видим, что функция u удовлетворяет условию (8.8) и поэтому выпукла.

Поскольку $u(0)\leqslant 0$, то тангенс угла наклона $t^{-1}u(t)$ прямой, соединяющей начало координат с точкой $(t,\ u(t))$, меняется монотонно и, следовательно, $(\mathbf{E}(|\mathbf{X}|^t))^{1/t}$ *является неубывающей функцией от* $t>0$.

Неравенство Гёльдера

Пусть $p>1$, $q>1$ *и* $p^{-1}+q^{-1}=1$. *Тогда для неотрицательных функций* φ *и* ψ *имеет место неравенство*

$$\mathbf{E}(\varphi\psi)\leqslant(\mathbf{E}(\varphi^p))^{1/p}(\mathbf{E}(\psi^q))^{1/q}, \tag{8.11}$$

если только математические ожидания существуют.

---

[1]) Каждая функция u, удовлетворяющая условию (8.8), либо выпукла, либо она колеблется на каждом интервале от $-\infty$ до ∞. См. Харди Г. Г., Литтльвуд Дж. Е. и Полиа Г., Неравенства, ИЛ, М., 1948, особенно стр. 114.

(Неравенство (8.1) является частным случаем этого неравенства, получающимся из него при $p=q=1/2$. Для неравенств (8.2) и (8.3) имеются обобщения, аналогичные (8.11).)

Доказательство. При $x>0$ функция $u=\log x$ вогнута, т. е. она удовлетворяет неравенству, обратному (8.7). Потенцируя это неравенство, приходим к соотношению

$$x_1^{1-t}x_2^t \leqslant (1-t)x_1 + tx_2, \tag{8.12}$$

справедливому при $x_1, x_2>0$. Как при доказательстве неравенства Шварца (второе доказательство), чтобы установить (8.11), достаточно ограничиться случаем, когда $\mathbf{E}(\varphi^p)=\mathbf{E}(\psi^q)=1$. Положим $t=q^{-1}$ и $1-t=p^{-1}$. Подставив в (8.12) $x_1=\varphi^p$, $x_2=\psi^q$ и взяв математические ожидания, получаем $\mathbf{E}(\varphi\psi)\leqslant 1$, что и требовалось доказать. ▶

Неравенство Колмогорова

Неравенство Колмогорова рассматривалось в **1**, IX, 7, и данное там его доказательство переносится без изменений на случай произвольных распределений. Усиленный вариант этого неравенства (для мартингалов) приводится в гл. VII, 8.

§ 9. Простые условные распределения. Смеси

В гл. III, 2 мы определили «условную плотность случайной величины **Y** при фиксированном значении другой случайной величины **X**» в том случае, когда совместное распределение **X** и **Y** имеет непрерывную плотность. Не претендуя на общность, мы введем теперь аналогичное понятие для более широкого класса распределений. (Систематически теория условных распределений и математических ожиданий развивается в § 10 и 10а.)

Для произвольной пары интервалов A, B на прямой положим

$$Q(A, B) = \mathbf{P}\{\mathbf{X}\in A, \mathbf{Y}\in B\}. \tag{9.1}$$

При этом обозначении для *частного распределения* **X** имеем

$$\mu\{A\} = Q(A, \mathscr{R}^1). \tag{9.2}$$

Если $\mu\{A\}>0$, то условная вероятность события $\{\mathbf{Y}\in B\}$ при условии $\{\mathbf{X}\in A\}$ равна

$$\mathbf{P}\{\mathbf{Y}\in B \mid \mathbf{X}\in A\} = \frac{Q(A, B)}{\mu\{A\}}. \tag{9.3}$$

(Если $\mu\{A\}=0$, то эта условная вероятность не определена.) Воспользуемся формулой (9.3), когда A есть интервал $A_h=[x, x+h)$. Пусть $h\to 0+$. При выполнении соответствующих

условий регулярности предел

$$q(x, B) = \lim_{h \to 0} \frac{Q(A_h, B)}{\mu\{A_h\}} \tag{9.4}$$

существует для всех x и B. В этом случае по аналогии с гл. III, 2 мы обозначаем

$$q(x, B) = \mathbf{P}\{\mathbf{Y} \in B \mid \mathbf{X} = x\} \tag{9.5}$$

и называем q «*условной вероятностью события* $\{\mathbf{Y} \in B\}$ *при условии* $\mathbf{X} = x$». Рассмотренная конструкция распространяет понятие условных вероятностей на случаи, когда «гипотезы» имеют нулевую вероятность. Никаких дальнейших трудностей не возникает, если функция q достаточно регулярна, однако мы не будем заниматься изучением соответствующих условий регулярности, поскольку в следующем параграфе рассматривается более общая схема. В частных случаях обычно оказывается достаточным простой наивный подход, и вид условного распределения нередко можно найти с помощью интуитивных рассуждений.

Пример. а) Пусть $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины с распределениями F и G соответственно. Предположим для простоты, что $\mathbf{X} > 0$ [т. е. $F(0) = 0$]. Рассмотрим произведение $\mathbf{Z} = \mathbf{XY}$. Имеем, очевидно,

$$\mathbf{P}\{\mathbf{Z} \leqslant t \mid \mathbf{X} = x\} = G\left(\frac{t}{x}\right). \tag{9.6}$$

Интегрируя (9.6) по F, получим функцию распределения U величины $\mathbf{Z}$ [см. гл. II, (3.1). Это утверждение представляет собой частный случай формулы (9.8), см. ниже]. В частности, когда величина $\mathbf{X}$ распределена равномерно на интервале $\overline{0, 1}$, получаем

$$U(t) = \int_0^1 G\left(\frac{t}{x}\right) dx. \tag{9.7}$$

Формула (9.7) может быть использована как удобная исходная точка в теории одновершинных распределений [1]). ▶

---

[1]) Функция распределения U называется *одновершинной* с модой в нуле, если на интервале $\overline{-\infty, 0}$ она выпукла, а на интервале $\overline{0, \infty}$ — вогнута [см. (8, б)]. Нуль может быть точкой разрыва U, но вне нуля одновершинность предполагает существование у U *плотности и, которая монотонна в обоих интервалах* $\overline{-\infty, 0}$ *и* $\overline{0, \infty}$. (Интервалы постоянства не исключаются.)

Имеется хорошо известная характеризация одновершинных распределений, принадлежащая А. Хинчину. Как заметил Л. Шепп, (9.7) представляет собой упрощенный вариант формулы Хинчина, допускающий простую вероятностную интерпретацию.

Теорема. *Функция распределения* U *одновершинна тогда и только тогда, когда ее можно представить в виде* (9.7). *Иными словами,* U *одно-*

Дальнейшие примеры содержатся в задачах 13—14. ▶

При выполнении соответствующих условий регулярности функция $q(x, B)$ при фиксированном x является вероятностным распределением по B и при фиксированном B представляет собой непрерывную функцию по x. При этом

$$Q(A, B) = \int_A q(x, B)\, \mu\{dx\}. \tag{9.8}$$

В самом деле, стоящий справа интеграл задает вероятностное распределение на плоскости, и, дифференцируя его по типу (9.4), получаем $q(x, B)$. Формула (9.8) показывает, как заданное распределение на плоскости можно представить в терминах условного и частного распределений. Пользуясь терминологией гл. II, 5, можно сказать, что формула (9.8) дает представление заданного распределения в виде *смеси* зависящего от параметра x семейства распределений $q(x, B)$ и распределения μ, играющего роль распределения рандомизированного параметра.

вершинна тогда и только тогда, когда она является функцией распределения произведения $\mathbf{Z}=\mathbf{XY}$ *двух независимых случайных величин, из которых одна величина, скажем* $\mathbf{X}$, *распределена равномерно на* $\overline{0, 1}$.

Доказательство. Возьмем $h>0$ и обозначим U_h функцию распределения, график которой является кусочно-линейной функцией, совпадающей с U в точках $0, \pm h, \ldots$ [т. е. $U_h(nh)=U(nh)$ и функция U_h линейна в интервалах $\overline{nh, (n+1)h}$]. Из определения одновершинности непосредственно вытекает, что функция U одновершинна тогда и только тогда, когда одновершинны функции U_h. Далее U_h имеет плотность u_h, являющуюся ступенчатой функцией, и каждую ступенчатую функцию с разрывами в точках nh можно записать в виде

$$\sum p_n \frac{1}{|n|\,h} f\left(\frac{x}{nh}\right), \tag{*}$$

где $f(x)=1$ при $0<x<1$ и $f(x)=0$ в остальных точках. Функция (*) монотонна в интервалах $\overline{-\infty, 0}$ и $\overline{0, \infty}$ тогда и только тогда, когда $p_n \geqslant 0$ при всех n, и является плотностью, если $\Sigma p_n=1$. Но когда $p_n \geqslant 0$ и $\Sigma p_n=1$ (*) представляет собой плотность распределения произведения $\mathbf{Z}_h=\mathbf{XY}_h$ — двух независимых случайных величин $\mathbf{X}$ и $\mathbf{Y}_h$, из которых $\mathbf{X}$ распределена равномерно на $\overline{0, 1}$, а $\mathbf{Y}_h$ имеет арифметическое распределение $\mathbf{P}\{\mathbf{Y}_h=nh\}=p_n$. Мы доказали, таким образом, что функция U_h одновершинна тогда и только тогда, когда ее можно представить в виде (9.7) с функцией G, замененной на функцию G_h, сосредоточенную в точках $0, \pm h, \ldots$ Устремляя теперь h к нулю и используя теорему о монотонной сходимости, получаем нужный результат. (См. задачи 20—21 и задачу 8 из гл. XV, 9.)

В практике описанный процесс нередко обращается. В качестве *исходного* берут «*вероятностное* (*стохастическое*) *ядро*» q, т. е. функцию $q(x, B)$ от точки x и множества B, такую, что при фиксированном x она является вероятностным распределением, а при фиксированном B — бэровской функцией. При любом заданном вероятностном распределении μ интеграл в (9.8) определяет вероятности подмножеств плоскости вида (A, B) и тем самым определяет некоторое вероятностное распределение на плоскости. Обычно формулу (9.8) выражают в терминах функций точек. Рассмотрим семейство функций распределения $G(\theta, y)$, зависящих от параметра θ, и некоторое вероятностное распределение μ. Определим новую функцию распределения с помощью следующей формулы:

$$U(y) = \int_{-\infty}^{\infty} G(x, y)\,\mu\{dx\}. \tag{9.9}$$

[Эта формула представляет собой частный случай формулы (9.8); она получается из нее при $A = -\infty, \infty$ и $q(x, -\infty, y) = G(x, y)$.] Смеси подобного рода встречались нам уже в **1**, гл. V и обсуждались в гл. II, 5. В следующем параграфе будет показано, что q всегда можно интерпретировать как условное распределение вероятностей.

Примеры. б) Для всяких двух распределений F_1 и F_2 распределение $pF_1 + (1-p)F_2 (0<p<1)$ является их смесью и представляется формулой (9.9) с распределением μ, сосредоточенным в двух точках.

в) *Случайные суммы.* Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — независимые случайные величины с одинаковым распределением F. Пусть далее $\mathbf{N}$ — независимая от $\mathbf{X}_j$ случайная величина, принимающая значения $0, 1, \ldots$ с положительными вероятностями $p_0, p_1, \ldots$. Рассмотрим случайную величину $\mathbf{S}_\mathbf{N} = \mathbf{X}_1 + \ldots + \mathbf{X}_\mathbf{N}$. Условное распределение $\mathbf{S}_\mathbf{N}$ при условии $\mathbf{N} = n$ есть F^{n*}, и поэтому распределение $\mathbf{S}_\mathbf{N}$ дается формулой

$$U = \sum_{n=0}^{\infty} p_n F^{n*}, \tag{9.10}$$

представляющей собой частный случай (9.9). В рассматриваемом случае каждое условие $\mathbf{N} = n$ имеет положительную вероятность p_n, и поэтому мы имеем здесь дело с условными вероятностными распределениями в строгом смысле. Другие примеры имеются в гл. II, 5—7. (См. задачи 16 и 19.)

§ 10 *). Условные распределения

Вряд ли было бы разумным исследовать точные условия, при которых условные вероятности q могут быть получены с помощью процесса дифференцирования типа (9.4). Основные свойства условных вероятностей выражены в соотношении (9.8), представляющем вероятности множеств в терминах условных вероятностей, и проще всего использовать (9.8) в качестве *определения* условных вероятностей. Соотношение (9.8) не определяет q однозначно, поскольку, если для каждого множества B имеем $q(x, B) = \bar{q}(x, B)$ с точностью до множества μ-меры нуль, то (9.8) по-прежнему остается справедливым при замене q на $\bar{q}$. Однако указанная неопределенность неизбежна. Например, когда μ сосредоточено на интервале I, невозможно никакое естественное определение q для x, не лежащих в I. В силу самой природы вещей мы имеем дело со всем классом эквивалентных условных вероятностей и должны говорить скорее о *каком-то*, а не об определенном условном распределении вероятностей q. В конкретных же случаях требования регулярности диктуют обычно некоторый определенный выбор условного распределения вероятностей.

Для определенности мы будем рассматривать лишь события, задаваемые условиями типа $\mathbf{X} \in A$ и $\mathbf{Y} \in B$, где $\mathbf{X}$ и $\mathbf{Y}$ — фиксированные случайные величины и A, B — борелевские множества на прямой. Посмотрим сначала, как можно понимать фразу «условная вероятность события $\{\mathbf{Y} \in B\}$ относительно $\mathbf{X}$». Значение $\mathbf{X}$ может быть либо фиксированным числом, либо быть неопределенным. При второй из этих интерпретаций мы имеем функцию от $\mathbf{X}$, т. е. случайную величину, которую будем обозначать $\mathbf{P}\{B \mid \mathbf{X}\}$ или $q(\mathbf{X}, B)$ и т. д. Когда имеется в виду фиксированное значение $\mathbf{X}$, скажем x, будет использоваться обозначение $\mathbf{P}\{\mathbf{Y} \in B \mid \mathbf{X} = x\}$ или $q(x, B)$.

Определение 1. *Пусть B — фиксированное множество. «Условной вероятностью события $\{\mathbf{Y} \in B\}$ относительно $\mathbf{X}$» $\mathbf{P}\{\mathbf{Y} \in B \mid \mathbf{X}\}$ называется функция $q(\mathbf{X}, B)$, такая, что для любого подмножества A прямой*

$$\mathbf{P}\{\mathbf{X} \in A,\ \mathbf{Y} \in B\} = \int_A q(x,\ B)\,\mu\{dx\}, \tag{10.1}$$

где μ — частное распределение $\mathbf{X}$.

*) Этот параграф можно опустить при первом чтении.

Когда x является атомом распределения μ, условие $\mathbf{X}=x$ имеет положительную вероятность и условная вероятность $\mathbf{P}\{\mathbf{Y}\in B\,|\,\mathbf{X}=x\}$ была в этом случае определена нами раньше формулой (9.3), в которой под A следует понимать множество, состоящее из единственной точки x. При этом формула (10.1) сводится к (9.3), так что наши прежние и настоящие определения и обозначения согласуются друг с другом.

Из теоремы Радона — Никодима (§ 3) вытекает, что условная вероятность $\mathbf{P}\{\mathbf{Y}\in B\,|\,\mathbf{X}\}$ всегда существует. В самом деле, рассмотрим вероятность

$$\mathbf{P}\{\mathbf{X}\in A,\ \mathbf{Y}\in B\}$$

при фиксированном B как функцию множества A. Эта функция лишь нормировкой отличается от вероятностного распределения. Она абсолютно непрерывна по отношению к μ, поскольку при всех A вероятность $\mathbf{P}\{\mathbf{X}\in A,\ \mathbf{Y}\in B\}$ не превосходит $\mu\{A\}$. Следовательно, рассматриваемая функция множеств обладает плотностью q по отношению к μ, то есть имеет место (10.1).

До сих пор множество B было у нас фиксированным, но обозначение $q(x, B)$ было выбрано с тем расчетом, чтобы впоследствии сделать B переменным. Иными словами, мы хотим теперь рассматривать q как функцию от двух переменных — точки x и подмножества B прямой. При этом желательно, чтобы функция q была вероятностной мерой. Это означает, что $q(x, \mathscr{R}^1)=1$ и что для любой последовательности непересекающихся множеств $B_1, B_2, \ldots,$ в объединении дающих B, имеет место равенство

$$q(x,\ B)=\sum q(x,\ B_k). \tag{10.2}$$

Если стоящие справа члены представляют собой условные вероятности B_k, то их сумма является *некоторой* условной вероятностью множества B, однако у нас имеется дополнительное требование согласованности, состоящее в том, чтобы (10.2) выполнялось при нашем выборе q при *всех* x. [Отметим, что определение 1 не исключает абсурдных значений функции $q(x, B)$ в некоторых точках x типа $q(x, B)=17$.] Как нетрудно видеть, функцию $q(x, B)$, действительно можно выбрать так, чтобы выполнялись все эти требования [1]). Это означает, что *существует*

[1]) Проще всего построить такую функцию q, определяя *непосредственно* значения $q(x, B)$ лишь для множеств B, являющихся интервалами в диадическом подразбиении $\mathscr{R}^1$. Пусть, например, $B_1=\overline{0,\ \infty}$, $B_2=\overline{-\infty,\ 0}$. Возьмем в качестве $q(x, B_1)$ любую условную вероятность B_1, подчиненную очевидному требованию $0\leqslant q(x, B_1)\leqslant 1$. Теперь автоматически следует положить $q(x, B_2)=1-q(x, B_1)$. Разобьем далее B_1 на B_{11} и B_{12} и выберем $q(x, B_{11})$

условное распределение вероятностей **Y** *относительно* **X** в смысле следующего определения.

Определение 2. *Условным вероятностным распределением* **Y** *относительно* **X** *называется функция* q, *зависящая от двух переменных — точки* x *и множества* B, *такая, что*

i) *при фиксированном множестве* B

$$q(\mathbf{X}, B) = \mathbf{P}\{\mathbf{Y} \in B \mid \mathbf{X}\}, \tag{10.3}$$

т. е. $q(\mathbf{X}, B)$ *является условной вероятностью события* $\{\mathbf{Y} \in B\}$ *относительно* **X**,

ii) *при каждом фиксированном* x q *есть вероятностное распределение.*

В действительности условное вероятностное распределение представляет собой семейство обычных вероятностных распределений, и поэтому вся теория переносится на них без изменений. Так, *когда* q *задано*[1]), следующее определение вводит скорее новое обозначение, чем новое понятие.

Определение 3. *Условное математическое ожидание* $\mathbf{E}(\mathbf{Y} \mid \mathbf{X})$ *есть функция от* **X**, *принимающая в точке* x *значение*

$$\mathbf{E}(\mathbf{Y} \mid x) = \int_{-\infty}^{\infty} y q(x, dy). \tag{10.4}$$

При этом предполагается, что интеграл сходится (за исключением, возможно, x*-множества нулевой вероятности).*

Условное математическое ожидание $\mathbf{E}(\mathbf{Y} \mid \mathbf{X})$ является функцией от **X** и, следовательно, представляет собой случайную величину. Для большей ясности значение его в фиксированной точке x иногда предпочитают обозначать $\mathbf{E}(\mathbf{Y} \mid \mathbf{X}=x)$. Из определения $\mathbf{E}(\mathbf{Y} \mid \mathbf{X})$ непосредственно вытекает, что

$$\mathbf{E}(\mathbf{Y}) = \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{Y} \mid x) \mu\{dx\}. \tag{10.5}$$

так, чтобы выполнялись неравенства $0 \leqslant q(x, B_{11}) \leqslant q(x, B_1)$. Положим $q(x, B_{12}) = q(x, B_1) - q(x, B_{11})$. Процесс этот продолжается далее аналогично с бесконечным измельчением подразбиений. Требование аддитивности (10.2) позволяет затем определить $q(x, B)$ для всех открытых множеств B и, следовательно, для всех борелевских множеств.

Описанная конструкция основывается лишь на существовании так называемой *сети*, т. е. разбиении пространства на конечное число непересекающихся множеств, каждое из которых в свою очередь разбивается на конечное число непересекающихся множеств, причем процесс продолжается таким образом, что каждая точка пространства является единственным пределом стягивающейся последовательности множеств, входящих в последовательные разбиения. Рассматриваемое утверждение имеет место, следовательно, в $\mathcal{R}^2$ и во многих других пространствах.

[1]) Более гибкое общее определение приводится в § 10 а.

Многомерный случай.
Повторные математические ожидания

Введенное для простоты предположение одномерности случайной величины $\mathbf{X}$ не играло серьезной роли в предыдущих рассмотрениях, и они сохраняют силу для векторной случайной величины в $\mathscr{R}^n$. При $n=2$ мы получим условное математическое ожидание $\mathbf{E}(\mathbf{Z}|\mathbf{X}, \mathbf{Y})$.

Условное математическое ожидание $\mathbf{E}(\mathbf{Z}|\mathbf{X}, \mathbf{Y})$ представляет собой функцию от двух случайных величин $\mathbf{X}$ и $\mathbf{Y}$, и мы можем образовать ее условное математическое ожидание относительно $\mathbf{X}$, т. е. $\mathbf{E}(\mathbf{E}(\mathbf{Z}|\mathbf{X}, \mathbf{Y})|\mathbf{X})$. Естественно ожидать, что это повторное математическое ожидание совпадает с $\mathbf{E}(\mathbf{Z}|\mathbf{X})$. Из определения (10.6) (см. ниже) непосредственно вытекает, что совпадение действительно имеет место. Для частного случая. нормального распределения эти формулы уже обсуждались в гл. III.

10 * а. Условные математические ожидания

Условные математические ожидания были определены нами в терминах условных вероятностных распределений. Этот подход вполне удовлетворителен, когда имеют дело лишь с одной парой случайных величин $\mathbf{X}$, $\mathbf{Y}$. В теории случайных процессов, однако, приходится иметь дело с целыми семействами случайных величин, и неединственность условных распределений приводит там к серьезным трудностям. Оказывается, что можно построить более гибкую теорию, определяя условные математические ожидания независимо от условных распределений. Для понимания этой теории лучше всего начать с более внимательного рассмотрения формулы (10.5).

Пусть A — борелевское множество *на прямой* и $\mathbf{1}_A(\mathbf{X})$ — случайная величина, равная 1 при $\mathbf{X}\in A$ и нулю в остальных случаях. Согласно (10.4),

$$\mathbf{E}(\mathbf{Y}\mathbf{1}_A(X)) = \int_A \mathbf{E}(\mathbf{Y}|x)\,\mu\{dx\}. \tag{10.6}$$

Поскольку каждую ограниченную функцию v можно равномерно приблизить линейными комбинациями индикаторов, то применение соответствующего предельного перехода позволяет вывести из (10.6) более общую формулу

$$\mathbf{E}(\mathbf{Y}\cdot v(\mathbf{X})|\mathbf{X}) = v(\mathbf{X})\,\mathbf{E}(\mathbf{Y}|\mathbf{X}). \tag{10.7}$$

*) Содержание этого пункта будет использоваться лишь в связи с мартингалами в гл. VI, 12 и гл. VII, 8.

Грубо говоря, эта формула означает, что при известном $\mathbf{X}$ множитель $v(\mathbf{X})$ ведет себя как постоянная и может быть вынесен за знак интеграла.

Формула (10.6) заключает в себе все основные свойства условного математического ожидания. Исходя из нее как из *определения*, можно совсем отказаться от более сложного понятия условного распределения. Можно пойти еще дальше и придать понятию условного математического ожидания еще большую гибкость, отказавшись даже от задающей условие случайной величины $\mathbf{X}$. В самом деле, величина $\mathbf{X}$ использовалась нами главным образом для определения событий вида $\{\mathbf{X}\in A\}$, где A — борелевское множество на прямой. Пока мы имеем дело с парой случайных величин $(\mathbf{X}, \mathbf{Y})$, мы не теряем общности, беря в качестве выборочного пространства плоскость и рассматривая $\mathbf{X}$ и $\mathbf{Y}$ как координатные переменные. Событие $\{a<\mathbf{X}<b\}$ представляет собой в этом случае бесконечную полосу, а $\{\mathbf{X}\in A\}$ есть множество, состоящее из прямых, параллельных оси игреков. Класс всех таких множеств в плоскости образует σ-алгебру $\mathfrak{B}$, содержащуюся в σ-алгебре всех плоских борелевских множеств. Эта σ-алгебра $\mathfrak{B}$ является наименьшей σ-алгеброй, содержащей все множества вида $\{\mathbf{X}\leqslant a\}$, и называется σ-алгеброй, *порожденной величиной* $\mathbf{X}$. (Данное определение применимо ко всем случайным величинам.) Положим $B=\{\mathbf{X}\in A\}$. Тогда произведение $\mathbf{Y}\cdot\mathbf{1}_A(\mathbf{X})$ совпадает, очевидно, с произведением $\mathbf{Y}\cdot\mathbf{1}_B$. Формула (10.6) означает теперь, что случайная величина $\mathbf{U}=\mathbf{E}(\mathbf{Y}|\mathbf{X})$ удовлетворяет соотношению

$$\mathbf{E}(\mathbf{Y}\cdot\mathbf{1}_B)=\mathbf{E}(\mathbf{U}\cdot\mathbf{1}_B) \tag{10.8}$$

для любого множества B из σ-алгебры, порожденной $\mathbf{X}$.

До сих пор мы занимались лишь переформулировкой соотношения (10.6), которое было получено с помощью интегрирования по условному вероятностному распределению. Но соотношение (10.8) имеет смысл в более общей ситуации, и мы можем использовать его в качестве *определения* условного математического ожидания произвольной случайной величины $\mathbf{Y}$ относительно произвольной σ-алгебры $\mathfrak{B}$ множеств. Так определенное условное математическое ожидание является случайной величиной и будет обозначаться $\mathbf{E}(\mathbf{Y}|\mathfrak{B})$. Схема нового подхода состоит в следующем.

Пусть имеется произвольное вероятностное пространство с σ-алгеброй $\mathfrak{A}$ и $\mathfrak{B}$ — произвольная σ-алгебра, *содержащаяся в* $\mathfrak{A}$. *Пусть далее* $\mathbf{Y}$ — *случайная величина, имеющая математическое ожидание. Случайная величина* $\mathbf{U}$ *называется условным математическим ожиданием* $\mathbf{Y}$ *относительно* $\mathfrak{B}$ *и обозначается*

$\mathbf{U}=\mathbf{E}(\mathbf{Y}\mid\mathfrak{B})$, *если она* $\mathfrak{B}$*-измерима*[1]) *и удовлетворяет соотношению* (10.8) *при всех* $B\in\mathfrak{B}$.

Условие $\mathfrak{B}$-измеримости заменяет собой прежнее требование, чтобы $\mathbf{U}$ было функцией от $\mathbf{X}$. Если $\mathfrak{B}$ есть σ-алгебра, порожденная $\mathbf{X}$, то новое определение согласуется с (10.6), однако даже в этом случае у него есть то преимущество, что оно применимо при любом вероятностном пространстве и нет необходимости отображать вероятностное пространство в плоскость.

Примеры. Если $\mathfrak{B}=\mathfrak{A}$, то можно взять $\mathbf{U}=\mathbf{Y}$. Если $\mathfrak{B}$ содержит лишь все пространство и пустое множество, то $\mathbf{U}=\mathbf{E}(\mathbf{Y})$. Когда величина $\mathbf{Y}$ $\mathfrak{B}$-измерима, можно положить $\mathbf{U}=\mathbf{Y}$. ▶

Случайную величину $\mathbf{U}$ можно менять на любом множестве $N\in\mathfrak{B}$ нулевой вероятности, и обозначение $\mathbf{E}(\mathbf{Y}\mid\mathfrak{B})$ обычно относится к некоторому произвольному, но фиксированному представителю класса всех условных математических ожиданий.

Существование условного математического ожидания доказывается совсем просто. Если $\mathbf{Y}$ — неотрицательная случайная величина, то левая часть (10.8) определяет меру $\Phi(B)$ на $\mathfrak{B}$, для которой, очевидно, $\Phi(B)=0$ всякий раз, когда $\mathbf{P}\{B\}=0$. Отсюда, согласно абстрактному варианту теоремы Радона — Никодима, вытекает, что мера Φ имеет плотность U, а соотношение (10.8) именно это и означает. Случай произвольной случайной величины $\mathbf{Y}$ сводится к только что рассмотренному с помощью обычного представления $\mathbf{Y}$ в виде разности двух неотрицательных случайных величин.

Подобно тому как из (10.6) вытекало (10.7), из (10.8) получается следующая общая формула

$$\mathbf{E}(\mathbf{YZ}\mid\mathfrak{B})=\mathbf{Z}\mathbf{E}(\mathbf{Y}\mid\mathfrak{B}), \tag{10.9}$$

для справедливости которой нужно лишь, чтобы величина $\mathbf{Z}$ была $\mathfrak{B}$-измерима и условные математические ожидания существовали. В частности, если $\mathbf{Z}$ $\mathfrak{B}$-измерима, то $\mathbf{E}(\mathbf{Z}\mid\mathfrak{B})=\mathbf{Z}$.

Другим важным свойством условных математических ожиданий является следующее. Если $\mathfrak{B}_0$ есть σ-алгебра, содержащаяся в $\mathfrak{B}$, и $\mathbf{U}_0=\mathbf{E}(\mathbf{Y}\mid\mathfrak{B}_0)$, то $\mathbf{E}(\mathbf{Y}\cdot\mathbf{1}_B)=\mathbf{E}(\mathbf{U}\cdot\mathbf{1}_B)=\mathbf{E}(\mathbf{U}_0\cdot\mathbf{1}_B)$ при всех $B\in\mathfrak{B}_0$. Следовательно, согласно самому определению, если $\mathfrak{B}_0\subset\mathfrak{B}$, то $\mathbf{E}(\mathbf{Y}\mid\mathfrak{B}_0)=\mathbf{E}(\mathbf{E}(\mathbf{Y}\mid\mathfrak{B})\mid\mathfrak{B}_0)$. Поскольку величина $\mathbf{U}_0$ $\mathfrak{B}$-измерима, то имеем также $\mathbf{E}(\mathbf{Y}\mid\mathfrak{B}_0)=\mathbf{E}(\mathbf{E}(\mathbf{Y}\mid\mathfrak{B}_0)\mid\mathfrak{B})$.

Главные свойства обычных математических ожиданий, такие, как линейность и справедливость основных неравенств для них,

---

[1]) Это означает, что множества $\{\mathbf{U}\leqslant a\}$ принадлежат $\mathfrak{B}$; см. гл. IV, 4.

тривиальным образом распространяются на условные математические ожидания.

Основы общей теории условных распределений и математических ожиданий были заложены Колмогоровым. В дальнейшем эта теория была усовершенствована Дубом, которому принадлежат последние определения.

§ 11. Задачи

1. Пусть $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины с функциями распределения F и G. Найти функции распределения следующих величин[1]: а) $\mathbf{X} \cup \mathbf{Y}$, б) $\mathbf{X} \cap \mathbf{Y}$, в) $2\mathbf{X} \cup \mathbf{Y}$, г) $\mathbf{X}^3 \cup \mathbf{Y}$.

2. *Смеси.* Пусть $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ — независимые случайные величины, $\mathbf{X}$ и $\mathbf{Y}$ имеют распределения F и G и $\mathbf{P}\{\mathbf{Z}=1\}=p$, $\mathbf{P}\{\mathbf{Z}=0\}=q\,(p+q=1)$. Найти функции распределения следующих величин: а) $\mathbf{ZX}+(1-\mathbf{Z})\mathbf{Y}$, б) $\mathbf{ZX}+(1-\mathbf{Z})(\mathbf{X} \cup \mathbf{Y})$, в) $\mathbf{ZX}+(1-\mathbf{Z})(\mathbf{X} \cap \mathbf{Y})$.

3. Показать, что для непрерывной функции распределения F

$$\int_{-\infty}^{\infty} F(x)\,F\{dx\} = \int_{0}^{1} y\,dy = \frac{1}{2},$$

а) используя само определение первого интеграла (разбиение $\overline{-\infty,\infty}$ на подинтервалы), и б) интерпретируя первый интеграл, как $\mathbf{E}(F(\mathbf{X}))$, и пользуясь тем, что $F(\mathbf{X})$ имеет равномерное распределение. Показать также, что имеет место более общая формула

$$\int_{-\infty}^{\infty} F^k(x)\,G\{dx\} = \frac{n}{n+k}, \quad \text{где} \quad G(x)=F^n(x).$$

4. *Двумерное распределение с заданными частными распределениями*[2]. Пусть F и G — функции распределения в $\mathscr{R}^1$ и

$$U(x,y)=F(x)\,G(y)[1+\alpha(1-F(x))\,(1-G(y))],$$

где $|\alpha|\leqslant 1$. Доказать, что U есть функция распределения в $\mathscr{R}^2$ с частными распределениями F и G и что U имеет плотность тогда и только тогда, когда F и G имеют плотности.

Указание. Смешанные разности функции $w(x,y)=u(x)\,v(y)$ [определенные в (1,12)] имеют вид $\Delta u \Delta v$.

5. Внутри единичного квадрата положим $U(x,y)=x$ при $x\leqslant y$ и $U(x,y)=y$ при $x\geqslant y$. Показать, что U есть функция распределения и что соответствующее ей распределение сосредоточено на биссектрисе (и, следовательно, сингулярно).

---

[1]) Для двух чисел a и b $a \cup b=\max(a,b)$ обозначает наибольшее из них, а $a \cap b=\min(a,b)$ — наименьшее. Для функций $f \cup g$ обозначает функцию, которая в точке x принимает значение $f(x) \cup g(x)$ (см. гл. IV, 1). Таким образом, $\mathbf{X} \cup \mathbf{Y}$ и $\mathbf{X} \cap \mathbf{Y}$ являются случайными величинами.

[2]) Эта задача содержит новый пример ненормального распределения с нормальными частными распределениями (см. задачи 2 и 3 в гл. III, 9), принадлежащий Е. Дж. Гумбелу.

6. *Максимальное распределение Фреше с заданными частными распределениями.* Пусть F и G — функции распределения в $\mathcal{R}^1$ и $U(x, y) = F(x) \cap G(y)$. Доказать, что а) U есть функция распределения с маргинальными распределениями F и G, б) если V — произвольная функция распределения с маргинальными распределениями F и G, то $V \leqslant U$, в) распределение, отвечающее U, сосредоточено на множестве $\{(x, y) : F(x-) \cup G(y-) \leqslant F(x) \cap G(y)\}$ и, следовательно, сингулярно. (Задача 5 является частным случаем этой задачи.)

7. Пусть U — равномерное распределение на $\overline{-h, 0}$ и T — треугольное распределение на $\overline{-h, h}$ [см. гл. II, 4 (в)]. Показать, что *плотности* распределений $F * U$ и $F * T$ равны соответственно

$$h^{-1}[F(x+h) - F(x)] \quad \text{и} \quad h^{-2} \int_0^h [F(x+y) - F(x-y)]\, dy.$$

8. Пусть F — атомическое распределение с массами $p_1, p_2, \ldots$ в точках $a_1, a_2, \ldots$. Обозначим через p максимальное среди чисел $p_1, p_2, \ldots$. Используя лемму 1 § 4, показать, что

а) массы всех атомов распределения $F * F$ строго меньше p, за исключением того случая, когда F сосредоточено на конечном числе атомов одинаковой массы,

б) для симметризованного распределения F_0 нуль является атомом с массой $p' = \sum p_\nu^2$; массы остальных атомов ${}^0F$ строго меньше p'.

9. Если случайные величины $\mathbf{X}$ и $\mathbf{Y}$ имеют пуассоновские распределения с математическими ожиданиями α и β, то

$$\mathbf{P}\{\mathbf{X} - \mathbf{Y} = k\} = e^{-\alpha-\beta} \sqrt{\left(\frac{\alpha}{\beta}\right)^k}\, I_{|k|}(2\sqrt{\alpha\beta}),$$

где I_k — функция Бесселя, определенная в гл. II, (7.1).

10. Каждому распределению F, для которого интеграл

$$\varphi(\alpha) = \int_{-\infty}^{\infty} e^{\alpha x} F\{dx\}$$

существует при $-a < \alpha < a$, сопоставим новое распределение $F^\#$ посредством формулы $\varphi(\alpha) F^\#\{dx\} = e^{\alpha x} F\{dx\}$. Пусть F_1 и F_2 — два распределения, обладающие указанным выше свойством, и $F = F_1 * F_2$. Доказать, что (используются очевидные обозначения) $\varphi(\alpha) = \varphi_1(\alpha)\varphi_2(\alpha)$ и $F^\# = F_1^\# * F_2^\#$.

11. Пусть $\mathbf{X}$ и $\mathbf{Y}$ — случайные величины с плотностями f и g, такими, что $f(x) \geqslant g(x)$ при $x < a$ и $f(x) \leqslant g(x)$ при $x > a$. Доказать, что $\mathbf{E}(\mathbf{X}) \leqslant \mathbf{E}(\mathbf{Y})$. Далее если $f(x) = g(x) = 0$ при $x < 0$, то $\mathbf{E}(\mathbf{X}^k) \leqslant \mathbf{E}(\mathbf{Y}^k)$ для всех k.

12. Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — взаимно независимые случайные величины с одинаковой функцией распределения F. Пусть $\mathbf{N}$ — положительная случайная величина, принимающая целые значения и имеющая производящую функцию $P(s)$. Если $\mathbf{N}$ не зависит от $\mathbf{X}_j$, то случайная величина $\max[\mathbf{X}_1, \ldots, \mathbf{X}_\mathbf{N}]$ имеет распределение $P(F)$.

13. Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — взаимно независимые случайные величины с непрерывным распределением F. Положим $\mathbf{X}=\max[\mathbf{X}_1, \ldots, \mathbf{X}_n]$ и $\mathbf{Y}=\min[\mathbf{X}_1, \ldots, \mathbf{X}_n]$. Тогда

$$\mathbf{P}\{\mathbf{X}\leqslant x, \mathbf{Y}>y\}=(F(x)-F(y))^n \text{ при } y<x$$

и

$$\mathbf{P}\{\mathbf{Y}>y \mid \mathbf{X}=x\}=[1-F(y)/F(x)]^{n-1}.$$

14. При каждом фиксированном $k\leqslant n$ (в обозначениях предыдущей задачи)

$$\mathbf{P}\{\mathbf{X}_k\leqslant x \mid \mathbf{X}=t\}=\begin{cases}\dfrac{n-1}{n}\dfrac{F(x)}{F(t)} & \text{при } x<t,\\ 1 & \text{при } x\geqslant t.\end{cases}$$

Доказать это а) с помощью эвристического рассуждения, рассматривая событие $\{\mathbf{X}_k=\mathbf{X}\}$, и б) формально, используя (9.4).

15. *Продолжение.* Доказать, что

$$\mathbf{E}(\mathbf{X}_k \mid \mathbf{X}=t)=\frac{n-1}{n}\frac{1}{F(t)}\int_{-\infty}^{t} yF\{dy\}+\frac{t}{n}.$$

16. *Случайные суммы.* В примере (9. в) случайную величину $\mathbf{X}_k$ положим равной 1 и -1 с вероятностями p и $q=1-p$. Если $\mathbf{N}$ имеет пуассоновское распределение с математическим ожиданием α, то

$$\mathbf{P}\{\mathbf{S}_\mathbf{N}=k\}=e^{-\alpha}\sqrt{\left(\frac{p}{q}\right)^k}I_{|k|}(2\alpha\sqrt{pq}),$$

где I_k — функция Бесселя, определенная в гл. II, (7.1).

17. *Смеси.* Пусть распределение G в (9.9) имеет математическое ожидание $m(x)$ и дисперсию $\sigma^2(x)$. Показать, что математическое ожидание и дисперсия смеси U равны

$$a=\int_{-\infty}^{\infty} m(x)\,\mu\{dx\}, \qquad b=\int_{-\infty}^{\infty}\sigma^2(x)\,\mu\{dx\}+\int_{-\infty}^{\infty}(m^2(x)-a^2)\,\mu(dx).$$

18. Применяя очевидные обозначения, имеем $\mathbf{E}(\mathbf{E}(\mathbf{Y}\mid\mathbf{X}))=\mathbf{E}(\mathbf{Y})$, но $\mathrm{Var}(\mathbf{Y})=\mathbf{E}(\mathrm{Var}(\mathbf{Y}\mid\mathbf{X}))+\mathrm{Var}(\mathbf{E}(\mathbf{Y}\mid\mathbf{X}))$.
Задача 17 является частным случаем этой задачи.

19. *Случайные суммы.* В обозначениях примера (9. в) $\mathbf{E}(\mathbf{S}_\mathbf{N})=\mathbf{E}(\mathbf{N})\mathbf{E}(\mathbf{X})$ и $\mathrm{Var}(\mathbf{S}_\mathbf{N})=\mathbf{E}(\mathbf{N})\mathrm{Var}(\mathbf{X})+(\mathbf{E}(\mathbf{X}))^2\mathrm{Var}(\mathbf{N})$. Доказать эти соотношения непосредственно и показать, что они вытекают из результатов двух предыдущих задач.

Замечание. Следующие задачи относятся к композициям *одновершинных распределений*, определенных в сноске на стр. 197. Была выдвинута гипотеза, что композиция двух одновершинных распределений одновершинна. Противоречащий этой гипотезе пример был построен К. Л. Чжуном. Задача 20 содержит другой такой пример. Задача 21 показывает, что гипотеза верна для *симметричных* [1]) распределений (этот результат принадлежит А. Винтнеру).

[1]) С трудностями, возникающими в несимметричном случае, можно познакомиться в работе И. А. Ибрагимова, Теория вероятностей и ее применения, т. I (1956), стр. 283—288.

20. Пусть $u(x)=1$ при $0<x<1$ и $u(x)=0$ в остальных точках. Положим

$$v(x)=\frac{\varepsilon}{a}u\left(\frac{x}{a}\right)+\frac{1-\varepsilon}{b}u\left(\frac{x}{b}\right),$$

где $0<a<b$. Если ε и a малы, а b велико, то плотность v одновершинна, но плотность $w=v*v$ не одновершинна.

Указание как избежать вычислений. В результате композиции двух равномерных плотностей получается треугольная плотность. Следовательно, $w(a)>\varepsilon^2a^{-1}$, $w(b)>\varepsilon^2b^{-1}$ и интеграл от w в пределах от b до $2b$ больше $\frac{1}{2}(1-\varepsilon)^2$. Отсюда вытекает, что плотность w должна иметь минимум между a и b.

21. Пусть F — равномерное, а G — одновершинное распределения. С помощью простого дифференцирования показать, что если F и G симметричны, то композиция $F*G$ одновершинна. Вывести отсюда (без дальнейших вычислений), что тот же результат получается, когда F есть произвольная смесь симметричных равномерных распределений и что, следовательно, *композиция симметричных одновершинных распределений одновершинна.*

ГЛАВА VI

НЕКОТОРЫЕ ВАЖНЫЕ РАСПРЕДЕЛЕНИЯ И ПРОЦЕССЫ

Настоящая глава включена в книгу из-за прискорбной необходимости избегать повторений и взаимных ссылок в главах, предназначенных для независимого чтения. Например, теория устойчивых распределений будет изложена независимо на основе полугрупповых методов (гл. IX), анализа Фурье (гл. XVII), и, по крайней мере частично, на основе преобразований Лапласа (гл. XIII). Вынесение определений и примеров в отдельное место дает некоторую экономию и делает возможным тщательное исследование некоторых основных соотношений, не стесненное чистотой используемых методов.

Разнообразные темы, которым посвящена настоящая глава, не всегда логически связаны: теория очередей имеет мало общего с мартингалами или с устойчивыми распределениями. Материал главы не предназначен для последовательного чтения; лучше читать тот или иной параграф по мере появления необходимости. Параграфы 6—9 в известной степени связаны между собой, но не зависят от остальной части главы.

§ 1. Устойчивые распределения в $\mathscr{R}^1$

Устойчивые распределения играют все возрастающую роль в качестве естественного обобщения нормального распределения. Для их описания удобно ввести следующее сокращенное обозначение: если распределения двух случайных величин $\mathbf{U}$ и $\mathbf{V}$ совпадают, то мы будем писать

$$\mathbf{U} \stackrel{d}{=} \mathbf{V}. \tag{1.1}$$

В частности, соотношение $\mathbf{U} \stackrel{d}{=} a\mathbf{V}+b$ означает, что распределения величин $\mathbf{U}$ и $\mathbf{V}$ отличаются лишь параметрами расположения [см. определение 1 в гл. V, 2]. Всюду в этом параграфе символы $\mathbf{X}$, $\mathbf{X}_1$, $\mathbf{X}_2$, ... обозначают взаимно независимые случайные величины с одним и тем же распределением R, и $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$.

Определение 1. *Распределение R называется устойчивым (в широком смысле), если оно не сосредоточено в нуле и для*

каждого n существуют постоянные $c_n > 0$ и γ_n, такие, что [1])

$$\mathbf{S}_n \stackrel{d}{=} c_n\mathbf{X} + \gamma_n. \tag{1.2}$$

R устойчиво в узком смысле (строго устойчиво), если оно удовлетворяет соотношению (1.2) *с* $\gamma_n = 0$.

Примеры будут приведены в § 2. Весьма поучителен элементарный вывод некоторых основных свойств устойчивых распределений, и мы дадим его здесь несмотря на то, что дальше будут даны другие доказательства этих свойств. Систематическое изложение теории устойчивых распределений в гл. IX и XVII не опирается на нижеследующий материал.

Теорема 1. *Постоянные $c_n = n^{1/\alpha}$ являются единственно возможными нормирующими постоянными.*

Вскоре мы увидим, что $0 < \alpha \leqslant 2$. Постоянная α называется *характеристическим показателем* распределения R.

Доказательство. Доказательство значительно упрощается при использовании симметризации. Если распределение R устойчиво, то распределение ${}^0R$ величины $\mathbf{X}_1 - \mathbf{X}_2$ тоже устойчиво, причем с теми же самыми нормирующими постоянными c_n. Достаточно, следовательно, ограничиться случаем симметричного устойчивого распределения R.

Заметим сначала, что $\mathbf{S}_{m+n}$ есть сумма *независимых* случайных величин $\mathbf{S}_m$ и $\mathbf{S}_{m+n} - \mathbf{S}_m$, распределения которых совпадают соответственно с распределениями величин $c_m\mathbf{X}$ и $c_n\mathbf{X}$. Таким образом, для симметричных устойчивых распределений имеем

$$c_{m+n}\mathbf{X} \stackrel{d}{=} c_m\mathbf{X}_1 + c_n\mathbf{X}_2. \tag{1.3}$$

Сумму $\mathbf{S}_{rk}$ можно аналогичным образом разбить на r независимых частей по k членов в каждой, и поэтому $c_{rk} = c_r c_k$ при всех r и k. Отсюда по индукции заключаем, что

$$\text{если } n = r^\nu, \text{ то } c_n = c_r^\nu. \tag{1.4}$$

Докажем теперь, что *последовательность $\{c_n\}$ монотонна*. Поскольку случайные величины в (1.3) симметричны, то при $x > 0$ имеем

$$\mathbf{P}\{c_{m+n}\mathbf{X} > c_m x\} \geqslant \tfrac{1}{2}\mathbf{P}\{c_m\mathbf{X}_1 > c_m x\} = \tfrac{1}{2}\mathbf{P}\{\mathbf{X}_1 > x\}. \tag{1.5}$$

[1]) Ниже будет показано, что можно дать следующее (на первый взгляд более ограничительное) эквивалентное определение. Распределение R устойчиво тогда и только тогда, когда для любых постоянных c_1, c_2 существуют постоянные c и γ, такие, что $c_1\mathbf{X}_1 + c_2\mathbf{X}_2 \stackrel{d}{=} c\mathbf{X} + \gamma$.

По крайней мере при некотором фиксированном x правая часть в (1.5) является положительной постоянной, и поэтому отношение c_m/c_{m+n} должно оставаться ограниченным при $m \to \infty$, $n \to \infty$. Применим этот результат, когда $m = r^\nu$ и $m+n = (r+1)^\nu$, где r фиксированно и $\nu \to \infty$. Используя (1.4), видим, что отношение $(c_r/c_{r+1})^\nu$ остается ограниченным при $\nu \to \infty$ и, следовательно, $c_r \leqslant c_{r+1}$.

Рассмотрим теперь произвольную пару натуральных чисел j, k. Каждому ν отвечает единственное значение λ, такое, что

$$j^\lambda \leqslant k^\nu < j^{\lambda+1}. \tag{1.6}$$

Отсюда в силу монотонности последовательности $\{c_r\}$ и (1.4) имеем

$$c_j^\lambda \leqslant c_k^\nu < c_j^{\lambda+1}. \tag{1.7}$$

Из (1.6) и (1.7) получаем

$$\frac{\lambda}{\lambda+1}\frac{\log j}{\log j} < \frac{\log c_k}{\log k} < \frac{\lambda+1}{\lambda}\frac{\log c_j}{\log j}. \tag{1.8}$$

Так как λ можно сделать произвольно большим, то отношение $(\log c_k)/(\log k) = 1/\alpha$ не должно зависеть от k, и, следовательно, $c_k = k^{1/\alpha}$. ▶

Оказывается, что все устойчивые распределения имеют гладкие плотности, однако этот факт будет доказан лишь в гл. XVIII, 6. Здесь нам придется довольствоваться следующим результатом.

Лемма 1. *Все устойчивые (в широком смысле) распределения непрерывны.*

Доказательство. Пусть устойчивое распределение R имеет один или несколько атомов, и пусть p — максимальная масса атомов. Атомы распределения $R * R$ отличаются от атомов распределения R лишь расположением и имеют те же массы. Следовательно, распределение $R * R$ имеет атом массы p в противоречие с тем легко устанавливаемым фактом, что при композиции максимальный вес атомов убывает (см. задачу 8 из гл. V, 11). ▶

Теорема 2. *Если R — устойчивое в узком смысле распределение с характеристическим показателем α, то для любых положительных s и t имеем*

$$s^{1/\alpha}\mathbf{X}_1 + t^{1/\alpha}\mathbf{X}_2 \stackrel{d}{=} (s+t)^{1/\alpha}\mathbf{X}. \tag{1.9}$$

Доказательство. Для устойчивых в узком смысле распределений справедливо соотношение (1.3). Отсюда следует,

поскольку $c_n = n^{1/\alpha}$, что (1.9) выполняется при всех рациональных s и t. Переходя к пределу, убеждаемся в справедливости (1.9) и в общем случае. ▶

Для нормального распределения $\alpha = 2$ и (1.9) сводится просто к правилу сложения дисперсий. В общем случае из соотношения (1.9) вытекает, что *все линейные комбинации* $a_1\mathbf{X}_1 + a_2\mathbf{X}_2$ *принадлежат к одному и тому же типу.*

Покажем теперь, что требование устойчивости в узком смысле не так ограничительно, как это может показаться, поскольку всякое устойчивое распределение с характеристическим показателем $\alpha \neq 1$ всегда можно сдвинуть так, чтобы оно стало устойчивым в узком смысле. Случай характеристического показателя $\alpha = 1$ является весьма исключительным и особенно неприятным потому, что он не представляет особого интереса [1]).

Теорема 3. *Для всякого устойчивого распределения R с характеристическим показателем $\alpha \neq 1$ существует постоянная b, такая, что распределение $R(x+b)$ строго устойчиво.*

Доказательство: Случайные величины $\mathbf{X}'_k = \mathbf{X}_k - b$ удовлетворяют соотношению (1.2) с γ_n, замененным на $\gamma'_n = \gamma_n + (c_n - n)\gamma$. Следовательно, при $\alpha \neq 1$ можно выбрать b так, чтобы $\gamma'_2 = 0$. Отбросим теперь штрихи и предположим, что $\gamma_2 = 0$ и $\alpha \neq 1$, т. е. $c_2 \neq 2$. Заметим, что $\mathbf{S}_{2n}$ можно рассматривать как сумму n независимых слагаемых, каждое из которых распределено одинаково с $\mathbf{X}_1 + \mathbf{X}_2$, и как сумму двух независимых слагаемых, распределенных одинаково с $\mathbf{S}_n$. Имеем, следовательно,

$$\mathbf{S}_{2n} \stackrel{d}{=} c_2(\mathbf{X}_1 + \ldots + \mathbf{X}_n) \stackrel{d}{=} c_2 c_n \mathbf{X} + c_2\gamma_n$$

и

$$\mathbf{S}_{2n} \stackrel{d}{=} (c_n\mathbf{X}_1 + \gamma_n) + (c_n\mathbf{X}_2 + \gamma_n) \stackrel{d}{=} c_n c_2 \mathbf{X} + 2\gamma_n.$$

Поэтому должно быть $c_2\gamma_n = 2\gamma_n$, так что $\gamma_n = 0$. ▶

Та важная роль, которую играет нормальное распределение $\mathfrak{N}$ в теории вероятностей, в значительной мере основана на центральной предельной теореме. Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — взаимно независимые случайные величины с общим распределением F, нулевым математическим ожиданием и единичной дисперсией.

---

[1]) Оказывается, что при $\alpha = 1$ соответствующие центрирующие постоянные γ_n в (1.2) имеют вид $\gamma_n = \gamma n \log n$. Аналогом (1.9) является в этом случае соотношение

$$s(\mathbf{X}_1 + \gamma \log s) + t(\mathbf{X}_2 + \gamma \log t) \stackrel{d}{=} (s+t)(\mathbf{X} + \gamma \log (s+t)).$$

Положим $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. Центральная предельная теорема утверждает, что распределения случайных величин $\mathbf{S}_n n^{-1/2}$ сходятся к $\mathfrak{N}$. Если распределения не имеют дисперсий, то числа $n^{-1/2}$ в качестве нормирующих постоянных неудовлетворительны, но при ином выборе нормирующих постоянных предел попрежнему может существовать. Весьма интересным является тот факт, что в качестве таких пределов могут быть получены все устойчивые законы и только устойчивые законы. Введем следующую терминологию, которая облегчит дальнейшее обсуждение рассматриваемого вопроса.

Определение 2. *Скажем, что распределение F независимых случайных величин $\mathbf{X}_k$ принадлежит области притяжения распределения R, если существуют нормирующие постоянные $a_n > 0$, b_n, такие, что распределение величины $a_n^{-1}(\mathbf{S}_n - b_n)$ сходится к R.*

Приведенное выше утверждение можно теперь переформулировать следующим образом: *распределение R имеет область притяжения тогда и только тогда, когда оно устойчиво.* Докажем это утверждение. Из определения устойчивости непосредственно вытекает, что устойчивое распределение R принадлежит своей области притяжения. Тот факт, что никакое неустойчивое распределение не выступает в качестве предельного, можно обосновать с помощью рассуждения, использованного при доказательстве теоремы 1. Нормируем суммы $\mathbf{S}_n$ с индексами $n = rk$ двумя способами, положив

$$\mathbf{Y}_n = \frac{1}{a_n}(\mathbf{S}_n - b_n) \quad \text{и} \quad \mathbf{Z}_n = \frac{1}{a_r}(\mathbf{S}_n - kb_r). \tag{1.10}$$

Устремим далее $n = rk$ к бесконечности, оставляя k фиксированным. Согласно предположению распределения величин $\mathbf{Y}_n$ сходятся к R. С другой стороны, $\mathbf{Z}_n$ есть сумма k независимых случайных величин, каждая из которых распределена одинаково с $a_r^{-1}(\mathbf{S}_r - b_r)$. Распределение величин $a_r^{-1}(\mathbf{S}_r - b_r)$ сходятся к R, и поэтому интуитивно ясно, что распределения величин $\mathbf{Z}_n$ должны сходиться к R^{k*}. Так как распределения величин $\mathbf{Y}_n$ и $\mathbf{Z}_n$ отличаются лишь параметрами расположения, то же самое верно и для предельных распределений R и R^{k*}, что равносильно устойчивости R (последние утверждения могут быть строго доказаны на основе гл. VIII, 2 и теоремы 2 из гл. VIII, 3).

Применяя элементарные методы, можно показать (задача 3), что *устойчивое распределение с характеристическим показателем α имеет абсолютные моменты всех порядков $< \alpha$.* Поскольку распределение с конечной дисперсией принадлежит области при-

тяжения нормального закона, то отсюда следует, что *характеристический показатель* α *не превосходит* 2. Доказательство существования устойчивых распределений с характеристическими показателями $\alpha<2$ довольно сложно.

Полученные результаты имеют важные и неожиданные следствия. Рассмотрим, например, устойчивое распределение, удовлетворяющее соотношению (1.9) с $\alpha<1$. *Среднее* $(\mathbf{X}_1+\ldots+\mathbf{X}_n)/n$ имеет одинаковое распределение с величиной $\mathbf{X}_1 n^{-1+1/\alpha}$. Обратим внимание на то, что множитель $n^{-1+1/\alpha}$ стремится к бесконечности с ростом n. Выражаясь нестрого, можно сказать, что среднее n величин $\mathbf{X}_k$ оказывается значительно больше любого фиксированного слагаемого $\mathbf{X}_k$. Это возможно лишь тогда, когда *максимальный член* $\mathbf{M}_n=\max[\mathbf{X}_1, \ldots, \mathbf{X}_n]$ растет исключительно быстро и играет основную роль в сумме $\mathbf{S}_n$. Более детальный анализ подтверждает это заключение. В случае положительных случайных величин математическое ожидание отношения $\mathbf{S}_n/\mathbf{M}_n$ стремится к $(1-\alpha)^{-1}$ и это верно также для любой последовательности $\{\mathbf{X}_n\}$ с распределением, принадлежащим области притяжения рассматриваемого устойчивого закона (см. задачу 20 из гл. XIII, 11).

Исторические замечания. Основы общей теории устойчивых распределений были заложены П. Леви [1]), который нашел преобразования Фурье всех строго устойчивых распределений. (Устойчивые распределения в широком смысле первоначально назывались квазиустойчивыми. Как мы видели, они существенно отличаются от строго устойчивых распределений лишь при $\alpha=1$. Случай $\alpha=1$ был исследован совместно Леви и Хинчиным.) Более простой новый подход к теории устойчивых распределений стал возможным после открытия безгранично делимых распределений. Этот новый подход (тоже основанный на анализе Фурье) также принадлежит П. Леви (1937). Интерес к теории был стимулирован замечательным исследованием А. Дёблина (1939), посвященным областям притяжения. В его критериях впервые появились правильно меняющиеся функции. Современная теория по-прежнему носит отпечаток этой основополагающей работы, несмотря на то, что некоторыми авторами были внесены различные усовершенствования и получены новые результаты. Гл. VIII содержит изложение теории устойчивых распределений на основе методов Фурье, ставших теперь уже классическими. В гл. IX эта теория развивается с помощью прямого подхода, более близкого к современным методам теории марковских процессов. Значительные упрощения и объединение многих критериев оказались возможными благодаря систематическому использованию развитой Карамата теории правильно меняющихся функций. Усовершенствованный вариант этой теории излагается в гл. VIII, 8—9.

---

[1]) Преобразования Фурье *симметричных* устойчивых распределений были указаны Коши, однако оставалось неясным, действительно ли они отвечают вероятностным распределениям. При $\alpha<1$ этот вопрос был решен Пойа. Распределение Хольцмарка из примера (2. в) было известно астрономам, но не известно математикам. Читатели, интересующиеся историей развития теории и более старой литературой на эту тему, могут обратиться к книгам П. Леви (1925) и (1937).

§ 2. Примеры

а) *Нормальное распределение* с нулевым математическим ожиданием строго устойчиво с $c_n = \sqrt{n}$.

б) *Распределение Коши* с произвольными параметрами расположения, имеющее плотность

$$\frac{1}{\pi}\,\frac{c}{c^2+(x-\gamma)^2},$$

устойчиво в строгом смысле с характеристическим показателем 1. См. гл. II, (4, д).

в) *Устойчивое распределение с* $\alpha = \frac{1}{2}$. Распределение

$$F(x) = 2\left[1 - \Phi\left(\frac{1}{\sqrt{x}}\right)\right], \qquad x > 0, \tag{2.1}$$

имеющее плотность

$$f(x) = \frac{1}{\sqrt{2\pi x^3}}\, e^{-\frac{1}{2}x}, \quad x > 0 \tag{2.2}$$

[$f(x) = 0$ при $x < 0$], строго устойчиво с нормирующими постоянными $c_n = n^2$.

Проверка этого факта элементарными методами утомительна и неинтересна; здесь проще воспользоваться одной предельной теоремой. Рассмотрим симметричное случайное блуждание; пусть $\mathbf{X}_k$ есть число испытаний между $(k-1)$-м и k-м возвращением в начало координат. Величины $\mathbf{X}_k$, как известно, взаимно независимы и распределены согласно **1**, гл. III (4.2). Далее, величина $\mathbf{S}_k = \mathbf{X}_1 + \ldots + \mathbf{X}_k$ есть момент k-го возвращения в начало координат, и поэтому событие $\{\mathbf{S}_r > k\}$ означает, что «число возвращений за k шагов меньше r». Отсюда, в силу предельной теоремы **1**, гл. III, 6.2 имеем

$$\mathbf{P}\{\mathbf{S}_r \leqslant r^2 t\} \to F(t), \quad r \to \infty. \tag{2.3}$$

Таким образом, F имеет область притяжения и потому устойчиво (продолжение в примере (д)).

г) *Гравитационное поле звезд (распределение Хольцмарка).* На языке астрономии задача формулируется так: вычислить x-компоненту гравитационной силы, создаваемой звездной системой в случайно выбранной точке O. В основе этой постановки задачи лежит предположение, что звездная система представляет собой «случайное скопление» точек со «случайно распределенными массами». Это предположение можно было бы выразить строго математически в терминах распределений Пуассона и т. п., однако, к счастью, для рассматриваемой задачи подобных тонкостей не требуется.

Будем рассматривать плотность звездной системы как свободный параметр; пусть $\mathbf{X}_\lambda$ обозначает x-компоненту гравитационной силы звездной системы с плотностью λ. Нас интересуют возможные типы распределений $\mathbf{X}_\lambda$. Интуитивное понятие «случайное скопление звезд» предполагает, что два независимых скопления с плотностями s и t могут быть объединены в одно скопление с плотностью $s+t$. С вероятностной точки зрения это приводит к требованию, чтобы сумма двух независимых случайных величин, распределенных как $\mathbf{X}_t$ и $\mathbf{X}_s$, была распределена как $\mathbf{X}_{s+t}$. Иначе это можно выразить равенством

$$\mathbf{X}_s+\mathbf{X}_t \stackrel{d}{=} \mathbf{X}_{s+t}. \tag{2.4}$$

Считая, что замена единичной плотности плотностью λ приводит к изменению единичной длины на длину $1/\sqrt[3]{\lambda}$ и что гравитационная сила обратно пропорциональна квадрату расстояния, приходим к выводу, что величины $\mathbf{X}_t$ и $t^{2/3}\mathbf{X}_1$ должны иметь одинаковое распределение. Это означает, что распределения величин $\mathbf{X}_t$ отличаются лишь масштабным множителем и (2.4) сводится к (1.9) с $\alpha=\frac{3}{2}$. Иными словами, $\mathbf{X}_\lambda$ имеет симметричное устойчивое распределение с характеристическим показателем $\alpha=\frac{3}{2}$. В дальнейшем мы увидим, что (с точностью до тривиального масштабного множителя) имеется в точности одно такое распределение. Таким образом, мы решили нашу задачу, не прибегая к более глубокой теории. Астроном Хольцмарк получил равноспльный ответ, применяя другие методы (см. задачу 4), причем, что весьма замечательно, до появления работы П. Леви.

д) *Моменты первого прохождения броуновского процесса.* Введем сначала понятие одномерного диффузионного процесса. Диффузионный процесс — это такой процесс $\mathbf{X}(t)$, приращения которого $\mathbf{X}(s+t)-\mathbf{X}(s)$ за непересекающиеся промежутки времени независимы и имеют симметричное нормальное распределение с дисперсией t. Будем предполагать известным, что траектории одномерного диффузионного процесса *непрерывно* зависят от времени. При $\mathbf{X}(0)=0$ существует момент $\mathbf{T}_a$, *в который частица впервые достигает уровня* $a>0$. Для нахождения функции распределения $F_a(t)=\mathbf{P}(\mathbf{T}_a\leqslant t)$ заметим, что понятие аддитивного процесса заключает в себе полное отсутствие последействия (строго марковское свойство). Это означает, что приращение абсциссы $\mathbf{X}(t+T_a)-a$ между моментами времени $\mathbf{T}_a$ и $\mathbf{T}_a+t$ не зависит от течения процесса до момента $\mathbf{T}_a$. Далее,

чтобы достигнуть уровня $a+b>a$, частица должна сначала достигнуть уровня a, и поэтому время ожидания $\mathbf{T}_{a+b} - \mathbf{T}_a$ до достижения уровня $a+b$ после достижения a не зависит от $\mathbf{T}_a$ и имеет распределение, одинаковое с распределением $\mathbf{T}_b$. Иными словами, $F_a * F_b = F_{a+b}$. Так как вероятности перехода зависят лишь от отношения x^2/t, то $\mathbf{T}_a$ имеет то же распределение, что и $a^2\mathbf{T}_1$. Это означает, что распределения F_a отличаются лишь масштабным множителем и, следовательно, *устойчивы с характеристическим показателем* $\alpha = \frac{1}{2}$.

Проведенное рассуждение, основанное на соображениях размерности, доказывает устойчивость распределения момента первого прохождения уровня, но не дает его точной формы. Чтобы доказать *совпадение F с распределением примера* в), мы воспользуемся рассуждением, основанным на соображениях симметрии. В силу непрерывности траекторий событие $\{\mathbf{X}(t)>a\}$ может произойти лишь в том случае, когда уровень a был пересечен в некоторый момент $\mathbf{T}_a<t$. Если $\mathbf{T}_a=\tau<t$, то $\mathbf{X}(\tau)=a$, и из симметрии ясно, что вероятность события $\mathbf{X}(t) - \mathbf{X}(\tau)>0$ равна $\frac{1}{2}$. Следовательно,

$$\mathbf{P}\{\mathbf{T}_a < t\} = 2\mathbf{P}\{\mathbf{X}(t) > a\} = 2\left[\left(1 - \mathfrak{N}\left(\frac{a}{\sqrt{t}}\right)\right)\right], \qquad (2.5)$$

что и требовалось доказать. Только что проиллюстрированное использование соображений симметрии иногда называют методом Дезире Андре или принципом отражения.

е) *Точки достижения при двумерном броуновском движении.* Двумерное броуновское движение образуется парой $(\mathbf{X}(t), \mathbf{Y}(t))$ независимых одномерных броуновских движений. Нас интересует точка $(a, \mathbf{Z}_a)$, в которой траектория впервые достигает прямой $x=a>0$. Отметим, что подобно тому, как это было в предыдущем примере, траектория может достичь прямой $x=a+b>a$ лишь после пересечения прямой $x=a$. Если взять в качестве нового начала координат точку $(a, \mathbf{Z}_a)$, то станет ясным, что величина $\mathbf{Z}_{a+b}$ также распределена как сумма двух независимых случайных величин, распределенных одинаково с величинами $\mathbf{Z}_a$ и $\mathbf{Z}_b$. Далее, очевидные соображения подобия приводят к заключению, что величина $\mathbf{Z}_a$ распределена так же как величина $a\mathbf{Z}_1$, и поэтому $\mathbf{Z}_a$ имеет симметричное устойчивое распределение с характеристическим показателем $\alpha=1$. Единственным таким распределением является распределение Коши, так что *точка достижения $\mathbf{Z}_a$ имеет распределение Коши.*

Приведенное поучительное рассуждение, основанное на соображениях размерности, не определяет масштабного множителя. Для его вычисления заметим, что $\mathbf{Z}_a = \mathbf{Y}(\mathbf{T}_a)$, где $\mathbf{T}_a$ — момент первого достижения прямой $x=a$. Распределение величины $\mathbf{T}_a$ дается формулой (2.5), а $\mathbf{Y}(t)$ имеет нормальное распределение с дисперсией t. Следовательно, величина $\mathbf{Z}_a$ имеет плотность, равную [1])

$$\int_0^\infty \frac{1}{t^{1/2}\sqrt{2\pi}} e^{-\frac{1}{2}x^2/t} \frac{a}{t^{3/2}\sqrt{2\pi}} e^{-\frac{1}{2}a^2/t}\, dt = \frac{a}{\pi(a^2+x^2)}. \qquad (2.6)$$

[Здесь мы имеем пример подчиненности процессов, к которому мы еще вернемся в гл. X, 7.]

ж) *Устойчивые распределения в экономике.* Примененные в двух последних примерах рассуждения, основанные на соображениях размерности, были использованы, Б. Мандельбройтом для доказательства того, что различные экономическне процессы (в частности, распределения дохода) подчинены устойчивым (или «Леви — Парето») распределениям. До настоящего времени сила и ценность этой интересной теории, привлекшей внимание экономистов, проявились скорее в теоретической сфере, чем в прикладном аспекте. [По поводу кажущегося согласия хвостов распределения с многими эмпирическими явлениями, начиная от размера городов до частоты слов, см. гл. II, (4.з).]

з) *Произведения.* Существует много любопытных соотношений между устойчивыми распределениями с различными характеристическими показателями. Наиболее интересное из них можно сформулировать в виде следующего предложения. Пусть $\mathbf{X}$ и $\mathbf{Y}$ — независимые строго устойчивые случайные величины с характеристическими показателями α и β соответственно. Предположим, что величина $\mathbf{Y}$ положительна (при этом $\beta<1$). Тогда *произведение* $\mathbf{X}\mathbf{Y}^{1/\alpha}$ *имеет устойчивое распределение с характеристическим показателем* $\alpha\beta$. В частности, произведение нормальной случайной величины и квадратного корня из случайной величины с устойчивым распределением примера (в) имеет распределение Коши.

Мы не будем здесь доказывать это предложение, поскольку оно вытекает как простое следствие из теоремы о подчиненных

---

[1]) Подстановка $y = \frac{1}{2}(x^2+a^2)/t$ сводит подинтегральную функцию к функции e^{-y}.

процессах[1]) [пример, гл. X, (7. в)]. Кроме того, его легко проверить с помощью анализа Фурье [задача 8 из гл. XVII, 12] и для положительных случайных величин с помощью преобразований Лапласа (гл. XIII, (7.д) и задача 13].

§ 3. Безгранично делимые распределения в $\mathscr{R}^1$

Определение 1. *Распределение F безгранично делимо, если для каждого n существует распределение F_n, такое, что $F = F_n^{n*}$.*

Иными словами[2]), *F безгранично делимо, если при каждом n его можно представить как распределение суммы*

$$\mathbf{S}_n = \mathbf{X}_{1,\,n} + \ldots + \mathbf{X}_{n,\,n} \tag{3.1}$$

n независимых случайных величин с одним и тем же распределением F_n.

Это определение относится к распределениям в любом конечномерном пространстве, однако здесь мы ограничимся рассмотрением лишь одномерных распределений. Следует отметить, что безграничная делимость является свойством *типа*, т. е. вместе с F все распределения, отличающиеся от F лишь параметрами расположения, безгранично делимы. Устойчивые распределения безгранично делимы и выделяются среди безгранично делимых тем, что для них F_n отличаются от F лишь параметрами расположения.

Примеры. а) Все *гамма*-распределения (включая показательные) в силу их свойства гл. II, (2.3) безгранично делимы. Безграничная делимость дискретного варианта гамма-распределений — «*отрицательно биномиальных*» распределений (включая *геометрические* распределения) была доказана в 1, гл. XII, 2.

---

[1]) Для непосредственной проверки с минимальными вычислениями можно поступить следующим образом. Найдем распределение величины $\mathbf{Z} = \mathbf{X}_1 \overset{a}{\sqrt{\mathbf{Y}_1}} + \mathbf{X}_2 \overset{a}{\sqrt{\mathbf{Y}_2}}$, вычислив сначала условное распределение $\mathbf{Z}$ при условии $\mathbf{Y}_1 = y_1$, $\mathbf{Y}_2 = y_2$. Это условное распределение есть функция суммы $y_1 + y_2$, и замена переменных $u = y_1 + y_2$, $v = y_1 - y_2$ показывает, что распределение $\mathbf{Z}$ отличается от распределения $\mathbf{X}_1 \overset{a}{\sqrt{\mathbf{Y}_1}}$ лишь масштабным множителем. Аналогичные вычисления можно проделать в случае n таких слагаемых.

[2]) Следует иметь в виду, что случайные величины $\mathbf{X}_{k,\,n}$ вводятся только для простоты и наглядности. При фиксированном n случайные величины $\mathbf{X}_{1,\,n}, \ldots, \mathbf{X}_{n,\,n}$ предполагаются взаимно независимыми, но при $m \neq n$ величины $\mathbf{X}_{j,\,m}$ и $\mathbf{X}_{k,\,n}$ не обязаны даже быть определенными на одном и том же вероятностном пространстве (иначе говоря, совместное распределение величин $\mathbf{X}_{k,\,m}$ и $\mathbf{X}_{k,\,n}$ может не существовать). Это значение относится к схеме серий вообще.

б) Пуассоновское и обобщенное пуассоновское распределения безгранично делимы. Будет показано, что все безгранично делимые распределения являются пределами обобщенных пуассоновских распределений.

в) Связанное с *бесселевыми* функциями распределение гл. II, (7.13) безгранично делимо, однако это вовсе не очевидно; см. пример гл. XIII, (7. г).

г) *Безгранично делимое распределение не может быть сосредоточено на конечном интервале*, если оно не сосредоточено в одной точке. В самом деле, пусть F безгранично делимо. Если $|\mathbf{S}_n|<a$ с вероятностью единица, то $|\mathbf{X}_{k,n}|<an^{-1}$ и поэтому $\mathrm{Var}(\mathbf{X}_{k,n})<a^2n^{-2}$. Следовательно, дисперсия F, будучи меньше a^2n^{-1}, должна быть равна нулю. ▶

Возвращаясь к определению 1, посмотрим, что произойдет, если опустить требование одинаковой распределенности величин $\mathbf{X}_{k,n}$ и потребовать лишь, чтобы при каждом n существовало n распределений $F_{1,n}, \ldots, F_{n,n}$, для которых

$$F=F_{1,n}*\ldots*F_{n,n}. \tag{3.2}$$

Такая общность приводит к новому явлению, которое лучше всего проиллюстрировать на примерах.

Примеры. д) Если F — безгранично делимое распределение, а U — произвольное распределение, то распределение $G=U*F$ можно представить в виде (3.2) с $G_{1,n}=U$ и $G_{k,n}=F_{n-1}$ при $n>1$. Роль первой компоненты здесь совсем отлична от роли остальных компонент.

е) Рассмотрим сходящийся ряд $\mathbf{X}=\sum\mathbf{X}_k$ взаимно независимых случайных величин. Распределение F величины $\mathbf{X}$, являясь сверткой распределений величин $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1}$ и распределения остаточного члена $(\mathbf{X}_n+\mathbf{X}_{n+1}+\ldots)$, представлено в виде (3.2). Распределения такого рода называются *бесконечными свертками* и будут изучаться в дальнейшем. Пример гл. I, (11. в) показывает, что равномерное распределение является бесконечной сверткой. ▶

Отличительной чертой этих примеров является то обстоятельство, что в них вклад отдельной компоненты $\mathbf{X}_{1,n}$ в сумму $\mathbf{S}_n$ существен, в то время как в случае одинаково распределенных компонент вклад каждой отдельной компоненты стремится к нулю. Мы хотим теперь связать безгранично делимые распределения с типичными предельными теоремами для сумм «большого числа малых компонент». Для этого необходимо дополнить нашу схему требованием, чтобы отдельные компоненты $\mathbf{X}_{k,n}$

были асимптотически пренебрегаемыми в том смысле, что для любого $\varepsilon>0$ при достаточно больших n выполнялись бы неравенства

$$\mathbf{P}\{|\mathbf{X}_{k,n}|>\varepsilon\}<\varepsilon \qquad (k=1, \ldots, n). \tag{3.3}$$

В терминологии гл. VIII, 2 это означает, что величины $\mathbf{X}_{k,n}$ *стремятся по вероятности к нулю равномерно по* $k=1, \ldots, n$. Системы случайных величин такого типа встречаются довольно часто, и поэтому удобно дать им специальное название.

Определение 2. *Двойная последовательность случайных величин* $\mathbf{X}_{k,n}$ ($k=1, 2, \ldots, n$; $n=1, 2, \ldots$), *в которой образующие* n*-ю строку (серию) случайные величины* $\mathbf{X}_{1,n}, \ldots, \mathbf{X}_{n,n}$ *взаимно независимы, называется схемой серий*[1]).

Если выполняется (3.3), *то схема серий называется нулевой (или схемой серий с асимптотически пренебрегаемыми компонентами).*

Более общим образом можно было бы рассматривать схемы серий с r_n величинами в n-й серии, где $r_n\to\infty$, однако в действительности общность при этом не увеличивается[2]).

Пример. *Пусть* $\{\mathbf{X}_j\}$ — последовательность одинаково распределенных независимых случайных величин и $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$. Нормированная сумма $\mathbf{S}_n a_n^{-1}$ есть сумма величин n-й серии в схеме серий, в которой $\mathbf{X}_{k,n}=\mathbf{X}_k a_n^{-1}$. Эта последовательность серий является нулевой, если $a_n\to\infty$. Иного типа схема серий была рассмотрена при выводе пуассоновского распределения в **1**, гл. VI, 6. ▶

В главах IX и XVII будет доказан следующий замечательный факт: *предельное распределение для сумм* $\mathbf{S}_n$ *величин, образующих* n*-ю серию в нулевой схеме серий (если оно суще-*

---

[1]) В подлиннике triangular array — треугольная схема. — *Прим. перев.*

[2]) В том, что увеличение общности является здесь просто иллюзией, легко убедиться следующим образом. Можно увеличить число величин в каждой серии, добавив «случайные величины», тождественно равные нулю. Следовательно, не ограничивая общности, можно допустить, что число величин в n-й серии растет вместе с n. Пусть дана такая схема серий $\mathfrak{A}$. Мы построим теперь новую схему серий Первые r_1 серий возьмем произвольными с числом величин в серии равным номеру серии. Затем первую серию $\mathfrak{A}$ повторим r_2-r_1 раз, вторую — r_3-r_2 раз и т. д. В новой схеме серий n-я серия содержит не более n величин, и мы опять можем увеличить длину серий, добавив нули и сделав длину n-й серии равной n. Последовательность $\{\mathbf{S}'_n\}$ сумм величин n-серии в новой схеме серий отличается от $\{\mathbf{S}_n\}$ при больших n просто возможными повторениями, так что асимптотическое поведение обоих последовательностей $\{\mathbf{S}_n\}$ и $\{\mathbf{S}'_n\}$ одинаково.

ствует), безгранично делимо. Если выполнено условие асимптотической пренебрегаемости (3.3), то неважно, одинаково или неодинаково распределены компоненты $\mathbf{X}_{k,n}$, и в (3.2) знак равенства можно заменить пределом: класс безгранично делимых распределений совпадает с классом предельных распределений для сумм величин, образующих n-ю серию в нулевых схемах серий.

Примеры прикладного характера. з) *Дробовой эффект в вакуумных трубках.* Различные варианты и обобщения рассматриваемого ниже случайного процесса встречаются в физике и технике связи.

Рассмотрим флуктуации в электрическом токе, возникающие из-за случайных флуктуаций числа попадающих на анод электронов. Предполагается, что попадание электронов на анод образует пуассоновский процесс и что попавший электрон возбуждает ток, интенсивность которого x единиц времени спустя равна $I(x)$. Формально интенсивность тока в момент t можно записать в виде ряда

$$\mathbf{X}(t) = \sum_{k=1}^{\infty} I(t - \mathbf{T}_k), \tag{3.4}$$

где $\mathbf{T}_k$ — прошлые моменты попаданий электронов. (Иными словами, величины $t - \mathbf{T}_1$, $\mathbf{T}_2 - \mathbf{T}_1$, $\mathbf{T}_3 - \mathbf{T}_2$, ... взаимно независимы и имеют одинаковое показательное распределение.)

Нетрудно проанализировать сумму (3.4) непосредственно методами теории случайных процессов, однако простой подход с помощью схемы серий может помочь лучшему пониманию сути вещей. Разобьем интервал $-\infty$, t на маленькие подинтервалы с концами в точках $t_k = t - kh$ $(k = 0, 1, \ldots)$. В силу самого определения пуассоновского процесса вклад интервала $t_k t_{k-1}$ в сумму (3.4) можно приблизительно считать случайной величиной, принимающей значение 0 с вероятностью $1 - \alpha h$ и значение $I(t - t_k)$ с вероятностью αh. Математическое ожидание и дисперсия этой случайной величины равны соответственно $\alpha h I(kh)$ и $\alpha h(1 - \alpha h) I^2(kh)$. Положим $h = 1/\sqrt{n}$ и построим схему серий, в которой $\mathbf{X}_{k,n}$ есть вклад k-го интервала. Сумма величин, образующих отдельную серию, имеет математическое ожидание $\alpha h \sum I(kh)$ и дисперсию $\alpha h(1 - \alpha h) \sum I^2(kh)$. Если ряду (3.4) можно придать какой-то смысл, то распределения сумм по сериям должны будут сходиться к распределению $\mathbf{X}(t)$, и мы приходим к следующим соотношениям:

$$\mathbf{E}(\mathbf{X}(t)) = \alpha \int_0^{\infty} I(s)\, ds, \quad \operatorname{Var}(\mathbf{X}(t)) = \alpha \int_0^{\infty} I^2(s)\, ds. \tag{3.5}$$

Эти выводы легко подтвердить с помощью предельных теорем для схемы серий (см. задачу 5, гл. XVII, 12 и задачу 22, гл. VIII, 10). Соотношения (3.5) хорошо известны как *теорема Кэмпбелла*. С современной точки зрения это неглубокий результат, однако он был получен в 1909 г. за десятилетия до того, как была развита систематическая теория. Его справедливо считали замечательным результатом и передоказывали многими разными способами.

и) *Загруженность линий связи.* Приведем еще один сходный с предыдущим пример, иллюстрирующий характер возможных обобщений. Рассмотрим телефонную станцию с бесконечным числом линий. Предположим, что приходящие вызовы образуют пуассоновский процесс и что каждый приходящий вызов направляется на свободную линию. Времена обслуживания имеют одинаковое распределение F и как обычно предполагаются независимыми друг от друга и от процесса поступления вызовов. Число занятых линий в момент t является случайной величиной $\mathbf{X}(t)$, распределение которой можно найти, применяя метод предыдущего примера. В рассматриваемом случае величины $\mathbf{X}_{k,n}$, образующие серии, принимают лишь два значения 0 и 1, причем значение 1 имеет вероятность $\alpha h[1-F(kh)]$. В результате мы получаем, что *математическое ожидание числа занятых линий равно*

$$\mathbf{E}(\mathbf{X}(t)) = \alpha \int_0^\infty [1-F(s)]\,ds. \tag{3.6}$$

Отметим, что интеграл в (3.6) равен математическому ожиданию времени обслуживания. ▶

Исторические замечания. Понятие безграничной делимости восходит к Б. де Финетти (1929). Преобразования Фурье безгранично делимых распределений с конечными дисперсиями были найдены А. Колмогоровым (1932). В 1934 г. П. Леви нашел преобразования Фурье произвольных безгранично делимых распределений. Им же безгранично делимые распределения изучались с точки зрения случайных процессов. Все последующие исследования происходили под сильным влиянием фундаментальной работы П. Леви. Первые чисто аналитические выводы общей формулы были даны в 1937 г. независимо Феллером и Хинчиным. Этими авторами было также доказано, что предельные распределения для нулевых схем серий безгранично делимы.

§ 4. Процессы с независимыми приращениями

Безгранично делимые распределения тесно связаны со случайными процессами с независимыми приращениями. Под процессом с независимыми приращениями мы понимаем *зависящее от непрерывного параметра t семейство случайных величин* $\mathbf{X}(t)$,

такое, что для любого конечного множества значений параметра $t_1<t_2<\ldots<t_n$ *приращения* $\mathbf{X}(t_{k+1})-\mathbf{X}(t_k)$ *взаимно независимы.* На данном этапе нам не нужна теория случайных процессов; мы просто считаем, что *если* некоторое явление допускает вероятностное описание, то теория приведет к безгранично делимым распределениям. В этом смысле мы уже рассматривали некоторые процессы с независимыми приращениями в **1**, гл. XVII, 1 и в примере гл. III, (8.а). Ограничимся для простоты рассмотрением числовых случайных величин $\mathbf{X}(t)$, хотя теория переносится и на векторные величины.

Процесс имеет *стационарные приращения*, если распределение величины $\mathbf{X}(s+t)-\mathbf{X}(s)$ зависит лишь от длины t интервала и не зависит от s.

Разобьем интервал $s,\ s+t$ посредством $n+1$ равноотстоящих точек $s=t_0<t_1<\ldots<t_n=s+t$ и положим $\mathbf{X}_{k,\,n}=\mathbf{X}(t_k)-\mathbf{X}(t_{k-1})$. Приращение $\mathbf{X}(s+t)-\mathbf{X}(s)$ процесса со стационарными независимыми приращениями равно сумме n независимых одинаково распределенных случайных величин $\mathbf{X}_{k,\,n}$ и, следовательно, *имеет безгранично делимое распределение.* Мы увидим, что верно и *обратное.* Именно, однопараметрическое семейство вероятностных распределений Q_t, определенное при $t>0$, может задавать распределения приращений $\mathbf{X}(s+t)-\mathbf{X}(s)$ процесса со стационарными независимыми приращениями тогда и только тогда, когда

$$Q_{s+t}=Q_s*Q_t,\ s,\ t>0. \tag{4.1}$$

Семейство распределений, удовлетворяющее соотношению (4.1), называется *полугруппой* распределений (см. гл. VIII, 3). Каждое безгранично делимое распределение может быть взято в качестве элемента Q_t ($t>0$ произвольно) такой полугруппы.

Прежде чем переходить к нестационарному случаю, рассмотрим несколько типичных примеров.

Примеры. а) *Обобщенный пуассоновский процесс.* Пусть F — произвольное вероятностное распределение и $\alpha>0$. Обобщенный пуассоновский процесс задается формулой

$$Q_t=e^{-\alpha t}\sum_{k=0}^{\infty}\frac{(\alpha t)^k}{k!}F^{k*}. \tag{4.2}$$

Нетрудно убедиться, что соотношение (4.1) здесь действительно выполняется. Пусть $\mathbf{X}(t)$ — случайный процесс со стационарными независимыми приращениями, приращения $\mathbf{X}(t)-\mathbf{X}(0)$ которого имеют распределения Q_t. Когда распределение F сосредоточено в точке 1, этот процесс сводится к обычному

пуассоновскому процессу и (4.2) можно записать в виде

$$\mathbf{P}\{\mathbf{X}(t)-\mathbf{X}(0)=n\}=e^{-\alpha t}\frac{(\alpha t)^n}{n!}. \tag{4.3}$$

Общая модель (4.2) допускает следующую интерпретацию в терминах обычного пуассоновского процесса. Пусть $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ — независимые случайные величины с одинаковым распределением F и пусть $\mathbf{N}(t)$ — обычный пуассоновский процесс: $\mathbf{P}\{\mathbf{N}(t)=n\}=e^{-at}(at)^n/n!$, не зависящий от $\mathbf{Y}_k$. Распределение (4.2) совпадает с распределением случайной суммы $\mathbf{Y}_1+\ldots+\mathbf{Y}_{\mathbf{N}(t)}$. Иными словами, с n-м скачком пуассоновского процесса связывается эффект $\mathbf{Y}_n$, и приращение $\mathbf{X}(t)-\mathbf{X}(0)$ представляет собой сумму эффектов, соответствующих интервалу времени $\overline{0,\ t}$. Рандомизированное случайное блуждание, изучавшееся в гл. II, 7, является обобщенным пуассоновским процессом, отвечающим тому случаю, когда величины $\mathbf{Y}_k$ принимают лишь значения ± 1. Практические приложения будут проиллюстрированы в конце параграфа.

б) *Броуновское движение* или процесс Винера — Башелье. Так называется исходящий из начала координат процесс (т. е. $\mathbf{X}(0)=0$), приращения которого $\mathbf{X}(s+t)-\mathbf{X}(s)$ за непересекающиеся промежутки времени независимы и имеют нормальное распределение с нулевым математическим ожиданием и дисперсией t. Винером и Леви было показано, что траектории такого процесса с вероятностью 1 непрерывны [1]) и что это свойство характеризует нормальное распределение среди всех безгранично делимых распределений.

в) *Устойчивые процессы.* Соотношение (1.9) для строго устойчивых процессов равносильно соотношению (4.1) с $Q_t(x)=R(t^{-1/\alpha}x)$. Таким образом, распределение $R(t^{-1/\alpha}x)$ определяет переходные вероятности процесса со стационарными независимыми приращениями; при $\alpha=2$ получаем броуновское движение. ▶

Основная теорема теории безгранично делимых распределений утверждает (см. гл. IX и XVII), что *самое общее решение уравнения* (4.1), *и, следовательно, самое общее безгранично делимое распределение, можно представить как предел соответствующей последовательности обобщенных пуассоновских распределений.* Результат этот довольно удивителен, если принять во внимание большую формальную разницу между примерами (а) и (б).

[1]) Как всюду в этом томе, автор отвлекается от ограничений типа сепарабельности и т. п. — *Прим. перев.*

Распределение величины $\mathbf{X}(t+s) - \mathbf{X}(t)$ в случае процесса с нестационарными независимыми приращениями тоже представимо как распределение сумм по строкам в схеме серий $\mathbf{X}_{k,n}$, однако теперь для выполнения (3.3) нужно потребовать выполнения некоторых условий непрерывности. Пример (д) объясняет необходимость этого требования. Если наложить некоторые слабые ограничения, то *распределения величин* $\mathbf{X}(t+s) - \mathbf{X}(t)$ *будут необходимо безгранично делимы.*

Примеры. г) *Операционное время.* Простая замена времени нередко сводит общий процесс к более доступному изучению процессу со стационарными приращениями. Если φ — произвольная возрастающая непрерывная функция, то замена $t \to \varphi(t)$ переводит процесс $\mathbf{X}(t)$ в процесс $\mathbf{X}(\varphi(t))$. Свойство независимости приращений, очевидно, сохраняется, и при подходящем выборе φ приращения нового процесса могут оказаться стационарными. В практике выбор функции φ обычно подсказывается природой рассматриваемого процесса. Например, на *телефонной станции* вряд ли кто стал бы сравнивать ночной час с загруженным дневным часом — естественно измерять время в переменных единицах, таких, что ожидаемое число вызовов за единицу времени оставалось бы постоянным. Аналогично при растущем *страховом* бизнесе претензии будут поступать с возрастающей частотой, однако это отклонение от стационарности устраняется просто введением операционного времени для измерения частоты поступления претензий.

д) *Фиксированные точки разрыва.* Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — взаимно независимые случайные величины и $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. Определим процесс с (нестационарными) независимыми приращениями, положив $\mathbf{X}(t) = \mathbf{S}_n$ при $n \leqslant t < n+1$. Все изменения этого процесса происходят в моменты времени 1, 2, 3, ... Из этого примера видно, что для безграничной делимости распределений приращений необходимы некоторые условия непрерывности. Кроме того, этот пример наглядно иллюстрирует возможность разрывов в фиксированные моменты времени. В случае пуассоновского процесса траектории разрывны, однако при каждом *фиксированном* t вероятность того, что никаких изменений (скачков) не происходит за интервал времени $t, t+h$, равна $e^{-\alpha h}$ и стремится к единице при $h \to 0$. Иными словами, хотя число телефонных вызовов меняется скачками, вероятность того, что скачок произойдет в короткий промежуток времени, весьма мала. В противоположность этому в рассматриваемом примере при $t = 1, 2, \ldots$ скачки происходят с положительной вероятностью. ▶

Примеры *прикладного характера.* Огромное количество практических задач можно свести к обобщенным пуассоновским процессам. Приведем несколько типичных примеров. (i) Накопленный *убыток* за счет автомобильных катастроф, пожаров, молний и т. п. По поводу применений к *теории страхования* см. пример (5а). (ii) Общий улов рыболовецкого судна при поиске косяков рыбы (Дж. Нейман). (iii) Количество воды в *водохранилище* в зависимости от дождей и потребления воды. Аналогично рассматриваются другие задачи *хранения запасов.* (iv) *Камень, лежащий на дне реки,* находится без движения в течение столь долгих промежутков времени, что последовательные смещения допустимо считать практически мгновенными. Общее смещение камня за время $\overline{0, t}$ можно рассматривать как обобщенный пуассоновский процесс. (Впервые этот вопрос изучался другими методами Альбертом Эйнштейном младшим и Д. Пойа.) (v) *Телефонные вызовы,* или прибытия клиентов на пункт обслуживания. При соответствующих условиях общее время, необходимое для обслуживания требований, поступивших в интервале времени $\overline{0, t}$, описывается обобщенным пуассоновским процессом. Замечательной особенностью этого процесса является тот факт, что значение $\mathbf{X}(t)$ *не наблюдаемо в момент* t, поскольку оно определяется временами обслуживания, относящимися к будущему. (vi) Изменения энергии физических частиц при столкновениях рассмотрены в примере X, (1. б).

§ 5*). Обобщенные пуассоновские процессы и задачи о разорении

Пусть $\mathbf{X}(t)$ — обобщенный пуассоновский процесс. Это значит, что приращение $\mathbf{X}(t+s) - \mathbf{X}(s)$ за любой интервал времени длины t имеет распределение Q_t, задаваемое формулой (4.2). Пусть $c>0$ и $z>0$ фиксированы. *Разорением* мы называем событие

$$\{\mathbf{X}(t) > z + ct\}. \tag{5.1}$$

В дальнейшем c будет рассматриваться как постоянная, а $z>0$ как свободный параметр. *Обозначим через* $R(z)$ *вероятность того, что разорение никогда не произойдет.* Предполагая, что вся задача имеет смысл, мы выведем формально, что вероятность $R(z)$ должна быть неубывающим решением функциональ-

*) Этот параграф посвящен специальной теме. Он имеет большой практический интерес, но материал его нигде не будет использоваться в дальнейшем, за исключением примеров.

ного уравнения (5.2). Приведем сначала несколько примеров, показывающих степень разнообразия практических ситуаций, приводящих к рассматриваемой задаче.

Примеры. а) *Теория страхования*[1]). Значением $\mathbf{X}(t)$ здесь является накопленное количество претензий за интервал времени $\overline{0, t}$, предъявленных страховой компании. Предполагается, что поступление претензий описывается пуассоновским процессом и что отдельные претензии имеют распределение F. В принципе «претензии» могут быть как положительными, так и отрицательными (например, смерть клиента может освободить компанию от обязательства и увеличить ее резервы). В рассматриваемом примере z представляет собой исходный резерв компании, имеющийся в момент времени 0, а c — скорость возрастания резервов[2]) при отсутствии претензий. Общий резерв компании в момент t равен $z+ct-\mathbf{X}(t)$, а «разорение» означает достижение отрицательного резерва, т. е. банкротство.

б) *Хранение запасов.* Идеальное водохранилище наполняется реками и дождями с постоянной скоростью c. В случайные моменты времени расходуется часть воды в количестве $\mathbf{X}_1, \mathbf{X}_2, \ldots$. Рассматриваемая модель описывается обобщенным пуассоновским процессом, и если исходное количество воды в момент времени 0 равно z, то и в момент t теоретически количество воды будет равно $z+ct-\mathbf{X}(t)$, если только водохранилище не осушится до этого времени. Функция F сосредоточена здесь на полупрямой $\overline{0, \infty}$. Задачам подобного рода посвящена огромная литература (см. монографии, указанные в конце книги).

в) *Обслуживание больных*[3]). Будем рассматривать промежутки времени, которые врач тратит на обслуживание пациентов, как независимые случайные величины с показательным распределением и средней длительностью α^{-1}. Когда обслуживание происходит без перерывов, уход обслуженных пациентов подчинен

[1]) Этой теории (основанной Ф. Лундбергом) и, в частности, рассматриваемой нами задаче о разорении посвящена огромная литература. Обзор ее имеется в относительно недавней работе Г. Крамера (см. Cramér H., On some questions connected with mathematical risk, Univ. Calif. Publications in Statistics, vol. 2, № 5 (1954), 99—125). Отметим, что асимптотические оценки Крамера (полученные с помощью глубоких методов Винера — Хопфа) доказаны элементарными методами в примерах гл. XI, (7. а) и гл. XII (5. г).

[2]) На практике растущая компания будет использовать накопленную сумму полученных поступлений как операционное время [см. пример (4. г)].

[3]) Pyke R., The supremum and infimum of the Poisson process, *Ann. Math. Statist.*, **30** (1959), 568—576. Пайк рассматривает задачу о разорении лишь для чисто пуассоновского процесса, но зато получает более точные результаты (иными методами).

обычному пуассоновскому процессу. Пусть $\mathbf{X}(t)$ — число уходов пациентов за время $\overline{0, t}$. Предположим, что в момент 0 (начало работы врача) ожидают z пациентов и что последующие пациенты приходят в моменты времени c^{-1}, $2c^{-1}$, $3c^{-1}$, ... Врач не будет без дела, пока $\mathbf{X}(t) \leqslant z+ct$. ▶

Следующее формальное рассуждение приводит к уравнению, определяющему вероятность разорения $1-R$. Пусть *первый* скачок траектории происходит в момент τ и имеет величину x. Чтобы разорение никогда не произошло, необходимо, чтобы $x \leqslant z+c\tau$ и чтобы при всех $t>\tau$ приращения $\mathbf{X}(t)-x$ не превосходили $z-x+ct$. Поскольку приращения независимы, последнее из этих событий имеет вероятность, равную $R(z-x+c\tau)$. Интегрируя по всем возможным значениям τ и x, получаем

$$R(z)=\int_0^\infty \alpha e^{-\alpha\tau}\,d\tau \int_{-\infty}^{z+c\tau} R(z+c\tau-x)\,F\{dx\}. \tag{5.2}$$

Это и есть нужное нам уравнение, однако его еще можно упростить. Производя замену $s=z+c\tau$, имеем

$$R(z)=\frac{\alpha}{c}\int_z^\infty e^{-(\alpha/c)(s-z)}\,ds \int_{-\infty}^{s} R(s-x)\,F\{dx\}. \tag{5.3}$$

Из (5.3) следует, что функция R дифференцируема. Дифференцируя (5.3), получаем окончательное *интегро-дифференциальное уравнение*

$$R'(z)=\frac{\alpha}{c}R(z)-\frac{\alpha}{c}\int_{-\infty}^{z} R(z-x)\,F\{dx\}. \tag{5.4}$$

Отметим, что по определению $R(s)=0$ при $s<0$, и поэтому интеграл справа в (5.4) есть *свертка* $F*R$. Мы вернемся к уравнению (5.4) в примерах (9.г), гл. XI (7.а) и гл. XII (5.г).

§ 6. Процессы восстановления

Основные понятия теории восстановления были введены в 1, гл. XIII в связи с рекуррентными событиями. Мы увидим, что введение непрерывного времени связано скорее с изменениями в обозначениях, чем с изменениями по существу. Важнейшее свойство рекуррентных событий состоит в том, что последовательные времена ожидания $\mathbf{T}_k$ являются взаимно независимыми случайными величинами с общим распределением F;

момент n-го осуществления события равен

$$\mathbf{S}_n = \mathbf{T}_1 + \ldots + \mathbf{T}_n. \tag{6.1}$$

Принимается также соглашение, что $\mathbf{S}_n = 0$, и 0 считается осуществлением номер нуль.

Даже в вероятностных процессах, зависящих от непрерывного времени, нередко можно обнаружить одну или более последовательностей моментов времени, для которых имеет место (6.1). В таких случаях простыми методами удается получить удивительно точные результаты. Аналитически мы имеем дело просто с суммами независимых положительных случайных величин, и единственным оправданием введения термина «процесс восстановления» является его частое использование в связи с *другими* процессами, а также то обстоятельство, что при применении этого термина молчаливо подразумевается использование мощного средства — уравнения восстановления [1]).

Определение 1. *Последовательность случайных величин* $\{\mathbf{S}_n\}$ *образует процесс восстановления, если величины* $\mathbf{S}_n$ *имеют вид* (6.1), *где* $\mathbf{T}_k$ *— взаимно независимые случайные величины с общим распределением* F, *удовлетворяющим условию* [2]) $F(0) = 0$.

Поскольку рассматриваемые случайные величины положительны, то математическое ожидание $\mu = \mathbf{E}(\mathbf{T}_k)$ имеет смысл, даже если соответствующий интеграл расходится (в этом случае $\mu = \infty$). Математическое ожидание μ называется *средним временем возвращения*. Как обычно в подобных обстоятельствах, для нашего исследования неважно, появляются ли величины $\mathbf{T}_k$ в некотором вероятностном процессе или последовательность $\{\mathbf{T}_j\}$ сама определяет вероятностное пространство.

В большинстве (но не во всех) приложений величину $\mathbf{T}_j$ можно интерпретировать как «время ожидания», и суммы $\mathbf{S}_n$ являются тогда моментами *восстановления* (*или регенерации*).

Представляется интуитивно очевидным, что для любого конечного интервала $I = \overrightarrow{a, b}$ число попадающих в I моментов восстановления $\mathbf{S}_n$ конечно с вероятностью единица и, следовательно, представляет собой вполне определенную величину $\mathbf{N}$. Если событие $\{\mathbf{S}_n \in I\}$ называть «успехом», то $\mathbf{N}$ есть общее

---

[1]) Более сложное обобщение рекуррентных событий имеется в работе Kingman J. F. C., The stochastic theory of regenerative events, *Zeit. Wahrscheinlichkeitstheorie*, 2 (1964), 180—224. Русский перевод см. сб *Математика*, **11**:2 (1967), стр. 106—152.

[2]) Можно было бы допустить существование атома в нуле, однако это не привело бы ни к чему новому, и нам пришлось бы исключать случай распределения F, сосредоточенного в нуле.

число успехов в бесконечной последовательности испытаний с математическим ожиданием, равным

$$U\{I\} = \sum_{n=0}^{\infty} \mathbf{P}\{\mathbf{S}_n \in I\} = \sum_{n=0}^{\infty} F^{n*}\{I\}. \tag{6.2}$$

Для случайных величин, принимающих целочисленные значения, суммы (6.2) рассматривались в **1**, гл. XIII в связи с рекуррентными событиями. Для таких случайных величин достаточно было рассматривать интервалы, состоящие из одной целой точки, и события $\{\mathbf{S}_n \in I\}$ оказывались при этом взаимно исключающими. Поэтому было возможным рассматривать не математическое ожидание числа сумм $\mathbf{S}$, равных k, а вероятность u_k того, что одна из сумм $\mathbf{S}_n$ равна k. В рассматриваемой ситуации вовсе не очевидно, что ряд (6.2) сходится. Для доказательства его сходимости нужно доказать конечность функции

$$U(x) = \sum_{n=0}^{\infty} F^{n*}(x) \tag{6.3}$$

при всех x. Если функция U конечна, то функция интервала $U\{\overline{a, b}\} = U(b) - U(a)$ определяет сосредоточенную на $\overline{0, \infty}$ меру, имеющую единичный атом в начале координат.

Мы покажем конечность U и используем U для построения решения *уравнения восстановления*

$$Z = z + F * Z; \tag{6.4}$$

в этом уравнении Z и z — обращающиеся в нуль на $\overline{-\infty, 0}$ функции. Уравнение (6.4) является сокращенной записью следующего уравнения [1])

$$Z(x) = z(x) + \int_0^x Z(x-y) F\{dy\}, \qquad x > 0, \tag{6.5}$$

которое является аналогом уравнения восстановления, изучавшегося в **1**, гл. XIII, и не раз встретится нам в дальнейшем.

Теорема 1 [2]). *Функция $U(x)$ конечна при всех x. Если z — ограниченная функция, обращающаяся в нуль при $x<0$, то функция $Z = U * z$, определяемая равенством*

$$Z(x) = \int_0^x z(x-y) U\{dy\}, \tag{6.6}$$

[1]) Пределы интегрирования указаны лишь для удобства; интеграл не изменится, если интегрирование производить по всей прямой.

[2]) Более подробная теория развивается в § 10.

является решением уравнения восстановления (6.5). *Это решение единственно в классе функций, обращающихся в нуль на* $-\infty, 0$ *и ограниченных на конечных интервалах.*

Доказательство. Если $\{\mathbf{S}_n \leqslant x\}$, то $\mathbf{T}_j \leqslant x$ при $j=1, \ldots, n$. Следовательно, $\mathbf{P}\{\mathbf{S}_n \leqslant x\} \leqslant F^n(x)$. Отсюда вытекает, что ряд в (6.3) сходится со скоростью геометрической прогрессии при всех x, для которых $F(x)<1$. Но для каждого x существует целое число r, такое, что $F^{r*}(x)<1$, и, так как $F^{k*}(x)$ монотонно зависит от k, мы имеем

$$U(x) \leqslant r \sum_{k=0}^{\infty} F^{rk*}(x) < \frac{r}{1-F^{r*}(x)} < \infty. \tag{6.7}$$

Беря свертки обеих частей равенства (6.3) с z, убеждаемся в том, что Z удовлетворяет уравнению восстановления. Разность двух решений V удовлетворяет уравнению $V=F*V$ и, следовательно, всем уравнениям $V=F^{n*}*V$, $n=1, 2, \ldots$. Но $F^{n*}(x) \to 0$ при всех x, и поэтому $V(x)=0$. ▶

Отметим, что если $z(x)=1$ при $x>0$, то функция $Z(x)=U(x)$ является решением уравнения восстановления (6.5). Это можно показать непосредственно, используя следующее рассуждение.

Ожидаемое число моментов восстановления в замкнутом интервале $\overline{0, x}$ равно единице плюс ожидаемое число таких моментов в полуоткрытом интервале $\overline{0, x}$. Интервал $\overline{0, x}$ содержит моменты восстановления лишь когда $\mathbf{T}_1 \leqslant x$; при условии, что $\mathbf{T}_1 = y \leqslant x$, ожидаемое число моментов восстановления в интервале $\overline{0, x}$ равно $U(x-y)$. Интегрирование по распределению y приводит теперь к уравнению (6.5). Проведенное рассуждение является типичным при изучении процессов восстановления и нередко используется для нахождения различных распределений и математических ожиданий.

Имеются два полезных обобщения процесса восстановления. Рассмотрим сначала *обрывающиеся процессы*. Обрывающийся процесс может прекратиться в любой момент восстановления, причем это событие не зависит от прошлого и имеет фиксированную вероятность q, $0<q<1$ (такие процессы являются аналогом недостоверных рекуррентных событий). Выражаясь абстрактно, можно сказать, что к выборочному пространству добавляется одна точка, называемая *смертью*, и сумма $\mathbf{S}_n$ либо равна положительному числу, либо «погибает». Распределение F является теперь условным распределением времени возвращения $\mathbf{T}_j$ при условии, что процесс *не обрывается*. В качестве

безусловного распределения T_j имеет теперь распределение $(1-q)F$. Изменим обозначения и будем вместо $(1-q)F$ писать просто F, подразумевая при этом, что $F(\infty)<1$, т. е. что F есть *несобственное распределение*. Все рассмотренные выше формулы остаются тогда без изменений. Для удобства ссылок введем теперь следующее неформальное

Определение 2[1]). *Процесс, отличающийся от обычного процесса восстановления лишь тем, что функция распределения F — несобственная, называется обрывающимся или невозвратным процессом восстановления. Дефект $q=1-F(\infty)$ интерпретируется как вероятность обрыва.*

Ради согласованности 0 считается моментом восстановления номер нуль. Вероятность того, что процесс переживет n-й момент восстановления, равна $(1-q)^n$ и стремится к 0 при $n\to\infty$. Таким образом, *с вероятностью единица обрывающийся процесс обрывается за конечное время.* Общая масса распределения F^{n*} есть $(1-q)^n$, и поэтому *математическое ожидание числа моментов возвращения равно* $U(\infty)=q^{-1}<\infty$. Число $U(\infty)$ является, так сказать, ожидаемым числом поколений, до которого развивается процесс. Вероятность того что $\mathbf{S}_n\leqslant x$, и процесс обрывается на n-м моменте восстановления, равна $qF^{n*}(x)$. Нами доказана, таким образом,

Теорема 2. *Для обрывающегося процесса восстановления функция qU является собственным вероятностным распределением длительности жизни процесса (возраста в момент смерти).*

Второе обобщение состоит в том, что *исходному* времени ожидания разрешается иметь другое распределение. В этом случае добавляется величина $\mathbf{T}_j$ с $j=0$ и $\mathbf{S}_0=\mathbf{T}_0\neq 0$.

Определение 3. *Процессом восстановления с запаздыванием называется последовательность* $\mathbf{S}_0, \mathbf{S}_1, \ldots$ *вида* (6.1), *где* $\mathbf{T}_k$ *— взаимно независимые строго положительные (собственные или несобственные) случайные величины, причем* $\mathbf{T}_1, \mathbf{T}_2, \ldots$ (*исключая* $\mathbf{T}_0$) *имеют одинаковое распределение.*

Например, процесс восстановления становится процессом восстановления с запаздыванием, если его рассматривать лишь после некоторого момента τ.

§ 7. Примеры и задачи

Такие примеры, как самовосстанавливающиеся совокупности, счетчики и рост популяций, очевидным образом переносятся с

[1]) В примере 7(e) имеется иллюстрация этого определения, а в задаче 10 — его обобщение.

дискретного случая. Однако одна специальная задача приводит к интересным вопросам, которые будут рассматриваться в дальнейшем.

Пример. а) *Парадокс контроля.* В теории самовосстанавливающихся совокупностей рассматривается какой-либо прибор, например электрическая батарея, который устанавливается и работает до тех пор, пока не выходит из строя. Как только он выходит из строя, его сразу же заменяют другим аналогичным прибором, и процесс продолжается без перерыва. Моменты восстановления образуют процесс восстановления, в котором $\mathbf{T}_k$ есть срок действия k-й батареи.

Предположим теперь, что действительный срок службы участвующих в процессе батарей должен проверяться посредством систематического или выборочного контроля. Берется выборка батарей, участвующих в процессе в момент $t>0$, и наблюдается их полный срок службы. Поскольку срок службы *всех* батарей имеет распределение F, то можно ожидать, что срок службы наблюденных батарей тоже имеет распределение F. Но это не так. В самом деле, в случае экспоненциального распределения F рассматриваемая ситуация лишь словесно отличается от ситуации с парадоксом времени ожидания, изучавшейся в гл. I, 4, где длительность наблюдаемого промежутка имела совсем другое распределение. *Тот факт, что прибор исследуется в момент t, меняет распределение срока службы и удваивает его ожидаемую длительность.* Мы увидим в гл. XI, (3.6), что это обстоятельство типично для всех процессов восстановления. Из сказанного вытекают серьезные для приложений следствия. Беспристрастный на первый взгляд план контроля может вести к ложным заключениям, поскольку то, *что в действительности наблюдается, может не быть типичным для всей популяции.* Будучи замеченным, этот феномен легко получил свое объяснение (см. гл. I, 4), но он тем не менее указывает на возможные опасности при выведении следствий из наблюдений и на необходимость взаимосвязи теории с практикой. Отметим, что никаких трудностей не возникает, если наблюдения производить над первым прибором, который устанавливается после момента t. ▶

Введем три случайные величины, важные для теории восстановления. В рассмотренном выше примере все эти три величины относятся к сроку службы участвующего в процессе прибора в момент $t>0$ и могут быть описаны следующими, не нуждающимися в объяснении терминами: остающееся время службы, прошедшее время службы и полное время службы. Дадим теперь формальное определение.

Каждому $t>0$ поставим в соответствие однозначно определяемый (случайный) индекс $\mathbf{N}_t$, такой, что $\mathbf{S}_{\mathbf{N}_t} \leqslant t < \mathbf{S}_{\mathbf{N}_t+1}$. Тогда

а) *остающееся время ожидания* по определению равно $\mathbf{S}_{\mathbf{N}_t+1} - t$, т. е. равно времени, протекающему от t до следующего за t момента восстановления;

б) *прошедшее время ожидания* по определению равно $t - \mathbf{S}_{\mathbf{N}_t}$, т. е. равно времени, протекающему до t с первого предшествующего t момента восстановления;

в) сумма остаточного и проведенного времени ожидания, равная $\mathbf{S}_{\mathbf{N}_t+1} - \mathbf{S}_{\mathbf{N}_t} = \mathbf{T}_{\mathbf{N}_t+1}$, есть длина интервала возвращения, накрывающего t.

Введенную терминологию нельзя считать установившейся — она меняется от контекста к контексту. Например, в теории случайных блужданий остающееся время ожидания называют *точкой первого достижения* интервала t, ∞. В рассмотренном выше примере время ожидания называлось сроком службы. Все три введенные случайные величины будут изучаться в гл. XI, 3 и гл. XIV, 3.

Пуассоновский процесс был определен как процесс восстановления с показательным распределением промежутков времени ожидания $\mathbf{T}_j$. Во многих задачах массового обслуживания естественно считать, что входящий поток образует пуассоновский процесс. В некоторых других процессах промежутки времени между поступлениями постоянны. Для объединения этих двух случаев в теории массового обслуживания стало модным рассматривать общий процесс восстановления с произвольными промежутками времени между поступлениями [1]).

Перейдем теперь к некоторым задачам довольно общего характера, связанным с процессами восстановления. Соответствующее процессу распределение по-прежнему будем обозначать буквой F.

Начнем с того, что грубо можно было бы описать как «*время ожидания* $\mathbf{W}$ *до большого пропуска*». Процесс восстановления с временами ожидания $\mathbf{T}_j$ будем останавливать при первом появлении интервала времени длины ξ, в течение кото-

[1]) Общность эта в известной степени обманчива, поскольку трудно найти соответствующие практические примеры, за исключением примера автобуса, идущего без расписания по круговому маршруту. Иллюзия общности несколько умаляется тем печальным фактом, что непуассоновский вход обычно не является также и марковским.

рого не происходит восстановлений. Можно было бы дать такое определение обрывающегося процесса восстановления, чтобы оно включало рассматриваемый процесс (см. задачу 10), однако мы воспользуемся прямым рассуждением, приводящим к уравнению восстановления для *функции распределения V времени ожидания* **W**. Так как время ожидания **W** превышает ξ, то $V(t)=0$ при $t<\xi$. При $t\geqslant\xi$ рассмотрим две взаимно исключающие возможности: либо $\mathbf{T}_1>\xi$, либо $\mathbf{T}_1=y\leqslant\xi$. В первом случае $\mathbf{W}=\xi$. Во втором случае процесс начинается сначала, и условная вероятность того, что $\{\mathbf{W}\leqslant t\}$ при условии $\mathbf{T}_1=y$, равна $V(t-y)$. Суммируя по всем имеющимся возможностям, получаем

$$V(t)=1-F(\xi)+\int_0^{\xi+} V(t-y)\,F\{dy\}, \qquad t\geqslant\xi; \tag{7.1}$$

при $t<\xi$, как было отмечено выше, $V(t)=0$. Таким образом, функция V удовлетворяет обычному уравнению восстановления

$$V=z+G*V \tag{7.2}$$

с несобственной функцией распределения G и функцией z, определяемыми равенствами

$$G(x)=\begin{cases} F(x) & \text{при } x\leqslant\xi,\\ F(\xi) & \text{при } x\geqslant\xi; \end{cases} \tag{7.3}$$

$$z(x)=\begin{cases} 0 & \text{при } x<\xi,\\ 1-F(\xi) & \text{при } x\geqslant\xi. \end{cases} \tag{7.4}$$

Наиболее важным частным случаем является случай пропусков в пуассоновском процессе, т. е. случай, когда $F(t)=1-e^{-ct}$. Функция V здесь связана с теоремами о покрытии из гл. I, 9 [см. задачу 12 и пример гл. XIV, (2.а)].

Примеры *прикладного характера* б). *Пересечение потока автомобилей*[1]). Пусть автомобили движутся в один ряд с постоянной скоростью, причем последовательные пересечения автомобилями перехода (или переезда) образуют выборку из пуассоновского процесса (или из какого-либо другого процесса восстановления). Пешеход, появившийся на тротуаре (или автомобиль, подъехавший к перекрестку), начинает переход (переезд)

[1]) С относительно старой литературой и другими вариантами этой задачи (изучаемыми иными методами) можно познакомиться по работе Tanner J. C., The delay to pedestrians crossing a road, *Biometrica*, 38 (1951), 383—392.

после того, как убеждается, что в течение следующих ξ секунд времени, необходимых для перехода (переезда), не пройдет ни одного автомобиля. Обозначим через **W** время, необходимое для перехода (переезда), т. е. время ожидания на тротуаре (перекрестке) плюс ξ. Функция распределения V величины **W** удовлетворяет уравнению (7.1) с $F(t)=1-e^{-ct}$ [продолжение см. в примерах гл. XI, (7.б) и XIV, (2.а)].

в) *Счетчики Гейгера типа* II. Попадания частиц в счетчик образуют пуассоновский процесс, и каждая *попадающая* частица (зарегистрированная или нет) блокирует счетчик на фиксированное время ξ. Если в счетчик поступила частица, то он остается «мертвым» (блокированным) до появления интервала длины ξ, в течение которого нет попаданий новых частиц. Наша теория применима к распределению V *длительности мертвого периода* [см. 1, гл. XIII, 11, задача 10].

г) *Максимальное наблюденное время возвращения.* Рассмотрим процесс восстановления; пусть $\mathbf{Z}_t$ обозначает максимальную длину промежутков времени $\mathbf{T}_j$, наблюденных[1]) до момента t. Событие $\{\mathbf{Z}_t \leqslant \xi\}$ происходит тогда и только тогда, когда до момента t каждый интервал времени длины t содержит момент восстановления, и поэтому в наших обозначениях $\mathbf{P}\{\mathbf{Z}_t > \xi\} = V(t)$. ▶

Значительное число встречающихся в приложениях процессов восстановления может быть охарактеризовано как *процессы с двумя стадиями.*

Пользуясь терминологией, принятой для счетчиков частиц, мы будем описывать процессы такого типа в терминах «свободных» и «мертвых» периодов, хотя в некоторых физических приложениях вместо термина «мертвый» больше подходил бы «возбужденный». Мертвые и свободные периоды процесса чередуются, и полный цикл образует время возвращения.

Примеры. д) *Поломки с последующими простоями.* В качестве простейшего примера рассмотрим последовательные замены выходящей из строя детали в предположении, что каждая поломка сопровождается некоторым простоем (его можно интерпретировать как время, необходимое для обнаружения поломки или же как время, нужное для ремонта). Последовательные периоды работы $\mathbf{T}_1, \mathbf{T}_2, \ldots$ чередуются с последовательными

---

[1]) Более точно, пусть n — (случайный) индекс, такой, что $\mathbf{S}_{n-1} \leqslant t < \mathbf{S}_n$. Тогда $\mathbf{Z}_t = \max[\mathbf{T}_1, \ldots, \mathbf{T}_{n-1}, \xi]$. Случайные величины этого типа систематически изучались А. Ламперти.

мертвыми периодами $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$, и мы получаем собственный обычный процесс восстановления с временами возвращения $\mathbf{T}_j + \mathbf{Y}_j$. Этот процесс можно также рассматривать как процесс восстановления с запаздыванием с моментом первого восстановления, равным $\mathbf{T}_1$, и временами возвращения $\mathbf{Y}_j + \mathbf{T}_{j+1}$.

е) *Потерянные вызовы.* Рассмотрим одну телефонную линию; пусть поступающие на нее вызовы образуют пуассоновский процесс с распределением времен между вызовами, равным $F(t) = 1 - e^{-ct}$. Предположим, что длительности следующих за вызовами разговоров являются независимыми случайными величинами с общим распределением G. Линия может быть свободной или занятой (мертвой), и вызовы, поступающие в течение мертвых периодов, теряются. Мы имеем здесь процесс с двумя стадиями с распределением времени возвращения, равным $F * G$.

Найдем распределение V *времени ожидания* $\mathbf{W}$ *до первого потерянного вызова, предполагая, что в момент* 0 *линия была свободна.* Проще всего здесь применить метод обрывающихся процессов. Будем останавливать исходный процесс в момент первого теряющегося вызова. В полученном процессе свободные и мертвые периоды чередуются, однако мертвый период может теперь прекратиться по поступлении нового вызова. При этих условиях длительность мертвого периода равна наименьшей из двух независимых случайных величин с распределениями F и G соответственно. Несобственная функция распределения H длительности мертвого периода равна $\int_0^x [1 - F(y)]G\{dy\}$, и мы имеем обрывающийся процесс с несобственным распределением времени возвращения $F * H$. Время ожидания до первого потерянного вызова равно длительности жизни построенного обрывающегося процесса. Соответствующее уравнение восстановления и асимптотические оценки приводятся в гл. XI, 6 [1]).

ж) *Прибывший последним обслуживается первым.* Пусть имеется устройство для обработки информации и пусть время, необходимое для обработки вновь поступающей информации, является случайной величиной с распределением G. Предполагается, что выполнены обычные условия независимости. Если устройство свободно, то обработка начинается сразу же по поступлении новой информации. Бывают ситуации, когда интерес-

[1]) Поучительно сравнить использованный метод с обычным. Рассматривая все имеющиеся возможности, вплоть до момента остановки первого мертвого периода, можно вывести уравнение восстановления для V, которое аналогично уравнению задачи 13, хотя и сложнее его. Дальнейшее хорошее упражнение в формальных преобразованиях — доказать эквивалентность этого уравнения с более простым уравнением, предложенным в тексте.

на лишь последняя поступившая информация. В этих случаях она сразу же начинает обрабатываться, а эффект предыдущих поступлений полностью аннулируется. Опять мы получаем процесс восстановления, в котором свободные и мертвые периоды чередуются, однако распределение мертвых периодов здесь нужно вычислять с помощью метода, аналогичного методу примера (в) (задача 13).

з) *Случайные импульсы.* Первую модель предыдущего примера можно обобщить, сделав иные допущения относительно мертвых периодов. Пусть имеется произвольный исходный процесс восстановления с «прибытием» в момент 0 в течение свободного периода. После свободного начинается мертвый период. По истечении мертвого периода мы ждем до следующего прибытия в исходном процессе и предполагаем, что этот процесс начинается сначала. Иными словами, второй процесс является подпоследовательностью исходного процесса; время возвращения в нем состоит из мертвого периода со следующим за ним временем ожидания до следующего момента возвращения исходного процесса. Тот факт, что второй процесс является процессом восстановления, нужно либо предположить, либо доказать. Следующее условие является достаточным: длительности $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ мертвых периодов образуют последовательность независимых случайных величин с одинаковым распределением F, не зависящую от исходного процесса. Такое предположение независимости не является, однако, необходимым [пример (и)].

и) *Счетчики Гейгера.* В счетчиках типа I за каждой *регистрацией* следует мертвый период фиксированной длины ξ. Попадания частиц в этот период никак не отмечаются. Процесс здесь тот же, что и в примере (д), причем величины $\mathbf{T}_j$ имеют показательное распределение, а величины $\mathbf{Y}_j$ равны ξ. В счетчиках типа II нерегистрируемые частицы тоже приводят к блокировке и ситуация аналогична предыдущей с тем исключением, что распределения величин $\mathbf{Y}_j$ *зависят от исходного процесса* и должны вычисляться с учетом этого обстоятельства. Метод нахождения распределений $\mathbf{Y}_j$ описан в примере (в).

§ 8. Случайные блуждания

Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — взаимно независимые случайные величины с одинаковым распределением F. Обозначим, как обычно,

$$\mathbf{S}_0 = 0, \quad \mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n. \tag{8.1}$$

Будем говорить, что $\mathbf{S}_n$ есть положение в момент n частицы, совершающей случайное блуждание общего вида. Никаких новых

теоретических идей в этом параграфе не появляется [1]) — просто будет введена терминология для краткого и наглядного описания процесса $\{S_n\}$. Например, для всякого интервала I (или какого-либо другого множества) событие $\{S_n \in I\}$ называется *попаданием* в I, и изучение последовательных попаданий в заданный интервал I выявляет важные характеристики флуктуаций последовательности $S_1, S_2, \ldots$. Индекс n интерпретируется как временно́й параметр, и мы будем говорить о «моменте n». В этом параграфе некоторые удивительные особенности случайных блужданий описываются в терминах последовательных рекордных значений. Полезность полученных результатов будет видна из их применений в параграфе 9. Другой (независимый) подход будет рассмотрен в параграфе 10.

«Возложенные» процессы восстановления

Скажем, что в момент $n>0$ достигается рекордное значение, если

$$S_n > S_j, \quad j=0, 1, \ldots, n-1. \tag{8.2}$$

Такие индексы n могут существовать не для всех траекторий; если же они существуют, то образуют конечную или бесконечную упорядоченную последовательность. Поэтому можно говорить о первом, втором и т. д. осуществлении (8.2). Моменты этих осуществлений являются (возможно, несобственными) случайными величинами. Теперь мы в состоянии ввести важные случайные величины, на использовании которых будет основываться значительная часть исследования случайных блужданий.

Определение. *Назовем k-м (верхним) лестничным индексом момент k-го осуществления неравенств* (8.2). *Значение S_n в k-й лестничный момент будем называть k-й лестничной высотой. (Обе определенные случайные величины возможно являются несобственными).*

Нижние лестничные случайные величины определяются аналогично с помощью неравенств, обратных неравенствам (8.2) [2]).

Термин «верхний» обычно будет излишним и будет использоваться лишь при желании особо это подчеркнуть или достигнуть большей ясности.

---

[1]) Выборочные пространства бесконечных случайных блужданий уже рассматривались в первом томе, однако там нам приходилось определять такие понятия, как «вероятность разорения», с помощью очевидных предельных переходов. Теперь эти очевидные переходы к пределу обоснованы теорией меры (см. гл. IV, 6).

[2]) Заменяя в (8.2) строгие неравенства на неравенства $\geqslant$ и $\leqslant$, получаем слабые лестничные индексы. Это вызывающее некоторые осложнения различие несущественно, когда распределение F непрерывно. На рис. 1 слабые лестничные точки отмечены буквой w.

На графике траектории $(\mathbf{S}_0, \mathbf{S}_1, \ldots)$ лестничные точки — это те точки, в которых график доходит до не достигавшейся ранее высоты (рекордное значение). На рис. 1 (см. стр. 245) изображена траектория уходящего к минусу бесконечности случайного блуждания $\{\mathbf{S}_n\}$ с последним положительным положением при $n=31$. Здесь 5 верхних и 18 нижних лестничных точек этой траектории обозначены · и ∘ соответственно. Случайное блуждание с распределением Коши изображено на рис. 2 (см. стр. 259).

Пример. а) В «обычном» случайном блуждании F имеет атомы 1 и -1 с массами p и q. Верхние лестничные величины при $q>p$ будут несобственными, причем дефект равен $1-p/q$ [см. **1**, гл. XI, (3.9)]. Здесь k-я лестничная высота равна k и потому не упоминалась в первом томе; k-й лестничный индекс равен моменту *первого достижения* точки k. Распределение k-го лестничного индекса было найдено в **1**, гл. XI, 3, а для частного случая $p=\frac{1}{2}$ — в **1**, гл. III, 4. ▶

Первый лестничный индекс $\mathcal{T}_1$ равен моменту первого попадания в интервал $\overline{0, \infty}$, а *первая* лестничная высота $\mathcal{H}$ равна $\mathbf{S}_{\mathcal{T}_1}$. Продолжение случайного блуждания после момента $\mathcal{T}_1$ является вероятностной копией всего случайного блуждания, и поэтому число испытаний между первым и вторым лестничными индексами представляет собой случайную величину $\mathcal{T}_2$, не зависящую от $\mathcal{T}_1$ и имеющую одинаковое с $\mathcal{T}_1$ распределение. Аналогичное рассуждение приводит к следующему общему результату: *k-й лестничный индекс и k-ю лестничную высоту можно представить в виде*

$$\mathcal{T}_1+\ldots+\mathcal{T}_k,\quad \mathcal{H}_1+\ldots+\mathcal{H}_k,$$

где $\mathcal{T}_j$ и $\mathcal{H}_j$ — взаимно независимые одинаково распределенные случайные величины. Иными словами, лестничные индексы и высоты образуют (возможно, обрывающийся) *процесс восстановления.*

Интуитивно представляется очевидным, что в случае обрывающихся процессов суммы $\mathbf{S}_n$ уходят к $-\infty$ и с вероятностью единица достигается максимум последовательности $\{\mathbf{S}_n\}$. В следующем параграфе будет обнаружено, что лестничные случайные величины являются мощным аппаратом при исследовании одного класса процессов, имеющего значительный практический интерес.

Пример. б) *Точные выражения.* Пусть F имеет плотность

$$f(x)=\begin{cases}\dfrac{ab}{a+b}\,e^{ax}, & x<0,\\ \dfrac{ab}{a+b}\,e^{-bx}, & x>0.\end{cases} \tag{8.3}$$

Такое случайное блуждание обладает тем редким свойством, что все связанные с ним распределения могут быть точно вычислены. Это блуждание играет большую роль в теории массового обслуживания, поскольку функция f есть композиция двух экспоненциальных плотностей, сосредоточенных на $\overline{0,\infty}$ и $\overline{-\infty,0}$ соответственно. Иными словами, величины $\mathbf{X}_j$ могут быть представлены в виде разностей $\mathbf{X}_j=\mathscr{B}_j-\mathscr{A}_j$ двух *независимых положительных показательно распределенных случайных величин.* Не ограничивая общности, можем считать, что $a\leqslant b$.

В примере XII, (4.в) будет показано, что *верхняя лестничная высота* $\mathscr{H}_1$ *имеет плотность* ae^{-bx}; величина $\mathscr{H}_1$ будет несобственной тогда и только тогда, когда $a<b$, и если она несобственна, то дефект ее равен $(b-a)/b$. В гл. XVIII, 3 будет показано, что *верхний лестничный момент* $\mathscr{T}_1$ имеет производящую функцию $b^{-1}p(s)$, где

$$2p(s)=a+b-\sqrt{(a+b)^2-4abs}. \tag{8.4}$$

Дефект этой случайной величины тоже равен $(b-a)/b$.

Нижняя лестничная высота $\mathscr{H}_1^-$ при $x<0$ имеет плотность ae^{ax}, а производящая функция нижнего лестничного момента $\mathscr{T}_1^-$ равна [1]) $a^{-1}p(s)$.

При $a<b$ восходящий лестничный процесс является обрывающимся и величина

$$\mathbf{M}=\max[0,\ \mathbf{S}_1,\ \mathbf{S}_2,\ \ldots]$$

существует с вероятностью единица. Вероятность события $\{\mathbf{M}=0\}$ равна, очевидно, дефекту лестничных величин, т. е. равна $1-a/b$. В гл. XII, 5 будет показано, что при $x>0$ распределение величины $\mathbf{M}$ имеет плотность $\dfrac{b-a}{b}\,ae^{-(b-a)x}$. Производящая функция положения (индекса) максимума $\mathbf{M}$ равна $\dfrac{b-a}{b-p(s)}$. [Это вытекает из результатов примера гл. XVIII, (3.е).]

[1]) При $a=b$ производящая функция сводится к функции $1-\sqrt{1-s}$, известной нам из теории обычного случайного блуждания (бросаний монеты).

§ 9. Процессы массового обслуживания

Имеется чрезвычайно обширная литература [1]), посвященная различным вопросам, связанным с обслуживанием, управлением запасами, временем ожидания и т. п. Немало уже сделано для построения общей теории, однако наличие очень большого числа мало отличающихся вариантов затемняет ситуацию, так что трудно разглядеть лес за деревьями. Мощность новых общих методов до сих пор недооценивается. Начнем с формального введения случайного процесса, определяемого с помощью рекуррентной схемы, которая представляется на первый взгляд довольно искусственной. Дальнейшие примеры проиллюстрируют широкую применимость этой схемы; впоследствии мы увидим, что весьма точные результаты можно получить, используя удивительно простые методы (см. гл. XII, 5).

Определение 1. *Пусть* $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — *взаимно независимые случайные величины с одинаковым (собственным) распределением* F. *Индуцированным последовательностью* $\{\mathbf{X}_k\}$ *процессом ожидания назовем последовательность случайных величин* $\mathbf{W}_0, \mathbf{W}_1, \ldots$, *определяемую рекуррентно следующим образом: положим*

$$\mathbf{W}_{n+1} = \begin{cases} \mathbf{W}_n + \mathbf{X}_{n+1} & \text{при} \quad \mathbf{W}_n + \mathbf{X}_{n+1} \geqslant 0, \\ 0 & \text{при} \quad \mathbf{W}_n + \mathbf{X}_{n+1} \leqslant 0 \end{cases} \tag{9.1}$$

($n = 0, 1, \ldots$). *В сокращенной записи* (9.1) *означает, что* $\mathbf{W}_{n+1} = (\mathbf{W}_n + \mathbf{X}_{n+1}) \cup 0$.

---

[1]) Литературные ссылки можно найти в соответствующих монографиях, указанных в библиографии. Было бы трудно дать короткий очерк развития теории массового обслуживания со справедливым указанием авторства. Наиболее выдающиеся работы, в которых были заложены новые методы, считаются теперь устаревшими в силу того прогресса, который они стимулировали. [Примером может служить интегральное уравнение Д. В. Линдли в теории массового обслуживания (1952).] Другие работы заслуживают внимания благодаря исследованным в них (иногда очень сложным) частным задачам, однако нет возможности включить их в систематический обзор общей теории. В общем обширная литература в нескольких областях посвящена изучению различных вариантов и примеров, в то время как единой общей теории нет. Установление авторства затруднено также тем, что многие результаты были получены независимо разными авторами. [Например, решение одного простого интегрального уравнения мимоходом указано 1939 г. со ссылкой на неопубликованные лекции, читанные автором в 1934 г. Это решение было вновь найдено в различных контекстах многими авторами.] Читатели, интересующиеся историей развития теории, отсылаются к следующим двум обзорным статьям Д. Г. Кендалла, которые сами также оказали влияние на дальнейшие исследования (см., например, задачу в гл. XIV, 4): Some problems in the theory of queues; Some problems in the theory of dams, *J. Roy. Statist. Soc., Series B*, **13** (1951), 151—185, **19** (1957), 207—233.

Рис. 1 иллюстрирует это определение.

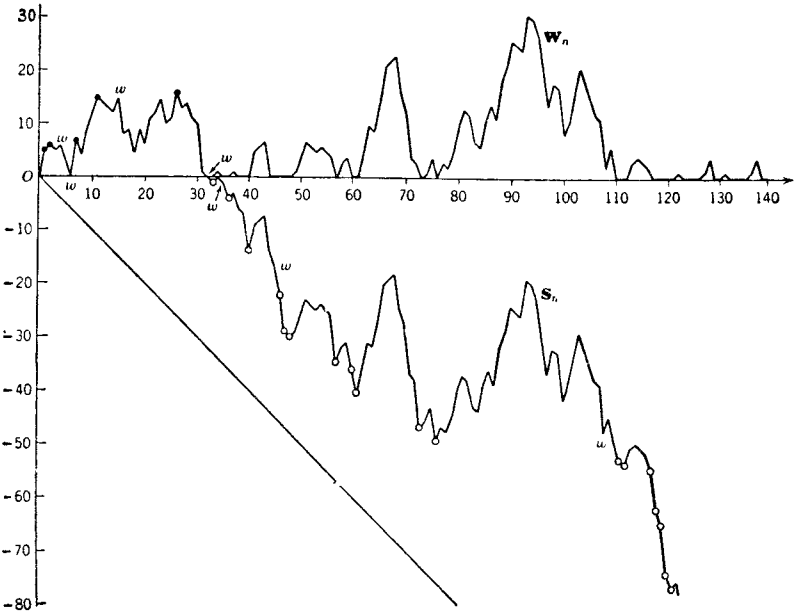


Рис. 1. Случайное блуждание и связанный с ним процесс ожидания.

Случайные величины $\mathbf{X}_n$ в случайном блуждании $\{\mathbf{S}_n\}$ имеют математическое ожидание -1 и дисперсию 16. Верхние и нижние лестничные точки обозначены ● и ○ соответственно. Пятая лестничная точка, имеющая координаты (26,16), с большой вероятностью представляет максимум всего случайного блуждания.

[Буква ω указывает те точки, где рекордные значения достигаются второй или третий раз; эти точки являются *слабыми* лестничными точками, определяемыми соотношениями (8.2) с заменой строгих неравенств на нестрогие $\geqslant$.]

Всюду на графике $\mathbf{S}_n$ *превосходит свое математическое ожидание, равное* $-n$. В действительности *первым индексом, для которого* $S_n \leqslant -n$, *является* $n=135$ (при этом $\mathbf{S}_{135} = -137$). Отмеченное обстоятельство согласуется с тем фактом, что математическое ожидание такого n равно бесконечности.

Случайные величины $\mathbf{X}_n$ имеют вид $\mathbf{X}_n = \mathscr{B}_n - \mathscr{A}_n$, где $\mathscr{B}_n$ и $\mathscr{A}_n$ — взаимно независимые случайные величины, распределенные равномерно на множествах 1, 3, 5, 7, 9 и 2, 4, 6, 8,10 соответственно. В примере (9. а) величины $\mathbf{W}_n$ представляют собой полное время ожидания n-го клиента, когда время между поступлениями требований принимает с одинаковой вероятностью значения 2, 4, 6, 8, 10, а время обслуживания равно каждому из чисел 1, 3, 5, 7 или 9 с вероятностью 1/5. Распределение величины $\mathbf{X}_n$ имеет массы $(5-k)/25$ в точках $\pm 2k-1$, где $k=0, 1, 2, 3, 4$.

Примеры. а) *Система с одним прибором.* Пусть на «прибор» поступают «требования», причем поступления требований образуют собственный процесс восстановления с *промежутками времени между поступлениями*[1]), равными $\mathscr{A}_1, \mathscr{A}_2, \ldots$ (требо-

[1]) Как правило, мы будем считать, что промежутки времени между поступлениями требований либо постоянны, либо имеют показательное распределение; см. сноску на стр. 236.

вания нумеруются числами 0, 1, 2, ... и поступают в моменты 0, $\mathscr{A}_1$, $\mathscr{A}_1+\mathscr{A}_2$, ...). Предположим, что на n-е требование тратится *время обслуживания* $\mathscr{B}_n$. Будем считать, что случайные величины $\mathscr{B}_n$ не зависят от моментов поступлений требований и друг от друга и имеют одинаковое распределение. Прибор может быть либо «свободным», либо «занятым», причем в начальный момент 0 он свободен. Схема обслуживания требований такова. Если требование поступает в момент, когда прибор свободен, то оно сразу же начинает обслуживаться. В противном случае оно становится в очередь и прибор непрерывно обслуживает требования в порядке их поступления [1]) до тех пор, пока очередь не исчезнет и прибор станет «свободным». Под *длиной очереди* мы понимаем число поступивших необслуженных требований (включая требование, которое обслуживается в данный момент). *Временем обслуживания* $\mathbf{W}_n$ n-го требования называется промежуток времени от его поступления до *начала* его обслуживания. Общее время, затраченное на обслуживание n-го требования, равно $\mathbf{W}_n+\mathscr{B}_n$ (например, если первые времена обслуживания суть 4, 4, 1, 3, ..., а времена между поступлениями требований суть 2, 3, 2, 3, то требования с номерами 1, 2, ... становятся в очередь длины 1, 1, 2, 1 соответственно и отвечающие им времена ожидания равны 2, 3, 2, 2, ...).

Чтобы избежать тривиальных сложностей типа поступления одного требования в момент окончания обслуживания другого, мы предположим, что распределения A и B случайных величин $\mathscr{A}_n$ и $\mathscr{B}_n$ непрерывны. В этом случае длина очереди в любой момент вполне определена.

Построим схему последовательного вычисления времен ожидания $\mathbf{W}_n$. По определению нулевое требование поступает в момент 0 на свободный прибор, и поэтому его время ожидания $\mathbf{W}_0$ равно нулю. Предположим теперь, что n-е требование поступает в момент t и что время его ожидания $\mathbf{W}_n$ известно. Обслуживание n-го требования начинается в момент $t+\mathbf{W}_n$ и завершается в момент $t+\mathbf{W}_n+\mathscr{B}_n$. Следующее $(n+1)$-е требование поступает в момент $t+\mathscr{A}_{n+1}$. В этот момент прибор свободен, если $\mathbf{W}_n+\mathscr{B}_n<\mathscr{A}_{n+1}$, а в противном случае время ожидания

[1]) Этот «порядок очереди» совершенно несуществен для длины очереди, длительности периодов занятости и аналогичных характеристик. Тот или иной порядок очереди имеет значение лишь с точки зрения обслуживания отдельных требований. Порядки очереди могут быть различными; «крайними случаями» являются следующие; требования обслуживаются в порядке их поступления, первым обслуживается последнее поступившее требование, обслуживается случайно выбранное из очереди требование. Вся картина меняется, если допустить возможность отказов.

$(n+1)$-го требования равно $\mathbf{W}_n + \mathscr{B}_n - \mathscr{A}_{n+1}$. Таким образом, *последовательность времен ожидания* $\{\mathbf{W}_n\}$ *совпадает с процессом ожидания, индуцированным независимыми случайными величинами*

$$\mathbf{X}_n = \mathscr{B}_{n-1} - \mathscr{A}_n, \quad n = 1, 2, \ldots \tag{9.2}$$

б) *Управление запасами.* Для наглядности мы рассмотрим *водохранилище*, однако используемая модель применима и в других ситуациях, связанных с управлением запасами. Количество воды в водохранилище зависит от ее притока и от ее использования. Приток происходит за счет рек и дождей, а использование определяется поступлением требований (требования, конечно, удовлетворяются лишь в том случае, когда водохранилище не пусто).

Рассмотрим теперь количества воды [1]) $0, \mathbf{W}_1, \mathbf{W}_2, \ldots$ в водохранилище в моменты $0, \tau_1, \tau_2, \ldots$. Пусть $\mathbf{X}_n$ есть разность между величиной действительного притока и величиной теоретического (идеального) требования, поступивших в интервал времени τ_{n-1}, τ_n. Будем считать, что все изменения происходят мгновенно и возможны лишь в моменты времени $\tau_1, \tau_2, \ldots$. В момент 0 имеем $\mathbf{W}_0 = 0$. В общем случае изменение $\mathbf{W}_{n+1} - \mathbf{W}_n$ содержимого бассейна равно $\mathbf{X}_{n+1}$, если только требование не превышает возможности. Следовательно, последовательность $\{\mathbf{W}_n\}$ должна удовлетворять соотношениям (9.1), т. е. *последовательные количества воды в бассейне образуют процесс ожидания, индуцированный последовательностью* $\{\mathbf{X}_n\}$, при условии, конечно, что случайные величины $\mathbf{X}_n$ независимы и одинаково распределены.

Задача (если не для прикладника, то во всяком случае для математика) состоит в том, чтобы найти условия, при выполнении которых величины $\mathbf{X}_k$ независимы и имеют одинаковое распределение F, а также найти возможные формы F. В практике величины τ_k либо находятся на одинаковом расстоянии друг от друга, либо представляют собой моменты последовательных скачков пуассоновского процесса, но для наших целей достаточно предположить, что τ_k образуют *процесс восстановления*, индуцированный, скажем, последовательностью $\mathscr{A}_1, \mathscr{A}_2, \ldots$. Наиболее часто используемые модели принадлежат одному из следующих двух классов.

(i) Приток поступает с постоянной скоростью c, требования $\mathscr{B}_n$ произвольны. При этом $\mathbf{X}_n = c\mathscr{A}_n - \mathscr{B}_n$, и мы должны предположить независимость $\mathbf{X}_n$ от «прошлого» $\mathbf{X}_1, \ldots, \mathbf{X}_{n-1}$

[1]) Для простоты мы начинаем с пустого водохранилища. Переход к произвольным начальным условиям не вызывает затруднений [см. пример (в)].

(обычное требование независимости $\mathscr{A}_n$ и $\mathscr{B}_n$ здесь излишне: нет оснований предполагать некоррелированность требования $\mathscr{B}_n$ с длительностью $\mathscr{A}_n$).

(ii) Использование запаса происходит с постоянной скоростью, приток произволен. Этот класс схож с предыдущим: отличие по существу лишь в том, что $\mathscr{A}_n$ и $\mathscr{B}_n$ поменялись ролями.

в) *Очереди на пригородные поезда*[1]). Пригородный поезд, имеющий r пассажирских мест, отправляется от станции каждый час. Будущие пассажиры появляются на станции и становятся в очередь. На каждый отходящий поезд садятся первые r человек, а остальные остаются в очереди. Предположим, что числа пассажиров, появляющихся на станции между последовательными отходами поездов, являются независимыми случайными величинами $\mathscr{A}_1, \mathscr{A}_2, \ldots$ с одинаковым распределением. Пусть число пассажиров, стоящих в очереди сразу же после n-го отправления, равно $\mathbf{W}_n$ и допустим, для простоты, что $\mathbf{W}_0=0$. Тогда $\mathbf{W}_{n+1}=\mathbf{W}_n+\mathscr{A}_{n+1}-r$, если $\mathbf{W}_n+\mathscr{A}_{n+1}-r>0$ и $\mathbf{W}_{n+1}=0$ в противном случае. Таким образом, последовательность $\{\mathbf{W}_n\}$ есть *процесс ожидания, индуцированный величинами* $\mathbf{X}_n=\mathscr{A}_n-r$. ▶

Поставим теперь себе задачу описать процесс ожидания $\{\mathbf{W}_n\}$ в терминах порожденного величинами $\mathbf{X}_k$ случайного блуждания. Как в параграфе 8, положим $\mathbf{S}_0=0$, $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ и будем придерживаться введенных там обозначений для лестничных величин. Для облегчения описания воспользуемся моделью примера (а).

Пусть ν — такой индекс, что $\mathbf{S}_1\geqslant 0$, $\mathbf{S}_2\geqslant 0, \ldots, \mathbf{S}_{\nu-1}\geqslant 0$, но $\mathbf{S}_\nu<0$. При этом требования с номерами $1, 2, \ldots, \nu-1$ имеют положительные времена ожидания $\mathbf{W}_1=\mathbf{S}_1, \ldots, \mathbf{W}_{n-1}=\mathbf{S}_{n-1}$, а ν-е требование является первым требованием (с положительным индексом), поступающим на свободный прибор (первое удачливое требование). С момента прибытия ν-го требования процесс идет как вероятностная копия всего процесса. С другой стороны, ν есть просто индекс первой отрицательной суммы, т. е. ν является первым нижним лестничным индексом. Обозначим его в соответствии с предыдущим через $\mathscr{T}_1^-$. Мы приходим, таким образом, к первому выводу: *нижние лестничные индексы соответствуют удачливым требованиям, находящим прибор сво-*

[1]) Boudreau P. E., Griffin J. S., Kac M., On elementary queuing problem, *Amer. Math. Monthly*, **69** (1962), 713—724. Эта статья имеет дидактический характер — она написана для неспециалистов, не знакомых с предметом. В ней используется другой подход, но приведенные там вычисления перекрываются вычислениями примера гл. XII, (4.в).

бодным. Отсюда вытекает, что моменты поступлений удачливых требований образуют процесс восстановления, у которого времена возвращений распределены одинаково с $\mathcal{T}_1^-$.

В практических случаях величина $\mathcal{T}_1^-$ должна быть собственной. В самом деле, в противном случае с вероятностью, равной дефекту $\mathcal{T}_1^-$, требование никогда не застает прибор свободным (исключая нулевое требование) и с вероятностью единица будет последнее удачливое требование, за которым будет следовать неоканчивающаяся очередь. Мы увидим, что $\mathcal{T}_1^-$ есть собственная случайная величина, если $\mathbf{E}(\mathscr{B}_k) < \mathbf{E}(\mathscr{A}_k)$.

Предположим теперь, что $(\nu-1)$-е требование поступает в момент τ. Ему приходится ждать время $\mathbf{W}_{\nu-1} = \mathbf{S}_{\nu-1}$, и обслуживание его завершается к моменту $\tau + \mathbf{W}_{\nu-1} + \mathscr{B}_{\nu-1}$. Первое удачливое требование (с номером ν) поступает в момент $\tau + \mathscr{A}_\nu$, и до этого момента прибор оставался свободным в течение времени

$$\mathscr{A}_\nu - \mathbf{W}_{\nu-1} - \mathscr{B}_{\nu-1} = -\mathbf{S}_{\nu-1} - \mathbf{X}_\nu = -\mathbf{S}_\nu.$$

Согласно определению ν, величина $\mathbf{S}_\nu$ есть первая нижняя лестничная высота $\mathscr{H}_1^-$. Поскольку с момента поступления ν-го требования процесс начинается сначала, то мы приходим ко второму выводу: *длительности свободных периодов являются независимыми случайными величинами, распределенными так же, как* $-\mathscr{H}_1^-$ *(т. е. как время возвращения для нижних лестничных высот)*. Таким образом, требование с номером $\mathcal{T}_1^- + \ldots + \mathcal{T}_r^-$ является r-м требованием, застающим прибор свободным. В момент его прибытия прибор был свободен в течение времени $-\mathscr{H}_r^-$.

Теперь должно быть ясно, что между последовательными лестничными моментами *сегменты графика траектории процесса массового обслуживания* $\{\mathbf{W}_n\}$ *конгруэнтны сегментам графика траектории случайного блуждания* $\{\mathbf{S}_n\}$, причем они сдвинуты вертикально так, чтобы начинаться с точки на временно́й оси (рис. 1). Для аналитического описания этого факта обозначим на время через $[n]$ *последний* нижний лестничный индекс $\leqslant n$; иными словами, $[n]$ есть (случайный) индекс, такой, что $[n] \leqslant n$ и

$$\mathbf{S}_{[n]} \leqslant \mathbf{S}_j, \quad j = 0, 1, \ldots, n. \tag{9.3}$$

Индекс $[n]$ определен однозначно с вероятностью 1 (напомним, что распределение величин $\mathbf{X}_j$ непрерывно). Имеем, очевидно,

$$\mathbf{W}_n = \mathbf{S}_n - \mathbf{S}_{[n]}. \tag{9.4}$$

Соотношение (9.4) приводит к очень важному выводу, если величины $\mathbf{X}_1, \ldots, \mathbf{X}_n$ рассмотреть в *обратном порядке*. Положим

для краткости $\mathbf{X}_1' = \mathbf{X}_n, \ldots, \mathbf{X}_n' = \mathbf{X}_1$. Частные суммы последовательности $\{\mathbf{X}_k'\}$ суть $\mathbf{S}_k' = \mathbf{X}_1' + \ldots + \mathbf{X}_k' = \mathbf{S}_n - \mathbf{S}_{n-k}$. Из (9.4) вытекает, что максимальный член последовательности $0, \mathbf{S}_1', \ldots, \mathbf{S}_n'$ имеет индекс $n - [n]$ и равен $\mathbf{W}_n$. Поскольку векторы $(\mathbf{X}_1', \ldots, \mathbf{X}_n')$ и $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ одинаково распределены, нами доказана важная

Теорема[1]). Случайные величины $\mathbf{W}_n$ *и*

$$\mathbf{M}_n = \max[0, \mathbf{S}_1, \ldots, \mathbf{S}_n] \tag{9.5}$$

распределены одинаково.

Следствия из этой теоремы будут обсуждаться в гл. XII. Здесь мы покажем, что она позволяет сводить некоторые задачи о разорении к задачам теории массового обслуживания (несмотря на внешнее несходство этих задач).

Примеры. г) *Задачи о разорении.* Пусть $\mathbf{X}(t)$ — обобщенный пуассоновский процесс, распределения приращений которого задаются формулой (4.2) с произвольным распределением F. В § 5 *разорение* было определено как осуществление при некотором t события $\{\mathbf{X}(t) > z + ct\}$, где z и c — фиксированные положительные числа. Обозначим $\tau_1, \tau_2, \ldots$ моменты последовательных скачков процесса $\mathbf{X}(t)$. Если разорение вообще происходит, то оно происходит также в некоторый момент τ_k, и поэтому достаточно рассматривать вероятность осуществления события $\mathbf{S}_n = \mathbf{X}(\tau_n) - c\tau_n > z$ при некотором n. Но по определению обобщенного пуассоновского процесса величина $\mathbf{X}(\tau_n)$ есть сумма n независимых случайных величин $\mathbf{Y}_k$ с одним и тем же распределением F, в то время как τ_n есть сумма n независимых показательно распределенных величин $\mathcal{A}_k$. Таким образом, по существу мы имеем дело со случайным блужданием, порожденным случайными величинами $\mathbf{X}_k = \mathbf{Y}_k - c\,\mathcal{A}_k$, *плотность* распределения которых дается формулой

$$\frac{\alpha}{c} \int_x^\infty e^{(\alpha/c)(x-y)} F\{dy\}. \tag{9.6}$$

Разорение происходит тогда и только тогда, когда в этом случайном блуждании при некотором n происходит событие $\{\mathbf{S}_n \geqslant z\}$. Нахождение вероятности разорения сводится, следовательно, к

[1]) Впервые этот результат был обнаружен, по-видимому, Ф. Полачеком в 1952 г. и использован (в ином контексте) Ф. Спицером, The Wiener — Hopf equation whose kernel is a probabillity density, *Duke Math. J.*, 24 (1947), 327—344. По поводу доказательства Спицера см. задачу 16.

отысканию распределений величин $\mathbf{W}_n$, образующих индуцированный величинами $\mathbf{X}_k$ процесс ожидания.

д) *Численная иллюстрация.* Наиболее важным процессом массового обслуживания является процесс, в котором *времена между поступлениями требований и времена обслуживания имеют показательные распределения* с математическими ожиданиями $1/a$ и $1/b$ соответственно, причем $a<b$. Характеристики этого процесса описаны в примере (8.б). *Время ожидания n-го требования имеет предельное распределение W с атомом веса $1-a/b$ в начале координат и плотностью* $\frac{b-a}{b}ae^{-(b-a)x}$ *при* $x>0$.

Математическое ожидание этого распределения равно $\frac{a}{b(b-a)}$.

Плотность распределения *свободных периодов* прибора совпадает с плотностью первой нижней лестничной высоты и равна ae^{-at}. В рассматриваемом случае свободные периоды и времена между поступлениями требований распределены одинаково (это, однако, не является общим фактом для процессов массового обслуживания).

Номер $\mathbf{N}$ *первого требования, застающего прибор свободным, имеет производящую функцию* $p(s)/a$, где $p(s)$ — функция, задаваемая формулой (8.4). Рассмотрим теперь *период занятости*, начинающийся в момент 0, т. е. время до первого момента, когда прибор оказывается свободным. Поскольку этот период начинается с поступления требования номер 0, то *случайная величина* $\mathbf{N}$ *равна числу требований, поступивших в течение первого периода занятости.* Простой подсчет показывает, что случайная величина $\mathbf{N}$ имеет математическое ожидание $b/(b-a)$ и дисперсию $ab(a+b)/(b-a)^3$.

Пусть, наконец, $\mathbf{T}$ есть длительность периода занятости. Точное выражение плотности $\mathbf{T}$ дается формулой гл. XIV, (6.15) с $cp=a$ и $cq=b$. Это выражение содержит бесселеву функцию и не поддается простым вычислениям. Тем не менее моменты величины $\mathbf{T}$ могут быть найдены с помощью ее преобразования Лапласа, полученного иными методами в примерах гл. XIV, (4.а) и XIV, (6.б). Они равны

$$\mathbf{E}(\mathbf{T})=\frac{1}{b-a} \quad \text{и} \quad \operatorname{Var}(\mathbf{T})=(a+b)\frac{1}{(b-a)^3}.$$

В рассматриваемом процессе ожидания периоды занятости чередуются со свободными периодами, и их математические ожидания равны соответственно $1/(b-a)$ и $1/a$. Следовательно, отношение $(b-a)/a$ есть мера *доли времени, в течение которого прибор остается свободным.*

Таблица

		$b=1$					
		$a=0{,}5$	$a=0{,}6$	$a=0{,}7$	$a=0{,}8$	$a=0{,}9$	$a=0{,}95$
Время ожидания (стационарный режим)	Математическое ожидание	1	1,5	2,3	4	9	19
	Дисперсия	3	5,3	10	24	99	399
Период занятости	Математическое ожидание	2	2,5	3,3	5	10	20
	Дисперсия	12	25	63	225	1900	15600
Число требований в течение периода занятости	Математическое ожидание	2	2,5	3,3	5	10	20
	Дисперсия	6	15	44	180	1710	14820

В таблице математическое ожидание времени обслуживания принято за единицу, так что a представляет собой *математическое ожидание числа требований, поступающих в течение времени обслуживания одного требования. Как видно из таблицы,* дисперсии периодов занятости очень велики. Следовательно, *нужно ожидать больших флуктуаций периодов занятости.* Таким образом, обычная доверчивость к математическим ожиданиям весьма опасна в практических приложениях. Для периода занятости с дисперсией 225 тот факт, что математическое ожидание его равно 5, мало о чем говорит.

Многомерный аналог рассмотренного нами процесса массового обслуживания аналитически гораздо более сложен. Основы его теории были заложены в фундаментальной работе Дж. Кифера и Дж. Вольфовица [On the theory of queues with many servers, *Trans. Amer. Math. Soc.*, 78 (1955), 1—18].

§ 10. Возвратные и невозвратные случайные блуждания

В настоящем параграфе дается классификация случайных блужданий. Эта классификация не зависит от материала § 8 и тесно связана с теорией восстановления, изложенной в параграфе 6. Каждой функции распределения F на прямой поставим формально в соответствие функцию интервалов

$$U\{I\} = \sum_{k=0}^{\infty} F^{k*}\{I\}. \tag{10.1}$$

Ряд в (10.1) тот же, что и в (6.2), но, если распределение F не сосредоточено на полупрямой, он может расходиться даже, ко-

гда I есть конечный интервал. Ниже будет показано, что сходимость или расходимость ряда (10.1) имеет глубокий смысл. Основные результаты настоящего параграфа просты, однако их формулировки, к сожалению, усложняются необходимостью отдельного рассмотрения арифметических распределений[1]).

Обозначим, для краткости, I_h и I_{h+t} интервалы $-h<x<h$ и $t-h<x<t+h$ соответственно.

Теорема 1. (i) *Если распределение F неарифметическое, то либо $U\{I\}<\infty$ для всех конечных интервалов, либо $U\{I\}=\infty$ для всех интервалов.*

(ii) *Если F — арифметическое распределение с шагом λ, то либо $U\{I\}<\infty$ для всех конечных интервалов, либо $U\{I\}=\infty$ для всякого интервала, содержащего точку вида $n\lambda$.*

(iii) *Если $U\{I\}<\infty$, то*

$$U\{I_h+t\}\leqslant U\{I_{2h}\} \tag{10.2}$$

при всех t и $h>0$.

Для облегчения ссылок на два возможных случая введем следующее определение (в этом определении F наделяется прилагательными, которые в действительности относятся к соответствующему случайному блужданию).

Определение. *Распределение F называется невозвратным, если $U\{I\}<\infty$ для всех конечных интервалов, и возвратным в противном случае.*

Имея чисто вероятностное значение, теорема 1 связана также с интегральным уравнением

$$Z=z+F*Z, \tag{10.3}$$

аналогичным уравнению восстановления (6.4). Мы используем интегральное уравнение (10.3) как исходную точку и докажем теорему 1 одновременно со следующим предложением.

Теорема 2. *Пусть z — непрерывная функция, такая, что $0\leqslant z(x)\leqslant\mu_0$ при $|x|<h$ и $z(x)=0$ вне интервала I_h.*

Если F — невозвратное распределение, то функция

$$Z(x)=\int_{-\infty}^{\infty} z(x-y)\,U\{dy\} \tag{10.4}$$

---

[1]) Распределение F арифметическое, если все его точки роста принадлежат множеству вида 0, $\pm\lambda$, $\pm 2\lambda$, ... Наибольшее число λ, обладающее этим свойством, называется *шагом* F (см. гл. V, 2).

является равномерно непрерывным решением уравнения (10.3), *удовлетворяющим условию*

$$0 \leqslant Z(x) \leqslant \mu_0 U\{I_{2h}\}. \tag{10.5}$$

Максимум функции Z *достигается в некоторой точке интервала* I_h.

Доказательство теорем 1 и 2. (i) Предположим, что $U\{I_\alpha\}<\infty$ при некотором $\alpha>0$. Возьмем $h<\frac{1}{2}\alpha$; и пусть функция z равна нулю вне I_h, но не обращается в нуль тождественно. Будем решать уравнение (10.3) методом последовательных приближений, положив $Z_0=z$ и

$$Z_n(x)=z(x)+\int_{-\infty}^{\infty} Z_{n-1}(x-y)\,F\{dy\}. \tag{10.6}$$

Обозначив

$$U_n\{I\}=F^{0*}\{I\}+\ldots+F^{n*}\{I\}, \tag{10.7}$$

имеем, очевидно,

$$Z_n(x)=\int_{-\infty}^{\infty} z(x-y)\,U_n\{dy\} \tag{10.8}$$

(интегрирование в действительности распространяется на интервал длины $\leqslant 2h$). Так определенная функция Z_n непрерывна. Покажем, пользуясь индукцией, что она достигает своего максимума μ_n в точке ξ_n, такой, что $z(\xi_n)>0$. Для функции $Z_0=z$ это верно тривиальным образом. Для перехода от $n-1$ к n достаточно заметить, что $z(x)=0$ влечет $Z_n(x)\leqslant\mu_{n-1}$, в то время как $\mu_n\geqslant Z_n(\xi_{n-1})\geqslant Z_{n-1}(\xi_{n-1})=\mu_{n-1}$.

Следовательно, интервал $I_h+\xi_n$ содержится в интервале I_{2h}, и поэтому в силу (10.8)

$$\mu_n \leqslant \mu_0 U\{I_{2h}\}. \tag{10.9}$$

Тем самым доказано, что функции Z_n равномерно ограничены. Поскольку $Z_0\leqslant Z_1\leqslant\ldots$, то отсюда вытекает, что $Z_n\to Z$, причем функция Z удовлетворяет неравенствам (10.5). Выбирая функцию z так, чтобы $z(x)=\mu_0$ при $|x|<\eta<h$, имеем

$$Z_n(x)\geqslant \mu_0 U_n\{I_h+x\}, \tag{10.10}$$

откуда следует, что $U\{I\}<\infty$ для каждого интервала длины $<h$. Разбивая произвольный конечный интервал I на малые интервалы, находим $U\{I\}<\infty$. Таким образом, распределение F невозвратно, и мы можем отбросить ограничение на h. Из (10.9)

и (10.10) вытекает (10.2). Наконец, беря разности значений Z в форме (10.4), получаем

$$|Z(x+\delta) - Z(x)| \leqslant U\{I_{2h}\} \cdot |z(x+\delta) - z(x)|.$$

Следовательно, функция Z равномерно непрерывна, и, таким образом, все утверждения, касающиеся невозвратного распределения F, доказаны.

(ii) Предположим теперь, что $U\{I_\alpha\} = \infty$ при *всех* $\alpha > 0$. Если $z \geqslant \mu_\theta$ в интервале I_η, то имеет место (10.10) и, следовательно, $z_n(x) \to \infty$ при всех x из некоторой окрестности нуля. Отсюда и из (10.6) вытекает, что если t — *точка* роста F, то $Z_n(x) \to \infty$ для всех x из некоторой окрестности t. Пользуясь индукцией, легко вывести, что то же самое верно для всех точек роста любой из функций распределения F^{2*}, F^{3*}, ... Предположим, что распределение F не арифметическое. Оно не может быть сосредоточенным на полупрямой, так как тогда мы имели бы $U\{I_\alpha\} < \infty$ (§ 6). В силу леммы 2 из гл. V, 4 точки роста функций F^{2*}, F^{3*}, ... образуют плотное множество на прямой и поэтому $Z_n(x) \to \infty$ при всех x. Отсюда вытекает, что $U_n\{I\} \to \infty$ для всех интервалов. С очевидными изменениями это же рассуждение применимо к арифметическим распределениям. Теоремы доказаны. ▶

Следствие. *Пусть* $Z \geqslant 0$ *— ограниченное решение уравнения восстановления* (10.3). *Если* $z \geqslant 0$ *и* $z \geqslant \mu > 0$ *в некотором интервале* I, *то распределение* F *невозвратно.*

Доказательство. В случае, когда I содержит начало координат, доказательство леммы содержится в доказательстве теоремы. В общем случае можно воспользоваться заменой x на $x - a$, которая не нарушает уравнения восстановления. ▶

Замечание *о единственности.* Разность ζ двух ограниченных решений уравнения восстановления (10.3) удовлетворяет уравнению $\zeta = F * \zeta$ В гл. XI, 2 будет показано, что если F — неарифметическое распределение, то $\zeta = \text{const}$, а если F — арифметическое распределение с шагом λ, то ζ — периодическая функция с периодом λ. За исключением этой неопределенности, решение уравнения (10.3) единственно. Частное решение, определяемое формулой (10.4), *минимально* и, следовательно, однозначно характеризуется тем свойством, что $\liminf_{|x| \to \infty} Z(x) = 0$, когда z обращается в нуль на бесконечности.

Мы еще вернемся к уравнению восстановления (10.3) в параграфах гл. XI.2 и гл. XI.9, а сейчас выведем из теоремы 1

ряд следствий для случайных блужданий. Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — независимые случайные величины с общим распределением F. Положим $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. Под «попаданием в I в момент n» мы имеем в виду событие $\mathbf{S}_n \in I$.

Теорема 3[1]). *Если распределение F невозвратно, то число попаданий в каждый конечный интервал конечно с вероятностью единица, а математическое ожидание числа таких попаданий равно $U\{I\}$.*

Если F — возвратное неарифметическое распределение, то число попаданий в каждый конечный интервал бесконечно с вероятностью единица. Если F — возвратное арифметическое распределение с шагом λ, то число попаданий в каждую точку вида $n\lambda$ бесконечно с вероятностью единица.

Доказательство. Пусть распределение F невозвратно. Вероятность попадания в I после момента n не превосходит n-го остаточного члена ряда в (10.1), и поэтому для любого $\varepsilon > 0$ при достаточно больших n вероятность более чем n попаданий не превосходит ε. Тем самым первое утверждение теоремы доказано.

Предположим теперь, что F — возвратное неарифметическое распределение. Обозначим через $\rho_h(t)$ вероятность хотя бы одного попадания в $I_h + t$. Достаточно показать, что $\rho_h(t) = 1$ при всех $h > 0$ и t, поскольку отсюда с очевидностью следует, что любое число попаданий в произвольный интервал имеет вероятность 1. Рассматривая возможные значения суммы $\mathbf{S}_1$, получаем

$$\rho_h(t) = F\{I_h + t\} + \int\limits_{|y-t| \geqslant h} \rho_h(t-y)\, F\{dy\}. \qquad (10.11)$$

Этому соотношению можно придать вид уравнения восстановления $\rho_h = z_h + F * \rho_h$, где

$$z_h(t) = \int\limits_{|y-t| < h} [1 - \rho_h(t-y)]\, F\{dy\}. \qquad (10.12)$$

[1]) Другое доказательство можно найти в работе Chung K. L., Fuchs W. H. J., On the distribution o values of sums of random variables, *Memoirs Amer. Math. Soc.*, **6** (1950), 1—12. Эта теорема непосредственно вытекает также из второго закона нуля и единицы (см. гл. IV, 6). Если $\varphi(I+t)$ есть вероятность бесконечного числа попаданий в интервал $I+t$, то при фиксированном I функция φ может принимать лишь два значения: 0 и 1. С другой стороны, рассматривая первый шаг случайного блуждания, видим, что $\varphi = F * \varphi$, откуда $\varphi = \text{const}$ (см. гл. XI, 2). Приводимое в тексте доказательство мы предпочли из-за его поучительности, а также с целью сделать содержание этого параграфа замкнутым.

Далее, при $|x|<\alpha$ интервал I_h+x содержится в интервале $I_{\alpha+h}$, и поэтому $\rho_h(x)\leqslant\rho_{h+\alpha}(0)$. Отсюда следует, что подинтегральное выражение в (10.12) не меньше $1-\rho_{3h}(0)$, и если бы было $1-\rho_{3h}(0)>0$, то функция $z_h(t)$ была бы равномерно больше нуля в некоторой окрестности каждой точки роста функции F. Так как это в силу следствия к теореме 2 невозможно, то $\rho_\alpha(0)=1$ при всех $\alpha>0$.

Предположим теперь, что $\rho_\alpha(-x)<1$ при некотором $x\neq0$. Если $|y+x|<\frac{1}{3}\alpha$ и $h<\frac{1}{3}\alpha$, то интервал I_h+y содержится в интервале $I_\alpha-x$, и поэтому $\rho_h(y)\leqslant\rho_\alpha(-x)$. Это означает, что при достаточно малых h функция $1-\rho_h$ от y строго положительна при y, принадлежащих некоторой окрестности точки $-x$. Рассмотрим теперь (10.11) при $t=0$. Если бы точка x была точкой роста F, то правая часть была бы меньше единицы, в то время как левая часть равна единице. Таким образом, если x есть точка роста F, то $\rho_\alpha(-x)=1$ при всех $\alpha>0$. Отсюда, опираясь на лемму 2 из гл. V, 4 и используя рассуждение из пункта (II) доказательства теоремы 2, выводим, что $\rho_\alpha(x)=1$ при всех x. Тем самым вторая часть теоремы доказана для неарифметических распределений F. С очевидными изменениями проведенное доказательство применимо и к арифметическим распределениям. ▶

При проверке, является ли заданное распределение F возвратным или невозвратным, иногда бывает трудно установить непосредственно, что $U\{I_n\}=\infty$. Задачу облегчает следующее свойство невозвратных распределений: *если распределение F невозвратно, то отношение $t^{-1}U\{I_t\}$ остается ограниченным при $t\to\infty$*. Свойство это немедленно вытекает из (10.2) (верхняя граница для $t^{-1}U\{I_t\}$ равна $U\{I_1\}$). Проиллюстрируем применение этого метода.

Теорема 4. *Распределение с математическим ожиданием μ является невозвратным, если $\mu\neq0$, и возвратным, если $\mu=0$.*

Доказательство[1]). Пусть $\mu>0$. Из усиленного закона больших чисел вытекает, что с вероятностью единица $n^{-1}\mathbf{S}_n>\frac{1}{2}\mu$ при всех достаточно больших n, и поэтому вероятность бесконечного числа попаданий в полупрямую $-\infty$, 0 равна нулю.

[1]) Это доказательство принадлежит К. Л. Чжуну и Д. Орнштейну. Используя вместо усиленного закона больших чисел центральную предельную теорему, тем же методом можно показать, что двумерное случайное блуждание с нулевыми математическими ожиданиями и конечными дисперсиями возвратно. Иное доказательство содержится в гл. XVIII, 7.

Пусть $\mu=0$. По закону больших чисел имеем $\mathbf{P}\{|\mathbf{S}_n|<\varepsilon n\}>\frac{1}{2}$ при $n>n_\varepsilon$. Отсюда следует, что $F^{k*}\{\overline{-a, a}\}>\frac{1}{2}$ при $n_\varepsilon<k<\varepsilon^{-1}a$, и поэтому $a^{-1}U\{\overline{-a, a}\}>\frac{1}{2}(\varepsilon^{-1}-n_\varepsilon a^{-1})$. При $a\to\infty$ правая часть последнего неравенства стремится к $\frac{1}{2}\varepsilon^{-1}$ и, поскольку ε произвольно, распределение F не может быть невозвратным. ▶

При возвратном случайном блуждании последовательность $\{\mathbf{S}_n\}$ меняет знак бесконечно часто, и поэтому соответствующие такому блужданию верхние и нижние лестничные процессы возвратны. Можно считать удивительным, что обратное утверждение неверно. *Даже при невозвратном случайном блуждании последовательность* $\{\mathbf{S}_n\}$ *может менять знак бесконечно часто.* Так как при невозвратном блуждании число попаданий в конечный интервал $\overline{-a, a}$ конечно, то это означает (очень грубо говоря), что перемены знаков происходят из-за отдельных скачков весьма большой величины: величина $|\mathbf{S}_n|$ перерастает все границы, но поразительное неравенство $\mathbf{X}_{n+1}<-\mathbf{S}_n-a$ осуществляется бесконечно часто, как бы велика ни была постоянная a.

Рис. 2 (см. стр. 259) иллюстрирует, как происходят большие скачки, однако он не вполне отражает рассматриваемое явление, поскольку распределение пришлось урезать, чтобы получить конечный график.

Пример. *Симметричные устойчивые распределения.* Симметричное устойчивое распределение F удовлетворяет условию $F^{n*}(n^{1/\alpha}x)=F(x)$. Рассмотрим случай $\alpha<1$ и будем считать известным, что F имеет плотность f, для которой $f(0)>0$ (см. гл. XVIII, 6). Имеем, очевидно,

$$F^{n*}\{\overline{-a, a}\}\approx 2af(0)n^{-1/\alpha},$$

и поэтому $U\{\overline{-a, a}\}<\infty$. Таким образом, F — *невозвратное* распределение. Нетрудно видеть, что верхний и нижний лестничные процессы не могут оба быть обрывающимися, и в рассматриваемом случае из соображений симметрии ясно, что оба эти процесса *возвратные.*

§ 11. Общие марковские цепи

Дискретные марковские цепи (1, гл. XV) довольно очевидным образом обобщаются на случай декартовых (и более общих) пространств состояний. В случае дискретных цепей вероятности перехода задавались стохастической матрицей с элемен-

тами p_{ij}, строки которой образовывали вероятностные распределения. Теперь нам придется рассматривать переходы из точки x в некоторый интервал или множество Γ в $\mathscr{R}^n$. Вероятность

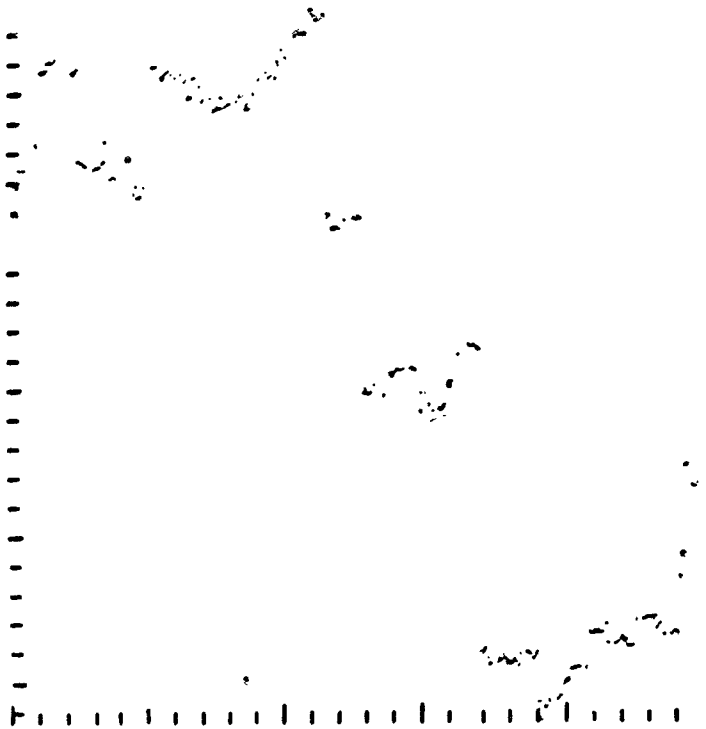


Рис. 2. *Случайное блуждание, порожденное распределением Коши.*

(Распределение было урезано с тем, чтобы устранить скачки, превышающие по величине размер графика.)

такого перехода мы обозначим $K(x, \Gamma)$. Чтобы обосновать существование потребующихся нам интегралов, мы должны наложить на K некоторые неограничительные условия регулярности. Требования непрерывности хватило бы для большинства практических целей, однако, накладывая это требование, мы ничего не выигрываем по сравнению с более общим случаем.

Определение 1. *Стохастическим ядром K называется функция от двух переменных, точки и множества, такая, что* (i) *при любом фиксированном x $K(x, \Gamma)$ является вероятностным*

распределением по Γ, *и* (ii) *для любого фиксированного интервала* Γ $K(x, \Gamma)$ *есть бэровская функция по* x.

Это определение относится к любому числу измерений (и даже к более общим пространствам).

Мы *не* требуем, чтобы ядро K было определено на всем пространстве. Если x и Γ являются подмножествами множества Ω, то мы говорим, что K *сосредоточено* на Ω. Иногда необходимо допустить *несобственные* (*дефектные*) распределения, и в этом случае ядро K называют *субстохастическим*. Нередко K можно представить в виде

$$K(x; \Gamma) = \int_{\Gamma} k(x, y)\, dy. \tag{11.1}$$

Функция k в (11.1) называется *плотностью стохастического ядра*. В соответствии с соглашением гл. V, (3.3) соотношение (11.1) кратко записывается в виде

$$K(x, dy) = k(x, y)\, dy.$$

[Строго говоря, k есть плотность по отношению к мере Лебега или длине; в случае произвольной меры m мы бы использовали запись $K(x, dy) = k(x, y)\, m\{dy\}$.]

Прежде чем давать формальное определение марковской цепи, подготовим соответствующий аналитический аппарат по аналогии со случаем дискретных цепей. Вероятность перехода из x в Γ за два шага положим по определению равной

$$K^{(2)}(x, \Gamma) = \int_{\Omega} K(x, dy) K(y, \Gamma). \tag{11.2}$$

Множество Ω в (11.2) либо есть все пространство, либо совпадает с тем множеством, на котором сосредоточено ядро K. Соотношение (11.2) выражает, конечно, тот факт, что первый шаг ведет из точки x в некоторую точку y, а второй шаг — из точки y в множество Γ. Основное допущение состоит в том, что при заданной промежуточной точке y прошлая история не влияет на дальнейшие переходы. Аналогичное можно сказать о $K^{(n)}$ — вероятности перехода за n шагов. Если положить $K^{(1)} = K$, то для произвольных целых положительных m и n должно быть

$$K^{(m+n)}(x, \Gamma) = \int K^{(m)}(x, dy) K^{(n)}(y, \Gamma). \tag{11.3}$$

Когда $m = n = 1$, эта формула сводится к (11.2). При $m, n = 1, 2, \ldots$ формула (11.3) *последовательно определяет вероятности перехода* $K^{(n)}$. Обозначим для согласованности символом $K^{(0)}$ вероятностное распределение, сосредоточенное в точке x (*кронекеров-*

ское ядро). Тогда формула (11.3) будет справедлива при всех $m \geqslant 0$, $n \geqslant 0$. Операция над ядрами (11.3), нередко встречающаяся и вне теории вероятностей, называется сверткой ядер. Она во всех отношениях схожа с перемножением матриц.

Вряд ли стоит специально отмечать, что ядра $K^{(n)}$ являются стохастическими. Если K имеет плотность, то все ядра $K^{(n)}$ имеют плотности, причем эти плотности связаны *формулами свертки*

$$k^{(m+n)}(x,\ z) = \int_{\Omega} k^{(m)}(x,\ y)\, k^{(n)}(y,\ z)\, dy. \tag{11.4}$$

Примеры. а) *Свертки.* Если $k(x, y) = f(x - y)$, где f—вероятностная плотность, то свертки (11.4) сводятся к обычным сверткам. Аналогичное справедливо и в общем случае, если ядро K однородно в том смысле, что

$$K(x,\ \Gamma) = K(x+s,\ \Gamma+s),$$

где $\Gamma+s$ обозначает сдвиг множества Γ на s. По поводу сверток на окружности см. теорему 3 в гл. VIII, 7.

б) *Потери энергии при столкновениях.* В физике последовательные столкновения частицы обычно рассматриваются как случайный процесс. Считают, что если энергия (масса) частицы до столкновения равна $x>0$, то после столкновения она представляется случайной величиной $\mathbf{Y}$, для которой $\mathbf{P}\{\mathbf{Y} \in \Gamma\} = K(x, \Gamma)$, где K — стохастическое ядро. Обычно предполагается, что возможны лишь потери и что отношение $\mathbf{Y}/x$ имеет функцию распределения, не зависящую от x. При этом $\mathbf{P}\{\mathbf{Y} \leqslant y\} = G(y/x)$ и функция $G(y/x)$ определяет стохастическое ядро.

В аналогичной задаче о звездной радиации [пример гл. X, (2.в)] Амбарцумяном был рассмотрен частный случай, когда $G(y) = y^{\lambda}$ при $0 \leqslant y \leqslant 1$, где λ — положительная постоянная. Плотность соответствующего ядра сосредоточена на множестве $0<y<x$ и равна $\lambda y^{\lambda-1} x^{-\lambda}$. Нетрудно проверить, что плотность вероятности перехода за n шагов выражается в рассматриваемом случае формулой

$$k^{(n)}(x,\ y) = \frac{\lambda^n}{(n-1)!}\, \frac{y^{\lambda-1}}{x^{\lambda}} \left(\log \frac{x}{y}\right)^{n-1}, \qquad 0 < y < x. \tag{11.5}$$

Значение $\lambda = 1$ соответствует равномерному распределению (*утерянная доля* «распределена случайно»), и (11.5) сводится при этом к гл. I, (8.2) [продолжение в примере гл. X, (1.а)].

в) *Случайные цепи.* Рассмотрим цепь (или ломаную линию) в $\mathscr{R}^3$ со звеньями длины единица и со случайными углами между соседними звеньями. Много конструкций подобного рода (часто

довольно сложных) встречается в химии полимеров. Мы ограничимся рассмотрением случая, когда последовательные углы являются независимыми случайными величинами.

Под *длиной* цепи с концевыми точками A и B мы понимаем расстояние между A и B. В результате добавления одного звена к цепи длиной x получается цепь длиной $\sqrt{x^2+1-2x\cos\theta}$, где θ — случайный угол между новым звеном и прямой, проходящей через точки A и B. Рассмотрим два распределения θ, особенно интересных в химии.

(i) Пусть θ принимает значения 60° и 120° с вероятностью $\frac{1}{2}$ каждое. Тогда $\cos\theta = \pm\frac{1}{2}$ и длина продолженной цепи определяется стохастическим ядром $K(x, \Gamma)$, наделяющим каждую из двух точек $\sqrt{x^2 \pm x + 1}$ вероятностью $\frac{1}{2}$. При фиксированном x распределение $K^{(n)}$ сосредоточено на 2^n точках.

(ii) Пусть направление нового звена выбирается «вполне случайно», т. е. предположим, что величина $\cos\theta$ распределена равномерно на интервале $\overline{-1, 1}$ [см. гл. I, 10]. Длина продолженной цепи не превосходит y тогда и только тогда, когда $\cos\theta \geqslant [x^2+1-y^2]/2x$. Отсюда следует, поскольку величина $\cos\theta$ распределена равномерно, что длина продолженной цепи определяется стохастическим ядром с плотностью

$$k(x, y) = y/2x, \quad |x-1| < y < x+1.$$

Длина L_{n+1} цепи с $n+1$ звеньями имеет плотность $k^{(n)}(1, y)$ (см. задачу 18).

г) *Дискретные цепи Маркова.* Стохастическую матрицу (p_{ij}) можно рассматривать как плотность $k(i, j) = p_{ij}$ определенного на множестве Ω целых положительных чисел стохастического ядра по отношению к мере m, наделяющей каждую точку множества Ω единичной массой. Эту плотность можно продолжить на всю прямую, положив $K(i, dy) = m\{dy\}$, где мера m равна нулю на дополнении Ω'. При $x \notin \Omega$ значения $K(x, \Gamma)$ можно определить произвольно, например с помощью линейной интерполяции. ▶

Абсолютные и стационарные вероятности

Последовательность случайных величин $\mathbf{X}_0, \mathbf{X}_1, \ldots$ подчинена вероятностям перехода $K^{(n)}$, если $K^{(n)}(x, \Gamma)$ является условной вероятностью события $\{X_{m+n} \in \Gamma\}$ при условии $\mathbf{X}_m = x$. *Вероятно-*

стное распределение величины $\mathbf{X}_n$ при заданном распределении γ_0 величины $\mathbf{X}_0$ *выражается формулой*

$$\gamma_n\{\Gamma\} = \int_\Omega \gamma_0\{dx\} K^{(n)}(x, \Gamma). \tag{11.6}$$

Определение 2. *Распределение* γ_0 *называется стационарным распределением ядра* K, *если* $\gamma_n = \gamma_0$ *при всех* n, *т. е. если*

$$\gamma_0\{\Gamma\} = \int_\Omega \gamma_0\{dx\} K(x, \Gamma). \tag{11.7}$$

Основные свойства стационарных распределений те же, что и в случае дискретных марковских цепей. При выполнении нежестких требований регулярности на K *существует единственное стационарное распределение; оно является асимптотическим распределением величин* $\mathbf{X}_n$ *при любом исходном распределении.* Иными словами, влияние начального состояния исчезает и система стремится к *устойчивому состоянию*, определяемому стационарным решением. Этот результат является одной из форм эргодической теоремы (см. гл. VIII, 7).

Примеры. д) *Процесс ожидания* $\{\mathbf{W}_n\}$, определенный в (9.1), является марковским процессом, сосредоточенным на замкнутой полупрямой $\overline{0, \infty}$. Вероятности перехода определены здесь лишь для $x, y \geqslant 0$ и при $\Gamma = \overline{0, y}$ имеем $K(x, \Gamma) = F(y - x)$. Существование стационарной меры будет доказано в гл. VIII, 7.

е) Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — взаимно независимые положительные случайные величины с непрерывной плотностью f, сосредоточенной на $\overline{0, \infty}$. Определим случайные величины $\mathbf{Y}_k$ последовательно посредством формул

$$\mathbf{Y}_1 = \mathbf{X}_1, \quad \mathbf{Y}_{n+1} = |\mathbf{Y}_n - \mathbf{X}_{n+1}|. \tag{11.8}$$

Последовательность $\{\mathbf{Y}_n\}$ образует сосредоточенный на $\overline{0, \infty}$ марковский процесс с плотностями вероятностей перехода

$$k(x, y) = \begin{cases} f(x-y) + f(x+y) & \text{при } 0 < y < x, \\ f(x+y) & \text{при } y > x > 0. \end{cases} \tag{11.9}$$

Стационарная плотность g определяется уравнением

$$g(y) = \int_0^\infty g(x+y) f(x)\, dx + \int_0^\infty g(x) f(x+y)\, dx. \tag{11.10}$$

Пусть F — отвечающая f функция распределения и предположим, что $\mathbf{X}_1$ имеет математическое ожидание μ. Тогда функция

$$g(y) = \mu^{-1}[1 - F(y)] \tag{11.11}$$

является *стационарной плотностью вероятности*. В самом деле, с помощью простого интегрирования по частям можно убедиться, что g удовлетворяет уравнению (10.10) [1]). С другой стороны, из гл. V, (6.3) следует, что g есть плотность вероятности (см. задачу 17).

ж) *Одно техническое приложение* [2]). Длинная линия передачи состоит из отдельных кусков кабеля, характеристики которых подвержены статистическим флуктуациям. Будем считать отклонения от идеального значения независимыми случайными величинами $\mathbf{Y}_1$, $\mathbf{Y}_2$, ... и предположим, что эффект их аддитивен. Перемена направления кабеля меняет знак его привзноса. Предположим, что отклонения $\mathbf{Y}_k$ симметричны, и обозначим $\mathbf{X}_k = |\mathbf{Y}_k|$. Эффективную конструкцию длинной линии передачи можно осуществить по следующему индуктивному правилу: $(n+1)$-й кусок кабеля подсоединяется так, что соответствующая ему ошибка имеет знак, противоположный знаку накопленной ошибки предыдущих n кусков. Накопленные ошибки подчиняются тогда правилу (11.8), и (11.11) является не только стационарной плотностью, но и предельным распределением: *распределение ошибки линии, состоящей из n кусков кабеля, близко (при больших n) к распределению с плотностью* (11.11). С другой стороны, если бы отдельные куски кабеля подсоединялись независимо, то была бы применима центральная предельная теорема и дисперсия ошибки росла бы линейно по n, т. е. линейно относительно длины кабеля. Таким образом, простой контроль знака ошибки не дает ей разрастаться.

В приведенных примерах марковская последовательность $\mathbf{X}_0, \mathbf{X}_1, \ldots$ определялась в терминах начального распределения γ_0 и вероятностей перехода K. Совместное распределение величин $(\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_n)$ имеет вид

$$\gamma_0\{dx_0\}K(x_0, dx_1) \ldots K(x_{n-1}, dx_n).$$

Распределения такого вида уже обсуждались в гл. III, 8 и **1**, гл. XV, 1. Мы имеем здесь типичный пример, показывающий преимущество определения абсолютных вероятностей в терминах условных вероятностей. Более систематичным был бы дру-

---

[1]) Как обнаружить такого рода результат? Предположив, что функции f и g имеют производные, можно формально продифференцировать (10.10). Интегрируя далее по частям, приходим к уравнению $g'(y) = -g(0)f(y)$, из которого вытекает, что функция g должна иметь вид (11.11). Непосредственная проверка показывает теперь, что (11.11) является решением (11.10) без предположений о дифференцируемости f и g.

[2]) Обобщение дискретной модели, использовавшейся фон Шеллингом, *Elektrische Nachr.-Technik*, **20** (1943), pp. 251—259.

гой подход, а именно — взять в качестве определения соотношение

$$\mathbf{P}\{\mathbf{X}_{n+1}\in\Gamma\,|\,\mathbf{X}_0=x_0,\ \ldots,\ X_n=x_n\}=K(x_n,\Gamma). \qquad (11.12)$$

Марковское свойство выражено здесь тем обстоятельством, что правая часть не зависит от $x_0, x_1, \ldots, x_{n-1}$, и поэтому «предыстория» не влияет на будущее. Недостаток этого определения состоит в том, что оно вынудило бы нас заняться бесполезным рассмотрением вопросов существования условных вероятностей, их единственности и т. п.

(Марковские процессы с непрерывным параметром изучаются в гл. X.)

§ 12*. Мартингалы

Для первой ориентировки можно рассмотреть случайный процесс $\{\mathbf{X}_n\}$, обладающий тем свойством, что совместное распределение величин $(\mathbf{X}_0,\ldots,\mathbf{X}_n)$ имеет строго положительную непрерывную плотность p_n $(n=0,\ 1,\ \ldots)$. При этом элементарный метод гл. III, 2 позволяет всюду определить условные плотности и условные математические ожидания. Предположим также, что рассматриваемые величины $\mathbf{X}_n$ и $\mathbf{Y}_n$ имеют математические ожидания.

Последовательность $\{\mathbf{X}_n\}$ назовем *абсолютно беспристрастной* (*безобидной*), если при $n=1,2,\ldots$

$$\mathbf{E}(\mathbf{X}_1)=0,\ \mathbf{E}(\mathbf{X}_{n+1}\,|\,\mathbf{X}_1,\ \ldots,\ \mathbf{X}_n)=0. \qquad (12.1)$$

Последовательность $\{\mathbf{Y}_n\}$ является *мартингалом*, если

$$\mathbf{E}(\mathbf{Y}_{n+1}\,|\,\mathbf{Y}_0,\ \ldots,\ \mathbf{Y}_n)=\mathbf{Y}_n. \qquad (12.2)$$

(Ниже будет дано более гибкое определение.)

Между этими двумя типами последовательностей имеется простая связь. Пусть $\{\mathbf{X}_n\}$ — абсолютно беспристрастная последовательность. Положим

$$\mathbf{Y}_n=\mathbf{X}_1+\ \ldots\ +\mathbf{X}_n+c, \qquad (12.3)$$

где c — постоянная и $n=1,2,\ldots$. Тогда

$$\mathbf{E}(\mathbf{Y}_{n+1}\,|\,\mathbf{X}_1,\ \ldots,\ \mathbf{X}_n)=\mathbf{Y}_n. \qquad (12.4)$$

Задающие условие величины $\mathbf{X}$ в (12.4) могут быть заменены на $\mathbf{Y}_k$, и поэтому (12.4) эквивалентно (12.2). С другой стороны, если последовательность $\{\mathbf{Y}_n\}$ — мартингал, то положим $\mathbf{X}_1=\mathbf{Y}_1-\mathbf{E}(\mathbf{Y}_1)$ и $\mathbf{X}_{n+1}=\mathbf{Y}_{n+1}-\mathbf{Y}_n$. Так построенная последова-

*) Мартингалы очень важны, и поэтому им посвящен отдельный параграф. Однако нигде в этой книге мартингалы не используются в качестве аппарата.

тельность $\{\mathbf{X}_n\}$ абсолютно беспристрастна и, кроме того, для рассматриваемых величин $\mathbf{X}_j$, $\mathbf{Y}_k$ выполняется (12.3) с $c=\mathbf{E}(\mathbf{Y}_1)$. Таким образом, *последовательность* $\{\mathbf{Y}_n\}$ *образует мартингал тогда и только тогда, когда она имеет вид* (12.3), *где* $\{\mathbf{X}_n\}$ *— абсолютно беспристрастная последовательность.*

Понятие мартингала было введено П. Леви. Затем Дж. Л. Дуб обнаружил неожиданные возможности аппарата мартингалов и развил их теорию [1]). В гл. VII, 8 будет показано, что при выполнении нежестких условий ограниченности образующие мартингал величины $\mathbf{Y}_k$ сходятся к пределу. Этот факт имеет важное значение в современной теории вероятностных процессов.

Примеры. а) В классических играх случайные величины $\mathbf{X}_k$ независимы и имеют нулевое математическое ожидание: $\mathbf{E}(\mathbf{X}_k)=0$. Такая игра абсолютно безобидна [2]), и частные суммы $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ образуют мартингал. Рассмотрим теперь обычную игру в орлянку, в которой игрок выбирает ставку по некоторому правилу, учитывающему его успехи в предыдущих турах. Последовательные выигрыши не являются уже независимыми, но игра по-прежнему абсолютно безобидна.

Смысл безобидной игры состоит в том, что знание прошлой истории игры не должно давать игроку возможности увеличить его капитал. Наглядно это означает, что абсолютно безобидная игра должна оставаться таковой при любой системе игры, т. е. при любых правилах пропуска отдельных туров. Мы увидим, что это именно так и есть.

б) *Урновая схема Пойа* [1, гл. V (2, в)]. Из урны, содержащей b черных и r красных шаров, выбирается наудачу один шар. Затем этот шар возвращается и, кроме того, в урну добавляется c шаров того же цвета, что и выбранный шар. Пусть $\mathbf{Y}_0=\frac{b}{b+r}$ и пусть $\mathbf{Y}_n$ есть доля черных шаров, оказавшихся в урне в результате n-го извлечения. Последовательность $\{\mathbf{Y}_n\}$ является мартингалом. Теорема о сходимости обеспечивает здесь существование предельного распределения [см. примеры гл. VII, (4.а) и гл. VII (8.а)].

в) *Конкордантные* [3]) *функции.* Пусть $\{\mathbf{X}_n\}$ — марковский процесс с вероятностями перехода, задаваемыми стохастическим

[1]) Более полно теория мартингалов излагается в книгах Дуба и Лоэва.

[2]) О практической ограниченности этого понятия говорилось в 1, гл. X, 3. Напомним, что существуют «безобидные» игры, в которых с вероятностью, большей $1-\varepsilon$, выигрыш игрока в n-й игре превосходит, скажем, $n/\log n$.

[3]) Этот термин был введен Дж. Г. Хантом.

ядром K. Существование математических ожиданий величин $\mathbf{X}_n$ не предполагается. Функция u называется конкордантной по отношению к K, если

$$u(x) = \int K(x, dy)\, u(y). \tag{12.5}$$

Рассмотрим случайные величины $\mathbf{Y}_k = u(\mathbf{X}_k)$ и предположим, что все моменты $\mathbf{Y}_n$ существуют (так будет, например, если функция u ограничена). Соотношение (12.5) означает, что $\mathbf{E}(\mathbf{Y}_{k+1} | \mathbf{X}_k = x) = u(x)$, откуда $\mathbf{E}(\mathbf{Y}_{k+1} | \mathbf{X}_k) = \mathbf{Y}_k$. Следовательно, поскольку последовательность $\{\mathbf{X}_k\}$ марковская, имеет место (12.4), а так как величины $\mathbf{Y}_k$ являются функциями от $\mathbf{X}_k$, то (12.4) влечет (12.2) [см. гл. V, 10.а]. Таким образом, последовательность $\{\mathbf{Y}_n\}$ является *мартингалом*. Этот результат имеет большое значение в граничной теории марковских процессов, поскольку существование предела последовательности $\{\mathbf{Y}_n\}$ обычно влечет существование предела исходной последовательности $\{\mathbf{X}_n\}$ [см. примеры (е) и гл. VII, (8.в)].

г) *Отношения правдоподобия.* Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — последовательность случайных величин и пусть известно, что совместное распределение величин $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ имеет либо плотность p_n, либо плотность q_n, однако неизвестно, какую именно. Чтобы прийти к какому-то решению, в статистике вводятся новые случайные величины

$$\mathbf{Y}_n = \frac{q_n(\mathbf{X}_1, \ldots, \mathbf{X}_n)}{p_n(\mathbf{X}_1, \ldots, \mathbf{X}_n)}. \tag{12.6}$$

Ясно, что если выполнены достаточно сильные условия регулярности и если истинными плотностями являются p_n, то наблюденные значения величин $\mathbf{X}_1, \ldots, \mathbf{X}_n$ в среднем будут группироваться вокруг тех точек, где p_n относительно велико. Когда дело обстоит именно так, величина $\mathbf{Y}_n$, вероятно, будет малой или большой смотря по тому, является ли истинной плотностью p_n или q_n. Таким образом, асимптотическое поведение величин $\{\mathbf{Y}_n\}$ представляет интерес в теории статистических решений.

Предположим для простоты, что плотности p_n строго положительны и непрерывны. Если истинными плотностями являются p_n, то условная плотность распределения величины $\mathbf{X}_{n+1}$ при заданных $\mathbf{X}_1, \ldots, \mathbf{X}_n$ равна отношению p_{n+1}/p_n и, следовательно,

$$\mathbf{E}(\mathbf{Y}_{n+1} | \mathbf{X}_1 = x_1, \ldots, \mathbf{X}_n = x_n) =$$

$$= \int_{-\infty}^{\infty} \frac{q_{n+1}(x_1, \ldots, x_n, y)}{p_{n+1}(x_1, \ldots, x_n, y)} \cdot \frac{p_{n+1}(x_1, \ldots, x_n, y)}{p_n(x_1, \ldots, x_n)}\, dy. \tag{12.7}$$

В подинтегральном выражении функции p_{n+1} сокращаются, а знаменатель второй дроби не зависит от y. Интегрирование q_{n+1} дает в результате частную плотность q_n. Таким образом, интеграл в (12.7) равен q_n/p_n и поэтому выполняется (12.4). Следовательно, при сделанных предположениях отношения правдоподобия $\mathbf{Y}_n$ образуют мартингал. ▶

Условия, относительно которых берутся условные математические ожидания в (12.2), не являются вполне удачными, поскольку нередко приходится заменять задающие условия величины $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ некоторыми функциями от них [примером этого является (12.4)]. Еще больший дефект такого задания условий обнаруживается в примере (а). Рассматриваемый процесс (скажем, игра в орлянку или рулетку) описывается последовательностью случайных величин $\mathbf{Z}_n$, и выигрыш игрока в $(n+1)$-м туре есть функция от $\mathbf{Z}_1, \ldots, \mathbf{Z}_{n+1}$, а также, возможно, от некоторых других величин. Наблюдаемое прошлое представляется величинами $(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$, которые могут содержать в себе больше информации, чем предыдущие выигрыши. Например, если игрок не участвует в турах с номерами 1, 3, 5, ..., то знание его выигрышей до момента $2n$ в лучшем случае эквивалентно знанию величин $\mathbf{Z}_2, \mathbf{Z}_4, \ldots, \mathbf{Z}_{2n}$. При этом дополнительное знание величин $\mathbf{Z}_1, \mathbf{Z}_3, \ldots$ в принципе может давать некоторое преимущество, и абсолютная безобидность в таком случае должна иметь в качестве условий величины $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. Таким образом, может оказаться необходимым рассматривать условные ожидания по отношению к различным множествам случайных величин, и, чтобы охватить все возможные ситуации, лучше всего использовать условные ожидания по отношению к произвольным σ-алгебрам событий.

Рассмотрим последовательность $\{\mathbf{Y}_n\}$ случайных величин на произвольном вероятностном пространстве и обозначим $\mathfrak{A}_n$ σ-алгебру событий, порожденную величинами $(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ [см. гл. V, 10.а]. Определяющее мартингалы соотношение (12.2) можно записать в виде $\mathbf{E}(\mathbf{Y}_{n+1} | \mathfrak{A}_n) = \mathbf{Y}_n$. Мы хотим дать более общее определение мартингалов в форме последнего соотношения с той разницей, что σ-алгебра $\mathfrak{A}_n$ заменяется более богатой событиями σ-алгеброй $\mathfrak{B}_n$. В большинстве случаев σ-алгебра $\mathfrak{B}_n$ будет порождаться величинами $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ и некоторыми другими величинами, зависящими от прошлого. Идея состоит в том, чтобы любая зависящая от прошлого случайная величина была измерима относительно $\mathfrak{B}_n$, и в этом смысле σ-алгебра $\mathfrak{B}_n$ представляет ту информацию, которая содержится в прошлой истории процесса. Поскольку информация о прошлом растет со временем, то мы предположим, что σ-алгебры

$\mathfrak{B}_n$ возрастают

$$\mathfrak{B}_1 \subset \mathfrak{B}_2 \subset \ldots . \tag{12.8}$$

Определение. *Пусть* $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ — *случайные величины, математические ожидания которых существуют, и пусть* $\mathfrak{B}_1, \mathfrak{B}_2, \ldots$ — *суть* σ-*алгебры событий, удовлетворяющие условию* (12.8).

Последовательность $\{\mathbf{Y}_n\}$ *образует мартингал по отношению к* $\{\mathfrak{B}_n\}$, *если*

$$\mathbf{E}(\mathbf{Y}_{n+1} \mid \mathfrak{B}_n) = \mathbf{Y}_n. \tag{12.9}$$

[В силу неединственности условных математических ожиданий равенство (12.9) следует понимать в следующем смысле: «существует вариант условного математического ожидания, для которого выполняется (12.9)». Это замечание относится и к дальнейшему]. Отметим, что из (12.9) вытекает $\mathfrak{B}_n$-измеримость величины $\mathbf{Y}_n$, и поэтому σ-алгебра $\mathfrak{B}_n$ содержит порожденную величинами $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ σ-алгебру $\mathfrak{A}_n$. Таким образом, *последовательность* $\{\mathbf{Y}_n\}$ *является также мартингалом по отношению к последовательности* σ-*алгебр* $\{\mathfrak{A}_n\}$, т. е. имеет место (12.2).

Пример. д) Пусть σ-алгебры $\mathfrak{B}_n$ удовлетворяют условию (12.8) и пусть $\mathbf{Y}$ — случайная величина, математическое ожидание которой существует. Положим $\mathbf{Y}_n = \mathbf{E}(\mathbf{Y} \mid \mathfrak{B}_n)$. Величины $\mathbf{Y}_n$ $\mathfrak{B}_n$-измеримы и поэтому удовлетворяют соотношению (12.9). Следовательно, последовательность $\{\mathbf{Y}_n\}$ есть *мартингал* относительно $\{\mathfrak{B}_n\}$. ▶

Возвратимся теперь к примеру (а) и покажем невозможность систем игры довольно общего типа. Пусть $\{\mathbf{Y}_n\}$ — мартингал по отношению к $\{\mathfrak{B}_n\}$. Для описания свободы игрока пропускать n-й тур введем *решающую функцию* $\boldsymbol{\varepsilon}_n$: $\boldsymbol{\varepsilon}_n$ есть $\mathfrak{B}_{n-1}$-измеримая[1]) случайная величина, принимающая лишь два значения 0 и 1. Если $\boldsymbol{\varepsilon}_n = 0$, то игрок пропускает n-й тур; при $\boldsymbol{\varepsilon}_n = 1$ он делает ставку, и в этом случае его выигрыш в n-м туре равен $\mathbf{Y}_n - \mathbf{Y}_{n-1}$. Обозначив накопленный выигрыш вплоть до n-го тура включительно через $\mathbf{Z}_n$, имеем

$$\mathbf{Z}_n = \mathbf{Z}_{n-1} + \boldsymbol{\varepsilon}_n [\mathbf{Y}_n - \mathbf{Y}_{n-1}]. \tag{12.10}$$

Из индуктивных соображений ясно, что величины $\mathbf{Z}_n$ имеют математические ожидания. Кроме того, все величины $\mathbf{Z}_{n-1}$, $\boldsymbol{\varepsilon}_n$ и $\mathbf{Y}_{n-1}$ $\mathfrak{B}_{n-1}$-измеримы, и поэтому [см. гл. V, (10.8)]

[1]) Это условие обеспечивает то обстоятельство, что решение принимается на основе прошлых наблюдений.

$$\mathbf{E}(\mathbf{Z}_n \mid \mathfrak{B}_{n-1}) = \mathbf{Z}_{n-1} + \varepsilon_n [\mathbf{E}(\mathbf{Y}_n \mid \mathfrak{B}_{n-1}) - \mathbf{Y}_{n-1}]. \quad (12.11)$$

Так как последовательность $\{\mathbf{Y}_n\}$ — мартингал, то выражение в квадратных скобках обращается в нуль, и, следовательно, последовательность $\{\mathbf{Z}_n\}$ тоже является мартингалом. Тем самым мы доказали теорему Халмоша, утверждающую

Невозможность систем игры. *Мартингал* $\{\mathbf{Y}_n\}$ *переводится всякой последовательностью решающих функций* $\varepsilon_1, \varepsilon_2, \ldots$ *снова в мартингал* $\{\mathbf{Z}_n\}$.

Наиболее важным частным случаем этого результата является случай так называемой *произвольной остановки.* Система с произвольной остановкой — это такая система, когда в первых $\mathbf{N}$ турах игрок принимает участие, а от всех последующих уклоняется; $\mathbf{N}$-й тур является последним. Случайная величина $\mathbf{N}$ (*момент остановки*) удовлетворяет условию $\{\mathbf{N}>k\} \in \mathfrak{B}_k$ (в обозначениях теоремы $\varepsilon_k = 1$ при $\mathbf{N} > k-1$ и $\varepsilon_k = 0$ при $\mathbf{N} \leqslant k-1$). Мы имеем стало быть такое

Следствие. *Произвольная остановка мартингал переводит в мартингал.*

Примеры. е) Простое случайное блуждание на прямой начинается в начале координат. Частица переходит вправо на один шаг с вероятностью p и влево на один шаг с вероятностью $q = 1-p$. Если $\mathbf{S}_n$ — положение частицы в момент n, то, как легко видеть, величины $\mathbf{Y}_n = (q/p)^{\mathbf{S}_n}$ образуют мартингал, причем $\mathbf{E}(\mathbf{Y}_n) = 1$ и $\mathbf{E}(\mathbf{Y}_0) = 1$ [частный случай примера (в)].

В задаче о разорении случайное блуждание останавливается в тот момент, когда оно впервые приводит в одну из точек $-a$, b (a, b — целые положительные числа). В таком модифицированном процессе $-a \leqslant \mathbf{S}_n \leqslant b$ и с вероятностью единица $\mathbf{S}_n$ становится равным и продолжает оставаться либо b, либо $-a$. Обозначим соответствующие вероятности x и $1-x$. Поскольку величины $\mathbf{S}_n$ ограничены,

$$\mathbf{E}(\mathbf{S}_n) = x\left(\frac{q}{p}\right)^b + (1-x)\left(\frac{q}{p}\right)^{-a}.$$

Но $\mathbf{E}(\mathbf{Y}_n) = 1$, так как математические ожидания образующих мартингал величин совпадают. Следовательно, $x(q/p)^b + (1-x)(q/p)^{-a} = 1$ и тем самым мы получили линейное уравнение для определения вероятности x остановки процесса в точке b, которая иными методами была вычислена в **1**, гл. XIV, 2. Использованный метод не работает при $p = q = 1/2$, однако в этом случае последовательность $\mathbf{S}_n$ образует мартингал и аналогич-

ное рассуждение показывает, что $x=a/(a+b)$. Хотя результат, к которому мы пришли, известен и весьма элементарен, использованный метод иллюстрирует возможные применения теории мартингалов.

ж) *О системах игры.* Пусть $\{\mathbf{X}_n\}$ — последовательность независимых случайных величин, такая, что $\mathbf{X}_n$ принимает значения $\pm 2^n$ с вероятностью $\frac{1}{2}$ каждое. Чтобы решить, участвовать ли в n-м туре, игрок бросает монету. Вероятность того, что впервые он примет участие в игре в n-м туре, равна 2^{-n}, и в этом случае его выигрыш равен $\pm 2^n$. При этом *выигрыш игрока при его первом вступлении в игру является случайной величиной без математического ожидания.* Теорема о системах игры связана, следовательно, с тем обстоятельством, что мы не меняем временно́го параметра. ▶

Нередко приходится иметь дело с абсолютными величинами и неравенствами, и в этих случаях важную роль играет понятие субмартингала. Последовательность $\{\mathbf{Y}_n\}$ называется *субмартингалом*[1]), если она удовлетворяет условиям определения мартингала с заменой знака равенства в (12.9) на $\geqslant$. На языке теории игр переход от мартингалов к субмартингалам есть переход от абсолютно безобидных игр к абсолютно благоприятным играм. Если $\{\mathbf{Y}_n\}$ — мартингал, то $\{|\mathbf{Y}_n|\}$ — субмартингал. Это вытекает из следующей более общей леммы.

Лемма. Пусть *u — выпуклая функция и последовательность $\{\mathbf{Y}_n\}$ является мартингалом. Тогда если математические ожидания величин $u(\mathbf{Y}_n)$ существуют, то последовательность $\{u(\mathbf{Y}_n)\}$ является субмартингалом.*

Утверждение леммы непосредственно вытекает из неравенства Иенсена [гл. V, (8.6)], которое применимо к условным математическим ожиданиям, так же как и к обычным. Согласно этому неравенству, имеем

$$\mathbf{E}(u(\mathbf{Y}_{n+1})\,|\,\mathfrak{B}_n) \geqslant u(\mathbf{E}(\mathbf{Y}_{n+1}\,|\,\mathfrak{B}_n)). \tag{12.12}$$

Остается заметить, что правая часть в (12.12) равна $u(\mathbf{Y}_n)$.

Аналогично устанавливается, что, если $\{\mathbf{Y}_n\}$ — субмартингал и u — выпуклая *неубывающая* функция, то последовательность $\{u(\mathbf{Y}_n)\}$ тоже является субмартингалом, если только математические ожидания величин $u(\mathbf{Y}_n)$ существуют.

---

[1]) Теперь обычно предпочитают этот термин более старому «нижний полумартингал».

§ 13. Задачи

1. Для устойчивости распределения F достаточно выполнения (1.2) при $n=2$ и 3 (П. Леви).

Указание Произведения вида $c_1^j c_2^k$, где $j, k=0, \pm 1, \pm 2, \ldots$, либо плотны в интервале $\overline{0,\infty}$, либо представляют собой степени некоторого фиксированного числа c. Нужно показать, что вторая возможность в рассматриваемом случае исключена.

Замечание. Весьма любопытно, что выполнение (1.2) лишь при $n=2$ еще *недостаточно* для устойчивости F; см. пример гл. XVII, (3.e) и задачу 9 из гл. IX, 10.

2. Если F и G — устойчивые распределения с одним и тем же характеристическим показателем α, то $F \ast G$ — тоже устойчивое распределение с характеристическим показателем α. Найти соответствующие нормирующие постоянные γ_n.

3. Из неравенства симметризации гл. V, (5.9) вытекает, что выражение $n[1-R(c_n x)]$, где R — симметричное устойчивое распределение, ограничено. Вывести отсюда, что R *имеет все моменты порядков меньше* α [использовать гл. V, (6.3)]. С помощью симметризации последнее утверждение переносится на несимметричные R.

4. *Иной вывод распределения Хольцмарка.* Рассмотрим шар радиуса r с центром в начале координат и n звезд (точек), расположенных в нем наудачу и независимо друг от друга. Пусть каждая звезда имеет единичную массу. Обозначим через $\mathbf{X}_1, \ldots, \mathbf{X}_n$ x-компоненты гравитационных сил, соответствующих отдельным звездам, и положим $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. Устремим r и n к бесконечности так, чтобы $\frac{4}{3}\pi r^3 n^{-1} \to \lambda$. Показать, что при этом распределение величины $\mathbf{S}_n$ стремится к симметричному устойчивому распределению с характеристическим показателем $\frac{3}{2}$.

5. Показать, что предыдущая задача по существу не изменится, если массу каждой звезды считать случайной величиной с единичным математическим ожиданием и массы различных звезд предполагать взаимно независимыми и не зависящими также от их расположения

6. *Распределение Хольцмарка в случае четырех измерений.* Четырехмерным аналогом распределения Хольцмарка является симметричное устойчивое распределение с характеристическим показателем $\frac{4}{3}$ (в четырехмерном пространстве гравитационная сила меняется обратно пропорционально кубу расстояния).

7. Пусть $\{\mathbf{X}_{k,n}\}$ — нулевая схема серий, причем величины $\mathbf{X}_{1,n}, \ldots, \mathbf{X}_{n,n}$ имеют одинаковое распределение F_n. Стремится ли к нулю вероятность $\mathbf{P}\{\max(|\mathbf{X}_{1,n}|, \ldots, |\mathbf{X}_{n,n}|) > \varepsilon\}$?

8. Найти плотность определяемой формулой (6.3) функции восстановления U, когда F имеет плотность а) $f(x)=e^{-x}$ и б) $f(x)=xe^{-x}$.

9. Рассматривается обрывающийся процесс восстановления с функцией распределения F, имеющей плотность pce^{-ct}. Найти распределения длительности жизни процесса и числа моментов восстановления.

10. *Обобщенный обрывающийся процесс восстановления.* Вместо того чтобы обрывать процесс мгновенно с вероятностью q, позволим ему (тоже с

вероятностью q) продолжаться в течение случайного времени с собственным распределением F_0 и *затем* остановим. Иными словами, моменты восстановления имеют вид $\mathbf{T}_1+\ldots+\mathbf{T}_n+\mathbf{Y}$, где величина $\mathbf{Y}$ имеет отличное от величин $\mathbf{T}$ распределение. Показать, что распределение V *длительности* процесса удовлетворяет уравнению восстановления

$$V=qF_0+F*V \quad (F(\infty)=1-q). \tag{*}$$

11. Показать, что процесс в задаче о времени ожидания до больших пропусков сводится к частному случаю процесса, рассмотренного в предыдущей задаче. Привести (7.1) к виду (*).

12. *Процесс Пуассона и теоремы о покрытии.* Напомним, что, согласно примеру гл. III, (3 г), в случае пуассоновского процесса условное распределение моментов восстановления на интервале $\overline{0,t}$ при условии, что их произошло ровно n, *равномерно*. Отсюда в силу теоремы о покрытиях (теорема 3 из гл. I, 9) следует, что вероятность $1-V(t)$ того, что все пропуски по длине меньше ξ, представима в виде

$$1-V(t)=e^{-ct}\sum_{n=1}^{\infty}\frac{(ct)^{n-1}}{(n-1)!}\varphi_n(t). \tag{**}$$

а) Проверить, что формула (**) действительно определяет решение уравнения (7.1) при $F(x)=1-e^{-cx}$.

б) Вывести теорему о покрытиях, считая известным, что определяемое формулой (**) решение уравнения (9.1) единственно. [Такое доказательство является примером доказательства с помощью рандомизации, см. **1**, гл. XII, задача 5]

13. Пусть V — распределение времени ожидания из примера (7. ж) («прибывший последним обслуживается первым»). Показать, что V удовлетворяет уравнению восстановления $V(t)=A(t)+B*V(t)$, где A и B — несобственные распределения, определяемые формулами $A\{dx\}=e^{-cx}G\{dx\}$ и $B\{dx\}=[1-G(x)]ce^{cx}dx$ (показать, что здесь мы имеем дело с обобщенным обрывающимся процессом в смысле задачи 10: процесс восстановления, порожденный распределением B, сопровождается ненарушаемым мертвым периодом, распределение которого пропорционально распределению A).

14. Малые пропуски в пуассоновском процессе [1]). Скажем, что происходит «совпадение», если два момента восстановления $\mathbf{S}_n$ и $\mathbf{S}_{n+1}$ с $n>0$ оказываются на расстоянии $\leqslant\xi$. Выразить распределение времени ожидания до первого совпадения в терминах распределения V из предыдущего примера.

15. *Обобщение предыдущей задачи о малых пропусках* [2]). Рассматриваете стандартный процесс восстановления с временами ожидания $\mathbf{T}_j$. Написать уравнение восстановления для распределения времени ожидания до первого осуществления события $\{\mathbf{T}_n\leqslant\mathbf{Y}_n\}$, где величины $\mathbf{Y}_n$ не зависят от процесса и друг от друга и имеют общее распределение G.

16. Пусть $a_1, \ldots, a_n$ — произвольные числа и $m=\max[0, a_1, a_1+a_2, \ldots, a_1+\ldots+a_n]$. Определим последовательно $v_1=a_n\cup 0$, $v_2=(v_1+a_{n-1})\cup 0$, ...,

[1]) По поводу других вариантов этой задачи (исследуемых иным методом) см. Gilbert E. N., Pollak H. O., Coincidences in Poisson patterns, *Bell System Tech. J.*, **36** (1957), 1005—1033.

[2]) Аналогично можно обобщить задачу о «больших пропусках». Ответ получается тоже аналогичным.

$v_n=(v_{n-1}+a_1)\cup 0$. Пользуясь индукцией, доказать, что $v_n=m$. Вывести отсюда теорему § 9.

17. Положить в примере (11.e) $f=1$ при $0<x<1$ и доказать, что функция $g(y)=2(1-y)$ является стационарной плотностью и что $k^{(n)}(x, y)=g(y)$ при $n\geqslant 2$. Если $f(x)=\alpha e^{-\alpha x}$, то $g(x)=f(x)$.

18. Определим сосредоточенную на $\overline{0,1}$ плотность стохастического ядра, положив

$$k(x, y)=\frac{1}{2}(1-x)^{-1} \quad \text{при} \quad 0<x<y<1$$

и

$$k(x, y)=\frac{1}{2}x^{-1} \quad \text{при} \quad 0<y<x<1.$$

Найти стационарную плотность (она удовлетворяет простому дифференциальному уравнению). Дать вероятностную интерпретацию.

19. Марковский процесс сосредоточен на $\overline{0, 1}$ и при $\mathbf{X}_n=x$ величина $\mathbf{X}_{n+1}$ равномерно распределена на $\overline{1-x, 1}$. Показать, что функция $2x$ является стационарной плотностью (Т. Угахери).

ГЛАВА VII

ЗАКОНЫ БОЛЬШИХ ЧИСЕЛ. ПРИМЕНЕНИЯ В АНАЛИЗЕ

В первой части этой главы показано, что некоторые известные глубокие теоремы анализа можно удивительно легко получить путем вероятностных рассуждений. В § 7 обсуждаются варианты закона больших чисел, а § 8 содержит частный случай теоремы о сходимости мартингалов.

§ 1. Основная лемма. Обозначения

Рассмотрим сначала одномерное распределение G с математическим ожиданием θ и дисперсией σ^2. Если $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — независимые случайные величины с распределением G, то их среднее арифметическое $\mathbf{M}_n = (\mathbf{X}_1 + \ldots + \mathbf{X}_n)n^{-1}$ имеет математическое ожидание θ и дисперсию $\sigma^2 n^{-1}$. При больших n эта дисперсия мала и $\mathbf{M}_n$, как правило, близко к θ. Отсюда следует, что для любой непрерывной функции u значение $u(\mathbf{M}n)$ будет, как правило, близко к $u(\theta)$. Это замечание объясняет суть (слабого) закона больших чисел. Оно несколько обобщается в приводимой ниже лемме, которая, несмотря на свою простоту, станет источником ценной информации.

Рассмотрим при $n = 1, 2, \ldots$ семейство распределений $F_{n,\theta}$ с математическими ожиданиями, равными θ, и дисперсиями $\sigma_n^2(\theta)$. Здесь θ — параметр, изменяющийся в конечном или бесконечном интервале. Для математических ожиданий введем обозначение

$$\mathbf{E}_{n,\theta}(u) = \int_{-\infty}^{+\infty} u(x) F_{n,\theta}\{dx\}. \tag{1.1}$$

Лемма 1. *Если* $\sigma_n^2(\theta) \to 0$, *то*

$$\mathbf{E}_{n,\theta}(u) \to u(\theta) \tag{1.2}$$

для любой непрерывной и ограниченной функции u. *Эта сходимость равномерна в любом подинтервале, где* $\sigma_n^2(\theta) \to 0$ *равномерно и функция* u *равномерно непрерывна.*

Доказательство. Очевидно, что

$$|\mathbf{E}_{n,\theta}(u) - u(\theta)| \leqslant \int_{-\infty}^{+\infty} |u(x) - u(\theta)| F_{n\theta}\{dx\}. \tag{1.3}$$

Существует окрестность $|x-\theta|<\delta$ точки θ, в которой подинтегральное выражение меньше ε. Вне этой окрестности подинтегральное выражение меньше некоторой константы M и по неравенству Чебышева [гл. V, неравенство (7.2)] вероятность попадания в область $|x-\theta|>\delta$ меньше, чем $\sigma_n^2(\theta)\delta^{-2}$. Таким образом, правая часть не превосходит $\varepsilon + M\sigma_n^2(\theta)\delta^{-2}$, что при достаточно большом n меньше 2ε. ▸

Ниже указаны три полезных и важных примера. В каждом из них $F_{n,\theta}$ является распределением среднего арифметического $(\mathbf{X}_1 + \ldots + \mathbf{X}_n)n^{-1}$. В примере (а) величины $\mathbf{X}_j$ принимают только значения 0 и 1, в примере (б) они имеют распределение Пуассона, а в примере (в) — гамма-распределение.

Примеры. а) Если $F_{n,\theta}$ — биномиальное распределение, то $\sigma_n^2(\theta) = \theta(1-\theta)n^{-1} \to 0$ и

$$\sum_{k=0}^{n} u\left(\frac{k}{n}\right)\binom{n}{k}\theta^k(1-\theta)^{n-k} \to u(\theta) \tag{1.4}$$

равномерно при $0 \leqslant \theta \leqslant 1$. Следствия этого соотношения обсуждаются в § 2.

б) Если $F_{n,\theta}$ приписывает вероятность $e^{-n\theta}(n\theta)^k/k!$ точке k/n, то $\sigma_n^2(\theta) = \theta/n$ и

$$e^{-n\theta}\sum_{k=0}^{\infty} u\left(\frac{k}{n}\right)\frac{(n\theta)^k}{k!} \to u(\theta) \tag{1.5}$$

равномерно в каждом конечном интервале значений θ. Эта формула верна и для n, отличных от целых чисел (продолжение см. в § 5 и 6).

в) Взяв в качестве $F_{n,\theta}$ гамма-распределение, получаем, что

$$\frac{1}{(n-1)!}\int_0^{\infty} u(x)\left(\frac{nx}{\theta}\right)^{n-1} e^{-nx/\theta}\frac{n\,dx}{\theta} \to u(\theta) \tag{1.6}$$

равномерно в каждом конечном интервале. Эта формула [с заменой $(n-1)!$ на $\Gamma(n)$] также верна для нецелых значений n.

В § 6 будет показано, что формула (1.6) представляет собой формулу обращения для преобразований Лапласа.

г) Статистики часто сталкиваются с положением, описанным в начале этого параграфа, но рассматривают математическое ожидание θ как неизвестный параметр, который нужно оценить по результатам наблюдений. На языке статистики (1.2) означает, что $u(\mathbf{M}n)$ является *асимптотически несмещенной* оценкой для неизвестного параметра $u(\theta)$ [оценка была бы несмещенной, если бы правая и левая части в (1.2) совпадали].

Обозначения для разностей

В последующих применениях мы будем использовать основные понятия и соотношения исчисления конечных разностей. Пусть дана последовательность $c_0, c_1 \ldots$. *Разностный оператор* Δ определяется равенством $\Delta c_n = c_{n+1} - c_n$. Применяя оператор Δ к последовательности $\{\Delta c_n\}$, мы получаем новую последовательность $\{\Delta^2 c_n\}$ и т. д. Таким образом, разности более высокого порядка определяются последовательно соотношением $\Delta^r = \Delta(\Delta^{r-1})$, где $\Delta^1 = \Delta$. Для удобства полагаем $\Delta^0 c_n = c_n$ для всех n. Простая индукция показывает, что

$$\Delta^r c_\nu = \sum_{k=0}^{r} \binom{r}{k} (-1)^{r+k} c_{\nu+k}. \tag{1.7}$$

Эта формула допускает обращение: можно выразить c_0 через разности, стоящие слева в (1.7). В самом деле, $c_\nu = -\Delta c_\nu + c_{\nu+1}$. Применяя это соотношение к Δc_ν и $c_{\nu+1}$, получаем $c_\nu = \Delta^2 c_\nu - 2\Delta c_{\nu+1} + c_{\nu+2}$. Продолжая таким же образом, приходим к тождеству

$$c_\nu = \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} \Delta^{r-k} c_{\nu+k}, \qquad r = 1, 2, \ldots \tag{1.8}$$

Мы будем применять операторы Δ^r к последовательностям вида $c_n = f(x+nh)$, получаемым с помощью какой-либо функции f при фиксированных x и шаге $h>0$. В этих случаях более удобно заменить Δ на *разностное отношение* $\underset{h}{\Delta} = h^{-1}\Delta$. Таким образом, при фиксированном $h>0$

$$\underset{h}{\Delta} f(x) = \frac{f(x+h) - f(x)}{h}. \tag{1.9}$$

Разностные отношения более высокого порядка снова определяются последовательно формулой $\underset{h}{\Delta}^{r+1} = \underset{h}{\Delta}\,\underset{h}{\Delta}^r$, где $\underset{h}{\Delta}^1 = \underset{h}{\Delta}$ и

$\Delta_h^0 f(x) = f(x)$. Формула (1.7) эквивалентна соотношению

$$\Delta_h^r f(x) = h^{-r} \sum_{k=0}^{r} \binom{r}{k} (-1)^{r+k} f(x + kh). \tag{1.10}$$

§ 2. Полиномы Бернштейна. Абсолютно монотонные функции

Каждая линейная комбинация полиномов вида $\binom{n}{k}\theta^k(1-\theta)^{n-k}$ называется *полиномом Бернштейна.* Для функции u, заданной на $0 \leqslant \theta \leqslant 1$, мы определим полином Бернштейна степени n формулой

$$B_{n,u}(\theta) = \sum_{k=0}^{n} u(kh) \binom{n}{k} \theta^k (1-\theta)^{n-k}, \qquad h = \frac{1}{n}. \tag{2.1}$$

Следующая теорема является лишь переформулировкой утверждения (1.4).

Теорема 1. *Если функция u непрерывна в замкнутом интервале $\overline{0, 1}$, то полиномы Бернштейна $B_{n,u}(\theta)$ равномерно сходятся к $u(\theta)$.*

Иными словами, для любого заданного $\varepsilon > 0$ и всех достаточно больших n

$$|B_{n,u}(\theta) - u(\theta)| < \varepsilon, \qquad 0 \leqslant \theta \leqslant 1. \tag{2.2}$$

Знаменитая *теорема Вейерштрасса об аппроксимации* утверждает возможность равномерной аппроксимации некоторыми полиномами. Приведенная выше теорема сильнее, так как в ней явно указаны аппроксимирующие полиномы. Доказательство принадлежит С. Н. Бернштейну.

Чтобы придать полиному (2.1) обычную форму записи по степеням θ, разложим биномы $(1-\theta)^{n-k}$ по степеням θ и заметим, что для фиксированного j коэффициент при θ^j в (2.1) равен

$$\sum_k u(kh) \binom{n}{k} \binom{n-k}{j-k} (-1)^{j-k} = \binom{n}{j} h^j \Delta_h^j u(0) \tag{2.3}$$

[равенство вытекает из (1.10)]. Таким образом, верна

Теорема 2. *Полиномы Бернштейна* (2.1) *допускают представление*

$$B_{n,u}(\theta) = \sum_{j=0}^{n} \binom{n}{j} (h\theta)^j \Delta_h^j u(0), \qquad h = \frac{1}{n}. \tag{2.4}$$

Мы используем это представление для того, чтобы просто обосновать один критерий для производящих функций. Функция u, определенная для $0 \leqslant x \leqslant 1$, называется *вероятностной производящей функцией, если*

$$u(x) = u_0 + u_1 x + u_2 x^2 + \ldots, \tag{2.5}$$

где $u_j \geqslant 0$ *и* $\sum u_j = 1$. Такая функция обладает неотрицательными производными всех порядков. Обратное утверждение вовсе не очевидно, если его рассматривать вне настоящего контекста.

Лемма. *Три следующих утверждения эквивалентны*:

(i) *u является вероятностной производящей функцией*;

(ii) *при* $0 < x < 1$ *и всех* n *существуют производные* $u^{(n)}$ *и* $u^{(n)}(x) \geqslant 0$; *кроме того,* $u(x) \to 1$ *при* $x \to 1$.

(iii) *функция u непрерывна при* $0 < x \leqslant 1$ *и*

$$\Delta_h^r u(0) \geqslant 0, \quad h = \frac{1}{n} \tag{2.6}$$

при $r \leqslant n$ *и* $n = 1, 2, \ldots$; *кроме того,* $u(1) = 1$.

Доказательство. Очевидно, что (i) влечет (ii). Из монотонности производных $u^{(n)}$ вытекает, что $\Delta_h u$ имеет неотрицательные производные всех порядков. По индукции заключаем, что (2.6) верно при всех n, т. е. (ii) влечет (iii). Если выполняется (2.6), то коэффициенты многочленов $B_{n,u}$ в (2.4) неотрицательны и в сумме дают $B_{n,u}(1)$. Из (2.1) следует, что $B_{n,u}(1) = u(1) = 1$, так что $B_{n,u}$ — *вероятностная производящая функция*. По теореме непрерывности из **1**, гл. XI, 6 (или из гл. VIII, 7 ниже) предел $u = \lim B_{n,u}$ снова является вероятностной производящей функцией. Доказательство закончено. ▶

Отбрасывая искусственное ограничение $u(1) = 1$, мы приходим к следующему определению, широко используемому вне теории вероятностей.

Определение. *Функция u называется абсолютно монотонной на* $\overline{a, b}$, *если* $u^{(n)}(x) \geqslant 0$ *при* $a < x < b$.

Пример функции $u(x) = (1 - x)^{-1}$ показывает, что неограниченная функция может быть абсолютно монотонной на $\overline{0, 1}$. Положим $v(x) = u(\alpha x)/u(\alpha)$, где $0 < \alpha < 1$. Если u абсолютно монотонна на $\overline{0, 1}$, то v является вероятностной производящей функцией, так что u можно разложить в степенной ряд, сходящийся при $0 < x < \alpha$; при этом α может быть выбрано сколь угодно близким к 1. Таким образом, нами доказана

Теорема 3. *Функция u абсолютно монотонна на $\overline{0,1}$ тогда и только тогда, когда ее можно разложить в степенной ряд с неотрицательными коэффициентами, сходящийся при $0<x<1$.*

Эта теорема имеет важные применения в теории преобразований Лапласа. Она была доказана впервые (другим методом) С. Н. Бернштейном.

§ 3. Проблема моментов

Пусть F — распределение вероятностей, сосредоточенное на *замкнутом* интервале $\overline{0,1}$. Обозначим через $\mathbf{E}(u)$ интеграл от u по отношению к F. *Момент* порядка k распределения F определяется равенством

$$\mu_k = \mathbf{E}(\mathbf{X}^k) = \int_0^1 x^k F\{dx\}. \tag{3.1}$$

Здесь подразумевается, что интервал интегрирования *замкнут* и $\mathbf{X}$, как обычно, обозначает координатную случайную величину. Вычисляя разности, видим, что $-\Delta\mu_k = \mathbf{E}(\mathbf{X}^k(1-\mathbf{X}))$, а по индукции

$$(-1)^r \Delta^r \mu_k = \mathbf{E}(\mathbf{X}^k(1-\mathbf{X})^r). \tag{3.2}$$

Поэтому математическое ожидание для полинома Бернштейна (2.1) равно

$$\mathbf{E}(B_{n,u}) = \sum_{k=0}^{n} u(kh) \cdot \binom{n}{k} \cdot (-1)^{n-k} \Delta^{n-k} \mu_k. \tag{3.3}$$

Обозначим через $p_k^{(n)}$ коэффициенты в правой части равенства (3.3)

$$p_k^{(n)} = \binom{n}{k} (-1)^{n-k} \Delta^{n-k} \mu_k, \qquad k = 0, \ldots, n. \tag{3.4}$$

Из (3.2) вытекает, что $p_k^{(n)} \geqslant 0$. Для постоянной $u=1$ мы имеем $B_{n,u}(\theta) = 1$ при всех θ, и (3.3) показывает, что при фиксированном n сумма коэффициентов $p_k^{(n)}$ равна единице. Поэтому мы можем определить *атомическое распределение* F_n, *приписывающее вес* $p_k^{(n)}$ *атому* k/n (*где* $k=0,1,\ldots,n$). Обозначим математическое ожидание по отношению к F_n через $\mathbf{E}_n$. Тогда величина (3.3) равна $\mathbf{E}_n(u)$. Из аппроксимационной теоремы предыдущего параграфа следует, что *если и непрерывна в замкну-*

том интервале $\overline{0, 1}$, *то при* $n \to \infty$

$$\mathbf{E}_n(u) \to \mathbf{E}(u). \tag{3.5}$$

Это соотношение, естественно, наводит на мысль, что и само распределение F можно выразить в терминах распределений F_n. Следующая теорема показывает, что это в самом деле так.

Теорема 1. *Формула обращения. В каждой точке непрерывности распределения* F

$$F_n(t) = \sum_{k \leqslant nt} \binom{n}{k} (-1)^{n-k} \Delta^{n-k} \mu_k \to F(t). \tag{3.6}$$

Доказательство. Введем ступенчатую функцию u_t, определив ее для фиксированного t формулой

$$u_t(\theta) = \begin{cases} 1 & \text{при } \theta \leqslant t, \\ 0 & \text{при } \theta > t. \end{cases} \tag{3.7}$$

Правая и левая части в (3.6) равны $\mathbf{E}_n(u_t)$ и $\mathbf{E}(u_t)$ соответственно, так что формально (3.6) является частным случаем соотношения (3.5). Здесь необходима некоторая осторожность, так как (3.5) опирается на теорему об аппроксимации, которая была доказана только для непрерывных функций. Однако прежнее доказательство показывает, что

$$B_{n,\, u_t}(\theta) \to u_t(\theta)$$

в любом замкнутом интервале, не содержащем точки, в которой u_t разрывна. Переходя к математическим ожиданиям, видим, что все возможные предельные значения левой части в (3.6) лежат между $F(t-\varepsilon)$ и $F(t+\varepsilon)$, где $\varepsilon > 0$ может быть выбрано произвольно малым. Полагая $\varepsilon \to 0$, получаем, что (3.6) верно в каждой точке t, где F непрерывна.

По счастливому стечению обстоятельств можно видоизменить это рассуждение так, чтобы получить простое доказательство следующей известной и важной теоремы, принадлежащей Ф. Хаусдорфу.

Теорема 2. *Последовательность чисел* $\mu_0, \mu_1, \ldots$ *является последовательностью моментов* (3.1) *некоторого распределения вероятностей* F, *сосредоточенного на* $\overline{0, 1}$, *тогда и только тогда, когда*

$$(-1)^r \Delta^r \mu_k \geqslant 0, \qquad \mu_0 = 1. \tag{3.8}$$

Доказательство. Моменты любого распределения на $\overline{0, 1}$ удовлетворяют (3.2), и потому условие (3.8) необходимо.

Допустим, что (3.8) выполняется. Определим числа $p_k^{(n)}$ по формуле (3.4). Используя формулу обращения (1.8) для разностей с $r=n-\nu$, получаем тождество

$$\sum_{k=0}^{n}\binom{k}{\nu}p_k^{(n)}=\binom{n}{\nu}\mu_\nu. \tag{3.9}$$

При $\nu=0$ выводим, что сумма всех $p_k^{(n)}$ дает $\mu_0=1$. Поэтому мы можем снова определить атомическое распределение вероятностей F_n, приписывающее вес $p_k^{(n)}$ атому k/n. Левая часть равенства (3.9) является математическим ожиданием случайной величины $\binom{n\mathbf{X}}{\nu}$ по отношению к этому распределению F_n. Умножая (3.9) на $\nu!\,n^{-\nu}$, мы видим, что при $n\to\infty$ математические ожидания величин $\mathbf{X}^\nu$ сходятся к μ_ν. Члены последовательности $\{\mu_\nu\}$ оказываются, таким образом, пределами соответствующих членов последовательностей моментов $\{\mathbf{E}_n(\mathbf{X}^\nu)\}$ и потому сами являются моментами. Это вытекает из простой теоремы, доказательство которой мы отложим до примера 1, г гл. VIII с тем, чтобы избежать повторений. Этим заканчивается доказательство. В то же время мы видим, что *распределение F с моментами μ_ν задается формулой* (3.6) ▶

Последовательность $\{\mu_n\}$, для которой $(-1)^r\Delta^r\mu_k\geqslant 0$ для всех пар r и k, называется *вполне монотонной*. Это название прочно установилось, хотя оно и нелогично. Мы вернемся к вполне монотонным последовательностям и функциям в гл. XIII, 4 и 11.

Чтобы избежать неправильных представлений, следует подчеркнуть, что положение существенно меняется при переходе к распределениям, не сосредоточенным ни на каком конечном интервале. В общем случае *распределение не определяется однозначно своими моментами.*

Пример. Элементарное, но утомительное интегрирование по частям показывает, что при $n=0,1,\ldots$

$$\int_0^\infty x^n\cdot e^{-\sqrt[4]{x}}\sin\sqrt[4]{x}\,dx=0. \tag{3.10}$$

При $0<\alpha<1$ отсюда следует, что функция $\varphi(x)=(1/24)e^{-\sqrt[4]{x}}\left(1-\alpha\sin\sqrt[4]{x}\right)$ есть плотность сосредоточенного на $\overline{0,\infty}$ распределения. *Все эти распределения имеют одну и ту же последовательность моментов.* Полагая $\varphi(x)=(1/2)\varphi_\alpha(x)$ при $x>0$ и $\varphi(x)=(1/2)\varphi_\beta(-x)$ при $x<0$, мы получаем плотность, которая *не является четной, хотя все ее нечетные моменты равны нулю.* ▶

Этот отрицательный результат не должен порождать необоснованного пессимизма, так как источник беспокойства легко устранить соответствующими условиями регулярности. Наилучший результат содержится в теореме

Карлемана, утверждающей, что распределение F на $\overline{-\infty, \infty}$ *однозначно определяется своими моментами, если*

$$\sum \mu_{2n}^{-1/2n} = \infty, \tag{3.11}$$

т. е. если ряд слева расходится. В этой книге мы докажем более слабое утверждение, а именно, что F однозначно определяется своими моментами, если степенной ряд $\Sigma \mu_{2n} t^n/(2n)!$ сходится в некотором интервале (см. § 6 и XV.4). Оба критерия ограничивают скорость роста μ_{2n}. Знание конечного числа моментов $\mu_0, \mu_1, \ldots, \mu_n$ даже в самой общей ситуации приводит к полезным неравенствам для F, аналогичным выведенным в гл. V, 7 для $n=1$ [1]).

§ 4*. Применение к симметрично зависимым случайным величинам

Мы выведем здесь один красивый результат, принадлежащий Б. де Финетти. Он может служить типичным примером того, сколь легко теорема 3.2 приводит к удивительным результатам.

Определение. *Случайные величины* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *называются симметрично зависимыми, если все* $n!$ *перестановок* $(\mathbf{X}_k, \ldots, \mathbf{X}_{k_n})$ *имеют одно и то же* n*-мерное распределение. Случайные величины, образующие бесконечную последовательность* $\{\mathbf{X}_n\}$*, называются симметрично зависимыми, если* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *симметрично зависимы при любом* n.

Как мы увидим из примеров, конечные и бесконечные последовательности существенно отличаются друг от друга. Мы рассмотрим здесь специальный случай *бесконечной последовательности* $\{\mathbf{X}_n\}$ *симметрично зависимых величин, принимающих только значения* 0 *и* 1. В следующей ниже теореме утверждается, что *распределение такого процесса* $\{\mathbf{X}_n\}$ *получается рандомизацией биномиального распределения.* Как обычно, мы положим $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ и будем называть событие $\{\mathbf{X}_k = 1\}$ «успехом».

Теорема. *Каждой бесконечной последовательности симметрично зависимых случайных величин соответствует такое*

[1]) Первые сильные результаты были получены Марковым и Стильтьесом около 1884 г. Современная литература на эту тему неисчерпаема. См., например, Wald A., *Trans. Amer. Math. Soc.*, **46** (1939), 280—306; Royden H. L., *Ann. Math. Statist.*, **24** (1953), 361—376 [приведены границы для $F(x) - F(-x)$]. Общий обзор см в монографии Shohat J. A., Tamarkin J. D., The problem of moments, New York, 1943 (Math. Surveys № 1). См. также Karlin S., Studden W. (1966).

*) Материал не используется в дальнейшем.

сосредоточенное на $\overline{0,\ 1}$ *распределение* F, *что*

$$\mathbf{P}\{\mathbf{X}_1=1,\ \ldots,\ \mathbf{X}_k=1,\quad \mathbf{X}_{k+1}=0,\ \ldots,\ \mathbf{X}_n=0\}=$$

$$=\int_0^1 \theta^k(1-\theta)^{n-k}\,F\{d\theta\} \qquad (4.1)$$

и

$$\mathbf{P}\{\mathbf{S}_n=k\}=\binom{n}{k}\int_0^1 \theta^k(1-\theta)^{n-k}\,F\{d\theta\}. \qquad (4.2)$$

Доказательство. Положим $\mu_0=1$ и при $n=1,\ 2,\ \ldots$

$$\mathbf{P}\{\mathbf{X}_1=1,\ \ldots,\ \mathbf{X}_n=1\}=\mu_n. \qquad (4.3)$$

Тогда при $n\geqslant 0$

$$\mathbf{P}\{\mathbf{X}_1=1,\ \ldots,\ \mathbf{X}_{n-1}=1,\ \mathbf{X}_n=0\}=\mu_{n-1}-\mu_n=-\Delta\mu_{n-1}, \qquad (4.4)$$

откуда

$$\mathbf{P}\{\mathbf{X}_1=1,\ \ldots,\ \mathbf{X}_{n-1}=1,\ \mathbf{X}_n=0,\quad \mathbf{X}_{n+1}=0\}=$$

$$=-\Delta\mu_{n-1}+\Delta\mu_n=(-1)^2\Delta^2\mu_{n-1}. \qquad (4.5)$$

Продолжая этот процесс, находим, что левая часть в (4.1) равна $(-1)^{n-k}\Delta^{n-k}\mu_k$. Следовательно, эти величины неотрицательны. Теорема 3.2 утверждает, что μ_k будет k-м моментом некоторого распределения вероятностей F. Утверждение (4.1) вытекает теперь из (3.2), а (4.2) эквивалентно (4.1), так как имеется ровно $\binom{n}{k}$ способов появления k успехов в n испытаниях. ▶

Обобщения. Теорема очевидным образом распространяется на случайные величины, принимающие только *конечное* число значений. Например, если $\mathbf{X}_j$ принимает значения a_1, a_2, a_3, то мы можем определить новые случайные величины $\mathbf{Y}_j$, полагая $\mathbf{Y}_j=0$ при $\mathbf{X}_j=a_1$ и $\mathbf{Y}_j=1$ в других случаях. К величинам $\mathbf{Y}_j$ применима доказанная теорема. Ее повторное применение приводит к аналогу соотношения (4.1). Эти рассуждения позволяют предположить, что в наиболее общем случае последовательность симметрично зависимых случайных величин получается из последовательности независимых случайных величин рандомизацией по некоторому параметру. В специальных случаях трудностей не возникает, однако общая задача трудна тем, что «параметры» не определены отчетливо и могут выбираться не-

постижимым образом. Несмотря на это, были доказаны весьма общие варианты этой теоремы [1]).

Эта теорема позволяет получить условие применимости закона больших чисел и центральной предельной теоремы к симметрично зависимым случайным величинам (см. задачу 21 в гл. VIII, 10). Первый из приводимых ниже примеров показывает, что в отдельных случаях теорема может приводить к удивительным результатам. Примеры (б) и (в) показывают, что теорема неверна для конечных последовательностей.

Примеры. а) *В схеме Пойа* (см. **1**, гл. V, 2) урна содержит первоначально b черных и r красных шаров. После каждого извлечения выбранный шар возвращается и в урну добавляется c шаров того же цвета. Таким образом, вероятность получить черный шар в каждом из n первых испытаний равна

$$\mu_n = \frac{b\,(b+c)\ldots(b+(n-1)\,c)}{(b+r)\ldots(b+r+(n-1)\,c)} = \frac{\Gamma\left(\frac{b}{c}+n\right)\Gamma\left(\frac{b+r}{c}\right)}{\Gamma\left(\frac{b+r}{c}+n\right)\Gamma\left(\frac{b}{c}\right)}. \quad (4.6)$$

Положим $\mathbf{X}_n=1$ или 0 в зависимости от того, появился при n-м испытании черный или красный шар. Легкие вычисления (см. **1**, гл. V, 2) показывают, что эти случайные величины симметрично зависимы и потому значение μ_n равно n-му моменту некоторого распределения F. По виду формула (4.6) напоминает бета-интеграл (2.5) из гл. II. Анализ показывает, что F *есть бета-распределение* (4.2) из гл. II *с параметрами* $\mu=b/c$ и $\nu=r/c$. Можно установить, используя снова бета-интеграл, что (4.1) согласуется с (2.3) из **1**, гл. V,а (4.2) с (2.4) из той же главы первого тома.

б) Рассмотрим 6 размещений двух различимых шаров в трех клетках. Припишем каждому вероятность $1/6$. Пусть $\mathbf{X}_i$ равно 1 или 0 в зависимости от того, занята i-я клетка или пуста. Величины $\mathbf{X}_i$ симметрично зависимы, но теорема к ним неприменима. В самом деле, из (4.3) получаем $\mu_0=1$, $\mu_1=1/2$, $\mu_2=1/6$, $\mu_3=0$, и на этом последовательность обрывается. Если бы она была начальным отрезком вполне монотонной последовательности $\{\mu_n\}$, мы должны были бы иметь $\mu_4=\mu_5=\ldots=0$. Но тогда было бы $\Delta^4\mu_1=-1/6<0$, что невозможно.

---

[1]) Hewitt E., Savage L. J., Symmetric measures on Cartesian products, *Trans. Amer. Math Soc.*, **80** (1956), 470—501. Изучение с применением мартингалов см. Loève (1963). См. также Bühlmann H., Austauschbare stochastische Variabeln und ihre Grenzwertsätze, Univ. of California Publications in Statistics, **3**, № 1 (1960), 1—36.

в) Пусть случайные величины $\mathbf{X}_1, \ldots, \mathbf{X}_n$ независимы, имеют одно и то же распределение вероятностей и $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. Положим $\mathbf{Y}_k = \mathbf{X}_k - n^{-1}\mathbf{S}_n$ для $k = 1, \ldots, n-1$. Случайные величины $\mathbf{Y}_k$ симметрично зависимы, но их совместное распределение нельзя представить в форме, соответствующей теореме де Финетти. ▶

§ 5 *. Обобщенная формула Тейлора и полугруппы

Мы видели, что полиномы Бернштейна из примера (1, а) можно записать в форме (2.4), содержащей конечные разности. Производя аналогичное преобразование в примере (1, б), мы приходим к принадлежащему Е. Хилле привлекательному варианту формулы Тейлора. Он может служить иллюстрацией некоторых основных положений теории полугрупп. Мы отправляемся от формулы (1.5), устанавливающей, что для всякой ограниченной и непрерывной на $\overline{0, \infty}$ функции u при $\lambda \to \infty$

$$e^{-\lambda\theta} \sum_{k=0}^{\infty} u\left(\frac{k}{\lambda}\right) \frac{(\lambda\theta)^k}{k!} \to u(\theta) \tag{5.1}$$

равномерно в каждом конечном интервале. Разлагая $e^{-\lambda\theta}$ в степенной ряд, мы можем представить левую часть в (5.1) в виде двойного ряда

$$\sum\sum \frac{(\lambda\theta)^{j+k}}{(j+k)!} \binom{j+k}{k} (-1)^j u\left(\frac{k}{\lambda}\right). \tag{5.2}$$

Этот ряд сходится абсолютно, и потому мы вправе переставить его члены. Положим $r = j+k$ и $\lambda^{-1} = h$. Сумма членов с одним и тем же r равна $h^r \overset{h}{\Delta}{}^r u(0)$ [см. обозначение (1.10)]. Таким образом, (5.1) представляется в виде

$$\sum_{r=0}^{\infty} \frac{\theta^r}{r!} \overset{h}{\Delta}{}^r u(0) \to u(\theta), \tag{5.3}$$

где $h \to 0^+$. Заменяя $u(\theta)$ на $u(t+\theta)$, приходим к соотношению

$$\sum_{r=0}^{\infty} \frac{\theta^r}{r!} \overset{h}{\Delta}{}^r u(t) \to u(t+\theta), \quad h \to 0, \tag{5.4}$$

которое напоминает разложение в ряд Тейлора. Единственное отличие состоит в том, что производные заменены разностями.

*) Этот параграф при первом чтении можно пропустить.

Мы рассматривали только случай $h>0$, $\theta>0$, но рассуждения сохраняются и при отрицательных h и θ. Поскольку область определения любой непрерывной функции u можно расширить до всей прямой, мы получаем следующую теорему.

Теорема. *Пусть функция u ограничена и непрерывна. Тогда, если h и θ имеют одинаковые знаки, то выполняется* (5.4). *При фиксированном t сходимость равномерна в каждом конечном интервале значений θ.*

Для аналитических функций левая часть приближается к ряду Тейлора, но теорема применима к недифференцируемым функциям. В этом смысле (5.4) представляет собой обобщение разложения в ряд Тейлора и открывает нам его новые черты.

Существует другое истолкование формулы (5.4), приводящее к так называемой экспоненциальной формуле теории полугрупп[1]) (см. теорему 2 в гл. X, 9). Левая часть в (5.4) содержит формальный экспоненциальный ряд, который естественно использовать для того, чтобы определить оператор $\exp \theta \underset{h}{\Delta}$. Соотношение (5.4) принимает при этом сокращенную форму

$$\exp \theta \underset{h}{\Delta} u(t) \to u(t+\theta). \tag{5.5}$$

Чтобы записать эту формулу в более удобном виде, введем *оператор сдвига*[2]) $T(\theta)$, переводящий функцию u в функцию u_θ, определяемую равенством $u_\theta(t)=u(t+\theta)$. Тогда $T(0)=1$, т. е. это единичный оператор и

$$\underset{h}{\Delta} = h^{-1}[T(h)-1]. \tag{5.6}$$

На операторном языке формула (5.4) [или (5.5)] принимает вид

$$e^{\theta h^{-1}[T(h)-1]} \to T(\theta). \tag{5.7}$$

Основная информация, доставляемая этой формулой, состоит в том, что все семейство операторов $T(\theta)$ определяется поведением $T(h)$ при малых h.

Оглядываясь назад, мы видим, что наш вывод формулы (5.7) применим к значительно более широкому классу операторов. Левая часть в (5.1) представляет собой просто линейную комбинацию значений u и может быть интерпретирована как интерполяционная формула для u. Аналогичное интерполяционное

[1]) Оставшуюся часть этого параграфа при первом чтении можно пропустить.

[2]) В стандартных обозначениях исчисления конечных разностей $T(\theta)=E^\theta$, но мы присоединяемся к обозначению, употребляемому в гл. X, 8 для общих полугрупп.

выражение имеет смысл для *любого* семейства операторов $\{T(\theta)\}$, определенных при $\theta>0$. В самом деле, при фиксированных θ и $h>0$ оператор

$$A_h(\theta)=e^{-\theta h^{-1}}\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{\theta}{h}\right)^k T(kh) \tag{5.8}$$

является линейной комбинацией операторов $T(kh)$. Коэффициенты (веса́) определяются в соответствии с распределением Пуассона и обладают тем свойством, что при $h\to 0$ преобладающую роль играют окрестности точки θ, а дополнительные множества имеют веса, стремящиеся к нулю. Это делает правдоподобным предположение, что при разумных определениях сходимости и непрерывности мы будем иметь $A_h(\theta)\to T(\theta)$ *для любого непрерывного семейства операторов* $T(\theta)$. В частности, если операторы $T(\theta)$ образуют *полугруппу*, то $T(kh)=(T(h))^k$ и интерполяционный оператор $A_h(\theta)$ имеет вид левой части в (5.7). Неудивительно поэтому, что «экспоненциальная формула» (5.7) *вообще верна для непрерывных полугрупп ограниченных операторов.* Мы вернемся к доказательству этого [1]) в гл. X, 9.

§ 6. Формулы обращения для преобразования Лапласа

Как неоднократно отмечалось, аппроксимационная формула (5.1) [см. пример (1, б)] имеет следующий вероятностный смысл Пусть $\mathbf{X}$ — случайная величина, распределенная по закону Пуассона с параметром $\lambda\theta$. Тогда при больших λ вероятность события $|\mathbf{X}-\lambda\theta|>\lambda\varepsilon$ мала. Поэтому для $\mathbf{P}\{\mathbf{X}\leqslant\lambda x\}$ при $\lambda\to\infty$

$$e^{-\lambda\theta}\sum_{k\leqslant\lambda x}\frac{(\lambda\theta)^k}{k!}\to\begin{cases}0, & \text{если} \quad \theta>x,\\ 1, & \text{если} \quad \theta<x.\end{cases} \tag{6.1}$$

Эта формула справедлива при фиксированных положительных x и θ и получается из (5.1), если u — ступенчатая функция. Использование этой формулы в анализе мы проиллюстрируем сейчас на примере преобразований Лапласа (систематически они изучаются в гл. XIII).

---

[1]) Это доказательство, принадлежащее М. Риссу, имеется (при несколько более общих условиях) в книге Хилле Э., Филлипс Р., Функциональный анализ и полугруппы, ИЛ, М., 1962. Как можно понять, авторы не учли, что ссылка на линейную интерполяционную формулу (5.8) как на пуассоновскую *рандомизацию параметра полугруппы* и ссылка на неравенство Чебышева упрощают изложение. Вероятностные соображения отмечались Д. Кендаллом и были в полной мере использованы в статье Chung K. L., On the exponential formulas of semi group theory, *Math. Scandinavica,* **10** (1962), 153—162.

Пусть F — распределение вероятностей, сосредоточенное на $0, \infty$. Преобразованием Лапласа функции F называется функция φ, определенная для $\lambda>0$ формулой

$$\varphi(\lambda)=\int_0^\infty e^{-\lambda\theta}F\{d\theta\}. \tag{6.2}$$

Очевидно, что производные $\varphi^{(k)}(\lambda)$ существуют и их можно вычислить формальным дифференцированием:

$$(-1)^k\varphi^{(k)}(\lambda)=\int_0^\infty e^{-\lambda\theta}\theta^k F\{d\theta\}. \tag{6.3}$$

Из этих равенств и из формулы (6.1) видно, что в каждой точке непрерывности распределения F

$$\sum_{u\leqslant\lambda x}\frac{(-1)^k}{k!}\lambda^k\varphi^{(k)}(\lambda)\to F(x). \tag{6.4}$$

Это очень полезная *формула обращения*. Она показывает, в частности, что *распределение F однозначно определяется своим преобразованием Лапласа.*

Такими же рассуждениями получается множество близких формул обращения, применимых при различных обстоятельствах. Например, (1.6) представляет собой формулу обращения для интегралов Лапласа типа

$$w(\lambda)=\int_0^\infty e^{-\lambda x}u(x)\,dx. \tag{6.5}$$

Проведем формальное дифференцирование, как в (6.3). Тогда (1.6) означает, что для *ограниченной и непрерывной функции u*

$$\frac{(-1)^{n-1}}{(n-1)!}\left(\frac{n}{\theta}\right)^n w^{(n-1)}\left(\frac{n}{\theta}\right)\to u(\theta) \tag{6.6}$$

равномерно в каждом конечном интервале.

[Эти формулы обращения справедливы при значительно более общих условиях. Однако сейчас представляется нежелательным «забивать» простоту рассуждений балластом новой терминологии. Абстрактный аналог соотношения (6.6) появится в гл. XIII, 9.]

Если распределение F обладает моментами $\mu_1, \ldots, \mu_{2n}$, то его *преобразование Лапласа удовлетворяет неравенствам*

$$\sum_{k=0}^{2n-1}\frac{(-1)^k\mu_k\lambda^k}{k!}\leqslant\varphi(\lambda)\leqslant\sum_{k=0}^{2n}\frac{(-1)^k\mu_k\lambda^k}{k!}. \tag{6.7}$$

Эти неравенства часто используются. Чтобы доказать их, мы будем исходить из хорошо известных неравенств [1])

$$\sum_{k=0}^{2n-1} \frac{(-1)^k t^k}{k!} < e^{-t} < \sum_{k=0}^{2n} \frac{(-1)^k t^k}{k!}, \qquad t > 0. \tag{6.8}$$

Заменяя t на λt и интегрируя относительно F, получим (6.7). Отсюда следует, в частности, что

$$\varphi(\lambda) = \sum_{k=0}^{\infty} \frac{(-1)^k \mu_k \lambda^k}{k!} \tag{6.9}$$

в любом интервале $0 \leqslant \lambda < \lambda_0$, *в котором сходится ряд, стоящий справа.* Если такой интервал существует, то, согласно теории аналитических функций, ряд в (6.9) однозначно определяет $\varphi(\lambda)$ при *всех* $\lambda > 0$. Следовательно, *последовательность моментов* $\mu_0, \mu_1, \ldots$ *однозначно определяет распределение* F, *если ряд в* (6.9) *сходится в некотором интервале* $|\lambda| < \lambda_0$. Этот полезный критерий справедлив и для распределений, не сосредоточенных на $0, \infty$. Но в общем случае доказательство основано на применении характеристических функций (см. гл. XV, 4).

§ 7*. Законы больших чисел для одинаково распределенных случайных величин

Всюду в этом параграфе мы используем обозначение $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. Самый старый вариант закона больших чисел состоит в утверждении, что для независимых случайных величин $\mathbf{X}_k$ с одним и тем же распределением и математическим ожиданием μ и конечной дисперсией [2])

$$\mathbf{P}\left\{\left|\frac{\mathbf{S}_n}{n} - \mu\right| > \varepsilon\right\} \to 0 \tag{7.1}$$

при $n \to \infty$ и любом фиксированном $\varepsilon > 0$. В начале этой главы отмечалось, что (7.1) вытекает из неравенства Чебышева. Чтобы получить более сильный результат, мы докажем, что неравенство (разновидность неравенства Чебышева) применимо и в случае, когда не существует математических ожиданий. Определим новые случайные величины $\mathbf{X}'_k$, получаемые «урезанием» $\mathbf{X}_k$

---

[1]) Простое дифференцирование и индукция показывают, что разность между любыми двумя членами в (6.8) является *монотонной* функцией от t.

* Тема этого параграфа связана с первым периодом развития теории вероятностей, однако никакой особой важности для остального содержания книги она не представляет. Мы останавливаемся на ней по причине ее исторического и методологического интереса, а также потому, что имеется много работ, посвященных обращению закона больших чисел.

[2]) В обозначениях, вводимых в гл. VIII, 2, соотношение (7.1) записывается проще: $n^{-1}\mathbf{S}_n - \mu \xrightarrow{\mathrm{p}} 0$, где знак $\xrightarrow{\mathrm{p}}$ означает сходимость по вероятности.

на произвольном, но фиксированном уровне $\pm s$, равенствами

$$\mathbf{X}_k = \begin{cases} \mathbf{X}_k, & \text{когда } |\mathbf{X}_k| \leqslant s, \\ 0, & \text{когда } |\mathbf{X}_k| > s. \end{cases} \tag{7.2}$$

Положим

$$\mathbf{S}'_n = \mathbf{X}'_1 + \ldots + \mathbf{X}'_n, \quad m'_n = \mathbf{E}(\mathbf{S}'_n). \tag{7.3}$$

Тогда очевидно, что

$$\mathbf{P}\{|\mathbf{S}_n - m'_n| > t\} \leqslant \mathbf{P}\{|\mathbf{S}'_n - m'_n| > t\} + \mathbf{P}\{\mathbf{S}_n \neq \mathbf{S}'_n\}, \tag{7.4}$$

так как из наступления события, написанного слева, вытекает наступление хотя бы одного из событий, стоящих справа. Неравенства (7.4) применимы и тогда, когда $\mathbf{X}_k$ зависимы. Можно получить полезное следствие, если оценить первый член в правой части по неравенству Чебышева. Наиболее важен случай, когда величины $\mathbf{X}_k$ независимы и одинаково распределены. Так как дисперсия никогда не превосходит второго момента, то мы получаем в этом случае неравенство

$$\mathbf{P}\{|\mathbf{S}_n - m'_n| > t\} \leqslant \frac{n}{t^2}\mathbf{E}(\mathbf{X}'^2_j) + n\mathbf{P}\{|\mathbf{X}_j| > s\}. \tag{7.5}$$

Из него, в предположении, что существует $\mu = \mathbf{E}(\mathbf{X}_j)$, легко вытекает (7.1). Это утверждение известно под названием *закона больших чисел* в форме Хинчина (именно в этой форме он доказан в **1**, гл. X, 2). Вместо того чтобы повторять прежнее доказательство, мы рассмотрим более сильное утверждение, указывающее необходимые и достаточные условия для (7.6). Они, очевидно, будут выполнены, когда $\mu = \mathbf{E}(\mathbf{X}_j)$ существует.

Теорема 1. (*Слабый закон больших чисел.*) *Пусть случайные величины $\mathbf{X}_k$ независимы и имеют одно и то же распределение F. Для того чтобы существовали такие постоянные μ_n, что*

$$\mathbf{P}\left\{\left|\frac{\mathbf{S}_n}{n} - \mu_n\right| > \varepsilon\right\} \to 0, \tag{7.6}$$

необходимо и достаточно, чтобы при $t \to \infty$

$$t[1 - F(t) + F(-t)] \to 0, \quad t > 0. \tag{7.7}$$

При выполнении (7.7) можно взять

$$\mu_n = \int_{-n}^{n} xF\{dx\}. \tag{7.8}$$

Доказательство. а) *Достаточность.* Мы используем неравенство (7.5), положив в нем $s = n$ и $t = n\varepsilon$. Тогда $m'_n = n\mu_n$ и

левые части соотношений (7.4) и (7.6) совпадают. Обозначим теперь левую часть в (7.7) через $\tau(t)$. Интегрирование по частям показывает, что правая часть в (7.5) не превосходит

$$\frac{4}{n\varepsilon^2}\int_0^n \tau(x)\,dx + \tau(n).$$

Из (7.7) следует, что эта величина стремится к нулю при любом фиксированном $\varepsilon>0$.

б) *Необходимость.* Допустим сначала, что распределение F симметрично. По неравенству симметризации (5.11) из гл. V в этом случае

$$\mathbf{P}\{|\mathbf{S}_n| \geqslant \eta n\} \geqslant \frac{1}{2}(1 - e^{-\tau(\eta n)/\eta}).$$

В силу симметрии мы можем предположить, что в (7.6) $\mu_n=0$. Тогда левая часть последнего неравенства должна стремиться к нулю при $n\to\infty$ и, таким образом, для симметричного распределения F условие (7.7) необходимо. Пусть F не является симметричным. Неравенство симметризации (5.6) из гл. V показывает, что (7.6) имеет место и для симметризованного распределения ${}^0F$ и, следовательно, $t[1 - {}^0F(t) + {}^0F(-t)]\to 0$. Неравенство симметризации (5.7) из гл. V показывает, что отсюда вытекает (7.7). Теорема доказана. ▶

Пример. Пусть F — симметричное распределение. Закон больших чисел *не* выполняется, если $1 - F(x) \sim x^{-1}$ (как у распределения Коши), и *выполняется*, если $1 - F(x) \sim (x \ln x)^{-1}$. ▶

Если F имеет математическое ожидание μ, то $\mu_n \to \mu$, и (7.6) равносильно (7.1). С другой стороны, предел $\mu = \lim \mu_n$ может существовать и при отсутствии математического ожидания. Например, $\mu_n \to 0$ для любого симметричного распределения F. Таким образом, *слабый закон больших чисел в его классической форме* (7.1) *применим и к некоторым случайным величинам, не имеющим математического ожидания.* В таких случаях левая часть в (7.1) будет мала для каждого отдельного большого числа n, но, как мы покажем, последовательность средних арифметических $\mathbf{S}_n/n$ с вероятностью единица содержит бесконечно много членов, лежащих вне любого наперед заданного конечного интервала. Это замечание (принадлежащее Колмогорову) показывает, что характер случайных флуктуаций сильно связан с фактом существования математического ожидания.

Теорема 2. (*Усиленный закон больших чисел и обратная теорема.*) *Пусть случайные величины* $\mathbf{X}_k$ *независимы и имеют*

одно и то же распределение F. Если они имеют математическое ожидание μ, *то* $n^{-1}\mathbf{S}_n \to \mu$ *с вероятностью единица. В противном случае последовательность* $\{n^{-1}\mathbf{S}_n\}$ *с вероятностью единица не ограничена.*

Доказательство. а) Для дискретных распределений первая часть утверждения содержится в теореме из 1, гл. X, 7. Ее доказательство применимо к произвольным распределениям и будет повторено (в усиленном варианте) в следующем параграфе.

б) Если какие-либо числа a_n остаются ограниченными, что тоже верно и для их разностей $a_n - a_{n-1}$. Для доказательства второй части нам достаточно установить, что при отсутствии математического ожидания для любого $a>0$ с вероятностью единица наступает бесконечно много событий $\{|\mathbf{X}_n| > an\}$. По второй лемме Бореля — Кантелли [см. 1, гл. VIII, 3] это будет так, если

$$\sum_{n=1}^{\infty} \mathbf{P}\{|\mathbf{X}_n| > an\} = \infty. \tag{7.9}$$

При фиксированном k интервал $ak<x\leqslant a(k+1)$ учитывается в левой части равенства (7.9) только при $n\leqslant k$. Следовательно, ряд в левой части равен

$$\sum_{k=1}^{\infty} k \int_{ak<|x|\leqslant a(k+1)} F\{dx\} \geqslant \frac{1}{2a} \sum_{k=1}^{\infty} \int_{ak<|x|\leqslant a(k+1)} |x| F\{dx\} =$$

$$= \frac{1}{2a} \int_{|x|>a} |x| F\{dx\}.$$

Последний интеграл по предположению расходится, поэтому (7.9) верно. ▶

Закон больших чисел неприменим в *Петербургской игре* (см. 1, гл. V, 4). Однако в ней применим слегка видоизмененный закон. Следующая ниже теорема обобщает этот пример и имеет одинаковую с ним интуитивную интерпретацию.

Теорема 3. *Пусть* $\mathbf{X}_k$ *— независимые положительные случайные величины с одним и тем же распределением F* [*таким, что* $F(0)=0$]. *Для того чтобы существовали такие постоянные* a_n, *что*

$$\mathbf{P}\left\{\left|\frac{\mathbf{S}_n}{a_n} - 1\right| > \varepsilon\right\} \to 0, \tag{7.10}$$

необходимо и достаточно, чтобы[1])

$$\frac{1}{t[1-F(t)]}\int_0^t xF\{dx\}\to\infty \tag{7.11}$$

при $t\to\infty$.
Излишне говорить, что (7.10) влечет $a_n\to\infty$.

Доказательство. а) *Достаточность.* Выберем произвольно η_n, $0<\eta_n<1$, и определим по η_n сначала ρ_n, а затем a_n формулами

$$\eta_n = n[1-F(\rho_n)],\quad a_n = n\int_0^{\rho_n} xF\{dx\}. \tag{7.12}$$

В предположении, что (7.11) выполняется, мы имеем $\eta_n\rho_n a_n^{-1}\to 0$, и потому можно выбрать $\eta_n\to 0$ так, чтобы $a_n\rho_n^{-1}\to\infty$. Произведя такой выбор, используем урезание (7.2) на уровне $s=\rho_n$. Тогда из (7.5) получаем

$$\mathbf{P}\{|\mathbf{S}_n-a_n|>\varepsilon a_n\}\leqslant\frac{n}{\varepsilon^2a_n^2}\int_0^{\rho_n}x^2F\{dx\}+$$

$$+\,n[1-F(\rho_n)]\leqslant\frac{\rho_n}{\varepsilon^2a_n}+\eta_n\to 0. \tag{7.13}$$

Поэтому (7.10) верно.

б) Пусть (7.10) выполняется. Положим

$$\eta_n = n[1-F(2a_n)],\quad \mu_n = n\int_0^{2a_n} xF\{dx\} \tag{7.14}$$

и используем урезание (7.2) с $s=2a_n$.

Из (7.5) с помощью оценок, примененных в (7.13), получаем

$$\mathbf{P}\{|\mathbf{S}_n-\mu_n|>\varepsilon a_n\}<\frac{2}{\varepsilon^2}\frac{\mu_n}{a_n}+\eta_n. \tag{7.15}$$

Все случайные величины $\mathbf{X}_k$ положительны. Поэтому из (7.10) необходимо следует, что вероятность $F^n(2a_n)$ того, что все n случайных величин $\mathbf{X}_1,\ \ldots,\ \mathbf{X}_n$ не превосходят $2a_n$, стремится

[1]) Мы увидим в гл. VIII, 8, что (7.11) эквивалентно *правильному изменению функции* $1-F(x)$ *с показателем, равным* -1. (Использование этого факта упростило бы доказательство.) Соотношение (7.10) означает, что $a_n^{-1}\mathbf{S}_n\xrightarrow{P}1$.

к единице. Отсюда вытекает, что $\eta_n \to 0$. Теперь становится очевидным, что (7.15) противоречило бы предположению (7.10), если бы отношения μ_n/a_n могли приближаться к нулю. Следовательно, отношения $\mu_n/(a_n\eta_n)$ стремятся к бесконечности. Это означает, что (7.11) выполняется, когда t пробегает последовательность $2a_1, 2a_2, \ldots$. Но из (7.10) легко вывести, что $a_{n+1}/a_n \to 1$, а потому (7.11) выполняется при $t \to \infty$. ▶

§ 8*. Усиленный закон больших чисел для мартингалов

Усиленный закон больших чисел для дискретных случайных величин обсуждался подробно в **1**, гл. X, 7. Все сказанное там переносится без изменений и на рассматриваемый сейчас более общий случай. Вместо того чтобы механически повторять теорию, мы распространим ее со случая сумм независимых величин на мартингалы. Используемые понятия будут при этом сложнее, но вычисления станут проще, а результаты лучше.

Предположим, что $\mathbf{E}(\mathbf{X}_j) = 0$ и обозначим

$$\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n, \tag{8.1}$$

$$\mathbf{M}_n = \max[|\mathbf{S}_1|, \ldots, |\mathbf{S}_n|]. \tag{8.2}$$

Напомним, что проведенный в **1**, гл. X, 7, анализ в значительной степени опирался на неравенство Колмогорова. Для независимых случайных величин с нулевыми математическими ожиданиями и конечными дисперсиями это неравенство дает оценку

$$\mathbf{P}\{\mathbf{M}_n > t\} \leqslant \frac{\mathbf{E}(\mathbf{S}_n^2)}{t^2}, \quad t > 0. \tag{8.3}$$

Оно значительно сильнее неравенства Чебышева, которое дает ту же самую оценку для «хвоста» распределения величины $|\mathbf{S}_n|$. Просматривая доказательство из **1**, гл. IX, 7, мы видим, что оно использует не независимость случайных величин $\mathbf{X}_j$, а тот факт, что последовательность $\{\mathbf{X}_n\}$ *абсолютно беспристрастна*, т. е.

$$\mathbf{E}(\mathbf{X}_n | \mathbf{X}_1, \ldots, \mathbf{X}_{n-1}) = 0 \tag{8.4}$$

для $n = 2, 3, \ldots$. Как было показано в гл. VI, 12, частичные суммы $\{\mathbf{S}_n\}$ такой последовательности образуют *мартингал*, и все мартингалы получаются таким путем.

Сформулируем теперь

Неравенство Колмогорова для мартингалов. *Если последовательность* $\{\mathbf{X}_n\}$ *абсолютно беспристрастна*

* Материал в дальнейшем не используется.

и величина S_n *определена по формуле* (8.1), *то верно неравенство* (8.3).

Мы можем сделать еще шаг дальше, замечая, что доказательство не зависит от того, что $\{S_n\}$ является мартингалом. По существу используется тот факт, что случайные величины $Y_n = S_n^2$ образуют *субмартингал*, т. е. что выполняются неравенства[1])

$$\mathbf{E}(Y_n \mid Y_1, \ldots, Y_{n-1}) \geqslant Y_{n-1}. \tag{8.5}$$

Мы докажем сейчас вариант неравенства Колмогорова, пригодный для произвольных субмартингалов (поучительно проверить, что приводимое ниже доказательство отличается от доказательства из 1, гл. IX, 7, лишь обозначениями).

Неравенство Колмогорова для субмартингалов. *Пусть* $\{Y_n\}$ — *последовательность случайных величин, подчиненных условию* (8.5), *и таких, что* $Y_j \geqslant 0$ *при всех* j. *Тогда при* $x>0$

$$\mathbf{P}\{\max[Y_1, \ldots, Y_n] > x\} \leqslant \frac{\mathbf{E}(Y_n)}{x}. \tag{8.6}$$

(Неравенство (8.3) для мартингалов получается при $Y_j = S_j^2$ и $x = t^2$.)

Доказательство. Пусть значение $x>0$ фиксировано. Обозначим через K *наименьший* индекс $j \leqslant n$, для которого $Y_j > x$, и положим $K=0$ в случае, когда такого индекса нет. Тогда K — случайная величина с возможными значениями $0, 1, \ldots, n$ и

$$\mathbf{P}\{\max[Y_1, \ldots, Y_n] > x\} = \mathbf{P}\{K \neq 0\} = \sum_{j=1}^{n} \mathbf{P}\{K = j\}. \tag{8.7}$$

Тогда

$$\mathbf{E}(Y_n) = \sum_{j=0}^{n} \mathbf{E}(Y_n \mid K = j) \cdot \mathbf{P}\{K = j\}. \tag{8.8}$$

Член с номером j равен интегралу от Y_n, взятому по множеству элементарных событий $\{K=j\}$ (и по отношению к совместному распределению величин $(Y_1, \ldots, Y_n)$). При $j>0$ событие $\{K=j\}$ зависит только от $Y_1, \ldots, Y_j$. Поэтому в соответствии с опреде-

---

[1]) Функция $u(x)=x^2$ выпукла. Следовательно, если $\{S_n\}$ — мартингал, то $\{S_n^2\}$ — субмартингал. В специальном случае независимых случайных величин с $\mathbf{E}(X_j)=0$, $\mathbf{E}(X_j^2)=\sigma_j^2$ тривиально получаем

$$\mathbf{E}(S_n^2 \mid S_1^2, \ldots, S_{n-1}^2) = \mathbf{E}((X_n + S_{n-1})^2 \mid S_{n-1}^2) = \sigma_n^2 + S_{n-1}^2.$$

лением условного математического ожидания величина $\mathbf{Y}_n$ в интеграле может быть заменена на $\mathbf{E}(\mathbf{Y}_n \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_j)$. Из неравенств (8.5) и определения величины $\mathbf{K}$ следует, что подинтегральное выражение не меньше $\mathbf{Y}_j > x$. Таким образом, при $j>0$ j-й член суммы (8.8) не меньше $x\mathbf{P}\{\mathbf{K}=j\}$. Из (8.7) и (8.8) с очевидностью вытекает справедливость неравенства (8.6). ▶

В качестве первого примера использования полученного результата мы докажем одну теорему, имеющую удивительно широкую область применений, в которую входят и частичные суммы (8.1) любых абсолютно беспристрастных последовательностей $\{\mathbf{X}_n\}$.

Теорема 1. *Теорема о сходимости мартингалов*[1]). *Пусть* $\{\mathbf{S}_n\}$ *— мартингал, определенный на некотором вероятностном пространстве. Если* $\mathbf{E}(\mathbf{S}_n^2) \leqslant \mathbf{M} < \infty$ *при всех* n, *то существует такая случайная величина* $\mathbf{Y}$, *что с вероятностью единица* $\mathbf{S}_n \to \mathbf{Y}$.

Доказательство. Начнем с замечания, что в силу свойств субмартингалов последовательность математических ожиданий $\mathbf{E}(\mathbf{S}_n^2)$ является неубывающей. Поэтому условие ограниченности эквивалентно условию

$$\mathbf{E}(\mathbf{S}_n^2) \to \mu < \infty. \tag{8.9}$$

По неравенству Колмогорова для мартингалов имеем

$$\mathbf{P}\{|\mathbf{S}_{m+k} - \mathbf{S}_m| > t \text{ при некотором } k \leqslant n\} \leqslant \leqslant \frac{\mathbf{E}\left((\mathbf{S}_{m+n} - \mathbf{S}_m)^2\right)}{t^2} = \frac{\mathbf{E}(\mathbf{S}_{m+n}^2) - \mathbf{E}(\mathbf{S}_m^2)}{t^2}. \tag{8.10}$$

[По поводу последнего равенства напомним, что $\mathbf{E}(\mathbf{S}_{m+n}\mathbf{S}_m \mid \mathbf{S}_m) = \mathbf{S}_m^2$.] В силу (8.9) при $m>m_0$ и всех n правая часть неравенства (8.10) меньше ε. Полагая в (8.10) $n \to \infty$, видим, что множество точек выборочного пространства, в которых $|\mathbf{S}_{m+k} - \mathbf{S}_m| > t$ для некоторого k и $m>m_0$, имеет вероятность меньшую, чем ε/t^2. В силу произвольной малости ε отсюда следует, что множество точек, в которых последовательность $\{\mathbf{S}_m\}$ не является фундаментальной, имеет нулевую вероятность. ▶

Примеры. а) *Урновая схема Пойа* уже рассматривалась в примере (12, б) из гл. VI и в примере (4, а) этой главы. Обозначим через $\mathbf{Y}_n$ долю черных шаров при n-м испытании. Уже было показано, что $\{\mathbf{Y}_n\}$ — мартингал. Теперь мы видим, что

[1]) Теорема верна при единственном предположении, что величины $\mathbf{E}(|\mathbf{S}_n|)$ ограничены. См. книги Дуба, Лоэва, Невё.

с вероятностью единица существует предел $\mathbf{Y}=\lim \mathbf{Y}_n$. С другой стороны, вероятность извлечь при n-м испытании черный шар получается рандомизацией биномиального распределения. Таким образом, если $\mathbf{S}_n$ — общее число черных шаров, извлеченных при первых n испытаниях, то распределение величины $n^{-1}\mathbf{S}_n$ сходится к бета-распределению F, найденному в примере (4, а). Поэтому предельная случайная величина $\mathbf{Y}$ имеет бета-распределение F.

б) *Ветвящиеся процессы.* В ветвящемся процессе, описанном *в* **1**, гл. XII, 5, число $\mathbf{X}_n$ индивидуумов n-го поколения имеет математическое ожидание $\mathbf{E}(\mathbf{X}_n)=\mu^n$ [см. **1**, гл. XII, формула (5.3)]. При условии, что $(n-1)$-е поколение содержит ν индивидуумов, математическое ожидание (условное) величины $\mathbf{X}_n$ равно $\mu\nu$ и не зависит от размера предшествующих поколений. Таким образом, последовательность $\{\mathbf{S}_n\}$, где $\mathbf{S}_n=\mathbf{X}_n/\mu^n$, образует мартингал. Нетрудно установить, что если $\mathbf{E}(\mathbf{X}_1^2)<\infty$, то $\mathbf{E}(\mathbf{S}_n^2)$ остается ограниченным в предположении $\mu>1$. (Случай $\mu\leqslant 1$ см. сноску на стр. 297; см. задачу 7 из **1**, гл. XII, 6.) Мы получаем удивительный результат, что $\mathbf{S}_n$ с вероятностью единица сходится к пределу $\mathbf{S}_\infty$. Отсюда следует, в частности, что распределение величины $\mathbf{S}_n$ сходится к распределению величины $\mathbf{S}_\infty$. Эти результаты принадлежат Т. Харрису.

в) *Гармонические функции.* Для ясности мы рассмотрим специальный пример, хотя последующие рассуждения применимы к более общим марковским цепям и согласованным с ними функциям (пример (12, в) из гл. VI).

Пусть D обозначает единичный круг, т. е. совокупность таких точек $x=(x_1, x_2)$, что $x_1^2+x_2^2\leqslant 1$. Для любой точки $x\in D$ обозначим через C_x наибольшую окружность с центром в x, содержащуюся в D. Рассмотрим марковский процесс в D, определенный следующим образом. При данном $\mathbf{Y}_n=x$ случайная величина $\mathbf{Y}_{n+1}$ распределена равномерно на C_x; начальное положение $\mathbf{Y}_0=y$ считается известным. Переходные вероятности определяются стохастическим ядром K, которое при фиксированном x сосредоточено на C_x и сводится там к равномерному распределению. Функция u, заданная на D, будет согласованной с процессом, если значение $u(x)$ равно среднему значению u на C_x. Рассмотрим теперь гармоническую функцию u, непрерывную в *замкнутом* круге D. Тогда $\{u(\mathbf{Y}_n)\}$ — ограниченный мартингал и потому $\mathbf{Z}=\lim u(\mathbf{Y}_n)$ существует с вероятностью единица. Так как координатные переменные x_j являются гармоническими функциями, то из сказанного следует, что с вероятностью единица $\mathbf{Y}_n$ сходится к пределу $\mathbf{Y}\in D$. Легко видеть, что процесс

не может сходиться к внутренней точке круга D. Поэтому $\mathbf{Y}_n$ *сходится* с вероятностью единица *к точке* $\mathbf{Y}$ *на границе круга* D.

В обобщенном виде подобного типа рассуждения используются при изучении асимптотических свойств марковских процессов и при доказательстве общих теорем о гармонических функциях, таких, как теорема Фату (теорема о существовании радиальных граничных значений почти всюду) [1]).

Следующая ниже теорема обобщает критерий Колмогорова из **1**, гл. X, 7, в двух направлениях: во-первых, она применима к мартингалам, а не только к независимым величинам; во-вторых, в ней используются любые нормирующие константы b_k.

Теорема 2. *Пусть случайные величины* $\{\mathbf{X}_k\}$ *удовлетворяют условию* (8.4). *Определим* $\{\mathbf{S}_n\}$ *по формуле* (8.1). *Если* $b_1 < b_2 < \ldots \to \infty$ *и*

$$\sum b_k^{-2}\mathbf{E}(\mathbf{X}_k^2) < \infty, \tag{8.11}$$

то с вероятностью единица

$$b_n^{-1}\mathbf{S}_n \to 0 \tag{8.12}$$

и величины

$$\mathbf{Y}_n = \sum_{k=1}^{n} b_k^{-1}\mathbf{X}_k \tag{8.13}$$

сходятся.

Доказательство. Очевидно, что $\{\mathbf{Y}_n\}$ — мартингал. Оценка, аналогичная примененной к правой части соотношения (8.10), показывает, что величина $\mathbf{E}(\mathbf{Y}_n^2)$ ограничена суммой ряда (8.11). По теореме 1 последовательность $\{\mathbf{Y}_n\}$ сходится с вероятностью единица и соотношение (8.12) выполняется в каждой точке выборочного пространства, где $\{\mathbf{Y}_n\}$ сходится [2]). Доказательство закончено. ▶

---

[1]) Brelot M., Doob J. L., Limites angulaires et limites fines, *Ann. Inst. Fourier*, **13** (1963), 395—415. [Русский перевод см. сб. *Математика*, **11**:3 (1967), 101—116.]

[2]) Это утверждение известно под названием *леммы Кронекера*. По этой лемме сходимость ряда $\sum b_k^{-1}x_k$ влечет соотношение $(x_1 + \ldots + x_n)b_n^{-1} \to 0$. Для доказательства обозначим остаток указанного ряда через ρ_n. Тогда $x_n = b_n(\rho_{n-1} - \rho_n)$ и, следовательно,

$$\frac{x_1 + \ldots + x_n}{b_n} = -\rho_n + \frac{1}{b_n}\sum_{k=0}^{n-1}\rho_k(b_{k+1} - b_k). \tag{*}$$

Пусть $|\rho_k| < \varepsilon$ для $k \geqslant r$. Так как $b_n \to \infty$, то поделенная на b_n сумма первых r членов справа стремится к нулю. В силу монотонности $\{b_n\}$ остальные члены дают величину, не превосходящую $\varepsilon(b_n - b_r)b_n^{-1} \leqslant \varepsilon$.

В качестве следствия мы получаем *усиленный закон больших чисел для мартингалов: если* $\sum n^{-2}\mathbf{E}(\mathbf{X}_n^2) < \infty$, *то* $n^{-1}\mathbf{S}_n \to 0$ *с вероятностью единица.* В специальном случае независимых и одинаково распределенных величин $\mathbf{X}_k$ урезание, описанное в **1**, гл. X, 7, приводит к более сильному результату, а именно: *если* $\mathbf{E}(\mathbf{X}_k) = 0$, *то с вероятностью единица* $n^{-1}\mathbf{S}_n \to 0$.

§ 9. Задачи

1. Если функция u непрерывна на $\overline{0, \infty}$, то при $n \to \infty$

$$\sum_{k=0}^{\infty} \binom{n+k}{k} \frac{t^k}{(1+t)^{n+k+1}} u\left(\frac{k}{n+1}\right) \to u(t).$$

равномерно в каждом конечном интервале.

Указание. Вспомните отрицательно-биномиальное распределение (**1**, гл. VI, 8). Не нужно никаких вычислений.

2. Если u имеет непрерывную производную u', то производные $B'_{n,u}$ полиномов $B_{n,u}$ равномерно сходятся к u'.

3. *Полиномы Бернштейна в* $\mathscr{R}^2$. Если $u(x, y)$ непрерывна в треугольнике $x \geqslant 0$, $y \geqslant 0$, $x+y \leqslant 1$, то имеет место равномерная сходимость

$$\sum u\left(\frac{j}{n}, \frac{k}{n}\right) \frac{n!}{j!\,k!\,(n-j-k)!} x^j y^k (1-x-y)^{n-j-k} \to u(x, y).$$

4. Функцию u, непрерывную на $\overline{0,1}$, можно равномерно аппроксимировать полиномами четной степени. Если $u(0) = 0$, то же самое верно для полиномов нечетной степени [1]).

5. Если u непрерывна в интервале $\overline{0, \infty}$ и $u(\infty)$ существует, то u может быть равномерно аппроксимирована линейными комбинациями функций e^{-nx}.

6. Ниже указаны три последовательности моментов. Найдите вероятности $p_k^{(n)}$, входящие в (3.4). Используя предельное соотношение (3.6), найдите соответствующее распределение вероятностей:

(a) $\mu_n = p^n, \qquad 0 < p < 1,$

(b) $\mu_n = \dfrac{1}{n+1},$

(c) $\mu_n = \dfrac{2}{n+2}.$

---

[1]) По известной теореме Мюнца равномерная аппроксимация линейными комбинациями степеней $1, x^{n_1}, x^{n_2}, \ldots$ возможна в том и только в том случае, когда ряд $\sum n_k^{-1}$ расходится.

7 [1]). Пусть p — полином степени ν. Покажите, что при $n>\nu$ разность $\Delta^n p$ тождественно равна нулю. Покажите, что $B_{n,p}$ — полином степени не более ν (несмотря на то что в записи он выглядит как полином степени $n>\nu$).

8. Покажите, что если F имеет плотность, то соотношение (6.4) можно получить интегрированием соотношения (6.6).

9. *Закон больших чисел для стационарных последовательностей.* Пусть $\{\mathbf{X}_k\}$ $(k=0, \pm 1, \pm 2, \ldots)$ — стационарная последовательность. Определим $\mathbf{X}'_k$ урезанием, таким же, как в (7.2). Если $\mathbf{E}(\mathbf{X}_k)=0$ и $\mathbf{E}\left(\mathbf{X}'_0\mathbf{X}'_n\right)\to 0$ при $n\to\infty$, то

$$\mathbf{P}\{n^{-1}\,|\,\mathbf{X}_1+\ldots+\mathbf{X}_n\,|>\varepsilon\}\to 0.$$

10. Распространите вариант неравенства Чебышева из примера (7, а) гл. V на мартингалы [2]).

---

[1]) Использование этого результата значительно упрощает классическое решение проблемы моментов (см книгу Шохата и Тамаркина). Это решение можно описать следующим образом. Для полинома $p(x)=p_0+\ldots+p_\nu x^\nu$ определим $\mathbf{E}(p)=p_0\mu_0+\ldots+p_\nu\mu_\nu$, где μ_k — предполагаемые моменты. Функционал $\mathbf{E}(p)$ — линейный функционал, определенный на множестве всех полиномов на $\overline{0,1}$. Из задачи 7 легко вывести, что неравенство $0\leqslant p\leqslant 1$ влечет неравенство $0\leqslant\mathbf{E}(p)\leqslant 1$. Поэтому функционал $\mathbf{E}(p)$ можно распространить, с использованием равномерной сходимости, на множество всех непрерывных функций. Теорема Рисса из гл IV, 5, показывает, что μ_k действительно являются моментами некоторого распределения вероятностей.

[2]) Marshall A. W., A one-sided analog of Kolmogorov's inequality, *Ann. Math. Statist.*, **31** (1960), 483—487.

ГЛАВА VIII

ОСНОВНЫЕ ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

Основные результаты этой главы содержатся в § 1, 3 и 6. Параграфы 4, 5 и 7 можно рассматривать как источник интересных примеров. Выбор примеров объясняется их важностью для тех или иных целей.

Два последних параграфа посвящены правильно меняющимся функциям (в смысле Карамата). Этот интересный вопрос приобретает все большее значение. Однако он не освещен в учебниках и его изложение не было приспособлено к случаю функций распределения. Огромное количество разрозненных вычислений в теории вероятностей может быть упразднено при использовании асимптотических соотношений § 9. Последние носят более «технический» характер по сравнению с простым § 8.

§ 1. Сходимость мер

Излагаемая ниже теория не зависит от размерности пространства. Для удобства формулировки в тексте даны для одномерного случая. Однако с учетом соглашений гл. III, 5 они без изменений переносятся на многомерный случай.

Два из приводимых ниже примеров типичны для явлений, с которыми нам придется иметь дело.

Примеры. а) Возьмем произвольное распределение вероятностей F и положим $F_n(x) = F(x - n^{-1})$. В точках x, в которых F непрерывна, имеем $F_n(x) \to F(x)$. В то же время в точках разрыва $F_n(x) \to F(x-)$. Тем не менее мы будем говорить, что $\{F_n\}$ сходится к F.

б) Положим на этот раз $F_n(x) = F(x+n)$, где F — непрерывная функция распределения. Теперь $F_n(x) \to 1$ при всех x; предел существует, но не является вероятностной функцией распределения. Аналогично если $F_n(x) = F(n^{-1}x)$, то $F_n(x) \to F(0)$. В подобных случаях мы будем говорить о несобственной сходимости.

в) В теореме 1 из гл. VII, 3 мы сталкивались с функциями распределения F_n и F, такими, что $F_n(x) \to F(x)$ в каждой точке x, где F непрерывна. В других точках это соотношение могло

и не выполняться. Формула (6.1) гл. VII иллюстрирует это же явление с распределением F, сосредоточенным в одной точке.

Основные понятия и обозначения

Нам будет нужно выделить три класса непрерывных функций. В одномерном случае[1]) $C(-\infty, \infty)$ обозначает класс всех ограниченных непрерывных функций; $C[-\infty, \infty]$ обозначает подкласс функций, имеющих конечные пределы $u(-\infty)$ и $u(\infty)$; наконец $C_0(-\infty, \infty)$ обозначает подкласс функций, *«исчезающих на бесконечности»*, т. е. функций с $u(\pm\infty)=0$.

Мы скажем, что I есть *интервал непрерывности* для распределения вероятностей F, если I — открытый интервал, концы которого не являются атомами F[2]). Вся прямая рассматривается как интервал непрерывности. В этом параграфе мы используем сокращенные обозначения

$$\mathbf{E}_n(u)=\int_{-\infty}^{+\infty} u(x)F_n\{dx\}, \quad \mathbf{E}(u)=\int_{-\infty}^{+\infty} u(x)F\{dx\}. \tag{1.1}$$

Всюду в этой главе, кроме § 9, мы будем рассматривать только последовательности распределений вероятностей. Однако даваемое ниже определение применимо к произвольным мерам[3]) F_n.

Определение 1. *Мы будем говорить, что последовательность $\{F_n\}$ сходится к F (в обозначениях $F_n \to F$) тогда и только тогда, когда*

$$F_n\{I\} \to F\{I\} \tag{1.2}$$

для каждого ограниченного интервала непрерывности F.

---

[1]) Для перехода к многомерному случаю заметим, что в одномерном случае $C[-\infty, \infty]$ — это класс непрерывных функций на прямой, «компактифицированной» добавлением точек $\pm\infty$ к $\mathscr{R}^1$. Для определения $C[-\infty, \infty]$ в $\mathscr{R}^2$ так пополняются обе координатные оси, что требуется существование при каждом x пределов $u(x, \pm\infty)$ и $u(\pm\infty, x)$. Сам по себе этот класс не очень интересен, но функции распределения входят в него. В определении $C_0(-\infty, \infty)$ требуется, чтобы $u(x, \pm\infty)=u(\pm\infty, x)=0$.

[2]) В многомерном случае требуется, чтобы граница I имела вероятность нуль.

[3]) Для читателей, интересующихся общей теорией меры, мы сделаем следующее замечание. Определения и теорема этого параграфа переносятся без изменений на произвольные локально компактные пространства, если заменить «интервалы непрерывности» на «открытые множества с границей меры нуль». Ограниченным интервалам соответствуют подмножества компактных множеств. Наконец, C_0 — класс всех непрерывных функций, исчезающих на бесконечности, т. е. $u \in C_0$, тогда и только тогда, когда u непрерывна и $|u|<\varepsilon$ вне некоторого компактного множества. Другие классы не играют никакой роли в этом параграфе.

Допустим, что F_n суть собственные распределения вероятностей и $F_n \to F$. Тогда $F\{I\} \leqslant 1$ для каждого ограниченного интервала, и F может быть как собственным, так и несобственным распределением вероятностей. Для несобственных распределений F соотношение (1.2) не выполняется, когда I совпадает со всей прямой. С другой стороны, если предельное распределение F собственное, то (1.2) выполняется и для всех неограниченных интервалов непрерывности I. В самом деле, всегда

$$\liminf F_n\{I\} \geqslant F\{I\} \tag{1.3}$$

(это неравенство становится очевидным, если заменить I на ограниченный подинтервал непрерывности J, такой, что $F(J)$ отличается от $F(I)$ менее чем на заданное число ε). Если F — собственное распределение, то, применяя соотношение (1.3) к дополнению I' интервала I, получим

$$\limsup F_n\{I\} \leqslant F\{I\}, \tag{1.4}$$

т. е. (1.2) верно. Соединим этот важный результат и

Определение 2. *Последовательность $\{F_n\}$ распределений вероятностей называется собственно сходящейся к F тогда и только тогда, когда $F_n \to F$ и F — собственное распределение вероятностей.*

Собственная сходимость равносильна тому, что (1.2) *выполняется для любого (ограниченного или неограниченного) интервала непрерывности F.*

Выбирая $I = (-\infty, x]$, мы видим[1]), что *если $F_n \to F$ в собственном смысле, то $F_n(x) \to F(x)$ в каждой точке непрерывности F.*

Вместо $F_n \to F$ мы будем также писать $F = \lim F_n$. Для ясности мы будем иногда говорить о несобственной сходимости, желая подчеркнуть, что предельное распределение F — несобственное.

Следующая ниже теорема является основной. Она описывает слабую сходимость в терминах математических ожиданий и в то же самое время дает критерий сходимости последовательности распределений.

---

[1]) Если сходимость несобственная, то последовательность $\{F_n(x)\}$ может расходится при любом x. Например, если $F_n(x) = F(x + (-1)^n n)$, то при всех x $F_{2k}(x) \to 1$ и $F_{2k+1}(x) \to 0$. В то же время $F_n(I) \to 0$ для всех ограниченных интервалов I. Промежуточный тип несобственной сходимости можно определить требованием сходимости $F_n(x) \to F(x)$ во всех точках непрерывности.

Теорема. *Пусть* $\{F_n\}$ — *последовательность собственных распределений вероятностей.*

а) *Для существования собственного или несобственного распределения* F, *такого, что* $F_n \to F$, *необходимо и достаточно, чтобы последовательность математических ожиданий* $\mathbf{E}_n(u)$ *сходилась для каждого* $u \in C_0(-\infty, \infty)$. *В этом случае*

$$\mathbf{E}_n(u) \to \mathbf{E}(u) \tag{1.5}$$

для $u \in C_0(-\infty, \infty)$.

б) *Если сходимость собственная, то* (1.5) *выполняется для всех* $u \in C(-\infty, \infty)$.

Замечания к терминологии. Если (1.5) выполняется для всех u из некоторого класса, то говорят, что F_n сходится к F слабо по отношению к этому классу. Введенная нами сходимость $F_n \to F$ может быть, таким образом, описана как слабая сходимость по отношению к $C_0(-\infty, \infty)$. Это определение сходимости имеет то преимущество, что оно применимо и к произвольным вероятностным пространствам. Однако наше элементарное определение больше соответствует интуиции. [Не нужно говорить, что все определения применимы к произвольным ограниченным мерам и нормировка вводится только ради удобства.]

Доказательство. (i) Пусть u — ограниченная и непрерывная функция. Допустим, что $|u| \leqslant 1$ и что $F_n \to F$ *в собственном смысле.* По заданному $\varepsilon > 0$ выберем интервал непрерывности A для F столь большой, что $F(A) > 1 - \varepsilon$. Разобьем A на интервалы непрерывности $I_1, \ldots, I_r$, предполагая их столь малыми, что колебание u на каждом из них меньше ε. Тогда существует функция σ, равная нулю вне A и постоянная на каждом из интервалов I_k, такая, что $|u(x) - \sigma(x)| < \varepsilon$ для $x \in A$. На дополнении A' к множеству A имеем $|u(x) - \sigma| \leqslant 1$. Следовательно,

$$|\mathbf{E}_n(u) - \mathbf{E}_n(\sigma)| \leqslant \varepsilon + F_n\{A'\}, \tag{1.6}$$

$$|\mathbf{E}(u) - \mathbf{E}(\sigma)| \leqslant \varepsilon + F\{A'\} < 2\varepsilon. \tag{1.7}$$

Теперь $\mathbf{E}_n(\sigma)$ — это линейная комбинация значений $F_n\{I_j\}$ с фиксированными коэффициентами. Следовательно, $\mathbf{E}_n(\sigma) \to \mathbf{E}(\sigma)$. Далее $F_n\{A'\} \to F\{A'\}$. Поэтому из (1.6) и (1.7) выводим, что $|\mathbf{E}_n(u) - \mathbf{E}(u)| < 10\varepsilon$ при всех достаточно больших n. Следовательно, (1.5) выполняется для всех $u \in C(-\infty, \infty)$.

В случае *несобственной* сходимости эти рассуждения непригодны, так как множество A' не ограничено и числа $F_n\{A'\}$ не обязаны сходиться к $F\{A\}$. Однако в этом случае мы рассматриваем только функции, обращающиеся в нуль на бесконечности,

и мы можем выбрать A столь большим, что $|u(x)|<\varepsilon$ для $x\in A'$. Тогда $|u-\sigma|<\varepsilon$ при всех x и левые части в (1.6) и (1.7) будут меньше ε. Следовательно, $\mathbf{E}_n(u)\to\mathbf{E}(u)$, во всяком случае для $u\in C_0(-\infty,\infty)$.

(ii) При доказательстве достаточности мы будем предполагать известной теорему о выборе (теорему 1, § 6). (Она совершенно элементарна, но ее предпочтительнее обсуждать вместе с ее вариантами и другими применениями).

В соответствии с этой теоремой каждая последовательность $\{F_n\}$ распределений вероятностей содержит подпоследовательность, сходящуюся (в собственном или несобственном смысле) к некоторому пределу F Из (1.5) и первой части теоремы мы заключаем, что $\mathbf{E}(u)$ совпадает с математическим ожиданием u по отношению к пределу F. Если (1.5) выполняется, то никакая подпоследовательность последовательности $\{F_n\}$ не может сходиться к другому пределу. Следовательно, $F_n\to F$, что утверждалось. ▶

Примеры. г) *Сходимость моментов.* Если распределения F_n сосредоточены на $\overline{0,1}$, то способ определения u вне этого интервала несуществен. Поэтому в формулировке теоремы достаточно предполагать функцию u непрерывной на $\overline{0,1}$. Каждая такая функция может быть равномерно аппроксимирована полиномами (см. гл. VII, 2). Поэтому теорема может быть сформулирована по-новому следующим образом.

Последовательность распределений F_n, сосредоточенных на $\overline{0,1}$, сходится к пределу F тогда и только тогда, когда при каждом r последовательность моментов $\mathbf{E}_n(\mathbf{X}^k)$ сходится к некоторому числу μ_k. В этом случае $\mu_k=\mathbf{E}(\mathbf{X}^k)$, т. е. μ_k будет k-м моментом распределения F, и сходимость будет собственной, так как $\mu_0=1$. (См. VI, 3.)

д) *Сходимость моментов* (продолжение). В общем случае математические ожидания F_n не обязаны сходиться, даже если $F_n\to F$ в собственном смысле. Например, если F_n приписывает вес n^{-1} точке n^2 и вес $1-n^{-1}$ точке 0, то F_n сходится к распределению, сосредоточенному в нуле, но $\mathbf{E}_n(\mathbf{X})\to\infty$. Однако имеется следующий полезный *критерий. Если $F_n\to F$ и при некотором $\rho>0$ математические ожидания $\mathbf{E}_n(|\mathbf{X}|^\rho)$ ограничены, то F является собственным распределением вероятностей.* В самом деле, часть $\mathbf{E}_n(|\mathbf{X}|^\rho)$, соответствующая области $|x|\geqslant a$, не меньше $a^\rho(1-F_n\{-a,a\})$. Последняя величина стремится к

пределу, не меньшему $a^{\rho}\eta$, где η — «дефект» F [1]). Так как a можно взять произвольно большим, то $\eta=0$. Слегка усиливая проведенные рассуждения, можно показать, что *абсолютные моменты* $\mathbf{E}_n(|\mathbf{X}|^{\alpha})$ *порядка* $\alpha<\rho$ *сходятся к* $\mathbf{E}(|\mathbf{X}|^{\alpha})$.

е) *Сходимость плотностей.* Если распределения вероятностей F_n имеют плотности f_n, то последние не обязаны сходиться, даже если $F_n \to F$ и F имеет непрерывную плотность. Пусть, например, $f_n(x)=1-\cos 2\pi n x$ при $0<x<1$ и $f_n(x)=0$ в других случаях. Тогда распределения F_n сходятся к равномерному распределению с плотностью $f(x)=1$ для $0<x<1$, но f_n не сходятся к f. С другой стороны, *если* $f_n \to f$ *и* f *— плотность вероятности, то* $F_n \to F$, где F — собственное распределение с плотностью f. В самом деле, лемма Фату [(2.9) гл. IV] влечет (1.3). Мы видели, что в свою очередь отсюда вытекает (1.4), а потому и (1.2). ▷

Пусть мы имеем дело с функциями u_t, зависящими от некоторого параметра t [такими, как $\sin tx$ или $v(t+x)$]. Тогда часто бывает полезно знать, что при достаточно больших n соотношение $|\mathbf{E}_n(u_t)-\mathbf{E}(u_t)|<\varepsilon$ выполняется одновременно для всех t. Мы докажем, что это свойство выполняется, если семейство функций u_t *равностепенно непрерывно* [2]).

Следствие. *Предположим, что* $F_n \to F$ *в собственном смысле. Пусть* $\{u_t\}$ *— семейство равностепенно непрерывных функций, зависящих от параметра* t. *Пусть при всех* t $|u_t|<M<\infty$, *где* M *— некоторая константа. Тогда* $\mathbf{E}_n(u_t) \to \mathbf{E}(u_t)$ *равномерно по* t.

Доказательство. Напомним, что в доказательстве (1.5) было использовано разбиение интервала A на интервалы, на каждом из которых колебание u меньше ε. В настоящих условиях это разбиение может быть выбрано одним и тем же для всех t, после чего утверждение становится очевидным. ▷

Пример. ж) Пусть $u_t(x)=\sin tx$. По теореме о среднем

$$|u_t(x_2)-u_t(x_1)|\leqslant|t|\cdot|x_2-x_1|,$$

так что при t, лежащем в конечном интервале $-\overline{a,a}$, рассматриваемое семейство равностепенно непрерывно. Следовательно, $\mathbf{E}_n(u_t) \to \mathbf{E}(u_t)$ равномерно в каждом конечном интервале. ▶

---

[1]) То есть разность $1-F\{-\infty,\infty\}$. — *Прим. перев.*

[2]) Семейство функций *равностепенно непрерывно*, если каждому $\varepsilon>0$ соответствует δ, не зависящее от t и такое, что $|u_t(x_2)-u_t(x_1)|<\varepsilon$ при $|x_2-x_1|<\delta$.

§ 2. Специальные свойства

В соответствии с определением 1 из гл. V, 2 два распределения U и V принадлежат одному и тому же *типу*, если одно от другого отличается только параметрами расположения, т. е. если

$$V(x)=U(Ax+B), \quad A>0. \tag{2.1}$$

Мы покажем теперь, что сходимость распределений сохраняет типы в том смысле, что изменения параметров расположения допредельных распределений не меняют типа предельного распределения. Именно этот факт дает возможность говорить об «асимптотических нормальных последовательностях» без точного указания соответствующих параметров. Более точно, имеет место

Лемма 1. *Пусть U и V — два распределения, ни одно из которых не сосредоточено в единственной точке. Пусть последовательность $\{F_n\}$ распределений вероятностей такова, что при некоторых константах $a_n>0$ и $\alpha_n>0$*

$$F_n(a_nx+b_n)\to U(x), \quad F_n(\alpha_nx+\beta_n)\to V(x) \tag{2.2}$$

во всех точках непрерывности. Тогда

$$\frac{\alpha_n}{a_n}\to A\neq 0, \quad \frac{\beta_n-b_n}{a_n}\to B \tag{2.3}$$

и выполняется (2.1). *Обратно, если верно* (2.3), *то каждое из соотношений* (2.2) *влечет за собой другое и соотношение* (2.1).

[Существенно, что ни одно из распределений не сосредоточено в единственной точке. Например, пусть V сосредоточено в нуле и $F_n=U$ при всех n. Тогда первое из соотношений (2.2) справедливо при $a_n=1$ и $b_n=0$, а второе — при любых $\alpha_n\to\infty$ и ограниченных β_n. Очевидно, что (2.3) не обязано выполняться]

Доказательство. Мы начнем со второго утверждения леммы. Допустим, что выполняется (2.3). Положим $y=Ax+B$. При заданном $\varepsilon>0$ и достаточно больших n имеем

$$F_n(a_n(y-\varepsilon)+b_n)\leqslant F_n(\alpha_nx+\beta_n)\leqslant F_n(a_n(y+\varepsilon)+b_n) \tag{2.4}$$

(так как аргументы удовлетворяют аналогичным неравенствам). Отсюда следует, что первое из соотношений (2.2) влечет второе. Точно так же устанавливается обратное утверждение.

Допустим теперь, что выполняются оба соотношения (2.2). Выберем две точки непрерывности x' и x'' распределения V, так что $x'<x''$ и $0<V(x')\leqslant V(x'')<1$. Тогда найдутся точки непрерывности распределения U, для которых $U(y')<V(x')$ и

$U(y'')>V(x'')$. При достаточно больших n из (2.2) вытекает

$$a_n y'+b_n \leqslant \alpha_n x'+\beta_n \leqslant \alpha_n x''+\beta_n \leqslant a_n y''+b_n.$$

Следовательно, $\alpha_n(x''-x') \leqslant a_n(y''-y')$ и отношения α_n/a_n и $(\beta_n-b_n)/a_n$ остаются ограниченнымн. По соображенням симметрии отношения a_n/α_n также должны быть ограничены. Поэтому можно выбрать последовательность $n_1, n_2, \ldots$, такую, что при n, пробегающем эту последовательность, выполняется (2.3). По второму утверждению леммы отсюда следует (2.1). Так как константы A и B определяются формулой (2.1) однозначно, то ясно, что наши пределы не зависят от снецнального выбора последовательности $\{n_k\}$. Иными словами, (2.3) выполняется нри n, стремящемся к бесконечности любым образом. Доказательство закончено. ▶

Два типа последовательностей $\{F_n\}$ распределений вероятностей встречаются так часто, что заслуживают специального названия. Для большей ясности обозначений мы дадим формальные определения в терминах случайных величин $\mathbf{X}_n$. Однако по существу эти понятия связаны только с соответствующими распределениями F_n. Стало быть, определения имеют смысл без всякого упоминания о каком бы то ни было вероятностном пространстве.

Определение 1. *Если*

$$\mathbf{P}\{|\mathbf{X}_n|>\varepsilon\}\to 0 \tag{2.5}$$

при любом $\varepsilon>0$, *то мы будем говорить, что случайные величины* $\mathbf{X}_n$ *сходятся по вероятности к нулю и будем обозначать это символом* $\mathbf{X}_n \xrightarrow{\mathrm{p}} 0$.

По определению $\mathbf{X}_n \xrightarrow{\mathrm{p}} \mathbf{X}$ *означает то же самое, что* $\mathbf{X}_n-\mathbf{X} \xrightarrow{\mathrm{p}} 0$.

Заметим, что (2.5) выполняется тогда и только тогда, когда распределения F_n сходятся к распределению, сосредоточенному в нуле. В общем случае, однако, сходимость $F_n \to F$ не позволяет сделать никаких заключений относительно сходимости последовательности $\mathbf{X}_1, \mathbf{X}_2, \ldots$. Например, если величины $\mathbf{X}_j$ взаимно независимы и имеют одно и то же распределение F, то $F_n \to F$, но последовательность $\{\mathbf{X}_n\}$ не сходится по вероятности.

Приводимая ниже простая лемма часто используется, хотя это и не всегда отмечается явно. (Так, метод «урезания» из 1, гл. X неявно опирается на эту лемму).

Лемма 2. *Обозначим распределения* $\mathbf{X}_n$ *и* $\mathbf{Y}_n$ *через* F_n *и* G_n *соответственно. Допустим, что* $\mathbf{X}_n-\mathbf{Y}_n \xrightarrow{\mathrm{p}} 0$ *и что* $G_n \to G$. *Тогда и* $F_n \to G$.

В частности, если $\mathbf{X}_n \xrightarrow{\mathrm{P}} \mathbf{X}$, то $F_n \to F$, где F — распределение $\mathbf{X}$.

Доказательство. Если $\mathbf{X}_n \leqslant x$, то или $\mathbf{Y}_n \leqslant x+\varepsilon$, или $\mathbf{X}_n - \mathbf{Y}_n \leqslant -\varepsilon$. Вероятность последнего события стремится к нулю. Следовательно, при всех достаточно больших n $F_n(x) \leqslant G_n(x+\varepsilon)+\varepsilon$. Точно так же устанавливается аналогичная оценка для F_n снизу. ▷

Определение 2. *Последовательности $\{\mathbf{X}_n\}$ и $\{F\}_n$ называются стохастически ограниченными, если для каждого $\varepsilon>0$ найдется такое a, что*

$$\mathbf{P}\{|\mathbf{X}_n|>a\}<\varepsilon \tag{2.6}$$

для всех достаточно больших n.

Это определение применимо и к случайным векторам $\mathbf{X}_n$ и к соответствующим им многомерным распределениям.

Последовательность, сходящаяся в собственном смысле, очевидно, будет стохастически ограниченной. В то же время несобственная сходимость исключает стохастическую ограниченность. Мы получаем, следовательно, простой, но полезный *критерий*: если распределения F_n сходятся, то предел F будет собственным распределением тогда и только тогда, когда последовательность $\{F_n\}$ стохастически ограничена.

Вместе с $\{\mathbf{X}_n\}$ и $\{\mathbf{Y}_n\}$ стохастически ограничена и последовательность $\{\mathbf{X}_n+\mathbf{Y}_n\}$. В самом деле, событие $|\mathbf{X}_n+\mathbf{Y}_n|>2a$ влечет событие: или $|\mathbf{X}_n|>a$, или $|\mathbf{Y}_n|>a$. Поэтому

$$\mathbf{P}\{|\mathbf{X}_n+\mathbf{Y}_n|>2a\}\leqslant\mathbf{P}\{|\mathbf{X}_n|>a\}+\mathbf{P}\{|\mathbf{Y}_n|>a\}. \tag{2.7}$$

Следующая ниже лемма может быть доказана многими способами (см. задачу 4). Она приведена здесь лишь для того, чтобы проиллюстрировать применение предыдущих лемм.

Лемма 3. *Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — независимые случайные величины с одним и тем же распределением F. Пусть $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ и случайные величины $a_n^{-1}\mathbf{S}_n - b_n$ имеют собственное предельное распределение U, не сосредоточенное в одной точке. Если $a_n>0$, то*

$$a_n \to \infty, \quad \frac{a_n}{a_{n-1}} \to 1. \tag{2.8}$$

Доказательство. Если $a_n \to \infty$, то $a_n^{-1}\mathbf{X}_n \xrightarrow{\mathrm{P}} 0$ и по последней лемме распределение $a_n^{-1}\mathbf{S}_{n-1} - b_n$ сходится к U. Следовательно, случайные величины $a_n^{-1}\mathbf{S}_{n-1} - b_n$ и $a_{n-1}^{-1}\mathbf{S}_{n-1} - b_{n-1}$

имеют одно и то же предельное распределение U. По лемме 1 отсюда вытекает, что $\frac{a_n}{a_{n-1}} \to 1$. Докажем, что $a_n \to \infty$. Если бы это было не так, то последовательность $\{a_n^{-1}\mathbf{S}_n - b_n\}$ не могла бы быть стохастически ограниченной. Последнее утверждение, как показывает метод симметризации, достаточно доказать для *симметричных* распределений F и $b_n = 0$. По неравенству V (5.9)

$$\mathbf{P}\{|\mathbf{S}_n| > t\} \geqslant \tfrac{1}{2}\mathbf{P}\{\max[|\mathbf{X}_1|, \ldots, |\mathbf{X}_n|] > t\}. \qquad (2.9)$$

Если распределение F не сосредоточено на конечном интервале, то вероятность в правой части (2.9) стремится к 1 при любом фиксированном $t > 0$. Поэтому последовательность $\{a_n^{-1}\mathbf{S}_n\}$ может быть стохастически ограниченной только при условии, что $a_n \to \infty$. Это рассуждение можно видоизменить так, чтобы оно было применимо и к распределениям, сосредоточенным на конечном интервале. Однако, как мы вскоре увидим, к этим распределениям применима центральная предельная теорема и потому для них по лемме 1 $a_n \sim a\sqrt{n}$. ▶

§ 3. Распределения как операторы

Свертка $U = F * u$ функции точки u и распределения F была определена в гл. V, 4. Вводя семейство функций u_t формулой $u_t(x) = u(t - x)$, мы можем выразить значение $U(t)$ как математическое ожидание:

$$U(t) = \int_{-\infty}^{+\infty} u(t - y)F\{dy\} = \mathbf{E}(u_t). \qquad (3.1)$$

Мы используем эту запись для того, чтобы вывести критерий собственной сходимости. В этом критерии рассматривается класс $C[-\infty, \infty]$ непрерывных функций, имеющих пределы $u(\pm\infty)$. Эти функции равномерно непрерывны.

Теорема 1. *Последовательность распределений вероятностей F_n сходится к распределению F в собственном смысле тогда и только тогда, когда для каждой функции $u \in C[-\infty, \infty]$ свертки $U_n = F_n * u$ равномерно сходятся к некоторому пределу U. В случае сходимости $U = F * u$.*

Доказательство. Условие необходимо, так как равномерная непрерывность u влечет равностепенную непрерывность семейства $\{u_t\}$ и по следствию из § 1 U_n равномерно сходится к $F * u$. Обратно, условие теоремы влечет сходимость математиче-

ских ожиданий $\mathbf{E}_n(u)$. Мы видели в § 1, что отсюда вытекает сходимость $F_n \to F$, нужно только доказать, что распределение F — собственное. Пусть u — непрерывная монотонная функция, возрастающая от 0 до 1. Тогда $U_n(-\infty) = 0$ для каждого n. Заменяя область интегрирования в (3.1) на $-\infty, -2x$, видим, что

$$U_n(-x) \geqslant u(x) F_n(-2x).$$

Так как $U_n \to U$ равномерно, то при всех достаточно больших x отсюда вытекает неравенство $F_n(-2x) < \varepsilon$. Таким образом, $F(-\infty) = 0$. Аналогично доказывается, что $1 - F(\infty) = 0$. ▶

Полезность этого критерия можно иллюстрировать важным примером.

Пример. а) *Теоремы об аппроксимации.* Пусть G — произвольное распределение вероятностей и $G_h(x) = G(h^{-1}x)$. При $h \to 0+$ распределение G_h стремится к распределению, сосредоточенному в нуле, и потому *для каждой функции* $u \in C[-\infty, \infty]$ *имеет место равномерная сходимость* $G_h * u \to u$ [1]).

Если G имеет плотность g, то свертка $G_h * u$ задается формулой

$$G_h * u(t) = \int_{-\infty}^{+\infty} u(y)\, g\left(\frac{t-y}{h}\right) \frac{1}{h}\, dy. \tag{3.2}$$

Если g имеет ограниченную производную, то этим же свойством обладает и G_h и можно дифференцировать (3.2) под знаком интеграла. Беря в качестве g нормальную плотность, мы получим следующую *лемму об аппроксимации: для каждой функции* $u \in C[-\infty, \infty]$ *найдется бесконечно дифференцируемая функция* $v \in C[-\infty, \infty]$, *такая, что для всех* x $|u(x) - v(x)| < \varepsilon$.

Подобные результаты часто используются и за пределами теории вероятностей. Доказательство, однако, всего проще в рассматриваемой ситуации. Типичное применение указано в задаче 10. ▶

В настоящем контексте желательно заменить символ свертки $*$ более простым обозначением, которое подчеркивало бы, что

---

[1]) Это можно проверить и непосредственно, отправляясь от формулы

$$G_h * u(t) - u(t) = \int_{-\infty}^{+\infty} [u(t-y) - u(t)]\, G\{dy/h\}.$$

Предположим $|u(x)| \leqslant 1$. По данному $\varepsilon > 0$ выберем δ так, чтобы колебания u на каждом интервале длины 2δ были меньше ε. Часть интеграла, взятая по множеству $|y| > \delta$, не превосходит выражение $2[1 - G(h^{-1}\delta) + G(-h^{-1}\delta)]$, которое стремится к нулю при $h \to 0$. Поэтому интеграл стремится к нулю равномерно по t.

в (3.1) распределение F играет роль оператора, переводящего u в U. Этот оператор мы будем обозначать буквой $\mathfrak{F}$. Мы условимся, что равенство $U=\mathfrak{F}u$ имеет тот же смысл, что и $U=F*u$. Истинное преимущество этого кажущегося педантичным подхода станет видимым только тогда, когда в том же контексте будут рассматриваться другие типы операторов. Тогда будет удобно видеть сразу, играет ли то или иное распределение свою первоначальную вероятностную роль или же оно служит только аналитическим оператором. После этого объяснения мы вводим

Соглашение об обозначениях. *С каждым распределением F мы связываем оператор $\mathfrak{F}$, отображающий $C[-\infty, \infty]$ на себя. Этот оператор ставит в соответствие функции и ее преобразование $\mathfrak{F}u=F*u$. Всюду, где это возможно, распределения и связанные с ними операторы будут обозначаться соответствующими латинскими и готическими буквами.*

Как обычно, $\mathfrak{F}\mathfrak{G}u$ обозначает результат применения $\mathfrak{F}$ к $\mathfrak{G}u$, так что $\mathfrak{F}\mathfrak{G}$ *обозначает оператор, соответствующий свертке* $F*G$ двух распределений вероятностей. В частности, $\mathfrak{F}^n$ *обозначает оператор, связанный с* F^{n*}, т. е. с n-кратной сверткой F с собой.

Пример. б) Пусть H_a — атомическое распределение, сосредоточенное в точке a. Тогда $\mathfrak{H}_a$ обозначает оператор *сдвига* $\mathfrak{H}_a u(x)=u(x-a)$. В частности $\mathfrak{H}_0$ — *тождественный* оператор: $\mathfrak{H}_0 u=u$. ▶

Примем обозначение [1])

$$\|u\|=\sup|u(x)| \tag{3.3}$$

и укажем легко проверяемое *неравенство треугольника:* $\|u+v\|\leqslant\|u\|+\|v\|$. Тогда соотношение $\|u_n-u\|\to 0$ означает *равномерную* сходимость u_n к u.

Оператор T называют ограниченным, если существует такая константа a, что $\|Tu\|\leqslant a\|u\|$. Наименьшее число с этим свойством называют *нормой* T и обозначают $\|T\|$. В этих обозначениях основные свойства линейных операторов, связанных с функциями распределения, формулируются следующим образом.

Они положительны, т. е. из $u\geqslant 0$ вытекает $\mathfrak{F}u\geqslant 0$. Они имеют норму 1, откуда следует

$$\|\mathfrak{F}u\|\leqslant\|u\|. \tag{3.4}$$

Наконец, они перестановочны, т. е. $\mathfrak{F}\mathfrak{G}=\mathfrak{G}\mathfrak{F}$.

[1]) $\|u\|$ — это *норма* u, как элемента пространства Банаха $C[-\infty, \infty]$. Мы не будем, однако, использовать теорию этих пространств.

Определение[1]). *Пусть* $\mathfrak{F}_n$ *и* $\mathfrak{F}$ *— операторы, связанные с распределениями вероятностей* F_n *и* F. *Мы пишем* $\mathfrak{F}_n \to \mathfrak{F}$ *в том и только том случае, когда*

$$\| \mathfrak{F}_n u - \mathfrak{F} u \| \to 0 \tag{3.5}$$

для каждого $u \in C[-\infty, \infty]$.

Теорему 1 можно теперь сформулировать иначе.

Теорема 1а. $F_n \to F$ *в собственном смысле тогда и только тогда, когда* $\mathfrak{F}_n \to \mathfrak{F}$.

Нижеследующая лемма является основной. Она содержит некоторое неравенство и показывает целесообразность нового обозначения.

Лемма 1. *Для операторов, связанных с распределениями вероятностей, имеем*

$$\| \mathfrak{F}_1 \mathfrak{F}_2 u - \mathfrak{G}_1 \mathfrak{G}_2 u \| \leqslant \| \mathfrak{F}_1 u - \mathfrak{G}_1 u \| + \| \mathfrak{F}_2 u - \mathfrak{G}_2 u \|. \tag{3.6}$$

Доказательство. Оператор в левой части равен $(\mathfrak{F}_1 - \mathfrak{G}_1)\,\mathfrak{F}_2 + (\mathfrak{F}_2 - \mathfrak{G}_2)\,\mathfrak{G}_1$. Поэтому (3.6) вытекает из неравенства треугольника и того факта, что $\mathfrak{F}_2$ и $\mathfrak{G}_1$ имеют нормы не больше 1. Отметим, что доказательство применимо и к несобственным распределениям вероятностей. ▶

Из (3.6) непосредственно следует

Теорема 2. *Пусть последовательности* $\{F_n\}$ *и* $\{G_n\}$ *распределений вероятностей сходятся в собственном смысле к* F *и* G *соответственно. Тогда*

$$F_n * G_n \to F * G. \tag{3.7}$$

(Сходимость является собственной в силу определения свертки $F * G$. Теорема неверна в случае несобственной сходимости: возьмите $F_n = G_n$ с массой $1/2$ в каждой из точек $\pm n$.)

В качестве второго применения докажем, что теорема 1 остается справедливой, если сузить рассматриваемый класс функций u до класса особенно «хороших» функций, имеющих производные всех порядков. На этом пути мы получим более гибкий

---

[1]) В терминологии банаховых пространств (3.5) называется *сильной сходимостью*. Заметим, что отсюда *не* вытекает сходимость в смысле $\| \mathfrak{F}_n - \mathfrak{F} \| \to 0$. Например, если F_n сосредоточено в точке $1/n$ и $\mathfrak{F}$ — тождественный оператор, то $\mathfrak{F}_n u(x) - \mathfrak{F} u(x) = u(x - n^{-1}) - u(x)$ и (3.5) верно. Однако $\| \mathfrak{F}_n - \mathfrak{F} \| = 2$, так как существует функция v с $|v| \leqslant 1$, такая, что $v(0) = 1$ и $v(-n^{-1}) = -1$.

Критерий 1. *Пусть F_n — распределение вероятностей. Если для каждой бесконечно дифференцируемой*[1]) $v \in C[-\infty, \infty]$ *последовательность* $\{\mathfrak{F}_n v\}$ *равномерно сходится, то существует такое распределение вероятностей F, что $F_n \to F$.*

Доказательство. Как было показано в примере (а), для любой $u \in C[-\infty, \infty]$ и любого $\varepsilon > 0$ найдется бесконечно дифференцируемая функция v, такая, что $\|u - v\| < \varepsilon$. По неравенству треугольника

$$\|\mathfrak{F}_n u - \mathfrak{F}_m u\| \leqslant \|\mathfrak{F}_n u - \mathfrak{F}_n v\| + \|\mathfrak{F}_n v - \mathfrak{F}_m v\| + \|\mathfrak{F}_m v - \mathfrak{F}_m u\|. \quad (3.8)$$

Первый и последний члены в правой части меньше ε, а средний член меньше ε при всех достаточно больших n и m. Таким образом, последовательность $\{\mathfrak{F}_n u\}$ сходится равномерно и по теореме 1 $F_n \to F$. ▷

Точно такие же рассуждения [в обозначениях (1.1)] дает

Критерий 2. *Пусть F_n и F — собственные распределения вероятностей. Если для всех бесконечно дифференцируемых и обращающихся в нуль на бесконечности функций v $\mathbf{E}_n(v) \to \mathbf{E}(v)$, то $F_n \to F$.*

Основное неравенство (3.6) по индукции распространяется на свертки более чем двух распределений. Для облегчения ссылок мы выделим этот очевидный результат в виде леммы.

Лемма 2. *Пусть $\mathfrak{A} = \mathfrak{F}_1 \dots \mathfrak{F}_n$ и $\mathfrak{B} = \mathfrak{G}_1 \dots \mathfrak{G}_n$, где $\mathfrak{F}_j$ и $\mathfrak{G}_j$ — операторы, соответствующие распределениям вероятностей. Тогда*

$$\|\mathfrak{A}u - \mathfrak{B}u\| \leqslant \sum_{j=1}^{n} \|\mathfrak{F}_j u - \mathfrak{G}_j u\|. \quad (3.9)$$

В частности,

$$\|\mathfrak{F}^n u - \mathfrak{G}^n u\| \leqslant n \cdot \|\mathfrak{F}u - \mathfrak{G}u\|. \quad (3.10)$$

(Применения см. в задачах 14 и 15.)

§ 4. Центральная предельная теорема

Центральная предельная теорема устанавливает условия, при которых суммы независимых случайных величин распределены асимптотически нормально. Ее роль и значение частично объяснялись в **1**, гл. X, 1. В нескольких случаях мы применяли ее

[1]) Под этим подразумевается, что все производные v существуют и принадлежат $C[-\infty, \infty]$.

[последний раз в примере (11, ж) гл. VI]. Она занимает почетное место в теории вероятностей благодаря своему «возрасту» и той роли, которую она играла в развитии теории вероятностей и все еще играет в применениях. Естественно поэтому использовать центральную предельную теорему для опробования различных методов, имеющихся в нашем распоряжении. По этой причине мы дадим несколько доказательств. Более систематическое изучение (включая необходимые и достаточные условия) будет проведено в гл. IX, XV и XVI. Теперешнее обсуждение отвлекает нас от развития основной темы. Его цель — показать преимущества операторной терминологии ярким и значительным примером. Кроме того, многим читателям понравится легкий путь к доказательству центральной предельной теоремы в простейшей ситуации. Мы начнем со специального случая, хотя это и приведет к некоторым повторениям.

Теорема 1. (*Одинаковые распределения в* $\mathscr{R}^1$.) *Пусть* $\mathbf{X}_1, \mathbf{X}_2, \ldots$ *— взаимно независимые случайные величины с одним и тем же распределением* F. *Допустим, что*

$$\mathbf{E}(\mathbf{X}_k) = 0, \quad \mathrm{Var}(\mathbf{X}_k) = 1. \tag{4.1}$$

При $n \to \infty$ *распределение нормированных сумм*

$$\mathbf{S}_n^* = \frac{\mathbf{X}_1 + \ldots + \mathbf{X}_n}{\sqrt{n}} \tag{4.2}$$

стремится к нормальному распределению $\mathfrak{N}$ *с плотностью* $n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

В чисто аналитических терминах это означает, что для распределения F с нулевым математическим ожиданием и единичной дисперсией

$$F^{n*}(x\sqrt{n}) \to \mathfrak{N}(x). \tag{4.3}$$

Мы покажем, что эта теорема немедленно вытекает из следующей леммы.

Лемма. *Если* $\mathfrak{F}_n$ *— оператор, соответствующий* $F_n(x) = F(x\sqrt{n})$, *то для каждой функции* $u \in C[-\infty, \infty]$, *имеющей три ограниченных производных,*

$$n[\mathfrak{F}_n u - u] \to \frac{1}{2} u'' \tag{4.4}$$

равномерно на всей прямой.

Доказательство *теоремы* 1. Пусть $\mathfrak{G}$ и $\mathfrak{G}_n$ — операторы, связанные с нормальным распределением $\mathfrak{N}$ и с $\mathfrak{N}(x\sqrt{n})$ соответственно. Так как (4.4) выполняется и для $\mathfrak{G}_n$, то по основному неравенству (3.10)

$$\|\mathfrak{F}_n^n u - \mathfrak{G} u\| = \|\mathfrak{F}_n^n u - \mathfrak{G}_n^n u\| \leqslant n\|\mathfrak{F}_n u - \mathfrak{G}_n u\| \to 0. \tag{4.5}$$

По критерию 1 $\mathfrak{F}_n^n \to \mathfrak{G}$. ▶

Доказательство *леммы.* Введем собственное распределение $F_n^{\#}$, определив его равенством

$$F_n^{\#}\{dy\} = ny^2 F_n\{dy\} = ny^2 F\{\sqrt{n}\,dy\}. \tag{4.6}$$

Замена переменных $\sqrt{n}y = s$ показывает, что $F_n^{\#}$ сходится к распределению, сосредоточенному в нуле. Ввиду (4.1) мы можем представить разность левой и правой частей (4.4) в виде

$$n[\mathfrak{F}_n u(x) - u(x)] - \frac{1}{2}u''(x) = \int_{-\infty}^{+\infty}\left[\frac{u(x-y) - u(x) + yu'(x)}{y^2} - \frac{1}{2}u''(x)\right]F_n^{\#}\{dy\}. \tag{4.7}$$

Из формулы Тейлора, примененной к числителю дроби под интегралом, видно, что при $|y|<\varepsilon$ подинтегральное выражение не превосходит $\frac{1}{6}|y|\cdot\|u'''\|<\varepsilon\|u'''\|$, а при всех y не превосходит $\|u''\|$. Так как $F_n^{\#}$ сходится к распределению, сосредоточенному в нуле, то при всех достаточно больших n абсолютная величина интеграла не превосходит $\varepsilon(\|u''\|+\|u'''\|)$. Таким образом, левая часть (4.7) равномерно стремится к нулю. ▶

Пример. а) *Центральная предельная теорема для случая бесконечной дисперсии.* С методологической точки зрения интересно отметить, что доказательство теоремы 1 применимо без изменений к некоторым распределениям с бесконечной дисперсией (в предположении, что нормирующие константы выбраны надлежащим образом). Например, если случайные величины $\mathbf{X}_k$ имеют плотность $f(x)$, такую, что $f(x) = 2|x|^{-3}\log|x|$ при $|x|\geqslant 1$ и $f(x)=0$ для $|x|\leqslant 1$, то суммы $(\mathbf{X}_1+\ldots+\mathbf{X}_n)/(\sqrt{n}\log n)$ имеют в пределе нормальное распределение (доказательство требует лишь очевидных видоизменений). Необходимые и достаточные условия сходимости к нормальному распределению указаны в гл. IX, 7. ▶

Приведенный выше метод доказательства получил широкое распространение. Хорошее упражнение предлагается в задаче 16. Здесь мы используем этот метод для доказательства центральной предельной теоремы в более общем случае. Сформулированная ниже теорема касается двумерного случая, но ее можно распространить и на $\mathscr{R}^r$.

Теорема 2. (*Многомерный случай.*) *Пусть* $\{\mathbf{X}_n\}$ — *последовательность взаимно независимых двумерных случайных величин с одним и тем же распределением* F. *Допустим, что математические ожидания равны нулю и что матрица ковариаций равна*

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \tag{4.8}$$

При $n \to \infty$ *распределение* $(\mathbf{X}_1 + \ldots + \mathbf{X}_n)/\sqrt{n}$ *сходится к двумерному нормальному распределению с нулевым математическим ожиданием и матрицей ковариаций* C.

Доказательство. Если использовать матричные обозначения гл III, 5, то почти без изменений применимо предыдущее доказательство. Так как нижние индексы уже заняты, то мы будем обозначать точки плоскости с помощью векторов-строк $x = (x^{(1)}, x^{(2)})$. В этих обозначениях $u(x)$ — функция двух переменных. Ее частные производные будут помечаться нижними индексами. Так $u' = (u_1, u_2)$ — вектор-строка и $u'' = (u_{jk})$ — симметричная (2×2)-матрица. Формула Тейлора принимает вид

$$u(x-y) = u(x) - yu'(x) + \frac{1}{2}\, yu''(x)\, y^T + \ldots, \tag{4.9}$$

где y^T получается транспонированием y, т. е. y^T — вектор-столбец с компонентами $y^{(1)}$ и $y^{(2)}$. Определим по аналогии с (4.6) собственное распределение вероятностей формулой

$$F_n^{\sharp}\{dy\} = nq(y)\, F\{\sqrt{n}\, dy\}, \quad \text{где} \quad 2q(y) = \frac{y_1^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2}.$$

Так же как и в предыдущем доказательстве, $F_n^{\sharp}$ сходится к распределению, сосредоточенному в начале координат. Соотношение (4.7) заменяется на

$$n[\mathfrak{S}_n u(x) - u(x)] - \frac{1}{2}\, m(x) = \int_{\mathscr{R}^2} \frac{u(x-y) - u(x) + yu'(x) - yu''(x)\, y^T}{q(y)}\, F_n^{\sharp}\{dy\}, \tag{4.10}$$

где [1])

$$m(x) = \mathbf{E}(yu''(x)\,y^T) = u_{11}(x)\,\sigma_1^2 + 2u_{12}(x)\,\rho\sigma_1\sigma_2 + u_{22}(x)\,\sigma_2^2. \quad (4.11)$$

(Здесь $\mathbf{E}$ обозначает математическое ожидание по отношению к F.) В силу (4.9) подинтегральное выражение стремится к нулю и так же, как и в предыдущей лемме, мы заключаем, что $n[\mathfrak{H}_n u - u] \to \frac{1}{2}\,m$ равномерно. Доказательство теоремы не требует изменений. ▶

Примеры. б) *Случайное блуждание в d-мерном пространстве.* Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — независимые случайные векторы с одним и тем же распределением, которое может быть описано следующим образом. Направление $\mathbf{X}_k$ случайно (I, 10), а длина $\mathbf{L}$ — случайная величина с $\mathbf{E}(\mathbf{L}^2) = 1$. По причинам симметрии матрица ковариаций C диагональна, с диагональными элементами $\sigma_j^2 = 1/d$. Распределение нормированной суммы $\mathbf{S}_n/\sqrt{n}$ сходится к нормальному распределению с матрицей ковариаций C. Распределение *длины вектора* $\mathbf{S}_n/\sqrt{n}$ сходится поэтому к распределению квадратного корня из суммы квадратов независимых нормальных величин. Как было показано в гл. II, 3, последнее распределение имеет плотность

$$w_d(r) = \frac{d^{\frac{1}{2}d}}{2^{\frac{1}{2}d-1}\,\Gamma\left(\frac{1}{2}d\right)}\,e^{-\frac{1}{2}dr^2}\,r^{d-1}. \quad (4.12)$$

Этот результат показывает влияние числа измерений; он применим, в частности, к примеру (10, д) гл. I.

в) *Случайное рассеивание популяций.* В качестве практического применения предыдущего примера рассмотрим распространение популяции дубов в доисторическое время. Если бы новые растения возникали только из семян, упавших с материнского дерева, то сеянцы были бы расположены около него. Тогда расстояние дерева n-го поколения от исходного было бы распределено приблизительно нормально. При этих условиях площадь, покрытая потомками некоторого дерева, была бы, грубо говоря, пропорциональна возрасту дерева. Наблюдения показывают, что действительное положение вещей несовместимо с этой гипотезой. Биологи приходят к выводу, что наблюдаемое

[1]) Очевидно, что $m(x)$ — след матрицы Cu'' (т. е. сумма ее диагональных элементов). В одномерном случае $m(x) = \sigma^2 u''(x)$.

рассеивание в большой степени определяется птицами, переносящими семена на большие расстояния [1]).

Обобщим теперь теорему 1 на случай разно распределенных величин. Условия могут произвести впечатление, что они введены искусственно с единственной целью сохранить путь доказательства. В действительности оказывается, что эти условия необходимы и для центральной предельной теоремы с классической нормировкой, использованной в (4.17) (см. гл. XV, 6).

Теорема 3. (*Линдеберг.*) [2]) *Пусть* $\mathbf{X}_1, \mathbf{X}_2, \ldots$ *взаимно независимые одномерные случайные величины с распределениями* $F_1, F_2, \ldots$. *Пусть*

$$\mathbf{E}(\mathbf{X}_k) = 0, \quad \mathrm{Var}(\mathbf{X}_k) = \sigma_k^2 \tag{4.13}$$

и

$$s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2. \tag{4.14}$$

Предположим, что при каждом $t > 0$

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|y| \geqslant ts_n} y^2 F_k\{dy\} \to 0, \tag{4.15}$$

или, что то же самое,

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|y| < ts_n} y^2 F_k\{dy\} \to 1. \tag{4.16}$$

Тогда распределение нормированной суммы

$$\mathbf{S}_n^* = \frac{\mathbf{X}_1 + \ldots + \mathbf{X}_n}{s_n} \tag{4.17}$$

сходится к нормальному распределению $\mathfrak{N}$ *с нулевым математическим ожиданием и единичной дисперсией.*

Условие Линдеберга (4.15) гарантирует, что дисперсии σ_k^2 отдельных слагаемых малы по сравнению с суммой s_n^2 в том

---

[1]) Skellam J. G., *Biometrika,* **38** (1951), 196—218.

[2]) Lindeberg J. W., *Math Zeit.,* **15** (1922), 211—235. Частные случаи были известны ранее, но Линдеберг дал первую общую формулировку, содержащую теорему 1. Необходимость условия Линдеберга при классической нормировке была доказана Феллером (там же, **40** (1935), см. гл. XV, 6).

Метод Линдеберга казался сложным и был практически заменен методом характеристических функций, развитым П. Леви. Троттер (*Archiv der Math.,* **9** (1959), 226—234) показал, что современные методы позволяют изложить доказательство Линдеберга простым и интуитивно понятным способом. Доказательства этого параграфа используют идеи Троттера.

смысле, что при любом $\varepsilon>0$ и всех достаточно больших n

$$\frac{\sigma_k}{s_n}<\varepsilon, \quad k=1, \ldots, n. \tag{4.18}$$

В самом деле, очевидно, что σ_k^2/s_n^2 меньше, чем t^2 плюс левая часть (4.15). Положив $t=\frac{1}{2}\varepsilon$, видим, что (4.15) влечет (4.18). Теорема 3 обобщается на многомерный случай способом, указанным в теореме 2. См. также задачи 17—20.

Доказательство. Поставим в соответствие каждому распределению F_k нормальное распределение G_k с нулевым математическим ожиданием и той же самой дисперсией σ_k^2. Распределение $F_k(xs_n)$ случайной величины $\mathbf{X}_k/s_n$ зависит и от k и от n. Соответствующий оператор мы обозначим $\mathfrak{F}_{k,n}$. Аналогично оператор $\mathfrak{G}_{k,n}$ соответствует нормальному распределению $G_k(xs_n)$. В силу (3.9) нам достаточно доказать, что

$$\sum_{k=1}^{n}\|\mathfrak{F}_{k,n}u-\mathfrak{G}_{k,n}u\|\to 0 \tag{4.19}$$

для каждой функции $u\in C[-\infty, \infty]$ с тремя ограниченными производными. Мы поступим так же, как при доказательстве теоремы 1. Однако теперь (4.7) заменится n соотношениями

$$\mathfrak{F}_{k,n}u(x)-u(x)-\frac{\sigma_k^2}{2s_n^2}u''(x)=\int_{-\infty}^{+\infty}\left[\frac{u(x-y)-u(x)+yu'(x)}{y^2}-\right.$$
$$\left.-\frac{1}{2}u''(x)\right]\cdot y^2F_k\{s_n\,dy\}. \tag{4.20}$$

Разбивая область интегрирования на части $|y|\leqslant\varepsilon$ и $|y|>\varepsilon$ и используя те же самые оценки, что и в (4.7), получаем

$$\left\|\mathfrak{F}_{k,n}u-u-\frac{\sigma_k^2}{2s_n^2}u''\right\|\leqslant\varepsilon\|u'''\|\frac{\sigma_k^2}{s_n^2}+$$
$$+\|u''\|\cdot\int_{|y|>\varepsilon}y^2F_k\{s_n\,dy\}. \tag{4.21}$$

Условие Линдеберга (4.15) с $t=\varepsilon$ гарантирует теперь, что при всех достаточно больших n

$$\sum_{k=1}^{n}\left\|\mathfrak{F}_{k,n}u-u-\frac{\sigma_k^2}{2s_n^2}u''\right\|\leqslant\varepsilon(\|u'''\|+\|u''\|). \tag{4.22}$$

Для нормальных распределений G_k условие Линдеберга (4.15) является простым следствием (4.18). Поэтому неравенство

(4.22) остается верным, если в нем $\mathfrak{S}_{k,n}$ заменить на $\mathfrak{S}_{k,n}$. Соединяя оба эти неравенства, получаем (4.19). Доказательство закончено. ▷

Примеры. г) *Равномерные распределения.* Пусть $\mathbf{X}_k$ распределено равномерно, т. е. с плотностью $1/(2a_k)$, на промежутке между $-a_k$ и a_k. Тогда $\sigma_k^2 = \frac{1}{3}a_k^2$. Легко видеть, что условия теоремы выполнены, если числа a_k ограничены, а $a_1^2 + \ldots + a_n^2 \to \infty$. В самом деле, в этом случае сумма (4.15) равна нулю при всех достаточно больших n. С другой стороны, если $\sum a_k^2 < \infty$, то s_n остаются ограниченными и (4.15) не может выполняться. В этом случае центральная предельная теорема *не применима* (вместо этого мы приходим к бесконечным сверткам, которые будут изучены в § 5).

Менее очевидный случай, где центральная предельная теорема *не применима*, — это случай $a_k^2 = 2^k$. Тогда $3s_n^2 = 2^{n+1} - 2 < 2a_n^2$ и, очевидно, левая часть (4.15) больше $^1/_2$ для $t < {}^1/_{100}$. Эти примеры показывают, что назначение условия (4.15) — обеспечить асимптотическую пренебрегаемость отдельных $\mathbf{X}_k$; вероятность того, что *какое-нибудь* из слагаемых $\mathbf{X}_k$ будет того же порядка, что и сумма $\mathbf{S}_n$, должна стремиться к нулю.

д) *Ограниченные величины.* Допустим, что $\mathbf{X}_k$ равномерно ограничены, т. е. что все распределения F_k сосредоточены на некотором конечном интервале $-a, a$ Условие Линдеберга (4.15) выполняется тогда и только тогда, когда $s_n \to \infty$. ▷

Методологически интересно заметить, что изложенный метод доказательства применим и к некоторым последовательностям случайных величин, *не имеющих математических ожиданий.* Конечно, нормирующие константы должны быть другими. Мы вернемся к этим вопросам в гл. XV, 6, где мы также подвергнем дальнейшему изучению природу условия Линдеберга. (См. задачи 19—20).

Мы закончим настоящий экскурс вариантом центральной предельной теоремы для *случайных сумм.* Идея состоит в следующем. Если в теореме 1 заменить число n слагаемых пуассоновской случайной величиной $\mathbf{N}$ с математическим ожиданием n, то кажется правдоподобным, что распределение $\mathbf{S}_{\mathbf{N}}$ будет асимптотически нормально. Сходные ситуации возникают в статистике и физике, когда число наблюдений не фиксируется заранее.

Мы рассмотрим только суммы вида $\mathbf{S}_{\mathbf{N}} = \mathbf{X}_1 + \ldots + \mathbf{X}_{\mathbf{N}}$, где $\mathbf{X}_j$ и $\mathbf{N}$ — взаимно независимые случайные величины. Мы пред-

положим, что $\mathbf{X}_j$ имеют одно и то же распределение F с нулевым математическим ожиданием и дисперсией 1. Используя обозначение § 2, мы покажем, что верна

Теорема 4[1]). (*Случайные суммы.*) *Пусть* $\mathbf{N}_1, \mathbf{N}_2, \ldots$ — *положительные целочисленные случайные величины, такие, что*

$$\frac{\mathbf{N}_n}{n} \xrightarrow{\text{p}} 1. \tag{4.23}$$

Тогда распределение $\mathbf{S}_{\mathbf{N}_n}/\sqrt{n}$ *сходится к* $\mathfrak{N}$.

Интересно отметить, что величина $\frac{1}{\sqrt{n}}\mathbf{S}_{\mathbf{N}_n}$ *не* обязана иметь единичную дисперсию. Действительно, теорема применима к случаям, где $\mathbf{E}(\mathbf{N}_n)=\infty$, и даже если математическое ожидание существует, из (4.23) *не* следует сходимость $\frac{1}{n}\mathbf{E}(\mathbf{N}_n)\to 1$. Нормировка, приводящая к единичной дисперсии, может быть невозможна, а если она и возможна, то усложняет доказательство.

Доказательство. Чтобы избежать двойных индексов, положим $\mathbf{P}\{\mathbf{N}_n=k\}=a_k$, подразумевая, что a_k зависит от n. Оператор, связанный с $\mathbf{S}_{\mathbf{N}_n}$, задается формальным степенным рядом $\sum a_k\mathfrak{F}^k$. Пусть $\mathfrak{F}$, как и в доказательстве теоремы 1, — оператор, соответствующий $F(x\sqrt{n})$. Так как $F_n^{n*}\to\mathfrak{N}$, то достаточно доказать, что

$$\sum_{k=1}^{\infty} a_k\mathfrak{F}_n^k u - \mathfrak{F}_n^n u \to 0 \tag{4.24}$$

равномерно для каждой функции $u\in C[-\infty,\infty]$ с тремя ограниченными производными.

Используя очевидное разложение на множители и основное неравенство (3.10), видим, что

$$\|\mathfrak{F}_n^k u - \mathfrak{F}_n^n u\| \leqslant \|\mathfrak{F}_n^{|k-n|}u - u\| \leqslant |k-n|\cdot\|\mathfrak{F}_n u - u\|. \tag{4.25}$$

В силу (4.23) сумма коэффициентов a_k с $|k-n|>\varepsilon n$ меньше ε при всех достаточно больших n. Для этих n норма левой части

[1]) Обобщение на зависимые $\mathbf{X}_j$ *см* у Billingsley P., Limit theorems for randomly selected partial sums, *Ann. Math. Statist.*, **33** (1963), 85—92. Если отбросить (4.23), то получаются предельные теоремы нового типа, см. Robbins H. E., The asymptotic distribution of the sum of a random number of random variables, *Bull. Amer. Math. Soc.*, **54** (1948), 1151—1161.

Обобщения центральной предельной теоремы на другие типы зависимых величин читатель может найти в книге М. Лоэва (о цепях Маркова см. **1**, гл. XV, 11; о симметрично зависимых величинах см. задачу 21).

(4.24) не превосходит

$$\sum_{k=1}^{\infty} a_k \| \mathfrak{F}_n^k u - \mathfrak{F}_n^n u \| \leqslant 2\varepsilon \cdot \| u \| + 2\varepsilon \cdot n \| \mathfrak{F}_n u - u \|. \qquad (4.26)$$

По лемме правая часть не превосходит $2\varepsilon\|u\| + 3\varepsilon\|u''\|$ при всех достаточно больших n, так что (4.24) выполняется в смысле равномерной сходимости. ▶

§ 5 *. Бесконечные свертки

Нижеследующие теоремы даны как по причине их внутреннего интереса, так и потому, что они хорошо иллюстрируют применение наших критериев. Более сильные варианты можно найти в гл. XI, 9 и XVII, 10.

Мы обозначаем через $\mathbf{X}_1, \mathbf{X}_2, \ldots$ взаимно независимые случайные величины с распределениями $F_1, F_2, \ldots$. Предполагается, что $\mathbf{E}(\mathbf{X}_j) = 0$ и что $\sigma_k^2 = \mathbf{E}(\mathbf{X}_k^2)$ существует.

Теорема. *Если* $\sigma^2 = \sum \sigma_k^2 < \infty$, *то распределения*[1]) G_n *частных сумм* $\mathbf{X}_1 + \ldots + \mathbf{X}_n$ *сходятся к распределению вероятностей* G *с нулевым математическим ожиданием и дисперсией* σ^2.

Доказательство. Чтобы установить существование собственного предела G, достаточно (по теореме 1 § 3) показать, что для любой бесконечно дифференцируемой функции $u \in C[-\infty, \infty]$ последовательность функций $\mathfrak{F}_1\mathfrak{F}_2 \ldots \mathfrak{F}_n u$ сходится равномерно при $n \to \infty$. Теперь при $n > m$ очевидное разложение на множители дает

$$\| \mathfrak{F}_1 \ldots \mathfrak{F}_n u - \mathfrak{F}_1 \ldots \mathfrak{F}_m u \| \leqslant \| \mathfrak{F}_{m+1} \ldots \mathfrak{F}_n u - u \|. \qquad (5.1)$$

Так как $\mathbf{E}(\mathbf{X}_k) = 0$, то верно тождество

$$\mathfrak{F}_k u(x) - u(x) = \int_{-\infty}^{+\infty} [u(x-y) - u(x) + yu'(x)] F_k\{dy\}. \qquad (5.2)$$

Из формулы Тейлора видим, что подинтегральное выражение не превосходит $\|u''\| y^2$. Поэтому $\| \mathfrak{F}_k u - u \| \leqslant \sigma_k^2 \| u'' \|$. По основному неравенству (3.9) величина (5.1) не превосходит $(\sigma_{m+1}^2 + \ldots + \sigma_n^2) \cdot \| u'' \|$. Следовательно, существует собственное распределение G, такое, что $G_n \to G$. Так как G_n имеет диспер-

*) Этот параграф не используется в дальнейшем.

[1]) Можно использовать здесь теорему о сходимости мартингалов (теорема 1 гл. VII, 8) и показать, что и *сами случайные величины* $\mathbf{S}_n$ *сходятся к пределу.*

сию $\sigma_1^2 + \ldots + \sigma_n^2$, то второй момент G существует и не больше σ^2. По критерию примера (1, д) отсюда вытекает, что распределение G имеет математическое ожидание, равное нулю. Наконец, G есть свертка G_n и предельного распределения для суммы $\mathbf{X}_{n+1} + \ldots + \mathbf{X}_{n+k}$. Следовательно, дисперсия G не может быть меньше дисперсии S_n. Этим заканчивается доказательство. ▶

Примеры. а) В примере 1, (11, в) случайный выбор точки между 0 и 1 осуществляется посредством последовательных бросаний монеты. В терминологии настоящего параграфа это означает представление равномерного распределения в виде бесконечной свертки. Пример 1 (11, г) показывает, что бесконечная свертка соответствующих распределений с четными номерами есть сингулярное распределение.

б) Пусть $\mathbf{Y}_k$ — независимые случайные величины с $\mathbf{E}(\mathbf{Y}_k) = 0$ и $\mathbf{E}(\mathbf{Y}_k^2) = 1$. Тогда распределения частных сумм ряда $\sum b_k \mathbf{Y}_k$ сходятся, если $\sum b_k^2 < \infty$. Этот факт был использован при обсуждении свойств нормальных стохастических процессов в гл. III, 7.

в) *Применение к процессам размножения.* Пусть $\mathbf{X}_n$ — положительная случайная величина с плотностью $\lambda_n e^{-\lambda_n t}$. Тогда $\mathbf{E}(\mathbf{X}_n) = [\mathrm{Var}(\mathbf{X}_n)]^{1/2} = \lambda_n^{-1}$, и в случае, когда $m = \sum \lambda_n^{-1} < \infty$, наша теорема применима к центрированным случайным величинам $\mathbf{X}_n - \lambda_n^{-1}$. Это замечание приводит к вероятностной интерпретации расходящегося *процесса чистого размножения* (описанного в 1, гл. XVII, 3—4). «Частица» движется последовательными скачками. Времена пребывания в отдельных состояниях, т. е. $\mathbf{X}_1, \mathbf{X}_2, \ldots,$ — независимые случайные величины, распределенные по показательному закону. Сумма $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ представляет собой момент n-го скачка. Если $\lim \mathbf{E}(\mathbf{S}_n) = m < \infty$, то распределение $\mathbf{S}_n$ сходится к собственному пределу G. Тогда величина $G(t)$ равна вероятности того, что до момента t произойдет бесконечно много скачков.

г) Применение к флуктуационным шумам, телефонным вызовам и тому подобное см. в задаче 22. ▶

§ 6. Теоремы о выборе

Один из стандартных методов доказательства сходимости числовой последовательности состоит в том, что сначала устанавливается существование хотя бы одной предельной точки, а затем — ее единственность. Аналогичный способ применим и к распределениям. Аналог утверждения о существовании

предельной точки дается следующей важной теоремой, приписываемой обычно Хелли. Как и все теоремы этого параграфа, она верна в пространстве любого числа измерений (специальный случай был использован в **1**, гл. XI, 6).

Теорема 1. (i) *Каждая последовательность $\{F_k\}$ распределений вероятностей в $\mathscr{R}^r$ содержит подпоследовательность $F_{n_1}, F_{n_2}, \ldots$, сходящуюся (в собственном или несобственном смысле) к некоторому пределу F.*

(ii) *Для того чтобы все эти пределы были собственными, необходимо и достаточно, чтобы последовательность $\{F_n\}$ была стохастически ограниченной (см. определение 2.2).*

(iii) *Для того чтобы $F_n \to F$, необходимо и достаточно, чтобы предел любой сходящейся подпоследовательности был равен F.*

Доказательство основано на следующей лемме.

Лемма. *Пусть $a_1, a_2, \ldots$ — произвольная последовательность точек. Каждая последовательность $\{u_n\}$ числовых функций содержит подпоследовательность $u_{n_1}, u_{n_2}, \ldots$, сходящуюся во всех точках a_j (может быть, к $\pm\infty$).*

Доказательство. Мы используем «диагональный метод» Кантора. Найдется последовательность номеров $\nu_1, \nu_2, \ldots$, такая, что последовательность значений $u_{\nu_k}(a_1)$ сходится. Чтобы избежать сложных индексов, положим $u_k^{(1)} = u_{\nu_k}$, так что $\{u_k^{(1)}\}$ — подпоследовательность $\{u_n\}$, сходящаяся в точке a_1. Из этой подпоследовательности выделим другую, сходящуюся в точке a^2. Продолжая по индукции, мы построим при каждом n последовательность $u_1^{(n)}, u_2^{(n)}, \ldots$, сходящуюся в точке a_n и содержащуюся в предыдущей последовательности. Рассмотрим теперь «диагональную» последовательность $u_1^{(1)}, u_2^{(2)}, u_3^{(3)}, \ldots$. Эта последовательность, за исключением ее $n-1$ первых членов, содержится в n-й последовательности $u_1^{(n)}, u_2^{(n)}, \ldots$ и потому сходится в a_n. Так как это верно при каждом n, то диагональная последовательность $\{u_n^{(n)}\}$ сходится во всех точках $a_1, a_2, \ldots$. Лемма доказана. ▶

Доказательство *теоремы* 1. Выберем в качестве $\{a_j\}$ какую-либо всюду плотную последовательность. По предыдущей лемме найдется подпоследовательность $\{F_{n_k}\}$ сходящаяся во всех точках a_j. Обозначим ее предел в точке a_j через $G(a_j)$. Мы можем доопределить функцию G при всех x, полагая $G(x)$ равным точной нижней грани значений $G(a_j)$ для $a_j > x$. Тогда G монотонна и $0 \leqslant G(x) \leqslant 1$ при всех x. Если x — точка непрерывности G, то найдутся a_i и a_j, такие, что $a_i < x < a_j$ и $G(a_j)$ —

$-G(a_i)<\varepsilon$. Тогда $F_{n_k}(a_i)\leqslant F_{n_k}(x)\leqslant F_{n_k}(a_j)$ и при $k\to\infty$ крайние члены стремятся к $G(a_i)$ и $G(a_j)$ соответственно. Поэтому верхний и нижний пределы последовательности $\{F_{n_k}(x)\}$ отличаются один от другого не более чем на ε. Следовательно, $F_{n_k}(x)\to G(x)$ во всех точках непрерывности G. Этим доказан пункт (i).

Напомним теперь, что сходящаяся последовательность распределений сходится собственно в том и только том случае, когда она стохастически ограничена. Коль скоро (i) установлено, другие утверждения почти очевидны. ▶

Теорема о выборе является крайне важной. Приводимая ниже известная теорема теории чисел может дать представление об ее удивительной силе. Она также служит напоминанием о том, что вероятностная терминология не должна затемнять более широкий характер развиваемой сейчас теории.

Примеры. а) *Теоретико-числовая теорема о равномерном распределении дробных долей* [1]). *Пусть α — иррациональное число и α_n — дробная часть $n\alpha$. Обозначим через $N_n(x)$ число членов среди $\alpha_1, \alpha_2, \ldots, \alpha_n$, не превосходящих x. Тогда $n^{-1}N_n(x)\to x$ при всех $0<x<1$.*

Доказательство. Мы будем рассматривать распределения и функции, заданные на *окружности* единичной длины. Иными словами, координаты складываются по mod 1. Эта идея уже была объяснена в гл. II, 8. (Заметим, что такое удобное понятие, как функция распределения, теряет на окружности смысл, однако понятие распределения как меры сохраняется.) Пусть F_n — атомическое распределение, сосредоточенное в n-точках $\alpha, 2\alpha, \ldots, n\alpha$ и приписывающее каждой из них вероятность $1/n$. По теореме о выборе существует последовательность номеров $n_1, n_2, \ldots$, такая, что $F_{n_k}\to F$, где F — *собственное* распределение вероятностей (так как окружность — ограниченное множество). Беря свертку F_{n_k} с произвольной непрерывной функцией u, получим

$$\frac{1}{n_k}[u(x-\alpha)+u(x-2\alpha)+\ \ldots\ +u(x-n_k\alpha)]\to v(x). \quad (6.1)$$

Ясно, что замена x на $x-\alpha$ не меняет асимптотического поведения левой части. Следовательно, $v(x)=v(x-\alpha)$ для всех x.

[1]) Эту теорему обычно приписывают Г. Вейлю, хотя она открыта независимо друг от друга Болем (Bohl) и Серпинским (Sierpiński). Из книги Hardy G. H., Wright E. M., Theory of numbers, Oxford, 1945, 378—381, можно понять трудности первоначального доказательства.

Отсюда вытекает в свою очередь, что $v(x)=v(x-k\alpha)$ для $k=1,2,\ldots$. По следствию леммы гл. V (4.2) точки $\alpha, 2\alpha, \ldots$ образуют всюду плотное множество. Поэтому $v=\text{const}$. Мы показали, таким образом, что для любой непрерывной функции u свертка $F*u$ является константой. Отсюда следует, что распределение F должно приписывать одну и ту же вероятность интервалам равной длины, т. е. $F(I)$ должно быть равно длине интервала I. Так как никакие другие предельные распределения существовать не могут, то вся последовательность $\{F_n\}$ сходится к этому распределению, что доказывает теорему. Мы назовем распределение F *равномерным распределением* на окружности.

б) *Сходимость моментов.* Пусть F_n и F — распределения, имеющие конечные моменты всех порядков, которые мы обозначим $\mu_k^{(n)}$ и μ^k соответственно. Как мы знаем из гл. VI, 3, различные функции распределения могут иметь одну и ту же последовательность моментов. Поэтому мы не всегда можем сделать по поведению $\mu_k^{(n)}$ вывод, что $F_n \to F$. Однако *если F — единственное распределение, имеющее моменты $\mu_1, \mu_2, \ldots$ и если $\mu_k^{(n)} \to \mu^k$ при $k=1,2,\ldots$, то $F_n \to F$.* В самом деле, результат примера (1, д) показывает, что всякая сходящаяся подпоследовательность $\{F_n\}$ имеет предел F.

в) *Сепарабельность.* Назовем для краткости распределение рациональным, если оно сосредоточено в конечном числе рациональных точек и приписывает каждой из них рациональный вес. Любое распределение F можно представить в виде предела последовательности $\{F_n\}$ рациональных распределений. Мы можем считать, что математическое ожидание F_n равно нулю, так как этого всегда можно добиться, добавляя новый атом и сколь угодно мало изменяя остальные веса. Но множество всех рациональных распределений счетно и его можно записать в форме последовательности $G_1, G_2, \ldots$. Таким образом, *существует последовательность $\{G_n\}$ распределений с нулевыми математическими ожиданиями и конечными дисперсиями, такая, что каждое распределение F является пределом некоторой подпоследовательности $\{G_{n_k}\}$.*

г) Доказательство того, что $F^{n*} \to 0$, намеченное в задаче 23, использует теорему о выборе. ▶

Теореме 1 была придана форма, наиболее полезная для теории вероятностей. Но эта форма излишне ограничена. Доказательство опиралось на тот факт, что последовательность $\{F_n\}$

монотонных функций с $F_n(-\infty)=0$, $F_n(\infty)=1$ содержит сходящуюся подпоследовательность. Но это утверждение остается верным, если условие $F_n(\infty)=1$ заменить менее жестким условием, а именно, что при любом x числовая последовательность $\{F_n(x)\}$ ограничена. Предел F будет при этом конечным, но он не обязан быть ограниченным. Индуцированная функцией F мера будет принимать конечные значения на интервалах $-\infty, x$, но может быть бесконечна на $-\infty, \infty$. Аналогично можно изменить условия на $-\infty$. В результате мы приходим к следующему обобщению теоремы 1, в котором символ $\mu_n \to \mu$ употребляется в том смысле, что это соотношение верно для всех конечных интервалов.

Теорема 2. *Пусть* $\{\mu_n\}$ *— последовательность мер, такая, что при каждом* x *числовая последовательность* $\mu_n\{-x, x\}$ *ограничена. Тогда существует мера* μ *и последовательность* $n_1, n_2, \ldots, \ldots$, *такие, что* $\mu_{n_k} \to \mu$.

Аналоги теоремы о выборе имеют место для многих классов функций. Особенно полезна следующая теорема (называемая обычно или теоремой Асколи, или теоремой Арцела).

Теорема 3. *Пусть* $\{u_n\}$ *— последовательность равностепенно непрерывных функций*[1]) *и* $|u_n| \leqslant 1$. *Тогда найдется подпоследовательность* $\{u_{n_k}\}$, *сходящаяся к непрерывному пределу* u. *Эта сходимость равномерна в каждом конечном интервале.*

Доказательство. Выберем снова всюду плотную последовательность точек a_j и подпоследовательность $\{u_{n_k}\}$, сходящуюся в точках $a_1, a_2, \ldots$. Обозначим ее предел в точке a_j через $u(a_j)$. Возьмем теперь произвольную точку x. По определению равностепенной непрерывности найдется точка a_j, такая, что $|u_n(x) - u_n(a_j)| < \varepsilon$ для всех n. Поэтому верхний и нижний пределы последовательности $\{u_{n_k}(x)\}$ отличаются один от другого не более чем на 2ε. Так как ε можно взять сколь угодно малым, отсюда следует, что последовательность $\{u_{n_k}(x)\}$ сходится к пределу $u(x)$. Более того, $|u_{n_k}(x) - u(x)| < 4\varepsilon$ для всех k, таких, что $|u_{n_k}(a_j) - u(a_j)| < \varepsilon$. Выбирая при любых данных $\varepsilon > 0$ и замкнутом интервале I надлежащее конечное множество точек a_j, мы видим, что сходимость равномерна в I. Отсюда вытекает непрерывность предела u. ▶

[1]) Для каждого $\varepsilon > 0$ найдется такое $\delta > 0$, что из $|x' - x''| < \delta$ вытекает $|u_n(x') - u_n(x'')| < \varepsilon$ для всех n.

§ 7*. Эргодические теоремы для цепей Маркова

Пусть K — стохастическое ядро, сосредоточенное на конечном или бесконечном интервале Ω (в соответствии с определением 1 гл. VI, 11 это означает, что K есть функция двух переменных, точки x и множества Γ; при фиксированном Γ K — бэровская функция x, а при фиксированном $x \in \Omega$ — распределение вероятностей на Ω). В многомерных пространствах интервал Ω может быть заменен областью более общего вида. При этом теория не требует никаких изменений.

В гл. VI, 11 было показано, что существует марковская цепь $(\mathbf{X}_0, \mathbf{X}_1, \ldots)$ с переходными вероятностями, определяемыми K. Распределение γ_0 начальной величины $\mathbf{X}_0$ может быть выбрано произвольно. После этого распределения $\mathbf{X}_1, \mathbf{X}_2, \ldots$ определяются последовательно по формуле

$$\gamma_n(\Gamma) = \int_{\Omega} \gamma_{n-1}(dx) K(x, \Gamma). \tag{7.1}$$

В частности, если γ_0 сконцентрировано в точке x_0, то $\gamma_n(\Gamma) = K^{(n)}(x_0, \Gamma)$ совпадает с вероятностью перехода из x_0 в Γ.

Определение 1. *Назовем меру α строго положительной на Ω, если $\alpha(I) > 0$ для любого открытого интервала $I \subset \Omega$. Назовем ядро K строго положительным, если $K(x, I) > 0$ для каждого x и каждого открытого интервала, входящего в Ω.*

Определение 2. *Назовем ядро эргодическим, если существует строго положительное распределение вероятностей α, такое, что $\gamma_n \to \alpha$ при любом начальном распределении γ_0.*

Это равносильно требованию сходимости

$$K^{(n)}(x, I) \to \alpha(I) > 0 \tag{7.2}$$

для каждого интервала непрерывности распределения α. Приведенное определение таково же, как и для дискретного случая (1, гл. XV). Его смысл был объяснен в примерах гл. VI, 11.

Так как с наиболее общими стохастическими ядрами могут быть связаны различные патологии, то мы ограничимся рассмотрением ядер, непрерывно зависящих от x. Простейший способ выразить это свойство связан с преобразованиями множества непрерывных функций, порождаемыми K. Пусть u — функция ограниченная и непрерывная на рассматриваемом интервале Ω.

*) Этот материал рассматривается как по причине его важности, так и потому, что он дает яркую иллюстрацию применений теоремы о выборе. Явным образом он в дальнейшем не используется.

Положим $u_0 = u$ и определим по индукции

$$u_n(x) = \int_\Omega K(x, dy)\, u_{n-1}(y). \tag{7.3}$$

Это преобразование функций двойственно с преобразованием мер (7.1). Отметим, что в обоих случаях индексы (всюду в этом параграфе) указывают на преобразование, индуцированное K.

Свойство регулярности, которое мы собираемся потребовать от K, состоит, грубо говоря, в том, что функция u_1 должна быть не хуже, чем u_0. Нижеследующее определение дает точное выражение нашему требованию, но выглядит довольно формальным. Примеры покажут, однако, что это определение применимо в типичных ситуациях.

Определение 3. *Назовем ядро K регулярным, если семейство преобразований u_k равностепенно непрерывно*[1]), *при равномерно непрерывной на Ω функции u_0.*

Примеры. *а) Свертки* дают пример преобразований, порождаемых регулярными стохастическими ядрами.

б) Пусть Ω — единичный интервал и пусть K имеет плотность k, непрерывную в замкнутом единичном квадрате. Тогда

$$|u_n(x') - u_n(x'')| \leqslant \int_0^1 |k(x', y) - k(x'', y)| \cdot |u_{n-1}(y)|\, dy. \tag{7.4}$$

По индукции заключаем, что если $|u_0| < M$, то и $|u_n| < M$ при всех n. В силу равномерной непрерывности k найдется такое δ, что

$$|k(x', y) - k(x'', y)| < \varepsilon/M \quad \text{при} \quad |x' - x''| < \delta.$$

Поэтому при всех n $|u_n(x') - u_n(x'')| < \varepsilon$. ▶

Условие строгой положительности в приводимой ниже теореме излишне ограничительно. Его основное назначение состоит в том, чтобы устранить осложнения, связанные с разложимостью и периодичностью цепей (с ними мы сталкивались в **1**, гл. XV).

Теорема 1. *Каждое строго положительное регулярное ядро K на ограниченном замкнутом интервале Ω является эргодическим.*

Теорема неверна для неограниченных Ω, так как в этом случае предел в (7.2) может быть тождественно равен нулю. Об-

[1]) См. сноску на стр. 329. Наша «регулярность» аналогична «полной непрерывности» в теории операторов в гильбертовом пространстве.

щий результат может быть сформулирован в терминах стационарных мер. Напомним, что мера α называется стационарной относительно K, если $\alpha_1=\alpha_2=\ \ldots\ =\alpha$, т. е. если все ее преобразования (7.1) одинаковы.

Теорема 2. *Строго положительное регулярное ядро K будет эргодическим тогда и только тогда, когда оно обладает строго положительным стационарным распределением вероятностей α.*

Доказательство *теоремы* 1. Пусть v_0 — непрерывная функция, и v_1 — ее преобразование (7.3). Доказательство использует тот очевидный факт, что для строго положительного ядра K максимум v_1 строго меньше максимума v_0 (исключая только случай постоянной v_0). Рассмотрим теперь последовательность преобразований u_n какой-либо непрерывной функции u_0. Так как Ω замкнуто, то функция u_0 равномерно непрерывна на Ω, и потому существует подпоследовательность $\{u_{n_k}\}$, равномерно сходящаяся к непрерывной функции v_0 (теорема 6.3). Тогда u_{n_k+1} сходится к преобразованию v_1 функции v_0. Числовая последовательность максимумов m_n функций u_n монотонна и потому $m_n \to m$. Так как сходимость к v_0 и v_1 равномерна, то и v_0 и v_1 имеют один и тот же максимум m. Следовательно, $v_0(x)=m$ при всех x. Так как этот предел не зависит от выбора подпоследовательности $\{u_{n_k}\}$, мы приходим к заключению, что $u_n \to m$ равномерно.

Пусть теперь γ_0 — произвольное распределение вероятностей на Ω. Обозначим через $\mathbf{E}_n$ математическое ожидание, взятое по отношению к γ_n [преобразование γ_n определяется по (7.1)]. Сравнение (7.1) и (7.3) показывает, что

$$\mathbf{E}_n(u_0)=\mathbf{E}_0(u_n)\to\mathbf{E}_0(m)=m.$$

Из сходимости $\mathbf{E}_n(u_0)$ для любой непрерывной функции u_0 вытекает существование вероятностной меры α, такой, что $\gamma_n \to \alpha$ (см. теорему § 1; сходимость будет собственной, так как все распределения γ_n сосредоточены на конечном интервале). Из (7.1) следует, что α — стационарное распределение. Строгая положительность α непосредственно вытекает из строгой положительности K. ▷

Доказательство *теоремы* 2. Обозначим через $\mathbf{E}$ математическое ожидание, взятое по отношению к данному стационарному распределению вероятностей α. Пусть u_0 — произвольная функция из $C[-\infty, \infty]$ и u_k — ее преобразование. В силу стационарности имеем

$$\mathbf{E}(u_0)=\mathbf{E}(u_1)=\ \ldots\ .$$

Далее, $\mathbf{E}(|u_k|)$ убывает с ростом k, так что $\lim \mathbf{E}(|u_k|) = m$ существует.

Так же как в предыдущем доказательстве, выберем такую подпоследовательность, что $u_{n_k} \to v_0$ и $u_{n_k+1} \to v_1$, где v_1 — преобразование v_0. По теореме об ограниченной сходимости отсюда следует, что $\mathbf{E}(u_{n_k}) \to \mathbf{E}(v_0)$ и $\mathbf{E}(|u_{n_k}|) \to \mathbf{E}(|v_0|)$. Таким образом,

$$\mathbf{E}(v_1) = \mathbf{E}(v_0) = \mathbf{E}(u_0) \quad \text{и} \quad \mathbf{E}(|v_1|) = \mathbf{E}(|v_0|) = m.$$

Из свойства строгой положительности K и последнего неравенства вытекает, что непрерывная функция v_0 не меняет знака. Когда $\mathbf{E}(u_0) = 0$, то $v_0(x) = 0$ тождественно. Из сказанного следует, что при любой начальной функции u_0 мы имеем $v_0(x) = \mathbf{E}(u_0)$ при всех x. Этим доказано, что $u_n(x) \to \mathbf{E}(u_0)$, а последнее соотношение равносильно сходимости $K^{(n)}(x, \Gamma) \to \alpha(\Gamma)$ на всех интервалах непрерывности. ▶

Применим теперь изложенную теорию к *сверткам на окружности* длины 1, т. е. к преобразованиям вида

$$u_{n+1}(x) = \int_0^1 u_n(x - y) F\{dy\}, \tag{7.5}$$

где F — распределение вероятностей на окружности, и сложение производится по mod 1 [см. гл. II, 8, и пример (6. а)].

Это преобразование можно записать в форме (7.3) с $\Omega = \overline{|0, 1|}$ и $K^{(n)}(x, \Gamma) = F^{n*}\{x - \Gamma\}$. Если F строго положительно, то применима теорема 1. Однако мы докажем следующий более общий аналог центральной предельной теоремы.

Теорема 3[1]). *Пусть F — распределение вероятностей на окружности. Допустим, что оно не сосредоточено в вершинах правильного многоугольника. Тогда F^{n*} сходится к распределению с постоянной плотностью.* ▶

Доказательство. Достаточно показать, что для произвольной непрерывной функции u_0 преобразования u_n сходятся к константе m (зависящей от u_0). В самом деле, как показывает вторая часть доказательства теоремы 1, из этого вытекает, что F^{n*} сходятся к некоторому распределению вероятностей α на окружности. Так как $\alpha * u_0$ равно постоянной для любой

[1]) Аналог этой теоремы для числовой прямой указан в задачах 23 и 24. Обобщение на случай различных распределений см. Lévy P., *Bull. Soc. Math. France*, **67** (1939), 1—41; Dvoretzky A. and Wolfowitz J., *Duke Math. J.*, **18** (1951), 501—507.

непрерывной функции u_0, то отсюда следует, что α совпадает с равномерным распределением.

Чтобы показать, что $u_n \to m$, мы используем первую часть доказательства теоремы 1. Мы только заново должны доказать, что максимум преобразования v_1 любой непрерывной функции v_0 строго меньше, чем максимум v_0 (за исключением случая, когда v_0 равно константе). Поэтому для доказательства теоремы достаточно установить справедливость следующего утверждения. *Если v_0 — непрерывная функция, такая, что $v_0 \leqslant m$ и $v_0(x) < m$ для всех x в интервале I длины $\lambda > 0$, то существует такое r, что $v_r(x) < m$ при всех x.*

Так как повороты не изменяют величины максимума, то мы можем без ограничения общности предположить, что 0 есть точка роста F. Если b — другая точка роста, то $0, b, 2b, \ldots, rb$ суть точки роста F^{r*}. Поэтому можно выбрать b и r таким образом, что каждый интервал длины λ содержит по крайней мере одну из этих точек (см. лемму 1 и следствие в гл. V, 4). По определению

$$v_r(x) = \int_0^1 v_0(x-y)\, F^{r*}\{dy\}. \tag{7.6}$$

Каждой точке x возможно поставить в соответствие такую точку роста функции F^{r*} (скажем, y), что $x-y$ входит в I. Тогда $v_0(x-y) < m$ и потому $v_r(x) < m$. Так как x произвольно, то требуемое утверждение доказано. ▶

Замечание. Доказательство легко видоизменить так, чтобы установить следующее предложение: *если F сконцентрировано в вершинах правильного многоугольника, одна из вершин которого расположена в точке 0, то F^{n*} сходится к атомическому распределению, атомам которого приписан один и тот же вес.* Сходимость *не* обязана иметь место, если 0 не находится в числе атомов.

Пример. в) Пусть F сосредоточено в двух иррациональных точках a и $a+1/2$. Тогда F^{n*} сосредоточено в точках na и $na+1/2$ и сходимость невозможна. ▶

§ 8. Правильно меняющиеся функции

Понятие правильного изменения функций (введенное Карамата в 1930 г.) оказалось во многих отношениях плодотворным и находит все возрастающее число применений в теории вероятностей. Причины этого отчасти объяснены в приводимой ниже

лемме, которая, несмотря на свою простоту, играет здесь основную роль. Примеры этого параграфа демонстрируют интересные вероятностные результаты, а задача 31 содержит известный результат, касающийся устойчивых распределений. Он элементарным путем выводится из леммы.

Мы часто сталкиваемся с монотонными функциями U, получающимися из некоторого распределения вероятностей F интегрированием функции $y^{\rho}F\{dy\}$ на $\overline{0, x}$ или $\overline{x, \infty}$ [см., например, (4.6), (4.15), (4.16)]. Обычной заменой параметров можно перейти от функции U к семейству функций вида $a_t U(tx)$. Мы должны изучить их поведение при $t \to \infty$. Если предел $\psi(x)$ существует, то достаточно брать нормирующие множители a_t вида $a_t = \psi(1)/U(t)$ (предполагая $\psi(1) > 0$). Следующая лемма является поэтому более общей, чем это может показаться с первого взгляда. Она показывает, что класс возможных пределов удивительно мал.

Лемма 1. *Пусть U — положительная монотонная функция на $\overline{0, \infty}$, такая, что при $t \to \infty$*

$$\frac{U(tx)}{U(t)} \to \psi(x) \leqslant \infty \tag{8.1}$$

на всюду плотном множестве A точек x. Тогда

$$\psi(x) = x^{\rho}, \tag{8.2}$$

где $-\infty \leqslant \rho \leqslant \infty$.

«Бессмысленный» символ x^{∞} вводится только для того, чтобы избежать исключений. Его следует интерпретировать, конечно, как ∞ при $x > 1$ и 0 при $x < 1$. Аналогично $x^{-\infty}$ равно ∞ или 0 в зависимости от того, будет $x < 1$ или $x > 1$ (см. задачу 25).

Доказательство. Тождество

$$\frac{U(tx_1x_2)}{U(t)} = \frac{U(tx_1x_2)}{U(tx_2)} \cdot \frac{U(tx_2)}{U(t)} \tag{8.3}$$

показывает, что если конечный положительный предел в (8.1) существует при $x = x_1$ и при $x = x_2$, то он существует и при $x = x_1x_2$ и

$$\psi(x_1x_2) = \psi(x_1)\psi(x_2). \tag{8.4}$$

Допустим сначала, что для некоторой точки x_1 $\psi(x_1) = \infty$. Тогда по индукции заключаем, что $\psi(x_1^n) = \infty$ и $\psi(x_1^{-n}) = 0$ при всех n. Так как ψ монотонна, то отсюда следует, что или $\psi(x) = x^{\infty}$, или $\psi(x) = x^{-\infty}$. Остается доказать лемму для случая конечных ψ (см. задачу 25). В силу монотонности ψ мы можем доопреде-

лить ее всюду по непрерывности справа. Тогда (8.4) выполняется при всех x_1 и x_2. Но это уравнение лишь обозначениями отличается от многократно использовавшегося нами характеристического уравнения для показательной функции. В самом деле, полагая $x=e^{\xi}$ и $\psi(e^{\xi})=u(\xi)$, мы преобразуем (8.4) к виду $u(\xi_1+\xi_2)=u(\xi_1)u(\xi_2)$. Из 1, гл. XVIII, 6, мы знаем, что все решения этого уравнения, ограниченные на каждом конечном интервале [1]), имеют вид $u(\xi)=e^{\varrho\xi}$. Это, однако, то же самое, что $\psi(x)=x^{\varrho}$. ▶

Пример. а) Все степени $|\log x|$ удовлетворяют (8.1), (8.2) с $\rho=0$. Аналогично если $U(x)$ стремится при $x\to\infty$ к *положительному конечному* пределу, то (8.1), (8.2) выполняются с $\rho=0$. Функции $(1+x^2)^p$ и e^x удовлетворяют (8.1), (8.2) с $\rho=2p$ и $\rho=\infty$ соответственно. Функция $2+\sin x$ не удовлетворяет (8.1). ▶

Пусть U — положительная функция, удовлетворяющая (8.1), (8.2) с *конечным* ρ. Положим

$$U(x)=x^{\rho}L(x). \tag{8.5}$$

Тогда при $t\to\infty$ и любом $x>0$

$$\frac{L(tx)}{L(t)}\to 1, \tag{8.6}$$

так что L удовлетворяет (8.1), (8.2) с $\rho=0$. Следующее ниже определение применимо и к немонотонным функциям.

Определение. *Заданная на* $\overline{0,\infty}$ *положительная функция* L *называется медленно меняющейся на бесконечности в том и только том случае, когда она удовлетворяет условию* (8.6). *Функция* U *называется правильно меняющейся с показателем* ρ *в том и только том случае, когда для нее выполняется* (8.5), *где* $-\infty<\rho<\infty$ *и* L *— медленно меняющаяся функция.*

По самому определению свойство функции быть правильно меняющейся не зависит от поведения U на конечных интервалах. Для облегчения ссылок мы перефразируем лемму 1, придав ей форму теоремы.

Теорема. *Монотонная функция* U *правильно изменяется на бесконечности тогда и только тогда, когда* (8.1) *выполняется*

[1]) Более общим образом любая конечная бэровская функция, удовлетворяющая (8.4), имеет вид x^{ρ}. Отсюда следует, что лемма 1 (за исключением случая $\rho=\pm\infty$) остается справедливой для произвольных (не обязательно монотонных) функций U, если предположить, что (8.1) выполняется во *всех* точках.

на всюду плотном множестве точек и предел ψ конечен и положителен в некотором интервале.

Полезный *критерий* дает следующая

Лемма 2. *Пусть*

$$\frac{\lambda_{n+1}}{\lambda_n} \to 1, \quad a_n \to \infty. \tag{8.7}$$

Если U — монотонная функция, такая, что предел

$$\lim \lambda_n U(a_n x) = \chi(x) \leqslant \infty \tag{8.8}$$

существует на всюду плотном множестве и функция χ конечна и положительна в некотором интервале, то U правильно меняется и $\chi(x) = cx^{\rho}$, где $-\infty < \rho < \infty$.

Доказательство. Мы можем предположить, что $\chi(1) = 1$ и что (8.8) выполняется при $x = 1$ (этого можно добиться тривиальным изменением масштаба). Для каждого данного t определим n как *наименьшее* целое число, для которого $a_{n+1} > t$. Тогда $a_n \leqslant t < a_{n+1}$ и для неубывающей функции U

$$\frac{U(a_n x)}{U(a_{n+1})} \leqslant \frac{U(tx)}{U(t)} \leqslant \frac{U(a_{n+1}x)}{U(a_n)}. \tag{8.9}$$

Для невозрастающей функции U верны обратные неравенства. Так как $\lambda_n U(a_n) \to 1$, то крайние члены стремятся к $\chi(x)$ в каждой точке, где выполняется (8.8). Поэтому утверждение леммы вытекает из последней теоремы. ▶

В качестве типичных примеров применений мы дадим известную теорему Р. А. Фишера — Б. В. Гнеденко и один новый результат.

Примеры. б) *Распределение максимума.* Пусть случайные величины $\mathbf{X}_k$ взаимно независимы и имеют одно и то же распределение F. Положим

$$\mathbf{X}_n^* = \max[\mathbf{X}_1, \ldots, \mathbf{X}_n].$$

Выясним, существуют ли нормирующие множители a_n, такие, что случайные величины $\mathbf{X}_n^*/a_n$ имеют предельное распределение G. Мы исключим при этом два тривиальных случая. Если F имеет наибольшую точку роста ξ, то распределение $\mathbf{X}_n^*$ тривиальным образом сходится к распределению, сосредоточенному в ξ. С другой стороны, всегда можно выбрать нормирующие множители a_n столь быстро возрастающими, что величины $\mathbf{X}_n^*/a_n$ сходятся по вероятности к нулю. Остальные случаи охватываются следующим *предложением.*

Пусть $F(x)<1$ при всех x. Для того чтобы при некотором выборе нормирующих постоянных a_n распределения G_n случайных величин $\mathbf{X}_n^*/a_n$ сходились к распределению G, не сосредоточенному в 0, необходимо и достаточно, чтобы функция $1-F$ была правильно меняющейся с показателем $\rho<0$. В этом случае

$$G(x)=e^{-cx^{\rho}} \tag{8.10}$$

при $x>0$ и $G(x)=0$ при $x<0$ (очевидно, $c>0$).

Доказательство. Если предельное распределение G существует, то

$$F^n(a_nx)\to G(x) \tag{8.11}$$

во всех точках непрерывности. Переходя к логарифмам и вспоминая, что $\log(1-z)\sim -z$ при $z\to 0$, получаем

$$n[1-F(a_nx)]\to -\log G(x). \tag{8.12}$$

Так как в некотором интервале $0<G(x)<1$, то последняя лемма гарантирует правильное изменение $1-F$. Обратно, если $1-F$ меняется правильно, то возможно определить постоянные a_n так, что $n[1-F(a_n)]\to 1$. В этом случае левая часть (8.12) стремится к x^{ρ} (см. задачу 26).

в) *Свертки.* В качестве второго примера мы докажем следующее *предложение.*

Пусть F_1 и F_2 — две функции распределения, такие, что при $x\to\infty$

$$1-F_i(x)\sim\frac{a_i}{x^{\rho}}L(x), \tag{8.13}$$

где L — медленно меняющаяся функция. Тогда свертка $G=F_1*F_2$ имеет правильно меняющийся «хвост», так что

$$1-G(x)\sim\frac{a_1+a_2}{x^{\rho}}L(x). \tag{8.14}$$

Доказательство. Пусть $\mathbf{X}_1$ и $\mathbf{X}_2$ — независимые случайные величины с распределениями F_1 и F_2. Для $t>1$ и $\varepsilon>0$, очевидно,

$$\mathbf{P}\{\mathbf{X}_1+\mathbf{X}_2>t\}\geqslant\mathbf{P}\{\mathbf{X}_1>t(1+\varepsilon)\}\cdot\mathbf{P}\{|\mathbf{X}_2|<t\varepsilon\}+ \\ +\mathbf{P}\{\mathbf{X}_2>t(1+\varepsilon)\}\cdot\mathbf{P}\{|\mathbf{X}_1|<t\varepsilon\}. \tag{8.15}$$

Так как $\mathbf{P}\{|\mathbf{X}_i|<t\varepsilon\}\to 1$ при $t\to\infty$, мы получаем

$$\liminf_{t\to\infty}\frac{1-G(t)}{t^{-\rho}L(t)}\geqslant\frac{a_1+a_2}{(1+\varepsilon)^{\rho}}. \tag{8.16}$$

С другой стороны, из события $\mathbf{X}_1+\mathbf{X}_2>t$ вытекает, что или хотя бы одна из величин $\mathbf{X}_i>t(1-\varepsilon)$, либо обе больше $t\varepsilon$. Следовательно,

$$\mathbf{P}\{\mathbf{X}_1+\mathbf{X}_2>t\}\leqslant\mathbf{P}\{\mathbf{X}_1>t(1-\varepsilon)\}+\\+\mathbf{P}\{\mathbf{X}_2>t(1-\varepsilon)\}+\mathbf{P}\{\mathbf{X}_1>t\varepsilon\}\cdot\mathbf{P}\{\mathbf{X}_2>t\varepsilon\}. \quad (8.17)$$

Деля обе части неравенства на $t^{-\rho}L(t)$ и полагая $t\to\infty$, мы получаем неравенство, противоположное (8.16). Из этих двух неравенств получаем (8.14). ▶

Индукцией по r получаем интересное

Следствие[1]). *Если* $1-F\sim x^{-\rho}L(x)$, то $1-F^{r\times}(x)\sim rx^{-\rho}L(x)$.

Это утверждение в тех случаях, где оно применимо, дополняет центральную предельную теорему, давая информацию о «хвостах» распределений.

§ 9*. Асимптотические свойства правильно меняющихся функций

Цель этого параграфа — изучить связь между «хвостами» и урезанными моментами распределений с правильно меняющимися «хвостами». Основной результат состоит в том, что если $1-F(x)$ и $F(-x)$ меняются правильно, то тем же свойством обладают урезанные моменты. Это утверждение содержится в теореме 2, которая дает бо́льшие сведения, чем то, что нам нужно для теории устойчивых распределений. Эту теорему можно доказать непосредственно, а можно рассматривать как следствие теоремы 1, которая включает в себя замечательную характеризацию правильно меняющихся функций, данную Карамата[2]).

Лучше всего, как нам кажется, дать полное изложение теории, в частности, потому, что теперь многие рассуждения могут быть значительно упрощены[3]).

---

[1]) Этот факт был отмечен С. Портом. Порт дал другое доказательство, в предположении, что F имеет правильно меняющуюся плотность и $F(0)=0$.

*) Материал этого параграфа используется только в теории устойчивых распределений. Однако применение теоремы 2 устранило бы многочисленные длинные вычисления, встречающиеся в литературе.

[2]) Karamata J., Sur un mode de croissance reguliére, *Mathematica (Cluj)*, 4 (1930), 38—53. Несмотря на частые ссылки на эту работу, более нового изложения ее результатов, по-видимому, не существует

[3]) Наше доказательство теоремы 1, будучи новым, использует тем не менее идеи Карамата.

Введем следующие сокращения:

$$Z_p(x)=\int_0^x y^p Z(y)\,dy, \quad Z_p^*(x)=\int_x^\infty y^p Z(y)\,dy. \tag{9.1}$$

Мы покажем теперь, что в случае правильно меняющейся функции Z эти функции асимптотически ведут себя по отношению к Z так же, как в простом случае $Z(x)=x^a$.

Асимптотическое поведение Z_p на бесконечности не зависит от поведения Z вблизи нуля. Поэтому мы можем, не ограничивая общности, допустить, что Z обращается в нуль всюду в некоторой окрестности нуля, так что интеграл, определяющий Z_p, имеет смысл при всех p.

Лемма. *Пусть $Z>0$ — медленно меняющаяся функция. Тогда интегралы в* (9.1) *сходятся на бесконечности при $p<-1$ и расходятся при $p>-1$.*

Если $p\geqslant -1$, то Z_p правильно меняется с показателем $p+1$. Если $p<-1$, то Z_p^* правильно меняется с показателем $p+1$. Последнее верно и для $p+1=0$, если только Z_{-1}^* существует.

Доказательство. При $x>0$ и $0<\eta<t$ имеем

$$Z_p(tx)=Z_p(\eta x)+x^{p+1}\int_\eta^t y^p Z(xy)\,dy. \tag{9.2}$$

Так как Z — медленно меняющаяся функция, то мы можем выбрать η (зависящее от x) так, что при $y\geqslant\eta$

$$(1-\varepsilon)Z(y)<Z(xy)<(1+\varepsilon)Z(y).$$

Тогда по (9.2)

$$x^{p+1}(1-\varepsilon)\leqslant\frac{Z_p(xt)-Z_p(xy)}{Z_p(t)-Z_p(y)}\leqslant x^{p+1}(1+\varepsilon) \tag{9.3}$$

при $t>y\geqslant\eta$. Фиксируя $x>1$ и применяя повторно (9.3) с $t=xy$ и $y=x^k\eta$ $(k=0,\ 1,\ \ldots)$, получаем

$$\sum_{k=0}^{\infty}(1-\varepsilon)^k x^{k(p+1)}\leqslant\frac{Z_p(\infty)-Z_p(\eta)}{Z_p(x\eta)-Z_p(\eta)}\leqslant\sum_{k=0}^{\infty}(1+\varepsilon)^k x^{k(p+1)}$$

Выбирая ε достаточно близким к 0, мы видим, что из $p+1>0$ вытекает расходимость Z_p, а из $p+1<0$ — сходимость Z_p. При $p+1=0$ возможно и то и другое.

Полагая в (9.3) $t\to\infty$, мы видим, что при расходимости Z_p эта функция меняется правильно с показателем $p+1$. Из сходи-

мости Z_p следует, что эта функция медленно меняется и в этом случае нечего доказывать.

Рассмотрим теперь функцию Z_p^* (которая определена тогда и только тогда, когда Z_p сходится). Полагая в (9.3) $t\to\infty$, получим

$$x^{p+1}(1-\varepsilon)\leqslant\frac{Z_p^*(xy)}{Z_p^*(y)}\leqslant x^{p+1}(1-\varepsilon) \tag{9.4}$$

при $x>0$ и $y>\eta$. Следовательно, функция Z_p^* меняется правильно с показателем $p+1$, если только она определена. ▶

Теорема 1. (i) *Если Z меняется правильно с показателем γ, то*

$$\frac{t^{p+1}Z(t)}{Z_p(t)}\to p+\gamma+1,\quad p+\gamma+1\geqslant 0. \tag{9.5}$$

Аналогично

$$\frac{t^{p+1}Z(t)}{Z_p^*(t)}\to|p+\gamma+1|,\quad p+\gamma+1<0. \tag{9.6}$$

Последнее верно и при $p+\gamma+1=0$, если только $Z_{-\gamma-1}^*$ существует.

(ii) *Обратно, если левая часть* (9.5) *или* (9.6) *стремится к некоторому пределу $\lambda>0$, то Z меняется правильно с показателем $\gamma=\lambda-p-1$ или $\gamma=-\lambda-p-1$ [так что* (9.5) *и* (9.6) *верны для всех допустимых p].*

Доказательство. Положим

$$\frac{y^pZ(y)}{Z_p(y)}=\frac{\eta(y)}{y}. \tag{9.7}$$

Пусть Z правильно меняется с показателем γ и пусть $p+\gamma+1\geqslant 0$. Тогда $y^pZ(y)$ правильно меняется с показателем $p+\gamma$ и по предыдущей лемме Z_p правильно меняется с показателем $p+\gamma+1$. Отсюда следует, что функция (9.7) правильно меняется с показателем -1. Поэтому η — *медленно меняющаяся* функция. Числитель левой части почти всюду совпадает с производной знаменателя. Выполняя интегрирование в пределах от t до tx, получаем

$$\log\frac{Z_p(tx)}{Z_p(t)}=\int_1^x\eta(ts)\cdot\frac{ds}{s}=\eta(t)\int_1^x\frac{\eta(ts)}{\eta(t)}\,\frac{ds}{s}. \tag{9.8}$$

Так как Z_p — правильно меняющаяся функция, то левая часть сходится к $(p+\gamma+1)\log x$. Пусть t пробегает последовательность $t_n\to\infty$, такую, что интеграл в правой части (9.8) сходится

к пределу $c \leqslant \infty$. Подинтегральное выражение сходится к s^{-1}, так как η — медленно меняющаяся функция. По теореме Фату [гл. IV, (2.9)] $c \geqslant \log x$. Следовательно, $\eta(t_n)$ сходится к конечному пределу $a \leqslant p+\gamma+1$. Таким образом, при любом выборе $\{t_n\}$ величины $\eta(t_n)$ остаются ограниченными и поэтому функция η *ограничена* на бесконечности. Теперь сходимость $\eta(t_n) \to a$ влечет сходимость $\eta(t_n s) \to a$ при каждом s. Так как эта сходимость является ограниченной, то средняя часть (9.8) сходится к $a \log x$. То есть из правильного изменения Z вытекает (9.5).

Обратное доказывается легче. Если левая часть (9.5) сходится к $\lambda \geqslant 0$, то $\eta(t) \to \lambda$. Следовательно, средняя часть (9.8) сходится к $\lambda \log x$, откуда выводим, что Z_p правильно меняется с показателем λ. Из предположения, что $Z(x) \sim \lambda x^{-p-1} Z_p(x)$, теперь заключаем, что Z правильно меняется с показателем $\lambda - p - 1$, что и утверждалось. Аналогичные рассуждения применимы к Z_p^*. Доказательство закончено. ▶

Отметим (хотя мы и не будем его использовать) интересное

Следствие. *Функция Z является медленно меняющейся тогда и только тогда, когда она может быть представлена в виде*

$$Z(x) = a(x) \exp\left(\int_1^x \frac{\varepsilon(y)}{y}\, dy \right), \tag{9.9}$$

где $\varepsilon(x) \to 0$ и $a(x) \to c < \infty$ при $x \to \infty$.

Доказательство. Достаточность тривиальным образом следует из определения. Для доказательства необходимости мы положим в (9.8) $t=1$ и подставим результат в (9.7). Тогда для Z получается явное выражение в терминах вспомогательной функции η. В нашем случае $p=0$ и $\eta(x) \to 1$. Полагая $\varepsilon(x) = \eta(x) - 1$, получаем (9.9). ▶

В применениях следующей ниже теоремы к функциям распределения можно рассматривать оба «хвоста» по отдельности. Иными словами, достаточно изучить распределения, сосредоточенные на $\overline{0, \infty}$.

Теорема 2. *Пусть F — распределение вероятностей, сосредоточенное на $\overline{0, \infty}$, и*

$$U_\zeta(x) = \int_0^x y^\zeta F\{dy\}, \tag{9.10}$$

где $\zeta > 0$ фиксировано и $U_\zeta(\infty) = \infty$.

(i) *Если или* U_ζ, *или* $1-F$ *меняются правильно, то существует предел*

$$\lim_{t\to\infty}\frac{t^\zeta[1-F(t)]}{U_\zeta(t)}=c, \quad 0\leqslant c\leqslant\infty. \tag{9.11}$$

(ii) *Обратно, пусть выполняется* (9.11) *с* $0<c<\infty$. *Представим* c *в форме*

$$c=\frac{\zeta-\alpha}{\alpha}. \tag{9.12}$$

Тогда

$$U_\zeta(x)\sim x^{\zeta-\alpha}L(x), \quad 1-F(x)\sim\frac{\zeta-\alpha}{\alpha}x^{-\alpha}L(x), \tag{9.13}$$

где L — *медленно меняющаяся функция.*

Далее из (9.11) с $c=0$ ($c=\infty$) *вытекает, что функция* U_ζ (*соответственно* $1-F$) *медленно меняющаяся.*

(Из медленного изменения U_ζ не следует правильного изменения $1-F$, см. задачу 29.)

Доказательство. Теорема симметрична по отношению к U_ζ и $1-F$. Поэтому доказательства для обеих функций одинаковы. Мы рассмотрим здесь лишь утверждения, касающиеся U_ζ (другие не будут использованы в этой книге).

Отправляясь от равенства

$$1-F(x)=\int_x^\infty y^{-\zeta}U_\zeta\{dy\}, \tag{9.14}$$

формальным интегрированием по частям получаем

$$1-F(x)=-x^{-\zeta}U_\zeta(x)+\zeta\int_x^\infty y^{-\zeta-1}U_\zeta(y)\,dy. \tag{9.15}$$

[Эта процедура законна. Применяя ее сначала на $\overline{x,\ t}$, убеждаемся, что интеграл в (9.15) сходится. В то же время из (9.10) видно, что $x^{-\zeta}U_\zeta(x)\to 0$ при $x\to\infty$.] Перепишем (9.15) в виде

$$\frac{x^\zeta[1-F(x)]}{U_\zeta(x)}=-1+\zeta\frac{1}{x^{-\zeta}U_\zeta(x)}\int_x^\infty y^{-\zeta-1}U_\zeta(y)\,dy. \tag{9.16}$$

Если функция U_ζ правильно меняется, то ее показатель необходимо $\leqslant\zeta$. Назовем его $\gamma=\zeta-\alpha$. Используя (9.6) с $Z=U_\zeta$ и $\rho=-\zeta-1$, мы видим, что правая часть (9.16) стремится к $-1+\zeta/\alpha$ (к ∞, если $\alpha=0$). Этим доказываются формулы (9.11) и (9.12). Первое из соотношений (9.13) верно по определению, а

при $\alpha<\zeta$ из него и (9.11) вытекает второе. Обратно, если выполняется (9.11), то по теореме 1 U_ζ — правильно меняющаяся функция с показателем γ, таким, что $\zeta/(\zeta-\gamma)=c+1$). ▶

(Теорема в форме, кажущейся несколько более общей, приведена в задаче 30.)

§ 10. Задачи

1. *Альтернативное определение сходимости.* Пусть F_n и F — некоторые распределения вероятностей. Покажите, что $F_n \to F$ (в собственном смысле) тогда и только тогда, когда для любых данных $\varepsilon>0$, $h>0$ и t существует $N(\varepsilon, h, t)$, такое, что при $n>N(\varepsilon, h, t)$

$$F(t-h)-\varepsilon<F_n(t)<F(t+h)+\varepsilon. \tag{10 1}$$

2. *Несобственная сходимость.* Если F — несобственное распределение, то из (10.1) вытекает, что $F_n \to F$ в несобственном смысле. Обратное *неверно*. Покажите, что *собственная* сходимость может быть *определена* требованием, что (10.1) выполняется для $n \geqslant N(\varepsilon, h)$ независимо от t.

3. Пусть $\{F_n\}$ сходится в собственном смысле к распределению, не сосредоточенному в одной точке. Тогда последовательность $\{F_n(a_n x+b_n)\}$ сходится к распределению, сосредоточенному в нуле в том и только том случае, когда $a_n \to \infty$ и $b_n=o(a_n)$.

4. *Другое доказательство леммы* 2.3. Используя теорему 3.2, покажите, что $a_{2n}/a_n \to \lambda$. Выведите отсюда, что $a_n \to \infty$ и что *предельное распределение необходимо устойчиво* (см гл. VI, 1). *Указание.* Достаточно ограничиться симметричными F.

5. Пусть $\{u_n\}$ — последовательность ограниченных монотонных функций, сходящаяся поточечно к ограниченной и *непрерывной* предельной функции (которая автоматически монотонна). Докажите, что сходимость равномерна на всех конечных интервалах. *Указание.* Разбейте ось абсцисс на интервалы, на каждом из которых колебание u меньше ε.

6. (К теореме 3.1.) Пусть F_n сосредоточено в точке n^{-1} и $u(x)=\sin(x^2)$. Тогда $F_n * u \to u$ поточечно, но неравномерно.

7. а) Если совместное распределение $(\mathbf{X}_n, \mathbf{Y}_n)$ сходится к совместному распределению $(\mathbf{X}, \mathbf{Y})$, то распределение $\mathbf{X}_n+\mathbf{Y}_n$ сходится к распределению $\mathbf{X}+\mathbf{Y}$.

б) Покажите, что теорема 3.2 — частный случай этого утверждения.

в) Сделанное утверждение, вообще говоря, неверно, если известна только сходимость маргинальных распределений $\mathbf{X}_n$ и $\mathbf{Y}_n$.

8. Пусть $F_n \to F$, где F — несобственное распределение. Если $u \in C_0(-\infty, \infty)$, то

$$F_n \overset{\times}{*} u \to F \overset{\times}{*} u$$

равномерно *в каждом конечном интервале* (это — обобщение теоремы 3.1).

10*). На плоскости каждая исчезающая на бесконечности функция может быть равномерно аппроксимирована конечными линейными комбина-

*) Задача 9 оказалась неверной и опущена. Нумерация последующих задач сохраняется. — *Прим. перев.*

циями вида $\Sigma c_k \varphi_k(x)\psi_k(y)$, где φ_k и ψ_k — бесконечно дифференцируемые функции.

Указание. Используйте теорему об аппроксимации из примера (3, а), полагая

$$G_k(x, y) = \mathfrak{N}_k(x)\,\mathfrak{N}_k(y),$$

где $\mathfrak{N}$ — нормальная плотность.

Метрики. Функция ρ называется *расстоянием* между распределениями вероятностей, если $\rho(F, G)$ определено для любой пары распределений и обладает следующими тремя свойствами: $\rho(F, G) \geqslant 0$ и $\rho(F, G) = 0$ тогда и только тогда, когда $F = G$; далее $\rho(F, G) = \rho(G, F)$; наконец, ρ удовлетворяет неравенству треугольника:

$$\rho(F_1, F_2) \geqslant \rho(F_1, G) + \rho(G, F_2).$$

11. *Метрика П. Леви.* Пусть F и G — два распределения вероятностей. Определим $\rho(F, G)$ как точную нижнюю грань всех $h > 0$, для которых при всех x

$$F(x-h) - h \leqslant G(x) \leqslant F(x+h) + h. \tag{10.2}$$

Проверьте, что ρ является расстоянием. Докажите, что $F_n \to F$ в собственном смысле тогда и только тогда, когда $\rho(F_n, F) \to 0$.

12. *Расстояние «по вариации».* Положим $\rho(F, G) = \sup \| \mathfrak{F}u - \mathfrak{G}u \|$, где $u \in C_0$ и $\| u \| = 1$. Покажите, что ρ — расстояние [1]). Если F и G — атомические распределения, приписывающие точке a_k веса p_k и q_k соответственно, то

$$\rho(F, G) = \sum | p_k - q_k |. \tag{10.3}$$

Если F и G имеют плотности f и g, то

$$\rho(F, G) = \int_{-\infty}^{+\infty} | f(x) - g(x) |\, dx. \tag{10.4}$$

[Достаточно доказать (10.4) для непрерывных f и g. Общий случай изучается с помощью надлежащей аппроксимации.]

13. *Продолжение.* Покажите, что из $\rho(F_n, G) \to 0$ вытекает собственная сходимость $F_n \to G$. Обратное неверно (рассмотрите нормальные функции распределения $\mathfrak{N}(nx)$ и распределения F_n, сосредоточенных в n^{-1}).

14. *Продолжение.* Если $U = F_1 * \ldots * F_n$ и $V = G_1 * \ldots * G_n$, то

$$\rho(U, V) \leqslant \sum_{k=1}^{n} \rho(F_k, G_k). \tag{10.5}$$

Последнее неравенство обобщает основное неравенство (3.9).

[*Указание.* Используйте (3.9) и функцию u, такую что $\| \mathfrak{U}u - \mathfrak{V}u \|$ близко к $\rho(U, V)$.]

---

[1]) Это определение может быть распространено на разности любых конечных мер и определяет для мер *«топологию, порождаемую нормой»*. Задача 13 показывает, что соответствующее понятие сходимости оказывается неестественным с точки зрения теории вероятностей.

15. *Аппроксимация пуассоновским распределением*[1]. Пусть F приписывает вес p точке 1 и вес $q=1-p$ точке 0. Если G — распределение Пуассона с математическим ожиданием p, то $\rho(F, G) \leqslant {}^9/_4 p^2$, где ρ — расстояние, определенное по (10.3). Выведите отсюда утверждение: если F — распределение числа успехов в n испытаниях Бернулли с вероятностями успеха p_1, p_2, ..., p_n и если G — распределение Пуассона с математическим ожиданием

$$p_1 + \ldots + p_n, \text{ то } \rho(F, G) \leqslant {}^9/_4 (p_1^2 + \ldots + p_n^2).$$

16. Закон больших чисел, установленный в гл. VII, утверждает, что для независимых одинаково распределенных величин $\mathbf{X}_k$ с $\mathbf{E}(\mathbf{X}_k)=0$,

$$(\mathbf{X}_1 + \ldots + \mathbf{X}_n)/n \xrightarrow{\text{p}} 0.$$

Докажите это методом, использованным в теореме 4.1.

17. Условие Линдеберга (4.15) выполняется, если при некотором $\delta>0$ существуют $\alpha_k = \mathbf{E}\left(|\mathbf{X}_k^{2+\delta}|\right)$ и $\alpha_1 + \ldots + \alpha_n = o\left(s_n^{2+\delta}\right)$ (условие Ляпунова).

18. Пусть F_k — симметричное распределение и $1 - F_k(x) = \dfrac{1}{2x^{2+k^{-1}}}$ для $x>1$. Покажите, что условие Линдеберга (4.15) выполняется

19. Пусть $\mathbf{X}_k = \pm 1$ с вероятностью ${}^1/_2(1-k^{-2})$ и $\mathbf{X}_k = \pm k$ с вероятностью ${}^1/_2 k^{-2}$. Покажите с помощью урезания, что величина $\mathbf{S}_n/\sqrt{n}$ ведет себя асимптотически так, как если бы $\mathbf{X}_k = \pm 1$ с вероятностью ${}^1/_2$. Таким образом, *распределение* $\mathbf{S}_n/\sqrt{n}$ *сходится к* $\mathfrak{N}$, *но* $\operatorname{Var}(\mathbf{S}_n/\sqrt{n}) \to 2$.

20. Видоизмените условие предыдущей задачи так, чтобы $\mathbf{E}\left(\mathbf{X}_k^2\right) = \infty$, но распределение $\mathbf{S}_n/\sqrt{n}$ сходилось к $\mathfrak{N}$.

21[2]. *Центральная предельная теорема для симметрично зависимых величин.* Пусть каждому значению θ соответствует распределение F_θ с нулевым математическим ожиданием и дисперсией $\sigma^2(\theta)$. Значение θ выбирается в соответствии с распределением вероятностей G. Затем рассматриваются взаимно независимые случайные величины $\mathbf{X}_n$ с одним и тем же распределением F_θ. Пусть a^2 — математическое ожидание σ^2, взятое по отношению к G. Покажите, что распределение $\mathbf{S}_n/(a\sqrt{n})$ сходится к распределению с плотностью

$$\int_{-\infty}^{+\infty} \frac{a}{\sigma(\theta)}\, \mathfrak{N}\left(\frac{ax}{\sigma(\theta)}\right) G\{d\theta\}.$$

Это распределение будет нормальным только в случае, когда G сосредоточено в одной точке.

22. *Шумы в электронных лампах и т. п. процессы.* Рассмотрим стохастический процесс из примера (3з) гл. VI с «дискретизированным» временем. Предполагая, что вероятность появления электрона в момент kh равна

[1] Задача подсказана неравенствами работы Le Cam L., An approximation theorem for the Poisson binomial distribution *Pacific J. Math.*, 10 (1960), 1181—1197.

[2] Blum J. R., Chernoff H., Rosenblatt M., and Teicher H., Central limit theorems for interchangeable proceses, *Canadian J. Math.*, **10** (1958), 222—229.

αh, покажите, что сила тока в дискретной модели определяется *бесконечной сверткой*. Переход к пределу при $h \to 0$ приводит к теореме Кемпбелла [(3,5), гл. VI].

Проделайте то же самое в примере с занятыми линиями [(3,и), гл. VI]. Обобщите модель на случай, когда последействие (after-effect) в момент kh есть случайная величина, принимающая значения $1, 2, \ldots$ с вероятностями $p_1, p_2, \ldots$.

23. *Для любого распределения вероятностей, не сосредоточенного в нуле,* $F^{n\star} \to 0$

Указание. Рассмотрите какую-либо сходящуюся подпоследовательность и обозначьте полную массу ее предела через p. Используя результат задачи 8, покажите, что $p^2 = p$. Невозможность существования *собственного* предела может быть легко установлена многими способами (например, из невозможности соотношения $G^{r\star} = G$ или из неравенств симметризации).

24. *Продолжение.* Тем не менее возможно, что при каждом x

$$\limsup_{n\to\infty} F^{n\star}(x) = 1, \quad \liminf_{n\to\infty} F^{n\star}(x) = 0.$$

В самом деле, можно выбрать две крайне быстро растущих последовательности целых чисел a_k и n_k так, что

$$(-1)^k \frac{1}{a_k} S_{n_k} \xrightarrow{\mathrm{p}} 1.$$

Указание. Возьмите распределение $\mathbf{P}\{X = (-1)^k a_k\} = p_k$. При подходящем выборе констант с подавляющей вероятностью один из членов последовательности $X_1, \ldots, X_{n_k}$ будет равен $(-1)^k a_k$ и ни один не превзойдет по абсолютной величине a_k. Тогда для четного k $S_{n_k} > a_k - n_k a_{k-1}$. Покажите, что для этой цели пригодны значения

$$n_k = (2k)!, \qquad p_k \sim \frac{1}{(2k-1)!}, \qquad a_k \sim (n_k)^k.$$

25. В доказательстве леммы 8.1 достаточно предположить, что множество A плотно в некотором открытом интервале.

26. *Распределение максимума.* Пусть $X_1, \ldots, X_n$ — независимые случайные величины с одним и тем же распределением F и $X_n^\star = \max(X_1, \ldots, X_n)$. Пусть G_n — распределенче величины $a_n^{-1} X_n^\star$.

а) Если $F(x) = 1 - e^{-x}$ и $a_n = n$, то G_n сходится к распределению сосредоточенному в точке 1. Покажите прямым подсчетом, что ни при каком выборе a_n не может получиться распределение, имеющее более одной точки роста.

б) Если F — распределение Коши с плотностью $\dfrac{1}{\pi(1+x^2)}$ и $a_n = n/\pi$, то при $x > 0$ $G_n(x) \to e^{-x^{-1}}$.

27. Если X и Y имеют одно и то же распределение F, такое, что $1 - F(x) \sim x^{-\rho} L(x)$, где L — медленно меняющаяся функция, то

$$\mathbf{P}\{X > t \mid X + Y > t\} \to \frac{1}{2}$$

при $t\to\infty$. Грубо говоря, большие значения суммы, как правило, возникают за счет больших значений одного из двух слагаемых [1]).

28. Пусть $v>0$ и $a>0$ на $\overline{0, \infty}$. Допустим, что предел

$$\lim_{t\to\infty}[a(t)v(tx)+b(t)x]=z(x)$$

существует и непрерывно зависит от x. Докажите, что при фиксированном $x_0>0$ функция $\dfrac{v(x_0x)}{x_0x}-\dfrac{v(x)}{x}$ правильно меняется. Установите, что или $z(x)=cx^\alpha$, или $z(x)=cx+c_1x\log x$.

29. Пусть F — атомическое распределение, приписывающее точке 2^n вес, пропорциональный $n^{-1}2^{-2n}$. Покажите, что функция U_2 [см. (9.10)] медленно меняется и $U_2(\infty)=\infty$, но $1-F$ не является правильно меняющейся функцией.

Указание. Для доказательства последнего утверждения достаточно рассмотреть величину скачков.

30. Пусть F — распределение, сосредоточенное на $\overline{0, \infty}$. Определим U_ζ, как в (9.10), и

$$V_\sigma(x)=\int_x^\infty y^{-\sigma}F\{dy\}. \qquad (*)$$

Теорема 9.2 допускает следующее обобщение. Пусть F имеет моменты порядка меньше α, но не имеет моментов порядка больше $\alpha>0$. Тогда U_ζ и V_σ или правильно меняются при *всех* $\zeta>\alpha$ и *всех* $\sigma>-\alpha$, или ни при одном из этих значений. В первом случае показатели равны $\zeta-\alpha$ и $-\sigma-\alpha$. Этот случай имеет место тогда и только тогда, когда

$$\frac{x^{\sigma+\zeta}V_\sigma(x)}{U_\zeta}\to\frac{\zeta-\alpha}{\sigma+\alpha} \qquad (**)$$

($0<\alpha<\zeta$). Если U_ζ медленно меняется, то соотношение (**) выполняется с $\alpha=\zeta$. Если V_σ медленно меняется, то левая часть стремится к ∞. [Специальный случай $\sigma=0$ сводится к (9.11).]

Указание. Можно повторить доказательство теоремы 2. Однако сформулированные утверждения можно получить из теоремы 2 одним лишь изменением обозначений.

31. Пусть G — симметричное устойчивое распределение, т. е. $G^r*(c_r x)=G(x)$ (см. гл. VI, 1). Выведите из последнего следствия § 8, что $1-G(x)\sim x^{-\alpha}L(x)$, где $\alpha<2$. При этом исключается случай $r[1-G(c_r x)]\to 0$, приводящий к нормальному распределению.

Указание. Неравенства симметризации гл. V, (5.10) показывают, что последовательность $r[1-G(c_r x)]$ ограничена. Дальнейшие рассуждения просты.

32. Обобщите эти утверждения на случай несимметричных устойчивых законов.

---

[1]) Это явление было впервые подмечено, кажется, Б. Мандельбройтом.

ГЛАВА IX

БЕЗГРАНИЧНО ДЕЛИМЫЕ РАСПРЕДЕЛЕНИЯ И ПОЛУГРУППЫ

Цель этой главы — показать, что основные теоремы, касающиеся безгранично делимых распределений, процессов с независимыми приращениями, устойчивых распределений и их областей притяжения, могут быть получены с помощью естественного обобщения рассуждений, использованных при доказательстве центральной предельной теоремы. Заметим, что *излагаемая теория будет развита заново* (*и в большем объеме*) *на основе анализа Фурье.* По этой причине здесь мы ограничимся лишь основными фактами. Настоящая глава представляет в значительной степени методологический интерес, связывая разбираемые вопросы с общей теорией марковских процессов. Метод Фурье, там где он применим, приводит к более сильным результатам. Чтобы облегчить понимание важных фактов, некоторые теоремы доказываются дважды. Так, общая теорема о строении рассматриваемых полугрупп распределений доказывается сначала для случая конечных дисперсий и т. п. В этом смысле § 1—4 дают замкнутое изложение основных фактов.

Полугрупповая операция, изучаемая в этой главе — свертка. Другие полугруппы будут изучены независимо новыми методами в следующей главе.

§ 1. Общее знакомство с темой

Предельные теоремы этой главы служат естественным обобщением центральной предельной теоремы, а безгранично делимые распределения тесно связаны с нормальным распределением. Чтобы понять это, стоит повторить доказательство теоремы гл. VIII, 4.1 в несколько иной обстановке.

Мы рассмотрим на этот раз схему серий $\{\mathbf{X}_{k,n}\}$, где при каждом n случайные величины[1]) $\mathbf{X}_{1,n}, \ldots, \mathbf{X}_{n,n}$ взаимно незави-

[1]) Схема серий была определена в гл. VI, 3. Следует помнить, что в действительности мы имеем дело с функциями распределения $F_{k,n}$; случайные величины $\mathbf{X}_{k,n}$ используются лишь для того, чтобы упростить обозначения. Соответственно случайные величины из различных строк не обязаны находиться в каких бы то ни было отношениях друг к другу (в частности, они могут быть заданы на разных вероятностных пространствах).

симы и имеют *одно и то же распределение* F_n. Обозначим суммы по строкам через $\mathbf{S}_n : \mathbf{S}_n = \mathbf{X}_{1,n} + \ldots + \mathbf{X}_{n,n}$. В гл. VIII мы имели дело с частным случаем $\mathbf{X}_{k,n} = \mathbf{X}_k a_n^{-1}$ и $F_n(x) = F(a_n x)$. Суммы по строкам обозначались $\mathbf{S}_n^*$.

Всюду в этой главе мы будем использовать операторные обозначения гл. VIII, 3. Так $\mathfrak{F}_n$ означает оператор, связанный с F_n, а $\mathfrak{F}_n^n$ — оператор, связанный с распределением $\mathbf{S}_n$. Наконец, $\|u\|$ обозначает максимум непрерывной функции u.

Примеры. а) *Центральная предельная теорема.* Допустим, что существуют такие числа $\varepsilon_n \to 0$, что

$$|\mathbf{X}_{1,n}| < \varepsilon_n, \quad \mathbf{E}(\mathbf{X}_{1,n}) = 0, \quad n\mathbf{E}(\mathbf{X}_{1,n}^2) \to 1. \tag{1.1}$$

Для любой функции u, имеющей три ограниченные производные, справедливо тождество

$$n[\mathfrak{F}_n u(x) - u(x)] = \int_{-\varepsilon_n}^{\varepsilon_n} \frac{u(x-y) - u(x) + yu'(x)}{y^2} \cdot ny^2 F_n\{dy\}. \tag{1.2}$$

Конечные меры $ny^2F_n\{dy\}$ сходятся по предположению к распределению вероятностей, сосредоточенному в нуле. Дробь под знаком интеграла непрерывно зависит от y и отличается от $\frac{1}{2}u''(x)$ не более чем на $\varepsilon_n\|u''\|$. Таким образом, равномерно по x

$$n[\mathfrak{F}_n u - u] \to \tfrac{1}{2}u''. \tag{1.3}$$

Допустим теперь, что $\mathfrak{G}_n$ — другая последовательность операторов и что $n[\mathfrak{G}_n u - u]$ равномерно сходится к $\frac{1}{2}u''$. Тогда

$$n(\mathfrak{F}_n^n u - \mathfrak{G}_n^n u) \to 0 \tag{1.4}$$

равномерно. По основному неравенству гл. VIII, (3.10) (оно будет постоянно использоваться в дальнейшем)

$$\|\mathfrak{F}_n^n u - \mathfrak{G}_n^n u\| \leqslant n\|\mathfrak{F}_n u - \mathfrak{G}_n u\|, \tag{1.5}$$

и правая часть стремится к нулю, как это следует из (1.4). Как мы уже видели при доказательстве теоремы 1 в гл. VIII, 4, в качестве $\mathfrak{G}_n$ можно выбрать оператор, связанный с симметричным нормальным распределением, имеющим дисперсию $1/n$. Тогда $\mathfrak{G}_n^n = \mathfrak{G}_1$, и потому $\mathfrak{F}_n^n \to \mathfrak{G}_1$. Мы доказали, таким образом, что *распределение сумм* $\mathbf{S}_n$ *сходится к нормальному распределению* $\mathfrak{N}$. ▶

Анализируя внимательно структуру доказательства, мы видим, что форма правой части (1.3) не играет никакой роли. Допустим, что мы имеем такую схему серий, что (в равномерном смысле)

$$n[\mathfrak{F}_n u - u] \to \mathfrak{A}u, \tag{1.6}$$

где $\mathfrak{A}$ — произвольный, но фиксированный оператор. Наши рассуждения позволяют нам сравнить любые две схемы серий, удовлетворяющих (1.6), и прийти к заключению, что их суммы по строкам ведут себя асимптотически одинаково. Если для *одной* из этих схем распределения $\mathbf{S}_n$ сходятся к пределу G, то это же будет верно и для *всех* других наших схем. Мы докажем это позже.

б) *Распределение Пуассона.* Допустим, что $\mathbf{X}_{1,n}$ равно единице с вероятностью p_n и нулю с вероятностью $1-p_n$. Если $np_n \to \alpha$, то

$$n[\mathfrak{F}_n u(x) - u(x)] = np_n[u(x-1) - u(x)] \to \alpha[u(x-1) - u(x)]. \tag{1.7}$$

На этот раз мы выберем в качестве $\mathfrak{G}_n$ оператор, связанный с распределением Пуассона с математическим ожиданием α/n. Простой подсчет показывает, что и $n[\mathfrak{G}_n u - u]$ сходится к правой части (1.7). Отсюда выводим, как и раньше, что $\mathfrak{F}_n^n \to \mathfrak{G}_1$. Таким образом, распределение $\mathbf{S}_n$ сходится к распределению Пуассона с математическим ожиданием α. [Правая часть (1.7) дает нам один из возможных примеров оператора $\mathfrak{A}$ в (1.6). Другой пример простой схемы серий см. в задаче 2]. ▷

В двух указанных выше примерах нам посчастливилось знать форму предельного распределения заранее. В общем случае схема серий может служить для определения предельного распределения, и таким путем мы можем получить новые функции распределения. Эта процедура была использована в **1**, гл. VI для *определения* распределения Пуассона как предела биномиальных распределений.

Напомним, что предельные распределения для сумм $\mathbf{S}_n$ были названы (см. гл. VI, 3) *безгранично делимыми*. Мы покажем, что такое предельное распределение существует всякий раз, когда выполняется соотношение типа (1.6), а также покажем, что это условие необходимо. Другой подход к рассматриваемой задаче основан на изучении мер $ny^2F_n\{dy\}$. В обоих примерах существовала предельная мера; в примере (а) она была сосредоточена в нуле, в примере (б) — в точке единица. В общем случае соотношение (1.6) тесно связано с существованием такой меры Ω, что $ny^2F_n\{dy\} \to \Omega\{dy\}$. При этом безгранично дели-

мые распределения могут быть описаны либо с помощью оператора $\mathfrak{A}$, либо с помощью меры Ω (которая может быть неограниченной).

Третий подход к разбираемой задаче использует уравнение свертки

$$Q_s * Q_t = Q_{s+t}, \tag{1.8}$$

в котором Q_t — распределение вероятностей, зависящее от параметра $t>0$. [Нормальное распределение и распределение Пуассона удовлетворяют (1.8) с t, пропорциональным дисперсии.] Схему серий мы получим, отождествляя F_n с $Q_{1/n}$. Тогда (1.6) равносильно

$$\frac{1}{t}[Q_t * u - u] \to \mathfrak{A}u \tag{1.9}$$

при t, пробегающем последовательность $1/2$, $1/3$ ... . Можно ожидать, что (1.9) будет выполняться при произвольном стремлении $t \to 0+$.

Отметим, что (1.8) есть не что иное, как основное уравнение для процессов со *стационарным и независимыми приращениями* (гл. VI, 4). Оно тесно связано с теорией полугрупп. В этом контексте $\mathfrak{A}$ выступает в роли «производящего оператора». Оказывается, что именно эта теория обеспечивает самый легкий подход к предельным теоремам и безгранично делимым распределениям. Поэтому мы с нее и начнем.

2. Полугруппы со сверткой

Пусть $Q_t (t>0)$ — семейство распределений в $\mathscr{R}^1$, удовлетворяющих (1.8), и $\mathfrak{Q}(t)$ — соответствующие операторы, т. е.

$$\mathfrak{Q}(t)\, u(x) = \int_{-\infty}^{+\infty} u(x-y)\, Q_t\{dy\}. \tag{2.1}$$

Тогда (1.8) эквивалентно

$$\mathfrak{Q}(s+t) = \mathfrak{Q}(s)\,\mathfrak{Q}(t). \tag{2.2}$$

Семейство операторов, удовлетворяющих (2.2), называется *полугруппой*. [Оно не будет, как правило, группой, так как оператор $\mathfrak{Q}(t)$ в общем случае не имеет обратного.] Операторы полугруппы могут иметь произвольную природу, и поэтому удобно ввести термин, подчеркивающий наше допущение о связи $\mathfrak{Q}(t)$ и вероятностных распределений.

Определение 1. *Полугруппа со сверткой* $\{\mathfrak{Q}(t)\}$ (*где* $t>0$) — *это семейство операторов, связанных с распределениями вероятностей и удовлетворяющих* (2.2).

В качестве области определения операторов мы возьмем $C_0[-\infty, \infty]$. Операторы $\mathfrak{Q}(t)$ являются операторами перехода, т. е. из $0 \leqslant u \leqslant 1$ вытекает $0 \leqslant \mathfrak{Q}(t)u \leqslant 1$, и, кроме того, $\mathfrak{Q}(t)1 = 1$.

Нам придется также иметь дело и с операторами [подобными d^2/dx^2 в (1.3)], которые определены не для всех непрерывных функций. Наши теперешние цели позволяют, к счастью, избежать утомительного обсуждения точной области определения этих операторов, так как нам достаточно рассмотреть только класс таких функций u, что $u \in C[-\infty, \infty]$; u имеет производные всех порядков, также принадлежащие $C[-\infty, \infty]$. Эти функции будут называться *бесконечно-дифференцируемыми*[1]), и их совокупность будет обозначаться C^∞. Пока мы ограничимся только операторами $\mathfrak{A}$, определенными для всех $u \in C^\infty$ и такими, что $\mathfrak{A}u \in C^\infty$, так что все встречающиеся операторы можно рассматривать как *операторы, отображающие* C^∞ в C^∞. Как мы видели в гл. VIII, 3, для операторов, связанных с распределениями вероятностей, сходимость $\mathfrak{F}_n \to \mathfrak{F}$ означает сходимость $\mathfrak{F}_n u \to \mathfrak{F}u$ для всех $u \in C^\infty$. Мы распространим теперь это определение сходимости на произвольные операторы.

Определение 2. *Пусть* $\mathfrak{A}_n$ *и* $\mathfrak{A}$ *— операторы, отображающие* C^∞ *в* C^∞. *Мы скажем, что* $\mathfrak{A}_n$ *сходится к* $\mathfrak{A}$, *в обозначениях* $\mathfrak{A}_n \to \mathfrak{A}$, *если при каждом* $u \in C^\infty$

$$\|\mathfrak{A}_n u - \mathfrak{A}u\| \to 0. \tag{2.3}$$

Соотношение (2.3) показывает, что $\mathfrak{A}_n u \to \mathfrak{A}u$ *равномерно*. Обратно, если для каждого $u \in C^\infty$ последовательность $\{\mathfrak{A}_n u\}$ сходится равномерно к пределу $v \in C^\infty$, то равенство $\mathfrak{A}u = v$ определяет некоторый оператор $\mathfrak{A}$ и, очевидно, $\mathfrak{A}_n \to \mathfrak{A}$.

Определение 3. *Полугруппа со сверткой* $\{\mathfrak{Q}(t)\}$ *называется непрерывной, если*

$$\mathfrak{Q}(h) \to \mathbf{1}, \qquad h \to 0+, \tag{2.4}$$

где $\mathbf{1}$ *— тождественный оператор. В этом случае мы полагаем* $\mathfrak{Q}(0) = \mathbf{1}$.

Из неравенства $\|\mathfrak{Q}(t)u\| \leqslant \|u\|$ и из определения 2.2 получаем при $h > 0$

$$\|\mathfrak{Q}(t+h)u - \mathfrak{Q}(t)u\| \leqslant \|\mathfrak{Q}(h)u - u\|. \tag{2.5}$$

[1]) Класс C^∞ введен только для того, чтобы избежать появления нового термина. Он мог бы быть заменен, скажем, классом функций с четырьмя ограниченными производными или (что еще проще) классом всех линейных комбинаций нормальных функций распределения с произвольными математическими ожиданиями и дисперсиями.

При достаточно малом h левая часть будет $<\varepsilon$, каково бы ни было t. В этом смысле полугруппа со сверткой является *равномерно непрерывной.*

Определение 4. *Говорят, что оператор* $\mathfrak{A}$, *отображающий* C^∞ *в* C^∞, *порождает полугруппу* $\{\mathfrak{Q}(t)\}$, *если при* $h \to 0+$

$$\frac{\mathfrak{Q}(h)-1}{h} \to \mathfrak{A}. \tag{2.6}$$

Оператор $\mathfrak{A}$ *будет называться производящим оператором*[1]).

Очевидно, что полугруппа, обладающая производящим оператором, автоматически непрерывна. Мы покажем в дальнейшем, что любая непрерывная полугруппа со сверткой имеет производящий оператор, но это утверждение ни в какой мере не очевидно.

Формально (2.6) определяет $\mathfrak{A}$ как производную $\mathfrak{Q}(t)$ при $t=0$. Существование $\mathfrak{A}$ влечет дифференцируемость и при $t>0$, так как

$$\frac{\mathfrak{Q}(t+h)-\mathfrak{Q}(t)}{h} = \frac{\mathfrak{Q}(h)-1}{h}\mathfrak{Q}(t) \to \mathfrak{A}\mathfrak{Q}(t) \tag{2.7}$$

при $h \to 0+$ и аналогично при $h \to 0-$.

Первые четыре из приводимых ниже примеров будут использованы в дальнейшем.

Примеры. а) *Обобщенная пуассоновская полугруппа.* Пусть

$$Q_t = e^{-at}\sum_{k=0}^{\infty}\frac{(at)^k}{k!}F^{k*} \tag{2.8}$$

— обобщенное пуассоновское распределение. Здесь

$$\mathfrak{Q}(h)u - u = (e^{-ah}-1)u + ahe^{-ah}\left[\mathfrak{F}u + \frac{ah}{2!}\mathfrak{F}^2u + \ldots\right]. \tag{2.9}$$

Деля обе части на h, видим, что (2.6) выполняется с $\mathfrak{A} = a(\mathfrak{F}-1)$. Таким образом, *обобщенная пуассоновская полугруппа* (2.8) *порождается оператором* $a(\mathfrak{F}-1)$ *и ее элементы мы будем записывать сокращенно в виде* $\mathfrak{Q}(t) = e^{a(\mathfrak{F}-1)t}$.

б) *Сдвиги.* Обозначим через T_a распределение, сконцентрированное в точке a, и через $\mathfrak{T}_a$ — соответствующий оператор. При фиксированном $\beta>0$ выполняется полугрупповое свойство

[1]) Мы ограничили область определения $\mathfrak{A}$ классом C^∞. Поэтому наша терминология слегка отличается от общепринятой, ср., например, книгу Хилле и Филлипса (1962).

$T_{\beta s} * T_{\beta t} = T_{\beta(t+s)}$. Кроме того, $\mathfrak{T}(\beta t)u(x) = u(x-\beta t)$. График функции $\mathfrak{T}(\beta t)u$ получается *сдвигом* графика u. Поэтому мы будем говорить о *полугруппе сдвигов*. *Производящий оператор равен* $-\beta\frac{d}{dx}$. Заметим, что он может быть представлен как предел при $h \to 0$ производящих операторов $\alpha(\mathfrak{F}-\mathbf{1})$, где $\alpha = \beta/h$ и F сосредоточено в h. Оператор $\alpha(\mathfrak{F}-\mathbf{1})$ — разностный оператор, и переход к пределу рассматривался в гл. VI, 5. Разумно обозначить эту полугруппу как $\mathfrak{T}(t) = \exp\left(-\beta t\frac{d}{dx}\right)$.

в) *Сложение производящих операторов.* Пусть $\mathfrak{A}_1$ и $\mathfrak{A}_2$ порождают полугруппы со сверткой $\{\mathfrak{Q}_1(t)\}$ и $\{\mathfrak{Q}_2(t)\}$. Тогда оператор $\mathfrak{A}_1+\mathfrak{A}_2$ *порождает полугруппу операторов* $\mathfrak{Q}(t) = \mathfrak{Q}_1(t)\mathfrak{Q}_2(t)$. [Такая полугруппа связана со сверткой распределений, соответствующих $\mathfrak{Q}_1(t)$ и $\mathfrak{Q}_2(t)$; см. теорему 2 гл. VIII, 3.] Высказанное утверждение очевидным образом следует из формулы

$$\frac{\mathfrak{Q}_1(h)\mathfrak{Q}_2(h)-\mathbf{1}}{h} = \frac{\mathfrak{Q}_1(h)-\mathbf{1}}{h} + \mathfrak{Q}_1(h)\mathfrak{A}_2 + \mathfrak{Q}_1(h)\left[\frac{\mathfrak{Q}_2(h)-\mathbf{1}}{h} - \mathfrak{A}_2\right]. \quad (2.10)$$

г) *Сдвинутые полугруппы.* Как частный случай предыдущего примера, отметим правило: если $\mathfrak{A}$ порождает полугруппу операторов $\mathfrak{Q}(t)$, связанных с распределениями Q_t, то оператор $\mathfrak{A}-\beta d/dx$ порождает полугруппу $\{\mathfrak{Q}^{\#}(t)\}$, такую, что $Q_t^{\#}(x) = Q_t(x-\beta t)$.

д) *Гамма-распределения.* Распределения с плотностями $e^{-x}x^{t-1}/\Gamma(t)$ образуют полугруппу. Здесь

$$\frac{\mathfrak{Q}(t)-\mathbf{1}}{t}u(x) = \frac{1}{\Gamma(t+1)}\int_0^\infty \frac{u(x-y)-u(x)}{y^{1-t}}e^{-y}\,dy. \quad (2.11)$$

При $t \to 0$ получаем

$$\mathfrak{A}u(x) = \int_0^\infty \frac{u(x-y)-u(x)}{y}e^{-y}\,dy. \quad (2.12)$$

Интеграл сходится, если u ограниченная и непрерывно дифференцируемая функция, так как в этом случае при фиксированном x подинтегральное выражение будет непрерывной функцией y, принимающей значение $-u'(x)$ в точке $y=0$.

(Дальнейшие примеры см. в задачах 3 и 4.)

Замечание *об уравнении Фоккера — Планка.* Рассмотрим семейство функций, определенных формулой $v(t, x) = \mathfrak{Q}(t) f(x)$. Соотношение (2.7) показывает, что для гладких f

$$\frac{\partial v}{\partial t} = \mathfrak{A} v. \tag{2.13}$$

Это — уравнение Фоккера — Планка для рассматриваемого процесса, и v — его единственное решение, удовлетворяющее начальному условию $v(0, x) = f(x)$. Уравнение (2.13) описывает процесс. Заметим, что традиционные попытки заменить (2.13) уравнением для самих переходных вероятностей Q_t приводят к ненужным усложнениям. Возьмем, например, сдвинутую обобщенную пуассоновскую полугруппу, порожденную оператором $\mathfrak{A} = \alpha(\mathfrak{F} - \mathbf{1}) - \beta \frac{d}{dx}$. Уравнение Фоккера—Планка (2.13) выполняется при любой начальной функции $f(x) = v(0, x)$, имеющей непрерывную производную. Его формальный аналог для переходных вероятностей имеет вид.

$$\frac{\partial Q_t}{\partial t} = -\beta \frac{\partial Q_t}{\partial x} - \alpha Q_t + \alpha F * Q_t. \tag{2.14}$$

Это уравнение имеет смысл, только если Q имеет плотность, и потому неприменимо к дискретным процессам. Обычные ссылки на (2.14) вместо (2.13) порождают лишь затруднения.

§ 3. Подготовительные леммы

В этом параграфе собрано вместе несколько простых лемм, на которые опирается вся теория. Несмотря на всю свою простоту, основную роль играет следующее неравенство.

Лемма 1. *Если операторы $\mathfrak{A}$ и $\mathfrak{A}^\#$ порождают полугруппы со сверткой $\{\mathfrak{Q}(t)\}$ и $\{\mathfrak{Q}^\#(t)\}$ соответственно, то при всех $t > 0$*

$$\|\mathfrak{Q}(t) u - \mathfrak{Q}^\#(t) u\| \leqslant t \|\mathfrak{A} u - \mathfrak{A}^\# u\|. \tag{3.1}$$

Доказательство. Из полугруппового свойства и основного неравенства (1.5) при $r = 1, 2, \ldots$ получаем

$$\|\mathfrak{Q}(t) u - \mathfrak{Q}^\#(t) u\| \leqslant r \left\|\mathfrak{Q}\left(\frac{t}{r}\right) u - \mathfrak{Q}^\#\left(\frac{t}{r}\right) u\right\| =$$

$$= t \left\| \frac{\mathfrak{Q}(t/r) - \mathbf{1}}{t/r} u - \frac{\mathfrak{Q}^\#(t/r) - \mathbf{1}}{t/r} u \right\|. \tag{3.2}$$

При $r \to \infty$ правая часть сходится к правой части (3.1), так что это неравенство верно. ▶

Следствие. *Различные полугруппы со сверткой не могут иметь один и тот же производящий оператор.*

Лемма 2. (*Сходимость.*) *Пусть при каждом n оператор $\mathfrak{A}_n$ порождает полугруппу со сверткой $\{\mathfrak{Q}_n(t)\}$.*

Если $\mathfrak{A}_n \to \mathfrak{A}$, *то* $\mathfrak{A}$ *порождает полугруппу со сверткой* $\{\mathfrak{Q}(t)\}$ *и* $\mathfrak{Q}_n(t) \to \mathfrak{Q}(t)$ *при каждом* $t>0$.

Доказательство. При каждом $t>0$ последовательность $\{\mathfrak{Q}_n(t)u\}$ сходится равномерно, так как по (3.1)

$$\|\mathfrak{Q}_n(t)u - \mathfrak{Q}_m(t)u\| \leqslant t\|\mathfrak{A}_n u - \mathfrak{A}_m u\|. \tag{3.3}$$

По критерию 1 гл. VIII, 3 существует оператор $\mathfrak{Q}(t)$, связанный с некоторым распределением вероятностей, такой, что $\mathfrak{Q}_n(t) \to \mathfrak{Q}(t)$. Тогда

$$\mathfrak{Q}_n(s)\mathfrak{Q}_n(t) \to \mathfrak{Q}(s)\mathfrak{Q}(t)$$

(см. теорему 2 гл. VIII, 3). Поэтому $\{\mathfrak{Q}(t)\}$ — полугруппа со сверткой. Покажем, что она порождается оператором $\mathfrak{A}$. Отметим с этой целью, что

$$\left\|\frac{\mathfrak{Q}(t)-1}{t}u - \mathfrak{A}u\right\| \leqslant \left\|\frac{\mathfrak{Q}_n(t)-1}{t}u - \mathfrak{A}u\right\| + \frac{\|\mathfrak{Q}(t)u - \mathfrak{Q}_n(t)u\|}{t}. \tag{3.4}$$

Полагая в (3.3) $m \to \infty$, видим, что

$$\|\mathfrak{Q}(t)u - \mathfrak{Q}_n(t)u\| \leqslant t\|\mathfrak{A}u - \mathfrak{A}_n u\|. \tag{3.5}$$

Следовательно, при фиксированном n и $t \to 0$ верхний предел левой части (3.4) не превосходит $2\|\mathfrak{A}u - \mathfrak{A}_n u\|$, что не превосходит ε при достаточно больших n. ▶

Последующая лемма делает по крайней мере правдоподобным то, что каждая непрерывная полугруппа имеет производящий оператор.

Лемма 3. *Пусть* $\{\mathfrak{Q}(t)\}$ — *непрерывная полугруппа со сверткой. Если для некоторой последовательности* $t_1, t_2, \ldots, \to 0$

$$\frac{\mathfrak{Q}(t_k)-1}{t_k} \to \mathfrak{A}, \tag{3.6}$$

то $\mathfrak{A}$ *порождает эту полугруппу.*

Доказательство. Обозначим левую часть $\mathfrak{A}_k$. Как было показано в примере (2, а), оператор $\mathfrak{A}_k$ порождает обобщенную пуассоновскую полугруппу и по последней лемме существует полугруппа $\{\mathfrak{Q}^{\#}(t)\}$, порождаемая $\mathfrak{A}$. Докажем, что $\mathfrak{Q}^{\#}(t) = \mathfrak{Q}(t)$. Поступая, как в (3.2), получим

$$\|\mathfrak{Q}(rt_k)u - \mathfrak{Q}^{\#}(rt_k)u\| \leqslant rt_k\left\|\mathfrak{A}_k u - \frac{\mathfrak{Q}^{\#}(t_k)-1}{t_k}u\right\|. \tag{3.7}$$

Пусть $k \to \infty$ и $r \to \infty$ так, что $rt_k \to t$. Правая часть стремится к нулю, а левая в силу (2.5) к $\|\mathfrak{Q}(t)u - \mathfrak{Q}^{\#}(t)u\|$. ▶

Эти леммы мы используем немедленно. Следующая лемма приведена здесь потому, что она представляет собой вариант леммы 2 и доказательство ее почти такое же. Мы воспользуемся лишь частным случаем $\nu_n = n$ и $t = 1$, который будет служить связующим звеном между схемами серий и полугруппами со сверткой.

Лемма 4. *Пусть при каждом n определен оператор $\mathfrak{F}_n$, соответствующий распределению вероятностей F_n. Если*

$$n(\mathfrak{F}_n - \mathbf{1}) \to \mathfrak{A}, \tag{3.8}$$

то $\mathfrak{A}$ порождает полугруппу со сверткой $\{\mathfrak{Q}(t)\}$. Если $n \to \infty$ и $\frac{\nu_n}{n} \to t$, то

$$\mathfrak{F}_n^{\nu_n} \to \mathfrak{Q}(t). \tag{3.9}$$

В частности, $\mathfrak{F}_n^n \to \mathfrak{Q}(1)$. Лемма остается верной, если n пробегает подпоследовательность $n_1, n_2, \ldots$.

Доказательство. Левая часть (3.8) порождает обобщенную пуассоновскую полугруппу [пример (2, а)], так что по лемме 2 $\mathfrak{A}$ будет производящим оператором. По основному неравенству (1.5) имеем

$$\left\| \mathfrak{F}_n^{\nu_n} u - \mathfrak{Q}\left(\frac{\nu_n}{n}\right) u \right\| \leqslant \frac{\nu_n}{n} \left\| n[\mathfrak{F}_n u - u] - n\left[\mathfrak{Q}\left(\frac{1}{n}\right) u - u\right] \right\|. \tag{3.10}$$

Если $u \in C^\infty$, то каждый из членов разности, стоящей под знаком нормы справа в (3.10), равномерно сходится к $\mathfrak{A}u$. ▶

§ 4. Случай конечных дисперсий

Полугруппы распределений с конечными дисперсиями особенно важны, а их теория столь проста, что заслуживает отдельного изложения. Многим читателям не будут интересны более общие полугруппы, а другим этот параграф даст полезный предварительный пример.

Рассмотрим полугруппу со сверткой $\{\mathfrak{Q}(t)\}$ и обозначим соответствующие распределения вероятностей через Q_t. Допустим, что Q_t имеет конечную дисперсию $\sigma^2(t)$. В силу полугруппового свойства $\sigma^2(s+t) = \sigma^2(s) + \sigma^2(t)$. Единственное положительное решение этого уравнения [1]) имеет вид $\sigma^2(t) = ct$. Предположим,

[1]) Уравнение $\varphi(s+t) = \varphi(s) + \varphi(t)$ называют уравнением Гамеля. Полагая $u(t) = e^{\varphi(t)}$, получаем $u(s+t) = u(s)u(t)$. В этой форме уравнение встречалось нам несколько раз и изучалось в I, гл. XVII, 6 Математическое ожидание Q_t также удовлетворяет уравнению Гамеля; следовательно, *или оно* имеет вид mt, или крайне причудливо; см. § 5 а.

что Q_t имеет нулевое математическое ожидание. Тогда имеет место тождество

$$\frac{\mathfrak{Q}(t)-1}{t}u(x)=c\int_{-\infty}^{+\infty}\frac{u(x-y)-u(x)+yu'(x)}{y^2}\Omega_t\{dy\}, \qquad (4.1)$$

где Ω_t — *распределение вероятностей*, определенное равенством

$$\Omega_t\{dy\}=\frac{1}{ct}y^2Q_t\{dy\}. \qquad (4.2)$$

Так как мы рассматриваем только $u\in C^\infty$, подинтегральное выражение (при фиксированном x) будет непрерывной функцией от y, обращающейся в нуль на бесконечности. По теореме о выборе существует последовательность $t_1, t_2, \ldots \to 0$, такая, что если t пробегает ее, то $\Omega_t\{dy\}\to\Omega\{dy\}$. Здесь Ω — распределение вероятностей, которое может быть и несобственным.

Определим оператор $\mathfrak{A}$ формулой

$$\mathfrak{A}u(x)=c\int_{-\infty}^{+\infty}\frac{u(x-y)-u(x)+yu'(x)}{y^2}\Omega\{dy\}. \qquad (4.3)$$

При t, пробегающем последовательность $t_1, t_2, \ldots$, правая часть (4.1) стремится к $\mathfrak{A}u(x)$, и эта сходимость равномерна по x, так как подинтегральное выражение равномерно непрерывно и равномерно мало на бесконечности. По лемме 3.3 полугруппа $\{\mathfrak{Q}(t)\}$ имеет производящий оператор $\mathfrak{A}$, определенный по (4.3).

Мы доказали, таким образом, существование производящего оператора и нашли для него представление (4.3). Докажем, что оно единственно. С этой целью отметим, что для функций вида

$$u(x)=1+\frac{x^2}{1+x^2}f(-x) \qquad (4.4)$$

из (4.3) вытекает

$$\mathfrak{A}u(0)=c\int_{-\infty}^{+\infty}\frac{f(y)}{1+y^2}\Omega\{dy\}. \qquad (4.5)$$

Поэтому знание $\mathfrak{A}u$ для всех $u\in C^\infty$ определяет однозначно меру $(1+y^2)^{-1}\Omega\{dy\}$, а вместе с тем и меру Ω.

Как следствие доказанной единственности представления получаем, что предельное распределение Ω не зависит от выбора последовательности $\{t_k\}$, т. е. $\Omega_t\{dy\}\to\Omega\{dy\}$ при любом стремлении $t\to 0$.

Мы покажем, что Ω — собственное распределение вероятностей и что каждый оператор вида (4.3) является производящим оператором. Доказательство использует анализ двух частных случаев, которые содержатся в следующих примерах.

Примеры. а) *Нормальные полугруппы.* При $Q_t(x)=\mathfrak{N}(x/\sqrt{t})$ распределения $\Omega_t\{dy\}$ сходятся к распределению вероятностей, сосредоточенному в нуле. При $y=0$ значение подинтегральной функции в (4.3) равно $\frac{1}{2}u''(x)$, так что в этом случае $\mathfrak{A}=\frac{1}{2}(d^2/dx^2)$.

б) *Обобщенные пуассоновские полугруппы.* Пусть c — положительная постоянная, и пусть Ω — мера с полной массой $\omega>0$, сосредоточенной на двух интервалах $|y|>\eta>0$. Существует единственная вероятностная мера F вида $\alpha F\{dy\}=cy^{-2}\Omega\{dy\}$. Обозначим два первых момента F через m_1 и m_2 соответственно. Оператор $\mathfrak{A}$ из (4.3) совпадает с $\alpha(\mathfrak{F}-\mathbf{1})+\alpha m_1\frac{d}{dx}$. Теперь, $\alpha(\mathfrak{F}-\mathbf{1})$ порождает полугруппу обобщенных пуассоновских распределений с математическим ожиданием $\alpha m_1 t$ и дисперсией $\alpha m_2 t$. Добавление члена $\alpha m_1\frac{d}{dx}$ к производящему оператору делает математическое ожидание равным нулю и не меняет дисперсии [пример (2, г)]. Таким образом, в перечисленных условиях оператор $\mathfrak{A}$ из (4.3) порождает полугруппу распределений Q_t с нулевым математическим ожиданием и дисперсией $\alpha m_2 t=c\omega t$. ▶

Теперь мы в состоянии сформулировать основную теорему.

Теорема. *Пусть Q_t имеет нулевое математическое ожидание и дисперсию ct. Полугруппа со сверткой $\{\mathfrak{Q}(t)\}$ имеет производящий оператор $\mathfrak{A}$ вида* (4.3), *где Ω — собственное распределение вероятностей. Представление* (4.3) *единственно. Обратно, каждый оператор такого вида порождает полугруппу распределений со сверткой, причем эти распределения имеют нулевое математическое ожидание и дисперсию ct.*

Доказательство. Мы установили существование производящего оператора вида (4.3), но относительно Ω знаем лишь, что соответствующая полная масса $\omega\leqslant 1$. Остается доказать, что если Ω имеет полную массу ω, то оператор $\mathfrak{A}$ из (4.3) порождает такую полугруппу, что Q_t имеет нулевое математическое ожидание и дисперсию $\leqslant c\omega t$.

Пусть $\mathfrak{A}_\eta$ — оператор, получаемый из (4.3) ограничением области интегрирования до интервалов $|y|>\eta$ плюс точку нуль (если в ней есть атом). Тогда $\mathfrak{A}_\eta$ представляется в виде суммы двух операторов, описанных выше типов, и по правилу сложения производящих операторов [пример (2, в)] оператор $\mathfrak{A}_\eta$ по-

рождает такую полугруппу $\{\mathfrak{Q}_\eta(t)\}$, что связанные с ней распределения $Q_{t,\eta}$ имеют нулевое математическое ожидание и дисперсию $\leqslant c\omega t$. Соответственно $\mathfrak{A}$ порождает полугруппу $\{\mathfrak{Q}(t)\}$, такую, что $Q_{t,\eta} \to Q_t$. Так как вторые моменты $Q_{t,\eta}$ ограничены константой $c\omega t$, то Q_t имеет нулевое математическое ожидание и дисперсию $\leqslant c\omega t$ [пример (1, д), гл. VIII]. Доказательство закончено. ▶

Пример. в) В примере (2, д) мы имеем $\Omega\{dy\} = ye^{-y}dy$. Для того чтобы придать оператору (2.12) каноническую форму (4.3), добавим член $yu'(x)$ к подинтегральному выражению и добавим $u'(x)$ перед интегралом. Мы видим, что Q_t имеет математическое ожидание t. ▶

§ 5. Основная теорема

В этом параграфе $\{\mathfrak{Q}(t)\}$ обозначает произвольную непрерывную полугруппу со сверткой. Соответствующие функции распределения будут снова обозначаться через Q_t. Предыдущее доказательство опиралось на существование такой меры Ω, что

$$\frac{1}{t}\, y^2 Q_t\{dy\} \to \Omega\{dy\}. \tag{5.1}$$

Такая мера существует всегда, но не обязана быть конечной. Вместо этого она удовлетворяет условию

$$\int_{-\infty}^{+\infty} \frac{1}{1+y^2}\, \Omega\{dy\} < \infty, \tag{5.2}$$

которое обеспечивает не слишком быстрый рост меры $\Omega\{\overline{-x, x}\}$.

Существование предела в (5.1) легко установить, допуская существование производящего оператора (задача 8), но сам этот факт доказать более трудно. Чтобы доказать его, вернемся к обозначениям § 1. Рассмотрим при каждом n сумму $\mathbf{S}_n = \mathbf{X}_{1,n} + \ldots + \mathbf{X}_{n,n}$ n взаимно независимых случайных величин с одним и тем же распределением F_n. Тогда $\mathfrak{F}_n^n$ будет оператором, связанным с суммой $\mathbf{S}_n$. Нас интересует в первую очередь случай $\mathfrak{F}_n = \mathfrak{Q}(1/n)$. Однако в связи с предельными теоремами мы будем использовать и произвольные $\mathfrak{F}_n$. Задача состоит в том, чтобы изучить асимптотическое поведение оператора $n[\mathfrak{F}_n - \mathbf{1}]$. Мы воспользуемся представлением, аналогичным (4.1), но при этом будем вынуждены ввести урезанные математические ожидания.

Определим при фиксированном $s>0$ *урезающую функцию* τ_s равенствами

$$\tau_s(x)=\begin{cases} x & |x|\leqslant s,\\ s & \text{при } x\geqslant s,\\ -s & x\leqslant -s.\end{cases} \tag{5.3}$$

Нашей отправной точкой будет *тождество*

$$n(\mathfrak{F}_n-\mathbf{1})u(x)=$$

$$=\int_{-\infty}^{+\infty}\frac{u(x-y)-u(x)+u'(x)\tau_s(y)}{y^2}\,ny^2F_n\{dy\}-b_nu'(x), \tag{5.4}$$

где

$$b_n=n\int_{-\infty}^{+\infty}\tau_s(y)\,F_n\{dy\}. \tag{5.5}$$

Для гладких u (которые мы только и рассматриваем) дробь под интегралом является (при фиксированном x) непрерывной функцией от y, принимающей значение $\frac{1}{2}u''(x)$ в нуле. Для $|y|<s$ числитель ограничен величиной $y^2\|u''\|$, а для $|y|\geqslant s$ — величиной $2\|u\|+s\|u'\|$.

Следующая ниже лемма легко выводится из леммы о компактности для схемы серий. Доказательство мы отложим до § 7.

Лемма. *Если случайные величины* $\{\mathbf{S}_n\}$ *имеют собственное предельное распределение, то существует мера* Ω, *удовлетворяющая* (5.2), *и последовательность* $n_1, n_2, \ldots$ *целых чисел, такие, что при* n, *пробегающем эту последовательность,*

$$ny^2F_n\{dy\}\to\Omega\{dy\}. \tag{5.6}$$

Сверх того, для каждого $\varepsilon>0$ *найдется такое число* a, *что при всех* n

$$n[1-F_n(a)+F_n(-a)]<\varepsilon. \tag{5.7}$$

Наконец, последовательность $\{b_n\}$ *ограничена.*

Так как $\{b_n\}$ ограничены, то, не уменьшая общности, мы можем предположить, что при n, пробегающем последовательность $n_1, n_2, \ldots$,

$$b_n\to b. \tag{5.8}$$

Определим оператор $\mathfrak{A}^{(0)}$ формулой

$$\mathfrak{A}^{(0)}u(x)=\int_{-\infty}^{+\infty}\frac{u(x-y)-u(x)+u'(x)\tau_s(y)}{y^2}\,\Omega\{dy\}. \tag{5.9}$$

[Интеграл сходится в силу (5.2) и неравенств, указанных после формулы (5.5).] В условиях леммы вклад области $|y|>a>s$ в интеграл (5.4) не превосходит $2\|u\|\varepsilon+s\|u'\|\varepsilon$, если только a достаточно велико. Вклад области $|y|\leqslant a$ сходится к соответствующему интегралу, взятому по отношению к Ω. Это во всяком случае верно, если Ω не имеет атомов в точках $\pm a$. Полагая $a\to\infty$, мы видим, что

$$n[\mathfrak{F}_n - \mathbf{1}]\,u \to \mathfrak{A}^{(0)}u - bu', \tag{5.10}$$

где сходимость равномерна. Как $\mathfrak{A}^{(0)}$, так и b зависят от постоянной s («высоты урезания»), однако левая часть не зависит от s. Поэтому изменение $\mathfrak{A}^{(0)}$, возникающее при изменении s, компенсируется соответствующим изменением b.

Теорема 1. *Непрерывная полугруппа со сверткой* $\{\mathfrak{Q}(t)\}$ *обладает производящим оператором* $\mathfrak{A}$, *причем* $\mathfrak{A}$ *имеет вид*

$$\mathfrak{A} = \mathfrak{A}^{(0)} - b\frac{d}{dx}, \tag{5.11}$$

где $\mathfrak{A}^{(0)}$ *определяется по* (5.3) *с мерой* Ω, *удовлетворяющей* (5.2).

Обратно, каждый оператор $\mathfrak{A}$ *такого вида порождает непрерывную полугруппу со сверткой* $\{\mathfrak{Q}(t)\}$. *Мера* Ω *единственна.*

Доказательство. Если $\mathfrak{F}_n = \mathfrak{Q}(1/n)$, то суммы $\mathbf{S}_\eta$ имеют распределение Q_1 при любом n. Поэтому применима предыдущая лемма и, согласно (5.10),

$$\frac{1}{t}[\mathfrak{Q}(t) - \mathbf{1}] \to \mathfrak{A} \tag{5.12}$$

при t, пробегающем последовательность $\frac{1}{n_1}, \frac{1}{n_2}, \ldots$. В силу леммы 3.3 $\mathfrak{A}$ служит производящим оператором для $\{\mathfrak{Q}(t)\}$.

Мы знаем, что каждый раз, когда мера Ω *конечна*, $\mathfrak{A}$ порождает полугруппу (с конечными дисперсиями). Из леммы о сходимости (лемма 3.2) мы выводим, что все наши операторы суть производящие операторы. Единственность получается так же, как и в § 4, поскольку для функций вида (4.4) верно (4.5) (без множителя c). Следовательно, знание $\mathfrak{A}u$ определяет Ω. Доказательство закончено. ▶

Мы в состоянии теперь сформулировать различные характеристические свойства безгранично делимых распределений, лежащие в основе всей их теории. Мы соберем их в одной теореме, в которой утверждение (iv) наиболее глубокое и неожиданное. Исторические замечания см. в гл. VI. 3.

Теорема 2. *Следующие классы распределений вероятностей совпадают.*

(i) *Распределения, связанные с операторами* $\mathfrak{Q}(t)$ *непрерывных полугрупп со сверткой (другими словами, распределения приращений в процессах со стационарными независимыми приращениями).*

(ii) *Пределы последовательностей обобщенных пуассоновских распределений.*

(iii) *Безгранично делимые распределения и их пределы.*

(iv) *Предельные распределения для сумм* $\mathbf{S}_n$ *в схемах серий* $\{\mathbf{X}_{k,n}\}$, *где* $X_{1,n}$, $\mathbf{X}_{2,n}$, ... *имеют одно и то же распределение.*

Доказательство. По теореме 1 каждая непрерывная полугруппа имеет производящий оператор $\mathfrak{A}$, и по самому определению $\mathfrak{A}$ является пределом операторов $t^{-1}(\mathfrak{Q}(t)-\mathbf{1})$. Как было показано в примере (2, а), эти операторы порождают обобщенные пуассоновские полугруппы Поэтому из леммы о сходимости 3.2 следует, что каждый элемент Q_t непрерывной полугруппы есть предел последовательности обобщенных пуассоновских распределений. Таким образом, (i) содержится в (ii), а (ii) (тривиальным образом) — в (iii). Далее, любое безгранично делимое распределение $G^{(n)}$ можно рассматривать как распределение суммы $\mathbf{S}_n=\mathbf{X}_{1,n}+\ldots+\mathbf{X}_{n,n}$ n одинаково распределенных независимых случайных величин. Отсюда вытекает, что предел G последовательности безгранично делимых распределений $G^{(n)}$ будет автоматически пределом распределений сумм по строкам $\mathbf{S}_n$ в подходящей схеме серий. Поэтому (iii) содержится в (iv). Наконец, если суммы по строкам $\mathbf{S}_n$ в некоторой схеме серий имеют предельное распределение G, то, согласно последней лемме, выполняется (5.10) (по крайней мере для значений n из некоторой подпоследовательности $n_1, n_2, \ldots$). По лемме 3.4 предел G совпадает с элементом Q_1 полугруппы, порождаемой оператором, стоящим в правой части (5.10). Таким образом, (iv) содержится в (i). Доказательство закончено. ▶

Из теоремы 1 следует, что с точностью до центрирующего слагаемого полугруппа $\{\mathfrak{Q}(t)\}$ определяется мерой Ω, которая в свою очередь определяется по (5.1). Иногда более удобно описывать полугруппу в терминах функций ψ^+ и ψ^-, задаваемых на правой и левой полуосях соответственно формулами

$$\psi^+(x)=\int_x^\infty \frac{1}{y^2}\,\Omega\{dy\},\quad \psi^-(-x)=\int_{-\infty}^{-x}\frac{1}{y^2}\,\Omega\{dy\},\quad x>0. \tag{5.13}$$

(Для определенности мы можем считать интервалы интегрирования замкнутыми.) Эти функции определяют Ω однозначно

с точностью до возможного атома в нуле. Из (5.1) следует, что

$$\frac{1}{t}[1-Q_t(x)]\to\psi^+(x),\quad \frac{1}{t}Q_t(-x)\to\psi^-(-x) \tag{5.14}$$

во всех точках непрерывности. Мы фиксируем этот результат как

Следствие. *Для каждой непрерывной полугруппы со сверткой выполняются соотношения* (5.1) *и* (5.14).

Пример. а) *Полугруппа Коши.* Пусть Q_t имеет плотность $\pi^{-1}t[t^2+x^2]^{-1}$. Полугрупповое свойство выполняется, и, очевидно,

$$\psi^+(x)=\psi^-(-x)=(\pi x)^{-1}.$$

Мера Ω определяется равенством $\Omega\{dy\}=\pi^{-1}dy$.

Применение к случайным процессам. Пусть случайные величины $\mathbf{X}(t)$ описывают процесс со *стационарными независимыми приращениями* (гл. VI, 4). Условимся интерпретировать Q_t как распределение приращения $\mathbf{X}(t+s)-\mathbf{X}(s)$. Рассмотрим временной интервал $s,\ s+1$ единичной длины. Разделим его точками $s=s_0<s_1<\ldots<s_n=s+1$ на подинтервалы длины n^{-1}. Тогда $\mathbf{P}\{\mathbf{X}(s_k)-\mathbf{X}(s_{k-1})>x_k\}=1-Q_{1/n}(x)$, так что $n[1-Q_{1/n}(x)]$ равно математическому ожиданию числа интервалов $s_{k-1},\ s_k$, приращения на которых $>x$. При $n\to\infty$ это математическое ожидание сходится к $\psi^+(x)$. Чтобы упростить дальнейшее обсуждение, предположим, что пределы $\mathbf{X}(t+)$ и $\mathbf{X}(t-)$ существуют при всех t и что $\mathbf{X}(t)$ лежит между ними. Пусть $s_{k-1},\ s_k$ — тот интервал нашего разбиения, который содержит точку t. Для достаточно больших n приращение $\mathbf{X}(s_k)-\mathbf{X}(s_{k-1})$ будет близко к скачку $\mathbf{X}(t+)-\mathbf{X}(t-)$ и интуитивно ясно, что предел $\psi^+(x)$ *представляет собой математическое ожидание числа моментов t (в единицу времени), для которых* $\mathbf{X}(t+)-\mathbf{X}(t-)>x$. Это рассуждение можно сделать строгим, но мы не будем входить в детали. Из сказанного следует, что математическое ожидание числа разрывов будет нулем, только если $\psi^+(x)=0$ и $\psi^-(-x)=0$ для всех $x>0$. В последнем случае Ω сосредоточено в нуле, т. е. приращения $\mathbf{X}(t+s)-\mathbf{X}(s)$ будут нормально распределены. Для такого процесса траектории непрерывны с вероятностью единица (теорема П. Леви́ и Н. Винера). Итак, *траектории непрерывны с вероятностью единица тогда и только тогда, когда процесс нормален.*

В качестве второй иллюстрации мы рассмотрим *обобщенный пуассоновский процесс* (2.8). Математическое ожидание числа скачков за единицу времени равно α, а вероятность получить скачок, превосходящий $x>0$, равна $1-F(x)$. Таким образом,

$\alpha[1-F(x)]$ есть математическое ожидание числа скачков $>x$ в полном согласии с нашими интуитивными доводами.

5 *а. Разрывные полугруппы

Естественно поставить вопрос о том, существуют ли разрывные полугруппы. Задача не имеет никакой практической ценности, но ответ до некоторой степени любопытен: *каждая полугруппа со сверткой* $\{\mathfrak{Q}(t)\}$ *лишь центрированием отличается от некоторой непрерывной полугруппы* $\{\mathfrak{Q}^{\#}(t)\}$. В частности, *если распределения* Q_t *симметричны, то полугруппа обязательно непрерывна.* В общем случае существует такая функция φ, что распределения $Q_t{}^{\#}$, определенные как $Q_t(x+\varphi(t))$, *связаны с некоторой непрерывной полугруппой.* Функция φ необходимо должна удовлетворять уравнению

$$\varphi(t+s)=\varphi(t)+\varphi(s). \tag{5.15}$$

Это известное уравнение Гамеля [1]), непрерывными решениями которого могут быть только функции вида ct. Более того, любая бэровская функция, удовлетворяющая (5.15), должна быть линейной. Другие решения слишком причудливы. Например, нелинейное решение принимает в любом интервале сколь угодно большие и сколь угодно малые значения. Его невозможно получить аналитически последовательными предельными переходами. Короче говоря, уместно спросить, что в точности означает его «существование».

Вернемся на землю и рассмотрим произвольную полугруппу со сверткой $\{\mathfrak{Q}(t)\}$ и схему серий $\{\mathbf{X}_{k,n}\}$, связанную с распределениями $Q_{1/n}$. Суммы по строкам имеют одно и то же распределение Q_1. Поэтому мы можем воспользоваться последней леммой и извлечь такую подпоследовательность $n_1, n_2, \ldots$, что когда n пробегает ее, $n[\mathfrak{Q}(1/n)-1]\to\mathfrak{A}^{\#}$, где $\mathfrak{A}^{\#}$ — производящий оператор некоторой непрерывной полугруппы $\{\mathfrak{Q}^{\#}(t)\}$. Мы можем выбрать n_k вида 2^{ν}. Неравенство (3.4) тогда показывает, что $\mathfrak{Q}(t)=\mathfrak{Q}^{\#}(t)$ при всех t, которые при каком-либо k кратны $1/n_k$, т. е. при всех t вида $t=a2^{-\nu}$ и ν — целые. Таким образом, *всегда существует непрерывная полугруппа* $\{\mathfrak{Q}^{\#}(t)\}$, *такая, что* $\mathfrak{Q}(t)=\mathfrak{Q}^{\#}(t)$ *при всех* t *из некоторого всюду плотного множества* Σ.

Мы можем теперь доказать исходное утверждение. Выберем $\varepsilon_n>0$ так, что $t+\varepsilon_n$ лежит в Σ. Тогда

$$\mathfrak{Q}^{\#}(t+\varepsilon_n)=\mathfrak{Q}(t+\varepsilon_n)=\mathfrak{Q}(t)\,\mathfrak{Q}(\varepsilon_n). \tag{5.16}$$

При $\varepsilon_n\to 0$ левая часть сходится к $\mathfrak{Q}^{\#}(t)$ и потому достаточно доказать, что если $\mathfrak{Q}(\varepsilon_n)\to\mathfrak{F}$, то распределение F сосредоточено в одной точке. Выберем точки h_n в Σ так, что $0<\varepsilon_n<h_n$ и $h_n\to 0$. Тогда $\mathfrak{Q}^{\#}(h_n)=\mathfrak{Q}(h_n-\varepsilon_n)\,\mathfrak{Q}(\varepsilon_n)$. Левая часть сходится к единичному оператору, так что F в самом деле может иметь только одну точку роста.

§ 6. Пример: устойчивые полугруппы

Полугруппа $\{\mathfrak{Q}(t)\}$ называется *устойчивой*, если соответствующие распределения имеют вид

$$Q_t(x)=G(\lambda_t(x-\beta_t)), \tag{6.1}$$

[1]) См. примечание на стр. 358.

где $\lambda_t \geqslant 0$ и β_t — постоянные, непрерывно зависящие от t, а G — фиксированное распределение. Очевидно, что G является устойчивым распределением в смысле гл. VI, 1. Теория устойчивых полугрупп излагается здесь главным образом как иллюстрация результатов предыдущего параграфа. Кроме того, мы выпишем их производящие операторы. Результаты этого параграфа независимо выводятся (обходным путем) в § 8.

В силу предположенной непрерывности λ_t и β_t полугруппа (если она существует) будет непрерывна. Кроме того, при $t \to 0$ $\lambda_t \to \infty$ и $\beta_t \to 0$. Первое из соотношений (5.14) принимает форму

$$\frac{1 - G(\lambda_t (x - \beta_t))}{t} \to \psi^+(x), \quad x > 0. \tag{6.2}$$

Так как функция G монотонна и $\beta_t \to 0$, то это соотношение остается верным, если отбросить β_t. Но после этого отбрасывания (6.2) сводится к соотношению, определяющему правильность изменения функции. Тогда по теореме из гл. VIII, 8

$$\psi^+(x) = c^+ x^{-\alpha}, \tag{6.3}$$

где $c^+ \geqslant 0$ — постоянная. Здесь $\alpha > 0$, так как $\psi^+(\infty) = 0$. Отсюда вытекает, что при $y > 0$ $\Omega\{dy\} = c^+ \alpha y^{-\alpha+1} dy$. Далее $\alpha < 2$, так как конечные интервалы должны иметь конечную меру.

Рассмотрим теперь каноническую форму (5.9) производящих операторов. Отвлекаясь от тривиального центрирования, видим, что вклад правой полуоси представляется оператором $c^+ \mathfrak{A}_\alpha^+$, где

$$\mathfrak{A}_\alpha^+ u(x) = \alpha \int_0^\infty \frac{u(x - y) - u(x)}{y^{\alpha+1}} dy \tag{6.4}$$

при $0 < \alpha < 1$, и

$$\mathfrak{A}_\alpha^+ u(x) = \alpha \int_0^\infty \frac{u(x - y) - u(x) + y u'(x)}{y^{\alpha+1}} dy \tag{6.5}$$

при $1 < \alpha < 2$. Наконец, при $\alpha = 1$

$$\mathfrak{A}_1^+ u(x) = \int_0^\infty \frac{u(x - y) - u(x) + u'(x)\, \tau_1(y)}{y^2} dy. \tag{6.6}$$

Соотношение типа (6.2) выполняются также и для левых хвостов и для суммы правого и левого хвостов. Ясно, что левому хвосту соответствует оператор $\mathfrak{A}_\alpha^-$, аналогичный $\mathfrak{A}_\alpha^+$, с тем же самым α. Комбинируя эти операторы, введем оператор

$$\mathfrak{A}_\alpha = c^+ \mathfrak{A}_\alpha^+ + c^- \mathfrak{A}_\alpha^-. \tag{6.7}$$

Функции ψ^+ и ψ^- определяют меру Ω с точностью до атома в нуле. Мы показали, что *производящие операторы устойчивых полугрупп необходимо имеют вид*

$$\mathfrak{A} = \mathfrak{A}_\alpha + \gamma \frac{d^2}{dx^2} + \beta \frac{d}{dx}. \tag{6.8}$$

При $c^+ = c^- = 0$ получим $\mathfrak{A}_\alpha = 0$, так что $\mathfrak{A}$ порождает *нормальную полугруппу*. Как легко вывести (это будет сделано в § 8), во всех других случаях $\gamma = 0$. Следовательно, если исключить нормальные полугруппы и произвольное центрирование, то (6.7) дает наиболее общую форму производящего оператора.

Остается установить, что полугруппы, порождаемые $\mathfrak{A}_\alpha$, в самом деле устойчивы. Мы покажем, что *при* $\alpha \neq 1$ *оператор* $\mathfrak{A}_\alpha$ *порождает строго устойчивую полугруппу вида*

$$Q_s(x) = G\left(\frac{x}{s^{1/\alpha}}\right), \tag{6.9}$$

а при $\alpha = 1$ — *устойчивую полугруппу вида* [1])

$$Q_t(x) = G\left(\frac{x}{t} - (c^+ - c^-)\log t\right). \tag{6.10}$$

Положим $Q^\#(x) = Q_t(x/\rho)$, где $\rho > 0$ — постоянная. Тогда Q_t и $Q_t^\#$ отличаются лишь масштабным параметром, так что $\{\mathfrak{Q}^\#(t)\}$ — снова непрерывная полугруппа. Полагая $u_\rho(x) = u(\rho x)$, имеем $\mathfrak{Q}^\#(t)\,u(x) = \mathfrak{Q}(t)\,u_\rho(x/\rho)$. Поэтому производящий оператор $\mathfrak{A}^\#$ получается простой заменой x на ρx в интегралах, определяющих $\mathfrak{A}$. [Таким образом, $u'(x) = \frac{du}{dx}$ заменяется на $(1/\rho)\cdot u'(\rho x)$]. При $\alpha \neq 1$ подстановка $y = z/\rho$ показывает, что $\mathfrak{A}_\alpha^\# = \rho^\alpha\,\mathfrak{A}_\alpha$. Оператор $\mathfrak{A}_\alpha^\#$ порождает полугруппу $\{\mathfrak{Q}(\rho^\alpha t)\}$, и ввиду единственности производящих операторов отсюда вытекает, что при всех t $\mathfrak{Q}^\#(\rho^{-\alpha}t) = \mathfrak{Q}(t)$. При $t = 1$ и $\rho^\alpha = s$ получаем (6.9).

При $\alpha = 1$ наличие дополнительного члена под интегралом приводит к видоизмененной форме $\mathfrak{A}_1^\# = \rho\mathfrak{A}_1 + \rho\gamma(\rho)\frac{d}{dx}$, где

$$\gamma(\rho) = (c^+ - c^-)\int_0^\infty [\rho\tau_1(y/\rho) - \tau_1(y)]\,y^{-2}\,dy = (c^+ - c^-) = \int_1^\rho y^{-1}\,dy. \tag{6.11}$$

Теперь рассуждения, использованные при $\alpha \neq 1$, приводят к (6.10).

[1]) Эти утверждения следуют и из результатов § 8, но их прямая проверка поучительна.

§ 7. Схемы серий

Пусть при каждом n величины $\mathbf{X}_{1,n}, \ldots, \mathbf{X}_{n,n}$ взаимно независимы, имеют одно и то же распределение F_n, и пусть $\mathbf{S}_n = \mathbf{X}_{1,n} + \ldots + \mathbf{X}_{n,n}$. Предельные теоремы для распределений сумм $\mathbf{S}_n$ опираются на лемму о компактности, которая уже была использована в § 5. Для ее доказательства мы используем *урезающую функцию* τ_s, как в (5.3), и положим

$$\mathbf{X}'_{k,n} = \tau_s(\mathbf{X}_{k,n}), \quad \mathbf{X}_{k,n} = \mathbf{X}'_{k,n} + \mathbf{X}''_{k,n}. \tag{7.1}$$

Здесь $\mathbf{X}'_{k,n}$ — уже знакомые нам величины, получающиеся урезанием на уровне $\pm s$, но зависимость от s в обозначениях не подчеркивается. Мы рассмотрим наряду с $\{\mathbf{X}_{k,n}\}$ вспомогательные схемы серий $\{\mathbf{X}'_{k,n}\}$ и $\{\mathbf{X}''_{k,n}\}$. Для них суммы по строкам мы обозначим $\mathbf{S}'_n$ и $\mathbf{S}''_n$, так что $\mathbf{S}_n = \mathbf{S}'_n + \mathbf{S}''_n$.

Напомним определение гл. VIII, 2, по которому последовательность $\{\mathbf{S}_n\}$ называется *стохастически ограниченной*, если каждому $\varepsilon > 0$ соответствует такое a, что при всех n

$$\mathbf{P}\{|\mathbf{S}_n| \geqslant a\} < \varepsilon. \tag{7.2}$$

Стохастическая ограниченность есть необходимое условие для *собственной* сходимости распределений $\mathbf{S}_n$. Для большей ясности мы установим интересующий нас критерий сначала для симметричных распределений.

Лемма 1. (*Симметричный случай.*) *Пусть распределения F_n симметричны. Последовательность $\{\mathbf{S}_n\}$ стохастически ограничена тогда и только тогда, когда существуют числа M_s и a_ε, такие, что*

$$n \operatorname{Var}(\mathbf{X}'_{j,n}) < M_s \tag{7.3}$$

и

$$n[1 - F_n(a) + F_n(-a)] < \varepsilon, \quad a > a_\varepsilon. \tag{7.4}$$

Доказательство. Согласно неравенству гл. V, (5.11), вероятность в (7.2) не меньше $\frac{1}{2}(1 - e^{-\eta})$, где η обозначает левую часть (7.4). Поэтому условие (7.4) необходимо для стохастической ограниченности $\{\mathbf{S}_n\}$. Начиная с этого места мы предположим, что оно выполнено.

При $s > a$ число членов среди $\mathbf{X}''_{1,n}, \ldots, \mathbf{X}''_{n,n}$, которые отличны от нуля, есть биномиальная случайная величина с математическим ожиданием $< \varepsilon$, и потому последовательность $\{\mathbf{S}''_n\}$ стохастически ограничена. Следовательно, $\{\mathbf{S}_n\}$ стохастически ограничена или не ограничена одновременно с $\{\mathbf{S}'_n\}$.

Положим $\sigma_n^2 = \operatorname{Var}(\mathbf{S}_n')$. Если $\sigma_n \to \infty$, то к величинам $\mathbf{X}_{k,n}'/\sigma_n$ применима центральная предельная теорема примера (1, а). В этом случае распределение $\mathbf{S}_n'/\sigma_n$ сходится к $\mathfrak{N}$ и последовательность $\{\mathbf{S}_n'\}$ *не* будет стохастически ограниченной. Это рассуждение, примененное к подпоследовательностям $\{\mathbf{S}_n'\}$, показывает, что последовательность $\{\sigma_n\}$ должна быть ограничена. Поэтому (7.3) необходимо при $s > a$, а, следовательно, и при всех меньших s. Обратно, если (7.3) верно, то по неравенству Чебышева $\mathbf{P}\{|\mathbf{S}_n'| > c\} < \varepsilon$ при всех c, для которых $c^2 > M_s/\varepsilon$. Поэтому последовательность $\{\mathbf{S}_n'\}$ стохастически ограничена. ▷

Из этой леммы мы выведем следующий общий критерий, в котором принято сокращение

$$\beta_n = \mathbf{E}(\mathbf{X}_{1,n}'). \tag{7.5}$$

Лемма 2. (*Компактность.*) *Предположим, что*[1]) $\mathbf{X}_{1,n} \xrightarrow{\mathrm{p}} 0$. *Для существования констант* b_n, *таких, что* $\{\mathbf{S}_n - b_n\}$ *стохастически ограничена, необходимы и достаточны условия леммы* 1. *Если они выполнены, то* $\{\mathbf{S}_n - n\beta_n\}$ *стохастически ограничена,* $\beta_n \to 0$ *и потому* $n^{-1}\mathbf{S}_n \xrightarrow{\mathrm{p}} 0$.

Доказательство. Как обычно, мы обозначим через ${}^0\mathbf{X}_{k,n}$ случайную величину, полученную симметризацией $\mathbf{X}_{k,n}$. Если $\{\mathbf{S}_n - b_n\}$ остается стохастически ограниченной, то это же верно и для сумм по строкам в симметризованной схеме серий. Так как $\mathbf{X}_{1,n} \xrightarrow{\mathrm{p}} 0$, то условие (7.4) для $\{\mathbf{X}_{k,n}\}$ эквивалентно аналогичному условию для $\{{}^0\mathbf{X}_{k,n}\}$ и потому необходимо.

Так же как и в предыдущем доказательстве, это условие обеспечивает стохастическую ограниченность сумм $\{\mathbf{S}_n''\}$. Таким образом, если $\{\mathbf{S}_n - b_n\}$ стохастически ограничена, то этим свойством обладает и $\{\mathbf{S}_n' - b_n\}$, и суммы по строкам в симметризованной схеме. Так как $\operatorname{Var}({}^0\mathbf{X}_{1,n}') = 2\operatorname{Var}(\mathbf{X}_{1,n}')$, то условие ограниченности (7.3) снова является необходимым. Если оно выполнено, то в силу неравенства Чебышева $\{\mathbf{S}_n' - \mathbf{E}(\mathbf{S}_n')\}$ стохастически ограничена. Этим доказано первое утверждение. Очевидно, что $\mathbf{X}_{1,n} \xrightarrow{\mathrm{p}} 0$ влечет $\mathbf{X}_{1,n}' \xrightarrow{\mathrm{p}} 0$ и вместе с тем $\beta_n \to 0$. ▷

Лемма 3. *Если* $\{\mathbf{S}_n\}$ *стохастически ограничена, то* $\mathbf{X}_{1,n} \xrightarrow{\mathrm{p}} 0$ *при* $n \to \infty$.

[1]) Это означает что $\mathbf{P}\{|\mathbf{X}_{1,n}| > \eta\} \to 0$ при каждом $\eta > 0$, см. гл. VIII, 2.

Доказательство. Допустим, что $\{\mathbf{S}_n\}$ стохастически ограничена и что μ_n — медиана $\mathbf{X}_{1,n}$. Суммы по строкам для симметризованной схемы серий $\{{}^0\mathbf{X}_{k,n}\}$ стохастически ограничены, и по лемме 1 ${}^0\mathbf{X}_{k,n} \xrightarrow{p} 0$. В силу неравенства симметризации V (5.8) отсюда следует, что $\mathbf{X}_{k,n} - \mu_n \xrightarrow{p} 0$. Поэтому к схеме серий $\{\mathbf{X}_{k,n} - \mu_n\}$ применима лемма 2. Таким образом, $n^{-1}\mathbf{S}_n - \mu_n \xrightarrow{p} 0$. Так как и $n^{-1}\mathbf{S}_n \xrightarrow{p} 0$, то $\mu_n \to 0$. Утверждение доказано. ▶

Условие (7.3) несколько неприятно для проверки. Однако *если* $n\beta_n^2 \to 0$, *то оно упрощается, принимая вид*

$$n \int_{-s}^{s} y^2 F_n\{dy\} < M_s. \tag{7.6}$$

Доказательство *леммы* § 5. Если $\{\mathbf{S}_n\}$ стохастически ограничена, то как показывают две последние леммы, величины $n\beta_n$ ограничены и потому верно (7.6). По второй теореме о выборе гл. VIII, 6 существует мера Ω, такая, что

$$ny^2F_n\{dy\} \to \Omega\{dy\} \tag{7.7}$$

при n, пробегающем некоторую подпоследовательность $\{n_k\}$. Хвост интеграла (5.2) меньше, чем предел выражения в левой части (7.4), и условие интегрируемости (5.2) тривиально верно. Наконец, (7.4) — это то же самое, что (5.7). Таким образом, все утверждения леммы проверены. ▶

Мы можем сформулировать теперь основную предельную теорему для схем серий. Допустим, что предельное распределение для $\{\mathbf{S}_n\}$ существует. Тогда найдется последовательность $\{n_k\}$, такая, что при n, пробегающем эту последовательность, выполняется (7.7) и $b_n = n\beta_n \to b$. Мы уже видели в § 5, что в этих обстоятельствах $n[\mathfrak{F}_n - 1] \to \mathfrak{A}$, где $\mathfrak{A}$ — оператор, описанный в (5.9) и (5.11). *Предельное распределение* $\mathbf{S}_n$ *совпадает с распределением* Q_1 *полугруппы, порожденной* $\mathfrak{A}$.

Мыслимо, конечно, что иной выбор подпоследовательности $\{n_k\}$ мог бы привести к другой паре $(\Omega^\#, b^\#)$ и, следовательно, к другой полугруппе $\{\mathfrak{T}^\#(t)\}$. В этом случае предельное распределение S_n было бы связано и с $\mathfrak{T}(1)$, и с $\mathfrak{T}^\#(1)$. Однако такие полугруппы не существуют. Это утверждение требует некоторого доказательства, но оно оказывается совершенно тривиальным в рамках теории интегралов Фурье и не играет никакой

роли в настоящем контексте[1]). Поэтому мы примем его без доказательства, хотя бы для того, чтобы избежать сбивающих с толку формулировок.

С чисто теоретической точки зрения мы решили задачу описания предельного распределения для сумм $\mathbf{S}_n$ (если оно существует) в терминах производящего оператора соответствующей полугруппы со сверткой. Этот результат часто бывает трудно применить, так как он предполагает сходимость распределений $\mathbf{S}_n$. В действительности центрирование случайных величин — это очень гибкое орудие, и вопрос можно поставить так: существуют ли постоянные b_n, такие, что распределение $\mathbf{S}_n - b_n$ сходится к предельному. Условимся по-прежнему обозначать через $\mathbf{X}'_{k,\,n}$ случайные величины, получаемые урезанием $\mathbf{X}_{k,\,n}$ на произвольном, но фиксированном уровне $\pm s$ [см. (7.1)]. Напомним, что $\{\mathbf{X}_{k,\,n}\}$ называется *нулевой схемой*, если $\mathbf{X}_{1,\,n} \xrightarrow{\mathbf{P}} 0$. Подобную схему можно центрировать так, что при $n \to \infty$

$$(\mathbf{E}(\mathbf{X}'_{1,\,n}))^2 = o\left(\mathbf{E}\left(\mathbf{X}'^2_{1,\,n}\right)\right). \tag{7.8}$$

Это условие безвредно, а без него критерии становятся сложными и потому бесполезными. Как и в (7.5), мы полагаем $\beta_n = \mathbf{E}(\mathbf{X}'_{1,\,n})$.

Теорема. *Пусть* $\{\mathbf{X}_{k,\,n}\}$ *— нулевая схема серий, удовлетворяющая условию* (7.8).

а) *Для существования постоянных* b_n, *таких, что распределение* $\mathbf{S}_n - b_n$ *сходится к собственному предельному распределению, необходимо и достаточно выполнения условий* (7.4) *и* (7.7).

б) *Если эти условия выполнены, то распределение* $\mathbf{S}_n - n\beta_n$ *сходится к распределению* Q_1, *связанному с оператором* $\mathfrak{Q}(1)$ *полугруппы с производящим оператором*

$$\mathfrak{A}^{(0)} = \lim n\left[\mathfrak{F}_n - \mathbf{1} + \beta_n \frac{d}{dx}\right], \tag{7.9}$$

точно описанным в (5.9).

Доказательство. Мы знаем, что из (7.4) и (7.7) следует и существование предела $\mathfrak{A}^{(0)}$ в (7.9), и его форма (5.9). Мы покажем теперь, что в этом случае

$$\mathfrak{A}^{(0)} = \lim n\left[\mathfrak{F}_n^{(0)} - \mathbf{1}\right], \tag{7.10}$$

[1]) Оно используется в следующем параграфе, но для устойчивых полугрупп единственность очевидна.

где $\mathfrak{T}_n^{(0)}$ оператор, связанный с $\mathbf{X}_{k,\,n} - \beta_n$. Очевидно, что

$$n[\mathfrak{T}_n u(x) - u(x) + \beta_n u'(x)] - n[\mathfrak{T}^{(0)} u(x - \beta_n) - u(x - \beta_n)] = \\ = n[u(x - \beta_n) - u(x) + \beta n u'(x)]. \quad (7.11)$$

В силу (7.8) $n\beta_n^2 \to 0$, и потому правая часть равномерно стремится к нулю. Поэтому утверждение теоремы содержится в лемме 3.4 и ее условия достаточны.

Для доказательства их необходимости отметим, что здесь применима лемма 2. В силу (7.8) дисперсии в (7.3) могут быть заменены вторыми моментами, так что из стохастической ограниченности $\{\mathbf{S}_n - b_n\}$ вытекает (7.7) (по крайней мере для n, пробегающего некоторую подпоследовательность $n_1, n_2, \ldots$). Теперь первая часть доказательства показывает, что предельное распределение для $\mathbf{S}_n - b_n$ может отличаться от Q_1 только центрированием. В силу единственности соответствующей полугруппы мера Ω определяется однозначно и мы заключаем, что (7.7) выполняется при любом стремлении n к ∞. ▷

Пример. Пусть $\mathbf{X}_{k,\,n}$ имеют нормальное распределение с математическим ожиданием $1/\sqrt[3]{n^2}$ и дисперсией $1/n$. Тогда $\mathbf{S}_n - \sqrt[3]{n}$ имеет стандартное нормальное распределение $\mathfrak{N}$, хотя $\{\mathbf{S}_n\}$ не ограничена стохастически. Наша теорема применима с $\mathfrak{A}^{(0)} = \frac{1}{2}\frac{d^2}{dx^2}$.

§ 8. Области притяжения

В этом параграфе $\mathbf{X}_1, \mathbf{X}_2, \ldots$ предполагаются независимыми случайными величинами с одним и тем же распределением F. В соответствии с определением 2 гл. VI, 1 распределение F *принадлежит области притяжения* G, если существуют постоянные $a_n > 0$ и b_n, такие, что распределение $\frac{1}{a_n}(\mathbf{X}_1 + \ldots + \mathbf{X}_n) - b_n$ сходится к G, где G — собственное распределение, *не сосредоточенное в одной точке*. Несмотря на уже известные предварительные результаты гл. VI, 1 и § 6, мы будем строить здесь теорию с самого начала.

Всюду в этом параграфе мы будем использовать обозначение

$$U(x) = \int_{-x}^{x} y^2 F\{dy\}, \qquad x > 0. \quad (8.1)$$

Напомним, что в соответствии с теорией правильной изменчивости из гл. VIII, 8 определенная на $0, \infty$ положительная

функция L называется *медленно меняющейся* (на ∞), если[1]) при всех $x>0$

$$\frac{L(sx)}{L(s)} \to 1 \qquad s \to \infty. \tag{8.2}$$

Теорема 1. *Распределение F принадлежит области притяжения некоторого распределения G тогда и только тогда, когда существует медленно меняющаяся функция L, такая, что*

$$U(x) \sim x^{2-\alpha} L(x) \qquad x \to \infty, \tag{8.3}$$

где $0<\alpha\leqslant 2$, и при $\alpha<2$

$$\frac{1-F(x)}{1-F(x)+F(-x)} \to p, \quad \frac{F(-x)}{1-F(x)+F(-x)} \to q \tag{8.4}$$

(при $\alpha=2$ не нужно никаких дополнительных условий).

Отметим, что при существовании конечной дисперсии (8.3) выполняется с $\alpha=2$. Однако U может быть медленно изменяющейся и при отсутствии дисперсии. Мы увидим, что случай $\alpha=2$ соответствует притяжению к нормальному закону.

По теореме гл. VIII, 9.2 *соотношение* (8.3) *эквивалентно*[2])

$$\frac{x^2[1-F(x)+F(-x)]}{U(x)} \to \frac{2-\alpha}{\alpha} \tag{8.5}$$

в том смысле, что каждое из этих соотношений влечет другое.

При $0<\alpha<2$ мы можем переписать (8.5) в виде

$$1-F(x)+F(-x) \sim \frac{2-\alpha}{\alpha} x^{-\alpha} L(x) \tag{8.6}$$

и (8.6) в свою очередь влечет (8.3) и (8.5). Мы приходим к новой формулировке теоремы, интуитивно более понятной, так как она описывает поведение отдельных хвостов (другие варианты см. в задаче 14).

---

[1]) Полезно иметь в виду, что при любом $\varepsilon>0$ $s^{\varepsilon}L(s)\to\infty$ и $s^{-\varepsilon}L(s)\to 0$. Если функция $x^{-\alpha}L(x)$ монотонна, то это утверждение легко получить, рассматривая отношение $L(2x)/L(x)$. [Для более общих L это утверждение содержится в гл. VIII, (9.9), но подобные функции не появятся в этом параграфе.]

[2]) Условие (8.4) налагает аналогичные ограничения на каждый хвост по отдельности:

$$\frac{x^2[1-F(x)]}{U(x)} \to p\frac{2-\alpha}{\alpha}, \quad \frac{x^2F(-x)}{U(x)} \to q\frac{2-\alpha}{\alpha}. \tag{*}$$

При $\alpha=2$ эти соотношения вытекают из (8.5), чем и объясняется отсутствие второго условия при $\alpha=2$. Теорема 1 могла бы быть сформулирована более сжато (но более искусственно) следующим образом: *F принадлежит некоторой области притяжения тогда и только тогда, когда выполняется* (*) *с* $0<\alpha\leqslant 2$, $p\geqslant 0$, $q\geqslant 0$, $p+q=1$.

Теорема 1а. (*Другая форма.*) (i) *Распределение F принадлежит области притяжения нормального распределения тогда и только тогда, когда U медленно меняется.*

(ii) *Оно принадлежит области притяжения, отличной от нормальной, тогда и только тогда, когда для некоторого $0<\alpha<2$ выполняются* (8.6) *и* (8.4).

Доказательство. Мы применим теорему § 7 к схеме серий, образованной величинами $\mathbf{X}_{k,\,n}=\mathbf{X}_k/a_n$ с распределениями $F_n(x)=F(a_nx)$. Суммы по строкам для схемы $\{\mathbf{X}_{k,\,n}\}$ равны

$$\mathbf{S}_n=\frac{\mathbf{X}_1+\ldots+\mathbf{X}_n}{a_n}. \tag{8.7}$$

Положим для краткости

$$v(x)=\int_{-x}^{x} yF\{dy\}. \tag{8.8}$$

Тогда условие (7.8) принимает форму

$$v^2(x)=o(U(x)). \tag{8.9}$$

Оно автоматически выполняется, если F не имеет математического ожидания [1]). Если математическое ожидание существует, то мы будем предполагать его равным нулю. Тогда $v(x)\to 0$ и (8.9) верно. Мы видим, что применима теорема § 7. Два ее условия: (7.7) и (7.4) принимают вид

$$ny^2F_n\{dy\}\to\Omega\{dy\} \tag{8.10}$$

и

$$n[1-F_n(\eta)+F_n(-\eta)]<\varepsilon \tag{8.11}$$

для достаточно больших η и всех n.

а) *Необходимость.* Из (8.10) с очевидностью следует, что

$$\frac{n}{a_n^2}U(a_nx)\to\Omega\{\overline{-x,\ x}\} \tag{8.12}$$

во всех точках непрерывности. При этом Ω не может быть тождественно равна нулю (так как в противном случае предельное распределение было бы сосредоточено в одной точке). По лемме 2 гл. VIII, 8 функция U меняется правильно [2]) и $\Omega\{-x,\ x\}=cx^{2-\alpha}$ с $0<\alpha\leqslant 2$. При $\alpha=2$ больше нечего доказы-

---

[1]) По неравенству Шварца $[v(x)-v(a)]^2\leqslant U(x)[1-F(a)+F(-a)]$ при $x>a$. Поэтому (8.9) верно при $U(x)\to\infty$.

[2]) Условие $a_{n+1}/a_n\to 1$ выполняется в силу леммы 3 § VIII, 2.

вать. Если $\alpha<2$, то нуль не будет атомом для Ω и вышеприведенные доводы применимы по отдельности к функции

$$U^+(x)=\int_0^x y^2F\{dy\} \tag{8.13}$$

и аналогичной ей функции U^-. Условие (8.4) следует из аналога (8.5) для U^+ и U^-.

б) *Достаточность.* Предположив, что (8.3) выполняется, можно выбрать числа a_n таким образом, что

$$\frac{n}{a_n^2}U(a_n)\to 1. \tag{8.14}$$

Тогда

$$\frac{n}{a_n^2}U(a_nx)\to x^{2-\alpha}. \tag{8.15}$$

Если $\alpha=2$, то отсюда вытекает (8.10), причем Ω будет распределением вероятностей, сосредоточенным в нуле. При $\alpha<2$ дополнительное соотношение (8.4) обеспечивает сходимость $na_n^{-2}U^+(a_nx)\to px^{2-\alpha}$ (и аналогично для U^-). Таким образом, (8.10) выполняется. Наконец, (8.11) непосредственно следует из (8.5). ▶

Это доказательство дает дополнительную информацию относительно предельного распределения. Допустим, что F удовлетворяет условиям теоремы при некоторых фиксированных значениях параметров. Из (8.6) следует, что при $\alpha>1$ существует конечное математическое ожидание, и в этом случае мы будем считать, что оно равно нулю. Положим $b_n=na_n^{-1}v(a_n)$. Тогда существует предельное распределение для $\mathbf{S}_n-b_n$. Теперь теорема 2 гл. VIII, 9 показывает, что при $\alpha\neq 1$ существует конечный предел $b=\lim b_n$. Следовательно, сами $\mathbf{S}_n$ имеют предельное распределение G. Таким образом, при $\alpha\neq 1$

$$F^{n*}(a_nx)\to G(x). \tag{8.16}$$

Наш выбор нормирующих констант a_n указан в (8.14). Можно выбрать и другие нормирующие константы a_n', но они необходимо удовлетворяют соотношению $a_n'\sim ca_n$ (лемма 1 из гл. VIII, 2) и дают предел $G(cx)$. Это означает, что для данного предельного распределения G в (8.16) выбор нормирующих констант по существу однозначен. Так как $a_{rn}/a_n\to r^{1/\alpha}$, то

$$G^{r*}(r^{1/\alpha}x)=G(x). \tag{8.17}$$

Таким образом, и само распределение G удовлетворяет условиям теоремы с $a_n = n^{1/\alpha}$. Эти коэффициенты должны удовлетворять (8.14). Для $0<\alpha<1$ и $1<\alpha<2$ отсюда следует, что при $x \to \infty$

$$x^\alpha [1 - G(x)] \to p\frac{2-\alpha}{\alpha}, \quad x^\alpha G(-x) \to q\frac{2-\alpha}{\alpha}. \tag{8.18}$$

При $\alpha=1$ центрирующие постоянные b_n не обязаны оставаться ограниченными, но тривиальное изменение наших рассуждений показывает, что предельное распределение для $\mathbf{S}_n - b_n$ удовлетворяет (8.18). Мы получаем, таким образом, следующее усиление теоремы 1.

Теорема 2. *Пусть F удовлетворяет условиям теоремы* 1. *При $\alpha>1$ допустим дополнительно, что F центрировано своим математическим ожиданием. Выберем числа a_n в соответствии с* (8.14).

(i) *Если $0<\alpha<1$ или $1<\alpha<2$, то верно* (8.16). *Предельное распределение G устойчиво и удовлетворяет* (8.18).

(ii) *Если $\alpha=2$, то верно* (8.16) с $G=\mathfrak{N}$, *т. е. со стандартным нормальным распределением.*

(iii) *Если $\alpha=1$, то предельное распределение для $\mathbf{S}_n - b_n$ устойчиво и удовлетворяет* (8.18).

Согласно определению устойчивости (см. гл. VI, 1), каждое устойчивое распределение принадлежит своей собственной области притяжения. Распределение, удовлетворяющее (8.17), строго устойчиво. Поэтому следующая далее теорема содержится в предыдущей.

Теорема 3. *Для каждого набора параметров существует в точности один тип устойчивых распределений, удовлетворяющих условиям теоремы* 1.

Для $\alpha=2$ это нормальные распределения.

Для $\alpha<1$ можно так выбрать масштабный параметр, что верно (8.18) [1].

Для $\alpha=1$ тип содержит ровно одно строго устойчивое распределение, удовлетворяющее (8.18).

Устойчивые распределения для случая $\alpha=1$ и $p\neq q$ удовлетворяют (6.10) и потому *не* будут строго устойчивыми.

Замечание о центральной предельной теореме. В теореме 2 содержится и характеризация области притяжения нормального распределения. Чтобы принадлежать ей, F должно иметь математическое ожидание (и даже конечные абсолютные моменты

[1] (8.18) не зависит от способа центрирования.

любого порядка <2). Если математическое ожидание F равно нулю, то *F принадлежит области притяжения $\mathfrak{N}$ тогда и только тогда, когда U — медленно меняющаяся функция.* Ясно, что конечность второго момента влечет $U(\infty)<\infty$ и потому служит достаточным условием, но пример (4, а) гл. VIII показывает, что оно не необходимо. Подходящие нормирующие постоянные каждый раз определяются из (8.14).

(Относительно так называемой области нормального притяжения для распределений, отличных от нормального, см. гл. XVII, 5.)

§ 9. Различные распределения. Теорема о трех рядах

На короткое время мы перейдем к общим схемам серий $\{\mathbf{X}_{k,n}\}$, в которых случайные величины[1]) $\mathbf{X}_{1,n}, \ldots, \mathbf{X}_{n,n}$, входящие в n-ю строку, взаимно независимы, но могут иметь произвольные распределения $F_{k,n}$. Как уже объяснялось в гл. VI, 3, для того чтобы сохранить характер наших предельных теорем, необходимо принять условие, означающее, что $\mathbf{X}_{n,k}$ в некотором смысле индивидуально пренебрегаемы. Наиболее удовлетворительное условие состоит в том, что для любых $\eta>0$ и $\varepsilon>0$ и всех достаточно больших n

$$\mathbf{P}\{|\mathbf{X}_{k,n}|>\eta\}<\varepsilon, \qquad k=1, \ldots, n. \tag{9.1}$$

Обозначения, использованные в § 7, не требуют изменений. Например, $\mathbf{X}'_{k,n}$ снова обозначает величину, получаемую урезанием $\mathbf{X}_{k,n}$ на уровне $\pm s$. Вся теория переносится на рассматриваемый случай с тем единственным изменением, что выражения, подобные $n\operatorname{Var}(\mathbf{X}'_{1,n})$, заменяются соответствующими суммами. Проверка требует стандартных приемов и может быть предоставлена читателю. Ниже приводится простой вариант леммы о компактности.

Теорема 1. (*Закон больших чисел.*) *Если выполняется условие* (9.1), *то* $\mathbf{S}_n - b_n \xrightarrow{\mathbf{P}} 0$ *при некоторых* b_n *в том и только том случае, когда*

$$\sum_{k=1}^{n}\mathbf{P}\{|\mathbf{X}_{k,n}|>\eta\}\to 0, \quad \sum_{k=1}^{n}\operatorname{Var}(\mathbf{X}'_{k,n})\to 0 \tag{9.2}$$

для каждого $\eta>0$ и каждого уровня урезания s. Если (9.2) *выполнено, то можно взять* $b_n=\mathbf{E}(\mathbf{S}'_n)$.

[1]) Относительно числа слагаемых n-й строки см. сноску 2 на стр. 222.

В качестве простого следствия получаем теорему 2, в которой штрихи снова означают урезание (5.3) на уровне $\pm s$. Специальные случаи и примеры были разобраны в гл. VIII, 5.

Теорема 2. *(Бесконечные свертки.) Пусть* $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ — *независимые случайные величины с распределениями* $G_1, G_2, \ldots$. *Для того чтобы распределения* $G_1 * G_2 * \ldots * G_n$ *сумм* $\mathbf{T}_n = \mathbf{Y}_1 + \ldots + \mathbf{Y}_n$ *сходились к собственному предельному распределению* G, *необходимо и достаточно, чтобы при каждом* $s > 0$

$$\sum \mathbf{P}\{|\mathbf{Y}_k| > s\} < \infty, \quad \sum \operatorname{Var}(\mathbf{Y}_k') < \infty \tag{9.3}$$

и

$$\sum_{k=1}^{n} \mathbf{E}(\mathbf{Y}_k') \to m. \tag{9.4}$$

Доказательство. Для данной возрастающей последовательности $\nu_1, \nu_2, \ldots$ целых чисел и $k = 1, \ldots, n$ положим $\mathbf{X}_{k,n} = \mathbf{Y}_{\nu_n + k}$. Распределения $G_1 * \ldots * G_n$ сходятся тогда и только тогда, когда *все* полученные таким образом схемы серий подчиняются закону больших чисел с центрирующими постоянными $b_n = 0$. Из теоремы 1 ясно видно, что для этого необходимы и достаточны условия (9.3) и (9.4). ▶

Так же как и в случае равных компонент, *предельными распределениями для сумм по строкам при условии* (9.1) *могут быть только безгранично делимые распределения* (см. задачу 11).

Теореме 2 можно придать следующую более яркую формулировку.

Теорема 3. *(Теорема Колмогорова о трех рядах.) Ряд* $\sum \mathbf{Y}_k$ *сходится с вероятностью единица, если выполняются соотношения* (9.3) *и* (9.4), *и сходится с вероятностью нуль в противном случае.*

Доказательство. Допустим, что имеют место (9.3) и (9.4). По теореме 1 гл. VII, 8 второе из условий (9.3) гарантирует сходимость с вероятностью единица ряда $\sum [\mathbf{Y}_k' - \mathbf{E}(\mathbf{Y}_k')]$. По лемме Бореля — Кантелли (см. **1**, гл. VIII, 3) из первого условия (9.3) следует, что лишь конечное число среди величин $\mathbf{Y}_k$ отличается от $\mathbf{Y}_k'$. Поэтому ряд $\sum \mathbf{Y}_k$ сходится с вероятностью единица.

Для доказательства необходимости наших условий напомним (см. гл. IV, 6), что вероятность сходимости равна или нулю, или единице. В последнем случае распределения частных сумм должны сходиться, и поэтому будут выполнены (9.3) и (9.4) ▶

Неоднородные процессы

Развитая в этой главе теория полугрупп представляет собой орудие, особенно пригодное для изучения процессов со стационарными и независимыми приращениями. При отказе от условия стационарности распределения приращений $\mathbf{X}(t)-\mathbf{X}(\tau)$ будут зависеть от двух параметров t и τ и нам приходится иметь дело с двупараметрическим семейством операторов $\mathfrak{T}(\tau, t)$, $0<\tau<t$, удовлетворяющих уравнению свертки

$$\mathfrak{T}(\tau, s)\,\mathfrak{T}(s, t)=\mathfrak{T}(\tau, t), \qquad \tau<s<t. \tag{9.5}$$

Будут ли распределения, связанные с такими операторами, безгранично делимы? Мы можем разбить τ, t на n интервалов $\overline{t_{k-1}, t_k}$ и рассмотреть случайные величины $\mathbf{X}(t_k)-\mathbf{X}(t_{k-1})$. Однако для применения теории, развитой для схем серий, мы должны принять ограничение (9.1), сводящееся к требованию равномерной непрерывности рассматриваемых распределений по двум временным параметрам. Но $\mathfrak{T}(\tau, t)$ не обязано зависеть от t непрерывно. В самом деле, частные суммы последовательности $\mathbf{X}_1$, $\mathbf{X}_2$, ... независимых случайных величин образуют процесс с независимыми приращениями, в котором все изменения происходят только в целочисленные моменты времени. Этот процесс разрывен. Однако в некотором смысле это есть единственный тип существенно разрывного процесса. Термин «существенно» необходим. В самом деле, мы показали в § 5а, что даже в обычных полугруппах искусственное центрирование может породить беспорядок. Он не имеет особенных последствий, но требует осторожности в формулировках. Для простоты мы ограничимся симметричными распределениями и докажем следующую лемму.

Лемма. *Если распределения, связанные с* $\mathfrak{T}(\tau, t)$, *симметричны, то при каждом* t *существует односторонний предел* $\mathfrak{T}(\tau, t-)$

[Аналогичное утверждение верно и для других пределов, например $(\tau+, t)$ и т. д.]

Доказательство. Пусть $\tau<t_1<t_2<\ldots$ и $t_n\to t$. Последовательность распределений, связанных с $\mathfrak{T}(\tau, t_n)$, стохастически ограничена, так что она содержит сходящуюся подпоследовательность. Чтобы избежать двойных индексов, допустим, что $\mathfrak{T}(\tau, t_n)\to\mathfrak{A}$ где $\mathfrak{A}$ соответствует собственному распределению U. Легко установить, что $\mathfrak{T}(t_n, t_{n+1})\to\mathbf{1}$ и потому $\mathfrak{T}(t_n, s_n)\to\mathbf{1}$ для любой последовательности моментов времени, такой, что $t_n<s_n<t_{n+1}$. В силу (9.5) это означает, что $\mathfrak{T}(\tau, s_n)\to\mathfrak{A}$, поэтому предел $\mathfrak{A}$ не зависит от выбора последовательности $\{t_n\}$. Лемма доказана. ▶

Следуя П. Леви, мы назовем момент t *фиксированным разрывом,* если пределы $\mathfrak{T}(\tau, t+)$ и $\mathfrak{T}(\tau, t-)$ различны. Из теоремы 2 о бесконечных свертках сразу следует, что *множество фиксированных разрывов не более чем счетно* Используя симметризацию, можно показать, что в общем случае разрывы обусловлены только центрированием (исключая не более чем счетное множество моментов времени) и что их можно устранить подходящим центрированием. Вклад $\mathfrak{T}_d(\tau, t)$ всех фиксированных разрывов в $\mathfrak{T}(\tau, t)$ представляет собой бесконечную свертку. Поэтому *процесс можно разложить на дискретную и непрерывную части.* Для схем серий, возникающих из непрерывных процессов, нетрудно показать, (используя теорему 2), что условие равномерной малости (9.1) автоматически выполнено. Мы приходим к заключению, что *распределения, связанные с непрерывными процессами, безгранично делимы.* П. Леви показал, что выборочные функции таких процессов «достаточно хороши» в том смысле, что с вероятностью единица они имеют при каждом t правые и левые предельные значения.

§ 10. Задачи

1. Покажите, что в примере (1, а) $\sum \mathbf{X}_{k,\,n}^2 \xrightarrow{\mathrm{P}} 1$ при $n\to\infty$. (*Указание.* Используйте дисперсии.)

2. Пусть $\mathbf{T}$ — момент первого прохождения $+1$ в обычном симметричном случайном блуждании. Другими словами, $\mathbf{T}$ — такая случайная величина, что

$$\mathbf{P}\{\mathbf{T}=2r-1\}=\frac{1}{2r-1}\binom{2r-1}{r}2^{-2r+1}.$$

Возьмите схему серий, в которой $\mathbf{X}_{k,\,n}$ распределено как $\mathbf{T}/n^2$. Используя элементарные методы § 1, покажите прямым вычислением, что

$$n\,[\mathfrak{T}_n u(x)-u(x)]\to\frac{1}{\sqrt{2\pi}}\int_0^\infty \frac{u(x-y)-u(x)}{\sqrt{y^3}}\,dy. \qquad (*)$$

Установите, что распределения сумм по строкам сходятся к устойчивому распределению F_1, определенному в гл. II, (4.7) и обладающему свойством (4.9). Дайте интерпретацию полученного результата в терминах случайного блуждания, в котором скачки равны $\pm 1/n$, а промежутки между последовательными скачками равны $1/n^2$.

3. Выведите из гл. II, (4.9), что распределения гл. II, (4.7) образуют полугруппу с производящим оператором, задаваемым правой частью (*) в предыдущей задаче.

4. Пусть распределения Q_t некоторой полугруппы сосредоточены на множестве целых чисел. Обозначим вес k через $q_k(t)$. Покажите, что

$$\mathfrak{A}u(x)=-q_0'(0)\,u(x)+\sum_{k\neq 0}q_k'(0)\,u(x-k).$$

Сравните с канонической формой (5.9). Рассмотрите в связи с этим производящие функции безгранично делимых распределений, найденные в **1**, гл. XII, 3.

5. Распространите понятия § 2 на полугруппы с несобственными распределениями. Покажите, что если $\mathfrak{A}$ порождает $\{\mathfrak{Q}(t)\}$, то $\mathfrak{A}-c\mathbf{I}$ порождает $\{e^{-ct}\mathfrak{Q}(t)\}$.

6. Обозначение $e^{t(\mathfrak{F}-1)}$ для обобщенной пуассоновской полугруппы наталкивает на мысль написать в общем случае $\mathfrak{Q}(t)=e^{t\mathfrak{A}}$. Для нормальной полугруппы это приводит к формальному операторному соотношению

$$\exp\left(\frac{1}{2}\,t\,\frac{d^2}{dx^2}\right)u(x)=\sum\frac{1}{n!}\left(\frac{t}{2}\right)^n u^{(2n)}(x).$$

Покажите, что оно верно, если ряд Тейлора для u сходится при всех x. *Указание.* Используйте моменты нормального распределения.

7. Распределения полугруппы имеют конечные математические ожидания тогда и только тогда, когда функция $\dfrac{1}{1+|x|}$ интегрируема относительно меры Ω, входящей в формулу для производящего оператора.

8. Покажите непосредственно, что если $n\,[\mathfrak{T}_n-\mathbf{I}]\to\mathfrak{A}$, то оператор $\mathfrak{A}$ обязательно имеет форму производящего оператора. [Используйте метод § 4,

рассматривая функции вида (4.4), но не привлекайте теорию полугрупп. Цель — вывести общий вид производящего оператора, не доказывая заранее его существование.]

9. Пусть F_k приписывает вероятности $^1/_2$ двум точкам $\pm\mu^k$. Тогда

$$\sum_{k=-\infty}^{+\infty} 2^{-k}\,(\mathfrak{F}_k - 1)$$

порождает полугруппу, такую, что $Q_{2t}(x) = Q_t\left(\frac{1}{2}\,x\right)$, но Q_t не являются устойчивыми распределениями. (П. Леви.)

10. Для любого распределения F и гладкой функции u

$$\|\,(\mathfrak{F}-1)\,u\,\| \leqslant 100\cdot(\|\,u\,\|+\|\,u''\,\|)\int_{-\infty}^{+\infty}\frac{x^2}{1+x^2}\,F\,\{dx\}+\|\,u'\,\|.$$

11. *Продолжение.* Схеме серий $\{\mathbf{X}_{k,\,n}\}$ с распределениями F_n соответствует другая схема серий $\left\{\mathbf{X}^{\#}_{k,\,n}\right\}$ с обобщенными пуассоновскими распределениями

$$\mathfrak{F}_n^{\#} = e^{\mathfrak{F}_n - 1}.$$

Покажите, что если $\{\mathbf{S}_n\}$ стохастически ограничена, то $n\left[\mathfrak{F}_n - \mathfrak{F}_n^{\#}\right]\to 0$. Отсюда видно, что *суммы по строкам* $\mathbf{S}_n$ и $\mathbf{S}_n^{\#}$ *асимптотически эквивалентны.* Так как распределение $\mathbf{S}_n^{\#}$ связано с $e^{n\,[\mathfrak{F}_n - 1]}$, мы получаем второй метод, пригодный для вывода основных теорем § 5. Этот метод применением и к схемам серий с различно распределенными величинами.

12. Положим в обозначениях § 5 $\mathbf{M}_n = \max\,[\mathbf{X}_{1,\,n},\ \ldots,\ \mathbf{X}_{n,\,n}]$. Если $\mathbf{S}_n$ имеет предельное распределение, то $\psi^+(x) = \lim \log \mathbf{P}\{\mathbf{M}_n > x\}$.

13. Если $\mathbf{S}_n$ имеет предельное распределение, то этим же свойством обладают суммы по строкам в схеме серий, образованной квадратами $\mathbf{X}^2_{k,\,n}$.

14. *Критерии для областей притяжения.* В теореме 8.1 использованы урезанные вторые моменты (функция U). Причина этого — традиция. (Множитель $2-\alpha$ — результат этого.) Теорема 2 гл. VIII, 9 позволяет нам заменить y^2 в (8.1) на $|y|^{\rho}$ с другим показателем ρ и при каждом ρ — заменить (8.3) и (8.5) равносильными соотношениями.

15. Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ независимые случайные величины с одним и тем же распределением F. Пусть $1-F(x)+F(-x)$ медленно меняется. Выведите из леммы О компактности, что в этом случае последовательность $\mathbf{S}_{n_k}/a_k + b_k$ не может иметь собственное предельное распределение, не сосредоточенное в одной точке. (Это можно выразить также, говоря, что F *не принадлежит никакой области частичного притяжения.*)

ГЛАВА X

МАРКОВСКИЕ ПРОЦЕССЫ И ПОЛУГРУППЫ

Эта глава открывается элементарным обзором наиболее распространенных типов марковских процессов, или точнее основных уравнений, которым подчиняются их переходные вероятности. Затем мы перейдем к введенному Бохнером понятию подчиненности процессов и к полугрупповой трактовке марковских процессов. Связующим звеном между этими вопросами служит так называемая показательная формула теории полугрупп. Существование производящих операторов будет доказано только в гл. XIII. Теоретически настоящее изложение могло бы охватывать в качестве частных случаев процессы и полугруппы предыдущей главы. Однако методы и применения столь различны, что последующая теория излагается замкнуто и *независимо от гл.* IX. Результаты будут усилены в гл. XIII, но *теория марковских процессов не используется* в остающейся части книги.

Эта глава по своему характеру — обзор, в котором не делается попытки достичь общности или полноты [1]). В частности, мы не будем обсуждать свойства выборочных функций, и всюду в этой главе факт существования рассматриваемых процессов будет рассматриваться как заранее данный. Наши интересы полностью концентрируются вокруг аналитических свойств переходных вероятностей и операторов.

Теория индуцированных полугрупп будет с достаточной степенью общности рассмотрена в § 8—9. В предшествующих им параграфах основное пространство есть интервал на прямой или вся прямая, хотя часть теории годится и в более общей обстановке. Чтобы избежать введения специальных обозначений, мы условимся считать, что если пределы не указаны, то *интеграл берется по фиксированному множеству* Ω, *служащему основным пространством.*

---

[1]) Более детально полугрупповой подход к марковским процессам описан в книге М Лоэва (1963) [*Примечание при корректуре*: общая теория марковских процессов и ее связи с полугруппами преобразований исчерпывающим образом рассмотрены в монографиях Е Б Дынкина (1965) Новая книга К Иосиды (1967) дает сжатое и ясное введение в аналитическую теорию полугрупп и ее применение к теории диффузии и эргодической теории.]

§ 1. Псевдопуассоновский тип

Всюду в этой главе мы ограничиваемся марковскими процессами со *стационарными переходными вероятностями* Q_t, определяемыми формулой

$$Q_t(x, \Gamma) = \mathbf{P}\{\mathbf{X}(t+\tau) \in \Gamma | \mathbf{X}(\tau) = x\} \tag{1.1}$$

в предположении, что правая часть не зависит от τ (см. гл. VI, 11).

Простое расширение понятия обобщенного пуассоновского процесса приводит к некоторому важному классу процессов, из которых все другие можно получить с помощью аппроксимации. Теория полугрупп опирается на аналитический аналог этого факта (§ 10).

Пусть K — стохастическое ядро и $\mathbf{S}_0, \mathbf{S}_1, \ldots$ — порожденная им марковская цепь (определение и примеры см. в гл. VI, 11). Пусть $\mathbf{N}(t)$ обозначает величины, образующие пуассоновский процесс, не зависящий от цепи $\{\mathbf{S}_k\}$. Случайные величины $\mathbf{X}(t) = \mathbf{S}_{\mathbf{N}(t)}$ образуют новый случайный процесс, который формально можно описать следующим образом. В промежутке между скачками пуассоновского процесса типичная выборочная функция остается постоянной. Переход из x в Γ может произойти за 0, 1, 2, ... шагов и, следовательно,

$$Q_t(x, \Gamma) = e^{-\alpha t} \sum_{n=0}^{\infty} \frac{(\alpha t)^n}{n!} K^{(n)}(x, \Gamma), \quad t > 0. \tag{1.2}$$

Обобщенное пуассоновское распределение гл. VI, (4.2) получается из этой формулы как частный случай, когда Ω совпадает со всей прямой и $\mathbf{S}_n$ представляет собой сумму независимых случайных величин с одним и тем же распределением F.

Правило композиции

$$Q_{t+\tau}(x, \Gamma) = \int Q_t(x, dy)\, Q_\tau(y, \Gamma) \tag{1.3}$$

($t, \tau > 0$), аналогичное гл. VI, (4.1), легко проверить аналитически[1]). Оно называется уравнением *Чепмена — Колмогорова* и утверждает, что переход из состояния x в момент 0 во множество Γ в момент $t+\tau$ происходит через некоторую точку y в момент τ и что последующее движение не зависит от прошлого[2]) [см. 1, гл. XVII, 9, а также гл. VI, (11.3)].

[1]) Аппроксимация Q_t и Q_τ их частными суммами с n членами показывает, что правая часть в (1.3) $\leqslant$ левой части, но $\geqslant$ n-й частной суммы $Q_{t+\tau}$.

[2]) Иногда объявляют (1.3) либо законом природы, либо формулой полной вероятности. Но (1.3) *не* выполняется для немарковских процессов; см. 1, гл. XV, 10.

Примеры. а) *Частицы, движущиеся со столкновениями.* Рассмотрим частицу, движущуюся с постоянной скоростью в однородной среде и время от времени подвергающуюся случайным соударениям. Каждое из них приводит к изменению энергии, описываемому стохастическим ядром K. Если число соударений образует пуассоновский процесс, то переходные вероятности для энергии $\mathbf{X}(t)$ имеют вид (1.2). Этот случай имеет место при хорошо знакомых нам предположениях пространственной однородности и отсутствия последействия.

Обычно предполагают, что *доля энергии*, теряемая при каждом соударении, не зависит от ее начальной величины. Это означает, что $K(x, dy) = V\{dy/x\}$, где V — распределение вероятностей, сосредоточенное на $\overline{0, 1}$. Имея в виду последующие применения, рассмотрим частный случай $V(x) = x^{\lambda}$. Тогда K имеет плотность

$$k(x, y) = \lambda x^{-\lambda} y^{\lambda-1}, \quad 0 < y < x. \tag{1.4}$$

При $\lambda = 1$ наше предположение означает, что доля теряемой энергии *распределена равномерно* [1]). Итерированные ядра $K^{(n)}$ были вычислены в гл. VI, (11.5). Подставляя их значение в (1.2), видим, что Q_t имеет атом веса $e^{-\alpha t}$ в нуле (соответствующий отсутствию соударений), а при $0 < y < x$ имеет плотность

$$q_t(x, y) = e^{-\alpha t} \sqrt{\lambda \alpha t} \frac{y^{\lambda-1}}{x^{\lambda} \sqrt{\log(x/y)}} I_1\left(2\sqrt{\alpha t \lambda \log(x/y)}\right), \tag{1.5}$$

где I_1 — функция Бесселя, введенная в гл. II, (7.1) [см. примеры 2а и 2б].

б) *Потеря энергии быстрыми частицами при ионизации* [2]). Поучительный вариант предыдущего примера получается в предельном случае, когда энергию частицы предполагают бесконечно большой. Потери энергии при последовательных соударениях будут представлять собой независимые случайные величины с одним и тем же распределением V, сосредоточенным на $\overline{0, \infty}$. Пусть $\mathbf{X}(t)$ обозначает полную величину потерянной за время $\overline{0, t}$ энергии. Тогда случайные величины $\mathbf{X}(t)$ образуют *обобщенный пуассоновский процесс*. Переходные вероятности для

[1]) Это предположение использовано Гейтлером и Яноши (Heitler W., Janossy L., Absorption of meson producing nucleons, *Proc. Phys. Soc., Ser.* A, **62** (1949), 374—385. В этой статье введено (но не решено) уравнение Фоккера — Планка (1.8).

[2]) См. статью: Ландау Л. Д., *J. Phys., USSR*, 8 (1944), 201—205. Ландау использует другую терминологию, но его предположения равносильны нашим. Ландау вывел прямое уравнение (1.6).

него определяются по (1.2) с заменой $K^{(n)}$ на свертку V^{n*} [см. гл. VI, (4.2)].

в) *Изменение направления.* Вместо энергии частицы мы можем рассмотреть направление ее движения и построить модель, подобную приведенной в примере (а). Основное различие будет в том, что направление в $\mathscr{R}^3$ определяется двумя величинами, так что плотность ядра k будет зависеть от четырех вещественных переменных.

г) *Рандомизированное случайное блуждание* примера гл. II, (7, в) представляет собой псевдопуассоновский процесс, состояниями которого служат целые числа. При фиксированном x ядро K приписывает вес $1/2$ каждой из двух точек $x \pm 1$. ▶

Из (1.2) легко получаем

$$\frac{\partial Q_t(x, \Gamma)}{\partial t} = -\alpha Q_t(x, \Gamma) + \alpha \int Q_t(x, dz) K(z, \Gamma). \tag{1.6}$$

Это — *прямое уравнение Колмогорова*[1]), которое (в более общей постановке) будет обсуждаться в следующем параграфе. Там будет показано, что (1.2) является *единственным* решением, удовлетворяющим естественным вероятностным требованиям. Если K имеет плотность k, то уравнение (1.6) принимает более привычную форму. В точке x распределение Q_t имеет атом веса $e^{-\alpha t}$, равный вероятности отсутствия изменений. Если исключить этот атом, то Q_t будет иметь плотность q_t, удовлетворяющую уравнению

$$\frac{\partial q_t(x, \xi)}{\partial t} = -\alpha q_t(x, \xi) + \alpha \int q_t(x, z) k(z, \xi)\, dz. \tag{1.6а}$$

Если μ_0 — распределение вероятностей в момент 0, то распределение в момент t определяется по формуле

$$\mu_t\{\Gamma\} = \int \mu_0\{dx\}\, Q_t(x, \Gamma) \tag{1.7}$$

и в силу (1.6)

$$\frac{\partial \mu_t\{\Gamma\}}{\partial t} = -\alpha \mu_t\{\Gamma\} + \alpha \int \mu_t\{dz\} K(z, \Gamma). \tag{1.8}$$

Этот вариант (1.6) известен физикам под названием уравнения *Фоккера — Планка* (или уравнения неразрывности).

Это уравнение будет проанализировано в § 3. В случае когда K и начальное распределение μ_0 имеют плотности, μ_t также

---

[1]) Относительно терминологии см. сноску в конце 1, гл. XVII, 8 (стр. 454 русск. изд). — *Прим. перев.*

имеет плотность m_t и уравнение Фоккера — Планка принимает вид

$$\frac{\partial m_t(\xi)}{\partial t} = -\alpha m_t(\xi) + \alpha \int m_t(z)\, k(z, \xi)\, dz, \tag{1.8a}$$

§ 2. Вариант: линейные приращения

В физике, теории очередей и других применениях встречается один вариант рассмотренного нами процесса. Предположения, касающиеся скачков, остаются теми же самыми, но *между скачками* $\mathbf{X}(t)$ *изменяется линейно со скоростью* c. Это означает, что величины $\mathbf{X}(t) - ct$ образуют уже описанный псевдопуассоновский процесс; обозначим переходные вероятности нового процесса через Q_t. Тогда вероятности $Q_t(x, \Gamma + ct)$ должны удовлетворять (1.6). Получающееся отсюда уравнение для Q_t имеет необычную форму, однако если существуют дифференцируемые плотности, то они удовлетворяют привычным уравнениям. Так, для m_t мы должны заменить ξ в (1.8а) на $\xi + ct$. После замены $y = \xi + ct$ получим уравнение Фоккера — Планка [1]).

$$\frac{\partial m_t(y)}{\partial t} = -c\frac{\partial m_t(y)}{\partial y} - \alpha m_t(y) + \alpha \int m_t(z)\, k(z - ct, y - ct)\, dz. \tag{2.1}$$

Уравнение, аналогичное (1.6а), получается простым добавлением члена $-c\partial q_t/\partial y$ в правой части.

В связи с теорией полугрупп мы придадим уравнению Фоккера — Планка более гибкую форму, не зависящую от неестественных условий дифференцируемости [см. пример (10б)].

Примеры. а). *Частицы, испытывающие соударения.* Пример (1.а) встречается в физической литературе обычно в другой постановке. Предполагается, что энергия рассеивается с постоянной скоростью благодаря поглощению или трению. Модель (2.1) подходит к этой ситуации. Если потеря энергии пропорциональна

[1]) Многие частные случаи уравнения Фоккера — Планка (1.8) были открыты независимо друг от друга. Много энергии было затрачено на обобщение (2.1). Общая теория уравнений Фоккера — Планка была развита Колмогоровым в его известной работе «Об аналитических методах в теории вероятностей» (*Math. Ann.*, **104** (1931), 415—458, перевод в *Успехах Матем. Наук*, **5** (1938), 5—41). В этой работе Колмогоров отмечает возможность добавить к правой части произвольный «диффузионный член»

$$\gamma \frac{\partial^2 m_t}{\partial y^2} - c\frac{\partial m_t}{\partial y}.$$

Формула (2.1) представляет собой лишь частный случай этой более общей формулы. Уже первые теоремы существования охватывали случай нестационарного общего уравнения [Feller W., *Math. Ann.*, **113** (1936).]

ее мгновенному значению, то логарифм энергии убывает с постоянной скоростью. Очевидно, что после соответствующей замены обозначений применима модель (2.1).

б) *Звездное излучение*[1]). В этой модели переменная t обозначает *расстояние* и $\mathbf{X}(t)$ — интенсивность светового луча, проходящего в пространстве. Предполагается, что (в экваториальной плоскости) каждый элемент объема порождает излучение постоянной интенсивности и, следовательно, $\mathbf{X}(t)$ возрастает линейно. Но в пространстве также размещены поглощающие черные облака, которые мы будем представлять себе в виде пуассоновского множества точек. Встречая облако, каждый луч теряет (случайную) часть интенсивности, так что мы оказываемся в точности в тех условиях, которые привели нас к (2.1). Кажется правдоподобным (и на самом деле это может быть доказано), что плотность m_t величины $\mathbf{X}(t)$ приближается к *стационарной плотности* m, которая не зависит от t и удовлетворяет (2.1) с левой частью, замененной нулем.

Амбарцумян предполагает, в частности, что потеря интенсивности при прохождении отдельного облака регулируется переходным ядром (1.4). В этом случае явный вид решения содержится в (1.5), но он представляет вторичный интерес. Более важен (и легко проверяется) тот факт, что (2.1) имеет независящее от времени (или стационарное) решение, задаваемое гамма-плотностью

$$m(y) = \left(\frac{\alpha}{c}\right)^{\lambda+1} \frac{1}{\Gamma(\lambda+1)} y^{\lambda} e^{-(\alpha/c)y}, \quad y > 0. \tag{2.2}$$

Этот результат показывает, что относящаяся к делу информация может быть получена из (2.1), даже минуя отыскание явного решения. Например, легко найти прямым интегрированием, что стационарное решение имеет математическое ожидание $(\mu+c)/\alpha$, где μ — математическое ожидание для распределения поглощаемой энергии V.

в) *Задача о разорении* из гл. VI, 5 представляет собой частный случай, в котором ядро k зависит от разности аргументов. Значения этого процесса получаются добавлением $-ct$ к значениям некоторого обобщенного пуассоновского процесса. Аналогичные «задачи о разорении» могут быть сформулированы для любого псевдопуассоновского процесса. Они приводят к уравнению (2.1).

---

[1]) Физические предпосылки взяты из статьи В. А. Амбарцумяна, О флуктуациях яркости Млечного Пути, *ДАН СССР*, **44** (1944), 223—226. В этой статье косвенным путем получен один вариант (2.1).

§ 3. Скачкообразные процессы

В псевдопуассоновском процессе время ожидания следующего скачка имеет фиксированное показательное распределение с математическим ожиданием $1/\alpha$. Мы приходим к естественному обобщению, допуская, что это распределение может зависеть от значения $\mathbf{X}(t)$ в данный момент (в примере (1а) это сводится к предположению, что вероятность столкновения зависит от энергии частицы). Марковский характер процесса требует, чтобы распределение времени до следующего скачка было показательным, но его математическое ожидание может зависеть от значения $\mathbf{X}(t)$ в данный момент. В соответствии со сказанным мы введем следующие

Основные постулаты. *При условии* $\mathbf{X}(t)=x$ *время ожидания следующего скачка имеет показательное распределение с математическим ожиданием* $1/\alpha(x)$ *и не зависит от прошлого. Вероятность того, что следующий скачок приводит к значению из некоторого множества* Γ, *равна* $K(x, \Gamma)$.

В аналитических терминах эти постулаты сводятся к интегральным уравнениям для вероятностей перехода $Q_t(x, \Gamma)$ рассматриваемого процесса (в предположении, что такой процесс на самом деле существует). Рассмотрим фиксированную точку x и фиксированное множество Γ, *не содержащее* x. Событие $\{\mathbf{X}(t) \in \Gamma\}$ может произойти только при условии, что первый скачок из x произошел в некоторый момент $s<t$. При этом условии условная вероятность события $\{\mathbf{X}(t) \in \Gamma\}$ получается интегрированием $K(x, dy)\,Q_{t-s}(y, \Gamma)$ по множеству Ω всех возможных значений y. Теперь момент первого скачка — это случайная величина с показательной плотностью $\alpha(x)e^{-\alpha(x)s}$. Интегрируя по отношению к этой плотности, мы получаем вероятность события $\{\mathbf{X}(t) \in \Gamma\}$ в форме

$$Q_t(x, \Gamma) = \alpha(x) \int_0^t e^{-\alpha(x)s}\, ds \int_\Omega K(x, dy)\, Q_{t-s}(y, \Gamma). \quad (3.1а)$$

Для множества Γ, содержащего x, мы должны добавить вероятность того, что до момента t не произойдет ни одного скачка. Мы получим

$$Q_t(x, \Gamma) = e^{-\alpha(x)t} + \alpha(x) \int_0^t e^{-\alpha(x)s}\, ds \int_\Omega K(x, dy)\, Q_{t-s}(y, \Gamma). \quad (3.1б)$$

Эти два уравнения, справедливые при $x \in \bar{\Gamma}$ и $x \in \Gamma$ соответственно, являются аналитическими эквивалентами основных постулатов. Эту пару уравнений можно заменить одним интегро-

дифференциальным уравнением. В самом деле, замена переменных $s=t-\tau$ делает дифференцирование по t тривиальным и приводит к уравнению

$$\frac{\partial Q_t(x,\Gamma)}{\partial t} = -\alpha(x)\,Q_t(x,\Gamma)+\alpha(x)\int_\Omega K(x,dy)\,Q_t(y,\Gamma). \quad (3.2)$$

Это — не что иное как *обратное уравнение Колмогорова*, которое служит отправной точкой для аналитического подхода к задаче, так как позволяет избежать раздражающего различия между двумя указанными случаями.

Обратное уравнение (3.2) допускает простую интуитивную интерпретацию, которая позволяет представить основные постулаты в более близких к практике терминах. В терминах разностных отношений (3.2) равносильно соотношению

$$Q_{t+h}(x,\Gamma)=[1-\alpha(x)\,h]\,Q_t(x,\Gamma)+ \\ +\alpha(x)\,h\int_\Omega K(x,dy)\,Q_t(y,\Gamma)+o(h). \quad (3.3)$$

Для интуитивной интерпретации последнего рассмотрим изменение состояния на временно́м интервале $\overline{0,t+h}$ как результат изменения за короткий интервал $\overline{0,h}$ и последующего изменения в интервале $\overline{h,t+h}$ длины t. Тогда, очевидно, (3.3) утверждает, что если $\mathbf{X}(0)=x$, то вероятность одного скачка в интервале $\overline{0,h}$ равна $\alpha(x)h+o(h)$, а вероятность большего числа скачков равна $o(h)$. Наконец, если на $\overline{0,h}$ скачок происходит, то условные вероятности тех или иных переходов определяются мерой $K(x,dy)$. Эти три постулата приводят к (3.3), а тем самым и к (3.2). По существу они повторяют основные постулаты [1]).

С вероятностной точки зрения обратное уравнение представляется несколько искусственным, так как в нем окончательное состояние Γ играет роль параметра и (3.2) описывает зависимость $Q_t(x,\Gamma)$ от начального положения x. Казалось бы более естественным получить уравнение для Q_{t+h}, разбивая интервал $\overline{0,t+h}$ на длинный начальный интервал $0,t$ и короткий конечный интервал $\overline{t,t+h}$. Вместо (3.3) мы получим *формально*

$$Q_{t+h}(x,\Gamma)=\int_\Gamma Q_t(x,dz)[1-\alpha(z)\,h]+ \\ +\int_\Omega Q_t(x,dz)\,\alpha(z)\,hK(z,\Gamma)+o(h) \quad (3.4)$$

[1]) Дифференцируемость Q_t по отношению к t и тот факт, что вероятность более чем одного скачка равна $o(h)$, теперь принимаются в качестве новых постулатов; в то же время они вытекают из прежней более сложной формулировки.

и, следовательно,

$$\frac{\partial Q_t(x, \Gamma)}{\partial t} = -\int_{\Gamma} Q_t(x, dz)\,\alpha(z) + \int_{\Omega} Q_t(x, dz)\,\alpha(z) K(z, \Gamma). \quad (3.5)$$

Это — *прямое уравнение Колмогорова* (известное физикам, как *уравнение неразрывности или уравнение Фоккера — Планка*). В частном случае постоянного α оно сводится к (1.6).

Формальный характер приведенного вывода мы подчеркнули потому, что на самом деле прямое уравнение не вытекает из наших основных постулатов. Это объясняется тем, что величина $o(h)$ в (3.3) зависит от z, и так как z служит в (3.4) переменной интегрирования, то эта величина должна была бы стоять под знаком интеграла. Но тогда возникает вопрос о том, сходятся ли интегралы в (3.5) и законен ли предельный переход при $h \to 0$. [Ни одно из этих затруднений не возникает в связи с обратным уравнением, так как начальное состояние x при этом фиксировано и интеграл в (3.2) существует в силу ограниченности Q_t.]

Прямое уравнение можно обосновать, добавляя к нашим основным постулатам необходимое ограничение на остаточный член в (3.3), но такой вывод перестал бы быть интуитивно привлекательным. Кроме того, кажется невозможным сформулировать условия, которые охватывали бы все типичные случаи, встречающиеся на практике. Если допустить, что функция α может быть неограниченной, то существование интегралов в (3.5) сомнительно и уравнение нельзя обосновать априори. С другой стороны, обратное уравнение есть необходимое следствие основных предположений. Поэтому в качестве отправной точки лучше использовать его, а затем исследовать, в каких условиях из него может быть выведено прямое уравнение.

Решение обратного уравнения легко построить, используя последовательные приближения, имеющие простой вероятностный смысл. Обозначим через $Q_t^{(n)}(x, \Gamma)$ вероятность перехода из $\mathbf{X}(0) = x$ в $\mathbf{X}(t) \in \Gamma$ *с не более чем n скачками*. Переход без скачков возможен, только если $x \in \Gamma$, и так как время пребывания в x имеет показательное распределение, мы имеем

$$Q_t^{(0)}(x, \Gamma) = e^{-\alpha(x)t} K^{(0)}(x, \Gamma) \quad (3.6)$$

(где правая часть равна 1 или 0 в зависимости от того, входит x в Γ или нет). Предположим затем, что первый скачок происходит в момент $s < t$ и приводит из x в y. Суммируя по всем возможным s и y, мы получаем [как и в (3.1)] рекурсивную

формулу

$$Q_t^{(n+1)}(x, \Gamma) = Q_t^{(0)}(x, \Gamma) + \int_0^t e^{-\alpha(x)s}\alpha(x)\,ds \int K(x, dy)\, Q_{t-s}^{(n)}(y, \Gamma), \tag{3.7}$$

верную при $n=0, 1, \ldots$. Очевидно, что $Q_t^{(0)} \leqslant Q_t^{(1)}$, и по индукции $Q_t^{(1)} \leqslant Q_t^{(2)} \leqslant \ldots$. Сверх того, все эти ядра ограничены сверху единицей. В самом деле, это тривиально верно при $n=0$. Если это верно при некотором n, то внутренний интеграл в (3.7) будет $\leqslant 1$, и потому второй член справа $\leqslant 1 - e^{-\alpha(x)t}$. В силу (3.6) отсюда следует $Q_t^{(n+1)}(x, \Gamma) \leqslant 1$. Мы видим теперь, что при каждых x и Γ существует предел

$$Q_t^{(\infty)}(x, \Gamma) = \lim_{n\to\infty} Q_t^{(n)}(x, \Gamma) \tag{3.8}$$

и что $Q_t^{(\infty)}$ является или стохастическим или субстохастическим ядром. Очевидно, что $Q_t^{(\infty)}$ удовлетворяет обратному уравнению [1]) в его интегральном варианте (3.1) и что любое другое положительное решение Q_t больше $Q_t^{(\infty)}$. По этой причине $Q_t^{(\infty)}$ называют *минимальным решением* обратного уравнения.

Если ядро $Q_t^{(\infty)}$ — строго стохастическое, то любое другое решение должно было бы приписывать всему пространству вероятность $\geqslant 1$. Отсюда следует, что *решение единственно, если минимальное решение строго стохастическое.* Следующий ниже пример показывает, что минимальное решение может быть несобственным. Как мы увидим в гл. XIV, 8, это соответствует тому случаю, когда за конечный промежуток времени может произойти бесконечно много скачков. По самому способу построения $Q_s^{(\infty)}$ равно вероятности перехода с конечным числом скачков или, что сводится к тому же самому, равно переходным вероятностям до первой точки накопления моментов скачков. Начиная с этого момента процесс может быть продолжен различными способами без нарушения наших основных постулатов.

Во избежание новторений мы отложим до гл. XIV, 7 доказательство того основного факта, что *минимальное решение* Q_t^{∞}

[1]) Детали этого доказательства в варианте, использующем преобразование Лапласа, даны в гл. XIV, 7. Там же обоснованы и другие утверждения, сформулированные в тексте. Обозначения, связанные с преобразованием Лапласа, более элегантны. Однако настоящее доказательство обладает тем преимуществом, что оно применимо и в *нестационарном случае* (т. е. когда α и K зависят от t). См. Feller W., *Amer. Math. Soc.*, **48** (1940), 488—515 [*поправки* в т. 58, стр. 474].

обратного уравнения автоматически удовлетворяет прямому уравнению и является минимальным также и для него. На этом пути прямое уравнение получается без всяких дополнительных предположений, и в то же самое время становится понятным, что непосредственный его вывод невозможен. Рассуждение, приводящее к (3.4), опирается на рассмотрение последнего скачка до момента $t+h$. Теперь из наших основных постулатов вытекает существование *первого* скачка после момента t, но не последнего скачка до t. Если происходит бесконечно много скачков, то последний не обязан существовать. Осуществляется или нет эта возможность в данной ситуации (т. е. при данных α и K) должно быть установлено с помощью анализа, а не на основе постулатов. Кроме того, как это ни удивительно, имеется много процессов, в которых за конечный промежуток времени происходит бесконечно много скачков, для которых прямое уравнение все же справедливо [хотя его вывод из (3.4) является абсурдом].

Эти замечания должны быть полезны для понимания роли обратного уравнения в диффузионных процессах и теории полугрупп.

Отметим в заключение, что (по контрасту с процессами § 2 и диффузионными процессами, в которых используется дифференцирование по x) чисто скачкообразные процессы *никоим образом не зависят от природы пространства состояний* и наши формулы применимы всякий раз, когда определено стохастическое ядро K.

Пример. *Счетные пространства состояний.* Если случайные величины $\mathbf{X}(t)$ положительны и целочисленны, то основное выборочное пространство (пространство состояний) состоит из чисел $1, 2, \ldots$. Теперь достаточно знать вероятности $P_{ik}(t)$ перехода от одного целого числа к другому. Все прочие переходные вероятности получаются суммированием по k. Теория подобных марковских процессов была намечена в **1**, гл. XVII, 9, где рассматривались и нестационарные переходные вероятности. Чтобы выделить внутри этой теории стационарный случай, следует считать коэффициенты c_i и вероятности p_{ik} постоянными (не зависящими от t). Тогда прежние предположения становятся равносильными принятым здесь и, как легко видеть, две системы уравнений Колмогорова, полученные в **1**, гл. XVII, 9 становятся частными случаями уравнений (3.2) и (3.5) [с заменой $\alpha(i)$ на c_i и $K(i, j)$ на p_{ik}]. Расходящийся процесс размножения из **1**, гл. XVII, 4 служит примером процесса с бесконечным числом скачков за конечный промежуток времени. Мы

вернемся к этому процессу в гл. XIV, 8, чтобы продемонстрировать возможность различных его продолжений. ▷

§ 4. Диффузионные процессы в $\mathscr{R}^1$

Покончив с процессами, в которых все изменения происходят скачками, мы обратимся к другому крайнему случаю, к процессам, у которых (с вероятностью единица) выборочные функции непрерывны. Их теория удивительно похожа на развитую в предыдущем параграфе, однако основные уравнения требуют более сложного анализа. Мы удовлетворимся поэтому выводом обратного уравнения и кратким резюме результатов, относящихся к минимальным решениям и другим проблемам. Прототипом диффузионных процессов служит броуновское движение (или винеровский процесс). Это процесс с независимыми нормально распределенными приращениями. Его переходные вероятности имеют плотности $q_t(x, y)$, которые нормальны с математическим ожиданием x и дисперсией at, где $a>0$ — некоторая постоянная. Эти плотности удовлетворяют обычному уравнению диффузии

$$\frac{\partial q_t(x, y)}{\partial t} = \frac{1}{2} a \frac{\partial^2 q_t(x, y)}{\partial x^2}. \tag{4.1}$$

Мы покажем теперь, что и другие переходные вероятности удовлетворяют соответствующим уравнениям в частных производных. Цель этого вывода — дать представление о типах процессов и о возникающих задачах, т. е. он должен служить начальным шагом. Вследствие этого мы не будем тратить усилий на достижение общности и полноты изложения.

Из предположенной нормальности распределений очевидно, что в броуновском движении приращения за короткий промежуток времени длины t обладают следующими свойствами: (i) при фиксированном $\delta>0$ вероятность смещения, превосходящего δ, равна $o(t)$; (ii) математическое ожидание смещения равно нулю; (iii) его дисперсия равна at. Мы сохраним первое требование, а два других приспособим к условиям неоднородной среды. Иначе говоря, мы допустим, что a зависит от x, и допустим ненулевое среднее смещение. В этой обстановке математическое ожидание и дисперсия смещения уже не будут в точности пропорциональны t, и мы можем лишь постулировать, что при данном $\mathbf{X}(\tau)=x$ смещение $\mathbf{X}(t+\tau)-\mathbf{X}(\tau)$ имеет математическое ожидание $b(x)t+o(t)$ и дисперсию $a(x)t+o(t)$. Моменты не обязаны существовать, но имея в виду первое требование, естественно рассмотреть урезанные моменты. Мы приходим к следующим постулатам.

Постулаты[1]) *относительно переходных вероятностей* Q_t. *При каждом* $\delta>0$ *при* $t\to 0$

$$\frac{1}{t}\int\limits_{|y-x|\geqslant\delta} Q_t(x,\,dy)\to 0, \tag{4.2}$$

$$\frac{1}{t}\int\limits_{|y-x|<\delta} (y-x)\,Q_t(x,\,dy)\to b(x), \tag{4.3}$$

$$\frac{1}{t}\int\limits_{|y-x|<\delta} (y-x)^2\,Q_t(x,\,dy)\to a(x). \tag{4.4}$$

Заметим, что если (4.2) верно при *всех* $\delta>0$, то асимптотическое поведение величин в левой части (4.3) и (4.4) не зависит от δ. Возможно поэтому в последних двух соотношениях заменить δ на 1.

Первое условие делает мало вероятными большие смещения и было введено в 1936 г. в надежде, что оно необходимо и достаточно для непрерывности выборочных функций[2]). Оно было названо именем Линдеберга ввиду сходства с его условием применимости центральной предельной теоремы. Можно показать, что при довольно слабых предположениях о регулярности переходных вероятностей существование пределов в (4.3) и (4.4) на самом деле вытекает из условия Линдеберга (4.2). Мы не будем обсуждать подобные детали, так как сейчас мы не интересуемся систематическим развитием теории[3]). Наша цель очень умерена — объяснить природу и эмпирический смысл диффузионных уравнений в простейшей ситуации. На пути к ней мы покажем формально, как из (4.2)—(4.4) можно вывести некоторые дифференциальные уравнения. Мы, однако, не будем обсуждать, когда существует решение этих уравнений. Поэтому мы допустим,

---

[1]) Первый вывод (4.1) из вероятностных предположений принадлежит Эйнштейну. Первый вывод обратного уравнения (4.6) и прямого уравнения (5.2) был дан в известной работе А. Н. Колмогорова 1931 г. (см. § 1). Улучшенный вариант постулатов, данный здесь, принадлежит Феллеру (1936). Феллер же доказал первые теоремы существования и исследовал связь между двумя уравнениями.

[2]) Это предположение было подтверждено Д. Рэем (D. Ray).

[3]) Современная теория полугрупп позволила автору получить наиболее общую форму обратного уравнения (или производящего оператора) для марковских процессов, удовлетворяющих требованию типа условия Линдеберга. Классические дифференциальные операторы заменяются при этом их модернизированными вариантами: роль коэффициента b играет «естественный масштаб» (natural scale), а роль a играет «мера скорости» (speed measure). Такие процессы были объектом плодотворных исследований Дынкина и его школы, с одной стороны, и К. Ито и Г. Маккина — с другой. Теория изложена в книгах этих авторов.

что коэффициенты a и b суть ограниченные непрерывные функции и $a(x)>0$.

В качестве пространства состояний мы возьмем конечный или бесконечный интервал I на прямой и, как и ранее, будем придерживаться следующего соглашения: если пределы не указаны, то интегрирование производится по всему I. Для упрощения записи и в качестве подготовительного шага для применения теории полугрупп мы введем преобразования

$$u(t,\ x)=\int Q_t(x,\ dy)\,u_0(y), \tag{4.5}$$

переводящие (при фиксированном t) ограниченную непрерывную «начальную функцию» u_0 в функцию [1]) со значениями $u(t,\ x)$.

Ясно, что, зная правую часть (4.5) для всех начальных функций u_0, мы можем однозначно определить Q_t. Теперь мы покажем, что при очень слабых условиях регулярности функция u должна удовлетворять *обратному уравнению*

$$\frac{\partial u}{\partial t}=\frac{1}{2}a\frac{\partial^2 u}{\partial x^2}+b\frac{\partial u}{\partial x}, \tag{4.6}$$

обобщающему стандартное уравнение диффузии (4.1). Мы ищем функцию u, удовлетворяющую (4.6), и такую, что $u(t,\ x)\to u_0(x)$ при $t\to 0$. В случае единственности это решение необходимо имеет вид (4.5) и Q_t называется функцией Грина данного уравнения. Случаи, когда единственности нет, будут рассмотрены в следующем параграфе [2]).

Чтобы получить обратное уравнение (4.6), мы отправимся от тождества

$$u(s+t,\ x)=\int Q_s(x,\ dy)\,u(t,\ y),\qquad s,\ t>0, \tag{4.7}$$

которое очевидным образом следует из уравнения Чепмена — Колмогорова (1.3). При $h>0$ из (4.7) выводим

$$\frac{u(t+h,\ x)-u(t,\ x)}{h}=\frac{1}{h}\int Q_h(x,\ dy)[u(t,\ y)-u(t,\ x)]. \tag{4.8}$$

Допустим теперь, что вероятности перехода Q_t достаточно регулярны, во всяком случае, что функция u в (4.5) имеет две ограниченные непрерывные производные по x (хотя бы для бесконечно дифференцируемой u_0). При фиксированных x и $\varepsilon>0$ най-

[1]) В терминах теории случайных процессов $u(t,x)$ есть условное математическое ожидание $u_0(\mathbf{X}(t))$ при гипотезе $\mathbf{X}(0)=x$.

[2]) Исследование диффузионных уравнений с помощью преобразования Лапласа дано в гл. XIV, 5.

дется в силу формулы Тейлора такое $\delta>0$, что

$$\left|u(t,\ y)-u(t,\ x)-(y-x)\frac{\partial u(t,\ x)}{\partial x}-\right.$$
$$\left.-\frac{1}{2}(y-x)^2\frac{\partial^2 u(t,\ x)}{\partial x^2}\right|<\varepsilon|y-x|^2 \qquad (4.9)$$

при всех $|y-x|\leqslant\delta$. Выбрав таким образом δ, рассмотрим в (4.8) отдельно интегралы по области $|y-x|>\delta$ и области $|y-x|\leqslant\delta$. Первый интеграл стремится к нулю в силу условия Линдеберга (4.2) и ограниченности u. Благодаря условиям (4.3) и (4.4) ясно, что при достаточно малых h второй интеграл отличается от правой части (4.6) менее, чем на $\varepsilon\cdot a(x)$. Так как ε произвольно, это означает, что при $h\to 0$ правая часть (4.8) сходится к правой части (4.6). В соответствии с этим, существует по крайней мере правая производная $\frac{\partial u}{\partial t}$ и она удовлетворяет (4.6). Главный вывод изложенной теории состоит, грубо говоря, в следующем. *Если переходные вероятности марковского процесса удовлетворяют условию непрерывности* (4.2), *то процесс определяется коэффициентами b и a.* Это утверждение звучит как чисто теоретическое, но в практических ситуациях именно коэффициенты b и a задаются априори, исходя из их эмпирического смысла и характера процесса.

Для объяснения смысла b и a забудем, что в определениях (4.3) и (4.4) участвуют урезанные моменты. Рассмотрим приращение $\mathbf{X}(t+\tau)-\mathbf{X}(\tau)$ за короткий промежуток времени, предполагая, что $\mathbf{X}(\tau)=x$. Если бы моменты в (4.3) и (4.4) не были бы урезаны, это приращение имело бы условное математическое ожидание $b(x)t+o(t)$ и условную дисперсию $a(x)t-b^2(x)t^2+$ $+o(t)=a(x)t+o(t)$. Таким образом, $b(x)$ служит мерой локальной средней скорости смещения (которая может быть нулем по причинам симметрии), а $a(x)$ служит мерой дисперсии. Мы будем называть b *инфинитезимальной скоростью* (сносом), а a — *инфинитезимальной дисперсией.*

Следующие примеры показывают, как определяются эти коэффициенты в конкретных ситуациях.

Примеры. а) *Броуновское движение.* Предположим, что пространство (ось x) однородно и симметрично. Тогда $a(x)$ не должно зависеть от x, а $b(x)$ должно быть нулем. Мы приходим к классическому уравнению диффузии (4.1).

б) *Процесс Орнштейна — Уленбека* получается при воздействии упругой силы на совершающую броуновское движение частицу. Аналитически это означает наличие сноса по направлению к началу координат, причем величина сноса пропорциональна

расстоянию от начала, т. е. $b(x) = -\rho x$. Так как это не изменяет *инфинитезимальной дисперсии*, то функция $a(x)$ остается постоянной, скажем, равной единице. Обратное уравнение принимает вид

$$\frac{\partial u(t, x)}{\partial t} = \frac{1}{2}\frac{\partial^2 u(t, x)}{\partial x^2} - \rho x \frac{\partial u(t, x)}{\partial x}. \tag{4.10}$$

Это уравнение решается удивительно легко. В самом деле, замена

$$v(t, x) = u(t, xe^{\rho t})$$

приводит его к виду

$$e^{2\rho t}\frac{\partial v}{\partial t} = \frac{1}{2}\frac{\partial^2 v}{\partial x^2} \tag{4.11}$$

и последующая замена

$$\tau = \frac{1 - e^{-2\rho t}}{2\rho} \tag{4.12}$$

превращает (4.11) в стандартное уравнение диффузии (4.1). Отсюда следует, что *переходные плотности $q_t(x, y)$ процесса Орнштейна — Уленбека нормальны с математическим ожиданием $xe^{-\rho t}$ и дисперсией τ, определяемой по* (4.12).

В примере гл. III, (8д) было показано, что процесс Орнштейна — Уленбека, описываемый уравнением (4.10) и нормальным начальным распределением, является единственным *стационарным нормальным марковским процессом* со стационарными переходными вероятностями. (Броуновское движение включается сюда как частный случай при $\rho = 0$.)

в) *Диффузия в генетике.* Рассмотрим популяцию, включающую различные поколения и имеющую постоянный размер (N). (Типичный пример — поле зерновых культур.) Имеется $2N$ генов, каждый из них принадлежит одному из двух генотипов. Обозначим $\mathbf{X}_n$ *долю* генов типа A. Если никакой ген не имеет преимуществ при отборе и если не рассматривается возможность мутаций, то гены $(n+1)$-го поколения могут рассматриваться как случайная выборка объема $2N$ из совокупности генов n-го поколения. Тогда $\mathbf{X}_n$ будет марковским процессом, $0 \leqslant \mathbf{X}_n \leqslant 1$, и при условии, что $\mathbf{X}_n = x$, распределение величины $2N\mathbf{X}_{n+1}$ будет биномиальным с математическим ожиданием $2Nx$ и дисперсией $2Nx(1-x)$. Изменение за одно поколение будет иметь математическое ожидание 0 и дисперсию, пропорциональную $x(1-x)$.

Допустим теперь, что мы прослеживаем огромное число поколений и вводим такую шкалу, что изменения представляются непрерывными. В этой аппроксимации мы будем иметь дело с марковским процессом, переходные вероятности которого удовлетворяют нашим основным условиям с $b(x) = 0$ и $a(x)$, пропор-

циональной $x(1-x)$. Множитель пропорциональности зависит от выбора единицы измерения времени. Его можно считать равным 1. Тогда (4.6) принимает вид

$$\frac{\partial u(t, x)}{\partial t} = x(1-x)\frac{\partial^2 u(t, x)}{\partial x^2}, \tag{4.13}$$

и на этот раз процесс протекает на конечном интервале $\overline{0, 1}$. Влияние селекции и мутаций породило бы снос и привело бы к появлению в уравнении (4.13) члена первого порядка.

Генетическую модель, математически равносильную нашей, предложили Р. Фишер и С. Райт, однако их отправной пункт и их рассуждения были другими [1]). Генетические выводы в какой-то степени сомнительны в силу предположенной неизменности размера популяции. Эффект этого допущения не всегда правильно оценивается. Правильная модель [2]) приводит к уравнению с двумя пространственными переменными (частота генов и размер популяции).

г) *Рост популяции.* Мы желаем описать рост большой популяции, в которой индивидуумы статистически независимы и скорость размножения не зависит от размера популяции. Для очень большой популяции процесс приближенно непрерывен, т. е. описывается диффузионным уравнением. Из независимости индивидуумов следует, что инфинитезимальные скорость и дисперсия пропорциональны размеру популяции. Таким образом, процесс описывается обратным уравнением (4.6) с $a=\alpha x$ и $b=\beta x$. Постоянные α и β зависят от выбора единиц измерения времени и размера популяции, и при подходящем выборе можно считать $\alpha=1$ и $\beta=1$, -1, или 0 (в зависимости от знака скорости).

В **1**, гл. XVII, (5.7) *этот же самый* рост популяции описывался дискретной моделью. Предполагалось, что при условии $\mathbf{X}(\tau)=n$ вероятности событий $\mathbf{X}(t+\tau)=n+1$, $n-1$ и n отличаются от λnt, μnt и $1-(\lambda+\mu)nt$ соответственно членами вида $o(t)$, так что инфинитезимальная скорость и дисперсия равны $(\lambda-\mu)n$ и $(\lambda+\mu)n$. Диффузионный процесс получается простым переходом к пределу. Можно показать, что его переходные вероятности получаются предельным переходом из вероятностей в дискретной модели.

Подобная аппроксимация дискретных процессов диффузионными процессами часто бывает весьма практичной. Типичным

[1]) Явное решение (4.13) (принадлежащее М. Кимура) см. в книге Баруча-Рида (A. T. Barucha-Reid, 1960).

[2]) Feller W., Second Berkeley Symposium on Math Statist. and Probability, 1951, 227—246.

примером может служить переход от обычных случайных блужданий к диффузионным процессам, описанным в **1**, XIV, 6 [продолжение см. в примере (5. а)].

д) *Преобразования.* Пусть Φ — непрерывная и возрастающая функция. Мы можем использовать ее для преобразования произвольного стохастического процесса $\{\mathbf{X}(t)\}$ в новый процесс $\mathbf{Y}(t)=\Phi(\mathbf{X}(t))$. Если функция Φ дважды дифференцируема и $\mathbf{X}$-процесс удовлетворяет условиям (4.2)—(4.4), использованным для вывода обратного уравнения, то им удовлетворяет и преобразованный процесс. Последний поэтому описывается обратным уравнением, которое легко вывести [1]). Это преобразование может быть использовано для того, чтобы описать некоторые сложные диффузионные процессы в терминах простого броуновского движения. В действительности, используя современные обобщенные диффузионные операторы, можно получить (не налагая никаких условий дифференцируемости) *все* диффузионные процессы преобразованием броуновского движения. Однако подобные преобразования требуют более тонкого подхода, опирающегося на понятие локального времени, введенное П. Леви. Этот метод широко использован в книге Ито и Маккина.

§ 5. Прямое уравнение. Граничные условия

Как объяснено в предыдущем параграфе, аналитическая характеризация переходных вероятностей необходимо опирается на обратное уравнение (4.6), хотя оно и описывает $Q_t(x, \Gamma)$ как функцию начального положения x при фиксированном Γ. Изучая зависимость от множества Γ, мы предположили для простоты, что Q_t обладает плотностью вероятности $q_t(x, y)$. При заданной плотности v_0 величины $\mathbf{X}(0)$ плотность $\mathbf{X}(s)$ равна

$$v(s, y)=\int v_0(x)\, q_s(x, y)\, dx \tag{5.1}$$

[это соотношение — «прямой» аналог (4.5)]. Многие известные факты, касающиеся дифференциальных уравнений в частных производных, делают правдоподобным заключение о том, что v должно удовлетворять уравнению

$$\frac{\partial v(s, y)}{\partial s}=\frac{1}{2}\frac{\partial^2}{\partial y^2}[a(y)\, v(s, y)]-\frac{\partial}{\partial y}[b(y)\, v(s, y)]. \tag{5.2}$$

[1]) Из формулы Тейлора для Φ видно, что при данном $\mathbf{Y}(t)=\Phi(\mathbf{X}(t))$ новая инфинитезимальная скорость равна $\Phi'(x)\, b(x)+\frac{1}{2}\Phi''(x)$, а дисперсия равна $\Phi'^2(x)a(x)$.

В теории вероятностей последнее известно под названием *прямого* уравнения или уравнения *Фоккера — Планка*.

Сейчас мы покажем, какого рода информацию можно получить из (5.2) легче, чем из обратного уравнения.

Пример. а) *Рост популяции.* Пример (4 г) приводит к прямому уравнению

$$\frac{\partial v(s, y)}{\partial s} = \alpha \frac{\partial^2 y v(s, y)}{\partial y^2} - \beta \frac{\partial y v(s, y)}{\partial y}. \tag{5.3}$$

Можно доказать, что при данной начальной плотности v_0 существует лишь одно решение. Хотя для него трудно получить явные формулы, массу информации о нем можно получить прямо из уравнения. Например, чтобы вычислить математическое ожидание размера популяции $M(s)$, можно умножить (5.3) на y и проинтегрировать обе части по y от 0 до ∞. Слева мы получим при этом производную $M'(s)$. Предполагая, что v стремится к нулю на бесконечности быстрее, чем $1/y^2$, и производя интегрирование по частям, мы увидим, что правая часть равна $\beta M(s)$. Таким образом,

$$M'(s) = \beta M(s),$$

так что $M(s)$ пропорционально $e^{\beta s}$. Аналогичные формальные выкладки показывают, что дисперсия пропорциональна $2\alpha\beta^{-1}e^{\beta s}(e^{\beta s} - 1)$ [сравните с аналогичным результатом для дискретного случая в **1**, гл. XVII, (5.10) и (11.9)]. Конечно, эти выкладки требовали бы обоснования, но и на эвристическом уровне их нельзя было бы получить из обратного уравнения, минуя явное вычисление q_t. ▶

Теперь задача состоит в том, чтобы понять, в какой степени прямое уравнение (5.2) является следствием [1]) обратного уравнения (4.6). Соответствующая теория в общем параллельна данной в § 3 (для скачкообразных процессов). Мы дадим здесь без доказательства краткое перечисление результатов.

Рассмотрим обратное уравнение (4.6) на открытом интервале x_1, x_2, который может быть конечным или бесконечным. Мы предполагаем, конечно, что $a>0$ и что коэффициенты a и b достаточно регулярны для того, чтобы (5.2) имело смысл. При

[1]) Более точно, задача состоит в том, чтобы отыскать «истинную» форму прямого уравнения, так чтобы оно стало следствием обратного уравнения. Отметим, что последнее определяет процесс, если $a>0$ и b непрерывны, в то время как обратное уравнение (5.2) требует существования производных a и b «Истинное» прямое уравнение может быть написано во всех случаях, но оно связано с обобщенными дифференциальными операторами, упомянутыми в сноске 3 на стр. 395.

этих условиях существует *единственное минимальное решение* Q_t, такое, что (4.5) дает решение обратного уравнения (4.6). При фиксированных t и x ядро $Q_t(x, \Gamma)$ может определять *несобственное распределение*. В любых обстоятельствах распределения Q_t имеют плотности и функция v, определяемая в (5.1), удовлетворяет прямому уравнению. На самом деле это решение будет минимальным в очевидном смысле слова (который был уточнен в § 3). В таком понимании прямое уравнение является следствием обратного уравнения. Однако эти уравнения определяют процесс единственным образом только в случае, когда минимальное решение собственно. Во всех других случаях природа процесса определяется дополнительными граничными условиями.

Характер граничных условий лучше всего может быть понят по аналогии с простым случайным блужданием на $\overline{0, \infty}$, обсуждавшимся в **1**, гл. XIV. Можно принять различные соглашения относительно течения процесса после его возвращения в нуль. В *задаче о разорении* процесс на этом останавливается. В этом случае говорят, что нуль является *поглощающим экраном*. С другой стороны, если нуль представляет собой *отражающий экран*, то частица мгновенно возвращается в точку 1 и процесс, таким образом, продолжается бесконечно. Следует подчеркнуть, что граничные условия нужны тогда и только тогда, когда граничная точка достижима. Событие «граничная точка x_2 достигается ранее момента t» вполне определено для диффузионных процессов ввиду непрерывности их траекторий. Оно тесно связано по своему характеру с событием «до момента t происходит бесконечно много скачков» в скачкообразных процессах.

В некоторых диффузионных процессах с вероятностью единица не достигается ни одна из граничных точек. Таково, например, броуновское движение [пример (4а)]. В подобных случаях минимальное решение является распределением вероятностей и не существует иных решений. В других ситуациях процесс описывается минимальным решением лишь до момента достижения границы. Оно соответствует поглощающим границам, т. е. описывает процесс, который останавливается по достижении границ. Это наиболее важный тип процессов, и не только потому, что все остальные получаются из него, но и потому, что все вероятности *первого прохождения* состояний могут быть вычислены искусственным введением поглощающих барьеров. Этот метод широко применим, но мы объясним его на простейшем примере (он уже был нами неявно использован при изучении случайных блужданий и в других местах, см., например, задачу 15 в **1**, гл. XVII).

Примеры. в) *Один поглощающий барьер. Время первого прохождения.* Рассмотрим броуновское движение на $\overline{0, \infty}$ с поглощающим барьером в нуле. Более точно, броуновское движение, начинающееся в точке $x>0$ в момент 0, останавливается в момент первого попадания в нуль. По причинам симметрии и обратное, и прямое уравнения принимают вид классического диффузионного уравнения

$$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}. \tag{5.4}$$

Нужное *граничное условие* таково: $q_t(0, y)=0$ при всех t (так же как и в примере случайных блужданий). Проверить это можно либо предельным переходом из формул **1**, гл. XIV, 6 или исходя из минимального характера соответствующего решения.

Найдем решение (5.4), определенное при $t\geqslant 0$, $x\geqslant 0$, и такое, что $u(0, x)=u_0(x)$ и $u(t, 0)=0$, где u_0 — заданная функция. Его построение опирается на *метод отражений*, введенный Кельвином[1]). Доопределим u_0 на левой полуоси равенством $\underline{u_0(-x)=-u_0(x)}$ и решим (5.4) с таким начальным условием на $-\infty, \infty$. По причинам симметрии решение должно удовлетворять условию $u(t, 0)=0$. Ограничиваясь теперь лишь положительными x, мы найдем требуемое решение. Оно равно интегралу по $\overline{0, \infty}$ от функции $u_0(y)q_t(x, y)$, где

$$q_t(x, y)=\frac{1}{\sqrt{2\pi t}}\left[\exp\left(-\frac{(y-x)^2}{2t}\right)-\exp\left(-\frac{(y+x)^2}{2t}\right)\right]. \tag{5.5}$$

Таким образом, *функции q_t совпадают с переходными вероятностями нашего процесса* ($t>0$, $x>0$, $y>0$). Легко видеть, что при фиксированном y q_t будет решением (5.4) с граничным условием $q_t(0, y)=0$ [другой вывод, использующий общий прием, см. в примере гл. XIV, 5а].

Интегрированием по y находим вероятность находиться в момент t в состоянии, отличном от нуля

$$\int_0^\infty q_t(x, y)\,dy = 2\mathfrak{N}\left(\frac{x}{\sqrt{t}}\right)-1, \tag{5.6}$$

где $\mathfrak{N}$ обозначает стандартное нормальное распределение. Иными словами, (5.6) *дает вероятность того, что процесс, начинающийся в точке $x>0$, не достигнет нуля до момента t.* Поэтому (5.6) можно истолковать и как распределение *времен первого*

[1]) См. задачи 7—10 в **1**, гл. XIV, где этот же метод применяется к разностным уравнениям.

прохождения в «свободном» броуновском движении (т. е. при отсутствии барьеров). Отметим, что функцию (5.6) можно охарактеризовать как решение дифференциального уравнения (5.4), определенное при $x>0$ и удовлетворяющее начальному условию $u(0, x)=1$ и граничному условию $u(t, 0)=0$.

[Читатель может узнать в (5.6) устойчивое распределение с показателем $\alpha=1/2$; этот же результат был получен в гл. VI, 2 предельным переходом от случайных блужданий к броуновскому движению.]

в) *Два поглощающих барьера.* Рассмотрим теперь броуновское движение при наличии двух поглощающих барьеров — в точках 0 и $a>0$. Это означает, что при фиксированном $0<y<a$ переходные плотности q_t должны удовлетворять дифференциальному уравнению (5.4) и граничным условиям $q_t(0, y)=q_t(a, y)=0$.

Легко видеть, что решение имеет вид [1])

$$q_t(x, y)=\frac{1}{\sqrt{2\pi t}}\sum_{k=-\infty}^{+\infty}\left\{\exp\left(-\frac{(y-x+2ka)^2}{2t}\right)-\right.$$
$$\left.-\exp\left(-\frac{(y+x+2ka)^2}{2t}\right)\right\}, \qquad (5.7)$$

где $0<x, y<a$. В самом деле, ряд (5.7) очевидным образом сходится и приведение подобных членов показывает, что $q_t(0, y)=q_t(a, y)=0$ при всех $t>0$ и $0<y<a$. [Преобразование Лапласа для (5.7) см. в гл. XIV, (5.17).]

Интегрируя (5.7) по $0<y<a$, найдем, что «непоглощенная» вероятностная масса в момент t равна

$$\lambda_a(t, x)=\sum_{k=-\infty}^{\infty}\left\{\mathfrak{N}\left(\frac{2ka+a-x}{\sqrt{t}}\right)-\mathfrak{N}\left(\frac{2ka-x}{\sqrt{t}}\right)-\right.$$
$$\left.-\mathfrak{N}\left(\frac{2ka+a+x}{\sqrt{t}}\right)+\mathfrak{N}\left(\frac{2ka+x}{\sqrt{t}}\right)\right\}. \qquad (5.8)$$

Эта формула дает вероятность того, что *частица, выходящая из точки x, не будет поглощена до момента t.*

[1]) Формула получается последовательными приближениями с использованием повторных отражений. В (5.5) указано решение, удовлетворяющее граничным условиям в точке 0 (но не в точке a). Отражение в a приводит к четырехчленному решению, удовлетворяющему граничному условию в a (но не в нуле). Чередующиеся отражения в 0 и в a приводят к пределу (5.7). Аналогичное решение для случайных блужданий дано в **1**, гл. XIV, (9.1), откуда (5.7) может быть получено предельным переходом (см. **1**, гл. XIV, 6).

Функция λ_a является решением дифференциального уравнения, стремящимся к 1 при $t \to 0$ и удовлетворяющим граничным условиям $\lambda_a(t, 0) = \lambda_a(t, a) = 0$. Это решение можно получить также стандартным применением метода Фурье. При этом оно имеет вид [1])

$$\lambda_a(t, x) = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \exp\left(-\frac{(2n+1)^2 \pi^2}{2a^2} t\right) \sin \frac{(2n+1)\pi x}{a}. \quad (5.9)$$

Возможность представить решение λ_a двумя рядами благоприятна для вычислений, так как ряд (5.8) быстро сходится при малых t, а ряд (5.9) — при больших t. Равенство выражений (5.8) и (5.9) может служить стандартным примером формулы суммирования Пуассона [см. гл. XIX, (5.8)]. Открытое первоначально в связи с теорией Якоби преобразования тэта-функций, это равенство приобрело некоторую историческую известность [2]).

Можно дать другую интерпретацию λ_a. Рассмотрим положение $\mathbf{X}(t)$ частицы, находящейся в свободном броуновском движении, начинающемся в нуле. Событие: частица за время $\overline{0, t}$ не выходит из $\overline{-\frac{1}{2}a, \frac{1}{2}a}$, равносильно событию: в процессе с поглощающими границами $\pm \frac{1}{2}a$, начинающемся в 0, поглощение не происходит до момента t. Таким образом, *величина* $\lambda_a(t, \frac{1}{2}a)$ *равна вероятности того, что в свободном броуновском движении, начинающемся в нуле,* $|\mathbf{X}(s)| < \frac{1}{2}a$ *при всех* s *из интервала* $0 < s < t$.

г) *Применение к предельным теоремам и критериям Колмогорова — Смирнова.* Пусть $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ — независимые случайные величины с одним и тем же распределением вероятностей,

$$\mathbf{E}(\mathbf{Y}_j) = 0, \quad \mathbf{E}(\mathbf{Y}_j^2) = 1. \quad \text{Положим } \mathbf{S}_n = \mathbf{Y}_1 + \ldots + \mathbf{Y}_n.$$

Пусть $\mathbf{T}_n = \max[|\mathbf{S}_1|, \ldots, |\mathbf{S}_n|]$. Центральная предельная теорема делает правдоподобным заключение, что асимптотическое поведение $\mathbf{T}_n$ такое же, как если $\mathbf{Y}_j$ были нормальны. В последнем же случае нормированные суммы $(1/\sqrt{n})\mathbf{S}_k$ можно интерпретировать как значения $\mathbf{X}(k/n)$, $k = 1, \ldots, n$ процесса

[1]) Аналогичная формула для случайных блужданий получена в **1**, гл. XIV, **5**, где, однако, граничные условия имели вид $\lambda_a(t, 0) = 1$ и $\lambda_a(t, a) = 0$.

[2]) См. теорему 277 в книге Ландау (Landau E., Verteilung der Primzahlen, 1909).

броуновского движения. Вероятность того, что все эти значения лежат в интервале $-\frac{1}{2}a, \frac{1}{2}a$, приблизительно равна, как следует из сказанного ранее, $\lambda_a\left(1, \frac{1}{2}a\right)$ Наши правдоподобные рассуждения приводят к гипотезе, что при $n \to \infty$

$$\mathbf{P}\{\mathbf{T}_n < z\} \to L(z), \tag{5.10}$$

где $L(z) = \lambda_{2z}(1, z)$ находится с помощью (5.8) и (5.9)

$$L(z) = 2\sum_{k=-\infty}^{\infty}\{\mathfrak{N}((4k+1)z) - \mathfrak{N}((4k-1)z)\} =$$

$$= \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}\exp\left(-\frac{(2n+1)^2\pi^2}{8z^2}\right). \tag{5.11}$$

Эта гипотеза была доказана в 1946 г. Эрдешем и Кацем. Лежащая в основе доказательства идея получила с тех пор название *принципа инвариантности*. Он устанавливает, грубо говоря, что предельное распределение некоторых функций от случайных величин нечувствительно к изменениям распределения этих величин и что оно может быть найдено построением надлежащего аппроксимирующего случайного процесса. Принцип инвариантности был усовершенствован М. Донскером, П. Биллингслеем, Ю. Прохоровым и другими, и стал мощным орудием доказательства предельных теорем.

По сходным причинам распределение (5.11) играет важную роль в обширной литературе о непараметрических критериях описанного в гл. I, 12 типа [1]).

д) *Отражающие экраны.* Вернемся к броуновскому движению. По аналогии с обычным случайным блужданием, при отражающем экране следует ввести граничное условие $\frac{\partial q_t(0, y)}{\partial y} = 0$. Легко проверить, что решение для интервала $\overline{0, \infty}$ получается заменой в (5.5) знака минус на плюс. Формальный вывод с применением метода отражений остается таким же, как и раньше, с той разницей, что теперь мы полагаем $u_0(-x) = u_0(x)$. В этом случае q_t будет плотностью собственного распределения вероятностей. Решение для интервала $\overline{0, a}$ при отражающих экранах

[1]) Эта тема сравнительно нова, но происхождение часто используемого тождества (5.11) уже успели, как кажется, забыть. Современный исследователь не может найти доказательства в соответствующей литературе. [См. Renyi A., On the distribution function, $L(z)$. Selected Translation in Math. Statist. and Probability, **4** (1963), 219—224. Новое доказательство Реньи опирается на классические рассуждения с привлечением тэта-функций и тем самым затемняет простой вероятностный смысл (5.11).]

в 0 и в a получается заменой знака минус на знак плюс в (5.7) (другое выражение для решения, получаемое применением метода Фурье или формулы суммирования Пуассона, дано в задаче 11 гл. XIX, 9). ▷

Аналогичные рассуждения, касающиеся граничных условий или времен первого прохождения, применимы и к диффузионным процессам, описываемым общим диффузионным уравнением (4.6). Однако получить явные решения при этом, как правило, очень трудно. Подход к этим задачам методом преобразований Лапласа описан в гл. XIV, 6.

§ 6. Диффузия в многомерном случае

Предыдущая теория легко переносится на двумерный случай. Чтобы не загромождать формулы индексами, мы будем обозначать координатные величины через $(\mathbf{X}(t), \mathbf{Y}(t))$, а *плотности* перехода через $q_t(x, y; \xi, \eta)$; здесь (x, y) — начальная точка и q_t — плотность в точке (ξ, η). Сохраняются постулаты § 4, с той разницей, что инфинитезимальная скорость $b(x)$ заменяется вектором, а дисперсия $a(x)$ — матрицей ковариаций. Вместо (4.6) мы получим тогда *обратное уравнение диффузии*

$$\frac{\partial u}{\partial t} = a_{11}\frac{\partial^2 u}{\partial x^2} + 2a_{12}\frac{\partial^2 u}{\partial x\,\partial y} + a_{22}\frac{\partial^2 u}{\partial y^2} + b_1\frac{\partial u}{\partial x} + b_2\frac{\partial u}{\partial y} \qquad (6.1)$$

с коэффициентами, зависящими от x и y. В случае двумерного броуновского движения мы налагаем условие круговой симметрии, что приводит (с точностью до несущественной константы) к уравнению

$$\frac{\partial u}{\partial t} = \frac{1}{2}\left[\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right]. \qquad (6.2)$$

Соответствующие переходные плотности нормальны с дисперсией t и центром в (x, y). Очевидное разложение этих плотностей на множители показывает, что $\mathbf{X}(t)$ и $\mathbf{Y}(t)$ статистически независимы.

Наиболее интересной из величин, связанных с этим процессом, является расстояние $\mathbf{R}(t)$ от начала координат ($\mathbf{R}^2 = \mathbf{X}^2 + \mathbf{Y}^2$). Интуитивно ясно, что величины $\mathbf{R}(t)$ образуют одномерный диффузионный процесс. Интересно сравнить различные способы вывода соответствующего диффузионного уравнения. Нормальные переходные плотности, отвечающие (6.2), в полярных координатах имеют вид

$$\frac{\rho}{2\pi t}\exp\left(-\frac{\rho^2 + r^2 - 2\rho r\cos(\theta - \alpha)}{2t}\right) \qquad (6.3)$$

(где $x=r\cos\alpha$ и т. д.). При данных r и α в момент 0 частная плотность $\mathbf{R}(t)$ получается интегрированием (6.3) по θ. Параметр α исчезает, и мы получаем [1]) *переходные плотности процесса* $\mathbf{R}(t)$

$$w_t(r,\rho)=\frac{1}{t}\exp\left(-\frac{r^2+\rho^2}{2t}\right)I_0\left(\frac{r\rho}{t}\right). \tag{6.4}$$

Здесь I_0 — функция Бесселя, определенная в гл. II, (7.1). Из самого вывода ясно, что при фиксированном ρ переходные вероятности w_t должны удовлетворять (6.2) в полярных координатах; иными словами, должно быть

$$\frac{\partial w_t}{\partial t}=\frac{1}{2}\left(\frac{\partial^2 w_t}{\partial r^2}+\frac{1}{r}\frac{\partial w_t}{\partial r}\right). \tag{6.5}$$

Это — *обратное уравнение для процесса* $\mathbf{R}(t)$. Оно просто выводится из (6.2) и требования круговой симметрии.

Из (6.5) видно, что бесконечно малая дисперсия для процесса $\mathbf{R}(t)$ такова же, как и для $\mathbf{X}(t)$, а скорость равна $\frac{1}{2}r^{-1}$. То что скорость имеет именно такую величину, можно объяснить (и даже вычислить) следующим образом. Рассмотрим частицу, совершающую двумерное броуновское движение, начиная от точки r на оси x. По причинам симметрии ее *абсцисса* в момент $h>0$ с одинаковой вероятностью будет $>r$ или $<r$. В первом случае, конечно, $\mathbf{R}(h)>r$. Однако это соотношение может выполняться и во втором случае. Таким образом, соотношение $\mathbf{R}(h)>r$ имеет вероятность $>\frac{1}{2}$ и в среднем $\mathbf{R}$ обязано возрастать.

Наш вывод формул для переходных вероятностей применим и в случае трех измерений, причем с одним существенным упрощением: якобиан ρ в (6.3) заменяется теперь на $\rho^2\sin\theta$ и интеграл элементарно вычисляется. Вместо (6.4) мы получаем следующую *переходную плотность для процесса* $\mathbf{R}(t)$ *в трех измерениях*

$$w_t(r,\rho)=\frac{1}{\sqrt{2\pi t}}\frac{\rho}{r}\left[\exp\left(-\frac{(\rho-r)^2}{2t}\right)-\exp\left(-\frac{(\rho+r)^2}{2t}\right)\right]. \tag{6.6}$$

§ 7. Подчиненные процессы

Пусть $\{\mathbf{X}(t)\}$ — данный марковский процесс со стационарными вероятностями перехода $Q_t(x,\Gamma)$. Исходя из $\{\mathbf{X}(t)\}$, можно построить множество новых марковских процессов, вводя то, что

[1]) Интеграл хорошо известен. Стандартная проверка результата состоит в разложении $e^{\cos\theta}$ в степенной ряд по $\cos\theta$.

называют *рандомизированным операционным временем.* Для первого знакомства рассмотрим семейство стохастических ядер

$$P_t(x, \Gamma) = \sum_{n=0}^{\infty} e^{-\alpha t} \frac{(\alpha t)^n}{n!} Q_n(x, \Gamma), \tag{7.1}$$

полученных в предположении, что временно́й параметр в Q_t распределен по закону Пуассона [1]). Обозначим $\mathbf{T}(t)$ процесс Пуассона. Пусть $\mathbf{X}(0) = x$, тогда (7.1) представляет собой распределение случайной величины $\mathbf{X}(\mathbf{T}(t))$. Эти величины определяют новый случайный процесс, в котором пуассоновские величины играют роль операционного времени. В такой же роли могут выступать многие другие процессы $\{\mathbf{T}(t)\}$, но в общем случае процесс $\{\mathbf{X}(\mathbf{T}(t))\}$ не будет марковским. Предположим, чтобы избежать усложнения в обозначениях, что $\mathbf{T}(t)$ принимает только значения 0, h, $2h$, ..., кратные фиксированному числу $h > 0$. Обозначим

$$\mathbf{P}\{\mathbf{T}(t) = nh\} = a_n(t).$$

При условии $\mathbf{X}(0) = x$ величина $\mathbf{X}(\mathbf{T}(t))$ имеет распределение

$$P_t(x, \Gamma) = \sum_{k=0}^{\infty} a_k(t)\, Q_{kh}(x, \Gamma). \tag{7.2}$$

Так как ядра $\{Q_t\}$ удовлетворяют уравнению Чепмена — Колмогорова, то

$$\int_{-\infty}^{+\infty} P_s(x, dy) P_t(y, \Gamma) = \sum_{j,\,k} a_j(s)\, a_k(t) \cdot Q_{(j+k)\,h}(x, \Gamma). \tag{7.3}$$

Отсюда видно, что ядра P_t удовлетворяют уравнению Чепмена — Колмогорова, если

$$a_0(s)\, a_n(t) + a_1(s)\, a_{n-1}(t) + \ \ldots\ + a_n(s)\, a_0(t) = a_n(s+t) \tag{7.4}$$

при всех $s > 0$ и $t > 0$. Последнее соотношение выполняется для *процесса* $\{\mathbf{T}(t)\}$ *со стационарными независимыми приращениями.* Общее решение (7.4) было изучено в 1, гл. XII, 3.

[1]) Отметим, что (7.1) определяет псевдопуассоновский процесс. Обобщение в терминах операторов будет дано в § 9. Там будет показано, что простое применение закона больших чисел, описанное в гл. VII, 5, приводит к так называемой *показательной формуле теории полугрупп.* Из этой формулы следует, что каждое семейство Q_t может быть аппроксимировано переходными вероятностями вида (7.1). Последующее изучение покажет, что распределение Пуассона в (7.1) может быть заменено другими и что понятие подчиненности приводит к пониманию истинного характера показательной формулы.

Пример. а). Если $\{a_n(t)\}$ имеет производящую функцию $\sum a_n(t)\zeta^n = (2-\zeta)^{-t}$, то (7.4) выполняется. При этом $a_n(t) = \binom{-t}{n}(-1)^n 2^{-n-t}$. ▶

Мы видели, что (7.4) обеспечивает равенство величины (7.3) вероятности $P_{s+t}(x, \Gamma)$. Это означает, что распределение $\mathbf{X}(\mathbf{T}(t+s))$ получается интегрированием $P_t(y, \Gamma)$ по распределению $\mathbf{X}(\mathbf{T}(s))$, так что по определению условных вероятностей

$$P_t(y, \Gamma) = \mathbf{P}\{\mathbf{X}(\mathbf{T}(t+s)) \in \Gamma \mid \mathbf{X}(\mathbf{T}(s)) = y\}. \tag{7.5}$$

Аналогичные рассуждения применимы и к переходным вероятностям более высокого порядка; тем самым устанавливается, что *процесс* $\{\mathbf{X}(\mathbf{T}(t))\}$ — *марковский*.

Формальный предельный переход при $h \to 0$ приводит нас к общей ситуации. Предположим, что случайные величины $\mathbf{T}(t)$ образуют процесс со стационарными независимыми приращениями. Распределение U_t величины $\mathbf{T}(t)$ сосредоточено на $[0, \infty)$ и удовлетворяет уравнению свертки

$$U_{t+s}(x) = \int_{0-}^{x} U_s(x-y)\, U_t\{dy\}, \qquad x > 0, \tag{7.6}$$

обобщающему (7.4). Сумма в (7.2) есть ни что иное, как интеграл по отношению к распределению $\mathbf{T}(t)$. Поэтому общая форма (7.2) имеет вид

$$P_t(x, \Gamma) = \int_{0-}^{\infty} Q_s(x, \Gamma)\, U_t\{ds\}. \tag{7.7}$$

Теперь каждое семейство распределений U_t, удовлетворяющих (7.6), может быть получено как предел при $\nu \to \infty$ последовательности арифметических распределений $\{a_n^{(\nu)}(t)\}$, удовлетворяющих (7.4). При весьма слабых условиях непрерывности [1]) на Q_t мы получим переходом к пределу следующий основной результат:

---

[1]) Достаточно потребовать, очевидно, чтобы для любой непрерывной функции f, обращающейся в нуль на бесконечности, интеграл $\int_{-\infty}^{+\infty} Q_t(x, dy) f(y)$ был непрерывной функцией по t и x. Тривиально устанавливается, что P_t — стохастическое ядро, так как оно получается рандомизацией. (*Прямая* проверка требует некоторого мастерства в проведении вычислений и может показаться сложной. Наш переход к пределу дает строгое обоснование этого результата без всяких затруднений и показывает, что наивный подход может быть наиболее усложненным.)

Пусть $\{\mathbf{X}(t)\}$ — марковский процесс с непрерывными переходными вероятностями Q_t и $\{\mathbf{T}(t)\}$ — процесс с неотрицательными независимыми приращениями [т. е. с безгранично делимыми переходными вероятностями U_t, удовлетворяющими (7.6)]. *Тогда* $\{\mathbf{X}(\mathbf{T}(t))\}$ *— также марковский процесс с переходными вероятностями* P_t, *определяемыми по* (7.7). Мы будем говорить, что этот процесс *подчинен* [1]) $\{\mathbf{X}(t)\}$ с использованием операционного времени $\mathbf{T}(t)$. Процесс $\{\mathbf{T}(t)\}$ называется *направляющим процессом.*

Наиболее интересен частный случай, когда процесс $\mathbf{X}(t)$ также имеет независимые приращения. В этом случае переходные вероятности зависят только от разностей $\Gamma - x$ и могут быть заменены соответствующими функциями распределения. Тогда (7.7) принимает более простой вид

$$P_t(x) = \int_0^\infty Q_s(x)\, U_t\{ds\}. \tag{7.8}$$

Все наши примеры будут именно этого типа.

Примеры. б) *Процесс Коши подчинен броуновскому движению.* Пусть $\{\mathbf{X}(t)\}$ — броуновское движение (винеровский процесс) с переходными *плотностями* $q_t(x) = (2\pi t)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2/t}$. В качестве $\{\mathbf{T}(t)\}$ мы возьмем устойчивый процесс с показателем $1/2$ и переходными плотностями

$$u_t(x) = \frac{t}{\sqrt{2\pi}\sqrt{x^3}}\, e^{-\frac{1}{2}t^2/x}.$$

Переход в (7.8) к плотностям и подстановка $s = 1/y$ превращает подинтегральное выражение в показательную функцию. Следовательно, P_t имеет плотность

$$p_t(x) = \frac{1}{\pi}\,\frac{t}{t^2 + x^2}, \tag{7.9}$$

соответствующую процессу Коши [сравните пример гл. VI, (2.e)].

Мы знаем, что U_t есть распределение времени первого прохождения через фиксированную точку $t > 0$ в обычном броуновском движении. На основании этого мы можем *следующим образом интерпретировать процесс Коши в терминах двух независимых броуновских движений* $\mathbf{X}(t)$ *и* $\mathbf{Y}(t)$. Случайная

[1]) Интересное понятие подчиненности введено С. Бохнером в 1949. [Дальнейшее развитие теории предствлено в его книге (1955).] Интерпретация в терминах операционного времени найдена Нелсоном, Троттером и Уоллом (Nelson, Trotter, Woll). [Систематическое изложение на «высшем уровне» дано в статье Nelson E., A functional calculus using singular Laplace integrale, *Trans. Amer. Math. Soc.*, 88 (1958), 400—413.]

величина $\mathbf{Z}(t)$ в процессе Коши равна значению $\mathbf{X}$ в момент $\mathbf{T}(t)$, когда $\mathbf{Y}(s)$ впервые достигает значения t.

в) *Устойчивые процессы.* Последний пример легко распространяется на произвольные устойчивые процессы $\{\mathbf{X}(t)\}$ и $\{\mathbf{T}(t)\}$ с показателями α и β соответственно. Здесь $\alpha \leqslant 2$, но так как величина $\mathbf{T}(t)$ должна быть положительна, то необходимо $\beta < 1$. Переходные вероятности Q_t и U_t имеют вид $Q_t(x) = Q(xt^{-1/\alpha})$ и $U_t(x) = U(xt^{-1/\beta})$, где Q и U — фиксированные устойчивые распределения. Мы покажем, что *подчиненный процесс* $\mathbf{X}(\mathbf{T}(t))$ *устойчив с показателем* $\alpha\beta$. Это утверждение эквивалентно соотношению $P_{\lambda t}(x) = P_t(x^{-1/\alpha\beta})$. Это соотношение тривиально выводится из определения Q_s и U_t и формулы (7.8), с применением подстановки $s = y\lambda^{1/\beta}$.

[Наш результат по существу эквивалентен формуле для произведений, установленной в гл. VI, (2.3). При $\mathbf{X}(t) > 0$ эта формула может быть переписана в терминах преобразований Лапласа, см. гл. XIII, (7.e). Вариант, использующий преобразование Фурье, см. в задаче 8 гл. XVII, 12.]

г) *Обобщенный пуассоновский процесс, направляемый гамма-процессом.* Пусть Q_t — обобщенное пуассоновское распределение, порождаемое распределением вероятностей F, и пусть U_t имеет гамма-плотность $e^{-x}x^{t-1}/\Gamma(t)$. Тогда (7.8) принимает вид

$$P_t = \sum_{n=0}^{\infty} a_n(t) F^{n*}, \tag{7.10}$$

где

$$a_n(t) = \int_0^{\infty} e^{-s}\frac{s^n}{n!} \cdot e^{-s}\frac{s^{t-1}}{\Gamma(t)}\, ds = \frac{\Gamma(n+t)}{n!\,\Gamma(t)} \cdot 2^{-n-t}. \tag{7.11}$$

Простые выкладки показывают, что $a_n(t)$ суть вероятности, встречавшиеся уже в примере (а). В частности, если F сосредоточено в точке 1, то $\{\mathbf{X}(t)\}$ будет процессом Пуассона с математическим ожиданием t. Мы видим, что *процесс примера* (а) *подчинен пуассоновскому процессу* при использовании гамма-процесса в качестве операционного времени.

д) *Гамма-процесс, направляемый законом Пуассона.* Рассмотрим теперь те же самые распределения, поменяв их ролями. Тогда операционное время будет целочисленным и нуль имеет вес e^{-t}. Отсюда следует, что результирующее распределение также имеет атом веса e^{-t} в нуле. Непрерывная часть имеет плотность

$$\sum_{n=1}^{\infty} e^{-x}\frac{x^{n-1}}{(n-1)!}\, e^{-t}\frac{t^n}{n!} = e^{-t-x}\sqrt{\frac{t}{x}}\, I_1(2\sqrt{xt}), \tag{7.12}$$

где I_1 — функция Бесселя (см. гл. II, (7.1)). Из сказанного следует, что это распределение безгранично делимо. Прямая проверка этого утверждения нелегка. ▶

§ 8. Марковские процессы и полугруппы

В гл. VIII были отмечены преимущества трактовки вероятностных распределений как операторов, действующих на непрерывные функции. Преимущества операторного подхода к стохастическим ядрам еще больше, и теория полугрупп приводит к цельной теории марковских процессов, что недостижимо при других методах. Пусть даны стохастическое ядро K в $\mathscr{R}^1$ и ограниченная непрерывная функция u. Тогда соотношение

$$U(x) = \int_{-\infty}^{+\infty} K(x, dy)\, u(y) \tag{8.1}$$

определяет нам новую функцию. Мы почти не теряем общности рассуждений, предположив, что функция U также непрерывна. Тогда мы могли бы перейти к изучению свойства ядра K в терминах порожденного им преобразования $u \to U$ семейства непрерывных функций. По двум приводимым ниже причинам целесообразно рассмотреть более общую ситуацию. Во-первых, преобразования вида (8.1) имеют смысл в произвольных пространствах, и было бы крайне не экономично развивать теорию, которая не охватывает простейший и очень важный частный случай, а именно процессы со счетным множеством состояний [когда (8.1) сводится к матричному преобразованию]. Во-вторых, даже в теории, ограничивающейся непрерывными функциями на прямой, различные типы граничных условий вынуждают нас вводить специальные классы непрерывных функций. С другой стороны, большая общность достигается без всяких затрат. Читатели в зависимости от настроения могут пренебречь общностью и рассматривать все теоремы в одной (или нескольких) из следующих типичных ситуаций: (i) пространство состояний Σ есть вещественная прямая и $\mathscr{L}$ есть класс ограниченных непрерывных функций, исчезающих на бесконечности; (ii) пространство Σ есть конечный замкнутый интервал I в $\mathscr{R}^1$ или $\mathscr{R}^2$ и $\mathscr{L}$ — класс непрерывных функций на нем; (iii) Σ состоит из целых чисел и $\mathscr{L}$ — из ограниченных последовательностей. В последнем случае последовательности лучше всего представлять себе в виде векторов-столбцов, а преобразования — в виде матриц.

Как и в гл. VIII *норма* ограниченной вещественной функции u определяется как $\|u\| = \sup |u(x)|$. Последовательность

функций u_n сходится равномерно к u тогда и только тогда, когда $\|u_n - u\| \to 0$.

Начиная с этого места $\mathscr{L}$ *будет обозначать семейство вещественных функций, определенных на некотором множестве* Σ *и обладающих следующими свойствами*: (i) если u_1 и u_2 принадлежат $\mathscr{L}$, то каждая линейная комбинация $c_1u_1+c_2u_2 \in \mathscr{L}$; (ii) если $u_n \in \mathscr{L}$ и $\|u_n - u\| \to 0$, то $u \in \mathscr{L}$; (iii) если $u \in \mathscr{L}$, то и u^+ и u^- принадлежат $\mathscr{L}$ (здесь $u = u^+ - u^-$ — обычное разложение u на положительную и отрицательную части) Другими словами, $\mathscr{L}$ замкнуто относительно образования линейных комбинаций, равномерного перехода к пределу и взятия абсолютной величины. Первые два свойства означают, что $\mathscr{L}$ — *банахово* пространство, последнее — что $\mathscr{L}$ есть структура.

Ниже приводятся стандартные определения. Линейное преобразование T называют *эндоморфизмом* на $\mathscr{L}$, если каждое $u \in \mathscr{L}$ имеет образ $Tu \in \mathscr{L}$, такой, что $\|Tu\| \leqslant m\|u\|$, где m — постоянная, не зависящая от u. Наименьшая постоянная с этим свойством называется *нормой* $\|T\|$ преобразования T. Преобразование T *положительно*, если из $u \geqslant 0$ вытекает $Tu \geqslant 0$. В этом случае $-Tu^- \leqslant Tu \leqslant Tu^+$. *Сжатием* называется положительный оператор T, такой, что $\|T\| \leqslant 1$. Если постоянная функция 1 принадлежит $\mathscr{L}$ и T — положительный оператор, такой, что $T1 = 1$, то T называется *оператором перехода* (он автоматически будет сжатием).

Пусть даны два эндоморфизма S и T на $\mathscr{L}$. Тогда их *произведение* ST есть эндоморфизм, отображающий u в $S(Tu)$. Очевидно, $\|ST\| \leqslant \|S\| \cdot \|T\|$. В общем случае $ST \neq TS$ в отличие от частного случая операторов свертки из гл. VIII, 3, которые перестановочны один с другим.

Мы всерьез интересуемся лишь преобразованиями вида (8.1), где K — стохастическое или, в крайнем случае, субстохастическое ядро. Операторы такого вида суть сжатия или операторы перехода. Им присуще также

Свойство монотонной сходимости. *Если* $u_n \geqslant 0$ *и* $u_n \uparrow u$ (*с* u_n *и* $u \in \mathscr{L}$), *то* $Tu_n \to Tu$ *в смысле поточечной сходимости.*

Практически во всех ситуациях сжатия, обладающие этим свойством, имеют вид (8.1). Два примера иллюстрируют это замечание.

Примеры. а) Пусть Σ — вещественная прямая и $\mathscr{L} = C$, где C — семейство всех ограниченных непрерывных функций на ней. Пусть $C_0 \subset \mathscr{L}$ — подкласс функций, обращающихся в нуль на $\pm\infty$. Если T — сжатие на $\mathscr{L}$, то при $u \in C_0$ значение $Tu(x)$ функции Tu в фиксированной точке x является положительным ли-

нейным функционалом на $\mathscr{L}$. По теореме Рисса о представлении таких функционалов существует (возможно несобственное) распределение вероятностей F, такое, что $Tu(x)$ равно математическому ожиданию u относительно F. Так как F зависит от x, то $F(\Gamma)$ мы обозначим $K(x, \Gamma)$. Тогда для $u \in C_0$

$$Tu(x) = \int_{-\infty}^{+\infty} K(x, dy)\, u(y). \tag{8.2}$$

Если оператор T обладает свойством монотонной сходимости, то это соотношение можно распространить на все ограниченные непрерывные функции.

При фиксированном x ядро K как функция от Γ является мерой. Если Γ — открытый интервал и $\{u_n\}$ — возрастающая последовательность непрерывных функций, такая, что $u_n(x) \to 1$ при $x \in \Gamma$ и $u_n(x) \to 0$ в других случаях, то $K(x, \Gamma) = \lim Tu_n(x)$ (в силу основных свойств интегралов). Так как функции Tu_n непрерывны, мы заключаем, что ядро K есть бэровская функция от x. Поэтому K обладает всеми свойствами стохастических или субстохастических ядер. То же самое распространяется на случай интервалов или $\mathscr{R}^n$.

б) Пусть Σ — множество целых чисел, а $\mathscr{L}$ — множество числовых последовательностей $\{u_n\}$ с $\|u\| = \sup |u_n|$. В частности, пусть $e^{(n)} = (0, 0, \ldots, 0, 1, 0, \ldots)$ — последовательность с единственным ненулевым членом, стоящим на n-м месте. Пусть T — сжатие. Обозначим i-й член последовательности $Te^{(k)}$ через p_{ik}. Если последовательность u имеет лишь конечное число членов, отличных от нуля, то ее можно представить как конечную линейную комбинацию $e^{(k)}$. Следовательно, i-й член Tu равен

$$(Tu)_i = \sum p_{ik} u_k, \tag{8.3}$$

где сумма на самом деле содержит лишь конечное число членов. В силу свойства монотонной сходимости представление (8.3) применимо, очевидно, ко всем u. Из того что T — сжатие, вытекает неравенство $p_{ik} \geqslant 0$. Кроме того, суммы по строкам $\leqslant 1$. Таким образом, матрица (p_{ik}) субстохастическая. Она будет строго стохастической, если T — оператор перехода. ▶

Эти примеры типичны [1]), и найти сжатие, не порождаемое стохастическим ядром, в самом деле трудно. Как бы то ни было, мы оправдали дальнейшее развитие общей теории сжимающих операторов и убедились в том, что применения к вероятностным задачам будут носить очевидный характер. (В действительности мы никогда не выйдем за рамки указанных примеров.)

[1]) В них используется только локальная компактность Σ.

Переходные вероятности марковского процесса образуют однопараметрическое семейство ядер, удовлетворяющих уравнению Чепмена — Колмогорова

$$Q_{s+t}(x, \Gamma) = \int Q_s(x, dy)\, Q_t(y, \Gamma) \tag{8.4}$$

($s>0$, $t>0$); интегрирование производится по всему пространству состояний. Каждое из этих ядер порождает оператор перехода $\mathfrak{Q}(t)$, определяемый соотношением

$$\mathfrak{Q}(t)\, u(x) = \int Q_t(x, dy)\, u(y). \tag{8.5}$$

Очевидно, что (8.4) эквивалентно

$$\mathfrak{Q}(s+t) = \mathfrak{Q}(s)\, \mathfrak{Q}(t), \quad s>0, \quad t>0. \tag{8.6}$$

Семейство эндоморфизмов с этим свойством называют *полугруппой*. Ясно, что $\mathfrak{Q}(s)\, \mathfrak{Q}(t) = \mathfrak{Q}(t)\, \mathfrak{Q}(s)$, т. е. *каждые два элемента полугруппы перестановочны. Последовательность* $\{T_n\}$ *эндоморфизмов на* $\mathscr{L}$ *называется сходящейся*[1]) *к эндоморфизму* T *тогда и только тогда, когда* $\|T_n u - Tu\| \to 0$ *при каждом* $u \in \mathscr{L}$. *Если последнее соотношение выполнено, мы пишем* $T_n \to T$.

Начиная с этого момента, мы сосредоточимся на полугруппах сжимающих операторов. При этом мы введем некоторые условия регулярности. Обозначим опять через $\mathbf{1}$ тождественный оператор $\mathbf{1}u = u$.

Определение. *Мы будем называть полугруппу сжимающих операторов* $\mathfrak{Q}(t)$ *непрерывной*[2]), *если* $\mathfrak{Q}(0) = \mathbf{1}$ *и* $\mathfrak{Q}(h) \to \mathbf{1}$ *при* $h \to 0+$.

При $0 \leqslant t' < t''$ мы имеем

$$\|\mathfrak{Q}(t'')\, u - \mathfrak{Q}(t')\, u\| \leqslant \|\mathfrak{Q}(t'' - t')\, u - u\|. \tag{8.7}$$

Для непрерывной полугруппы существует такое $\delta>0$, что правая часть $<\varepsilon$ при $t''-t'<\delta$. Таким образом, верно не только что $\mathfrak{Q}(t) \to \mathfrak{Q}(t_0)$ при $t \to t_0$, но и большее: формула (8.7) показы-

[1]) Такая сходимость называется *сильной*. Она введена в гл. VIII, 3. Из нее *не* вытекает сходимости $\|T_n - T\| \to 0$ (называемой равномерной). Более слабый тип сходимости определяется требованием $T_n u(x) \to Tu(x)$ при каждом x (но не обязательно равномерно); см. задачу 6 в гл. VIII, 10

[2]) Мы используем этот термин вместо стандартного «сильно непрерывна в нуле».

вает, что $\mathfrak{Q}(t)u$ при любом фиксированном u есть равномерно непрерывная функция t [1]).

Преобразование (8.i) — это, конечно, то же самое, что и (4.5); оно служило исходным пунктом при выводе обратного уравнения для диффузионных процессов. Теперь семейство марковских переходных вероятностей порождает также некоторую полугруппу преобразований *мер*. При этом мера μ переходит в меру $T(t)\mu$, приписывающую множеству Γ массу

$$T(t)\mu(\Gamma) = \int \mu\{dx\} Q_t(x, \Gamma). \tag{8.8}$$

Если Q_t имеет плотность q_t, то это преобразование совпадает с (5.1) и было использовано для вывода прямого уравнения. Так как теория вероятностей имеет в первую очередь дело с мерами, а не с функциями, то возникает вопрос — почему мы начали с полугруппы $\{\mathfrak{Q}(t)\}$, а не с $\{T(t)\}$. Ответ интересен и проливает новый свет на сложную связь прямого и обратного уравнений.

Причина состоит в том (и это подтверждают приведенные выше примеры), что в обычной обстановке непрерывные полугруппы сжимающих операторов на функциональном пространстве $\mathscr{L}$ порождаются переходными вероятностями: изучение наших полугрупп $\{\mathfrak{Q}(t)\}$ — это практически то же самое, что изучение марковских переходных вероятностей. Для полугрупп на пространстве мер это *неверно*. Существуют аналитически очень естественные полугруппы сжимающих операторов, не порождаемые марковскими процессами. Чтобы построить пример, рассмотрим любую марковскую полугруппу вида (8.8) на прямой, предполагая лишь, что абсолютно непрерывная мера μ переходит в абсолютно непрерывную меру $T(t)\mu$ [например, $T(t)$ может быть сверткой с нормальным распределением с дисперсией t]. Пусть $\mu = \mu_c + \mu_s$ — разложение μ на абсолютно непрерывную и сингулярную компоненты. Определим новую полугруппу $\{S(t)\}$, положив

$$S(t)\mu = T(t)\mu_c + \mu_s.$$

Эта полугруппа непрерывна и $S(0) = \mathbf{1}$, но легко показать, что она не связана ни с какой системой переходных вероятностей и что она с вероятностной точки зрения бессмысленна.

§ 9. «Показательная формула» в теории полугрупп

Псевдопуассоновские процессы § 1 являются, конечно, простейшими марковскими процессами. Теперь мы докажем, что практически все марковские процессы представимы как пределы псевдопуассоновских процессов [2]). Абстрактный вариант этого утверждения играет фундаментальную роль в теории

[1]) Существуют полугруппы, такие, что $\mathfrak{Q}(h)$ стремится к оператору $T \neq \mathbf{1}$, но они патологичны. Определим, например, эндоморфизм T равенством $Tu(x) = \frac{1}{2}u(0)[1 + \cos x] + \frac{1}{2}u(\pi)[1 - \cos x]$ и положим $\mathfrak{Q}(t) = T$ при всех $t \geqslant 0$.

[2]) Специальный случай, когда $\mathfrak{Q}(t)$ суть операторы свертки, был разобран в гл. IX.

полугрупп. Мы увидим сейчас, что он на самом деле есть следствие закона больших чисел.

Пусть T — оператор, порождаемый стохастическим ядром K. Тогда оператор $\mathfrak{Q}(t)$, индуцированный псевдопуассоновским распределением (1.2), принимает вид

$$\mathfrak{Q}(t) = e^{-\alpha t} \sum_{n=0}^{\infty} \frac{(\alpha t)^n}{n!} T^n, \tag{9.1}$$

где сумма ряда определяется как предел частных сумм. Из уравнения Чепмена — Колмогорова (1.3) следует, что эти операторы образуют полугруппу. Предпочтительнее, однако, доказать этот факт заново для произвольных сжимающих операторов T.

Теорема 1. *Если T — сжатие на $\mathscr{L}$, то операторы* (9.1) *образуют непрерывную полугруппу сжатий. Если T — оператор перехода, то это же верно и для $\mathfrak{Q}(t)$.*

Доказательство. Очевидно, что $\mathfrak{Q}(t)$ — положительный оператор и $\|\mathfrak{Q}(t)\| \leqslant e^{-\alpha t + \alpha t \|T\|} \leqslant 1$. Полугрупповое свойство легко проверяется формальным перемножением рядов для $\mathfrak{Q}(s)$ и $\mathfrak{Q}(t)$. Соотношение $\mathfrak{Q}(h) \to \mathbf{1}$ очевидно из (9.1). ▶

Мы будем записывать (9.1) сокращенно

$$\mathfrak{Q}(t) = e^{\alpha t (T - \mathbf{1})}. \tag{9.2}$$

Полугруппы сжатий такого вида будут называться псевдопуассоновскими[1]). *Мы скажем, что $\{\mathfrak{Q}(t)\}$ порождается оператором $\alpha(T - \mathbf{1})$.*

Рассмотрим теперь произвольную полугруппу сжатий $\mathfrak{Q}(t)$. Она ведет себя во многих отношениях так же, как вещественная непрерывная функция. В частности, применима без каких-либо серьезных изменений теория приближения функций, развитая в гл. VII на основе закона больших чисел. Возьмем формулу гл. VII, (5.1), использующую распределение Пуассона. Введем при фиксированном $h > 0$ операторы

$$\mathfrak{Q}_h(t) = e^{-th^{-1}} \sum_{n=0}^{\infty} \frac{(th^{-1})^n}{n!} \mathfrak{Q}(nh), \tag{9.3}$$

---

[1]) Существуют полугруппы сжатий вида e^{tS}, где S — *эндоморфизм, не* представимый в форме $\alpha(T - \mathbf{1})$. Таковы, например, полугруппы, связанные со скачкообразными процессами § 3, если $\alpha(x)$ неограничена.

которые получаются рандомизацией параметра t в $\mathfrak{Q}(t)$. Сравнение с (9.1) показывает, что операторы $\mathfrak{Q}_h(t)$ образуют псевдопуассоновскую полугруппу, порождаемую оператором $[\mathfrak{Q}(h)-1]/h$. Докажем теперь, что

$$\mathfrak{Q}_h(t)\to\mathfrak{Q}(t), \qquad h\to 0. \tag{9.4}$$

Отправным пунктом будет тождество

$$\mathfrak{Q}_h(t)\,u-\mathfrak{Q}(t)\,u=e^{-th^{-1}}\sum_{n=0}^{\infty}\frac{(th^{-1})^n}{n!}[\mathfrak{Q}(nh)\,u-\mathfrak{Q}(t)\,u]. \tag{9.5}$$

Математическое ожидание и дисперсия распределения Пуассона, встречающегося в этой формуле, равны th^{-1}. По неравенству Чебышева норма суммы всех членов с $|n-th^{-1}|>\eta th^{-1}$ не превосходит $\eta^{-2}t^{-1}h\cdot\Sigma\|u\|$. В силу свойства равномерной непрерывности (8 7) можно выбрать столь малое η, что

$$\|\mathfrak{Q}(nh)\,u-\mathfrak{Q}(t)\,u\|\leqslant\varepsilon \quad\text{при}\quad |nh-t|<\eta t. \tag{9.6}$$

Тогда норма оператора (9.5) не превосходит $\varepsilon+2\eta^{-2}t^{-1}\cdot\|u\|\cdot h$. Так как ε произвольно, то при $h\to 0$ эта норма стремится к нулю. Мы доказали, таким образом, знаменитую теорему, известную в теории полугрупп под названием *показательной формулы*.

Теорема 2. *Каждая непрерывная полугруппа сжатий* $\mathfrak{Q}(t)$ *является пределом* (9.4) *псевдопуассоновских полугрупп* $\{\mathfrak{Q}_h(t)\}$, *порождаемых эндоморфизмами* $h^{-1}[\mathfrak{Q}(h)-1]$.

Другие варианты можно получить, ставя на место распределения Пуассона другие безгранично делимые распределения. Из 1, гл. XII, 3 мы знаем, что производящая функция безгранично делимого распределения $\{u_n(t)\}$, сосредоточенного на множестве целых чисел $n\geqslant 0$, имеет вид

$$\sum_0^{\infty}u_n(t)\,\zeta^n=\exp\left(t\alpha[p(\zeta)-1]\right), \tag{9 7}$$

где

$$p(\zeta)=p_0+p_1\zeta+\ldots, \qquad p_j\geqslant 0, \quad \Sigma p_j=1. \tag{9.8}$$

Допустим, что распределение $\{u_n(t)\}$ имеет математическое ожидание bt и конечную дисперсию Заменяя в (9 3) распределение Пуассона на $\{u_n(t/b)\}$, получаем оператор

$$\mathfrak{Q}_h(t)=\Sigma u_n(t/bh)\,\mathfrak{Q}(nh)=\Sigma u_n(t/bh)\,\mathfrak{Q}^n(h). \tag{9.9}$$

Нетрудно понять, что при фиксированном $h>0$ $\{\mathfrak{Q}_h(t)\}$ есть псевдопуассоновская полугруппа, порождаемая оператором $ab^{-1}h^{-1}[p(\mathfrak{Q}(h))-1]$. Как и прежде, простое применение закона больших чисел показывает, что $\mathfrak{Q}_h(t)\to\mathfrak{Q}(t)$ при $h\to 0$. Таким путем мы получаем аналог «показательной

формулы», в котором распределение $\{u_n(t)\}$ играет роль распределения Пуассона [1]).

Чтобы понять вероятностное содержание и пути обобщения этих рассуждений, обозначим через $\{\mathbf{X}(t)\}$ марковский процесс, соответствующий полугруппе $\{\mathfrak{Q}(t)\}$, и через $\{\mathbf{T}(t)\}$ — процесс с независимыми приращениями, описываемый вероятностями перехода $\{u_n(t)\}$. Операторы (9.9) соответствуют переходным вероятностям процесса $\mathbf{X}(\mathbf{T}(t))$. Другими словами, мы ввели специальный процесс, подчиненный $\mathbf{X}(t)$. Закон больших чисел, примененный к $\mathbf{T}$-процессу, делает правдоподобным предположение о том, что при $h\to 0$ распределения нового процесса сходятся к распределениям исходного марковского процесса. Эта процедура аппроксимации никоим образом не связана с целочисленностью случайных величин $\mathbf{T}(t)$. В самом деле, мы можем взять в качестве $\mathbf{T}(t)$ любой процесс с положительными независимыми приращениями, такой, что $\mathbf{E}(\mathbf{T}(t))=bt$ и что дисперсия $\mathbf{T}(t)$ конечна.

Подчиненный процесс вида $\mathbf{X}(\mathbf{T}(t/h))$ *— это процесс псевдопуассоновского типа. Его полугруппа порождается некоторым эндоморфизмом* [2]) *и* $\mathfrak{Q}_h(t)\to\mathfrak{Q}(t)$ *при* $h\to 0$.

§ 10. Производящие операторы. Обратное уравнение

Рассмотрим псевдопуассоновскую полугруппу $\{\mathfrak{Q}(t)\}$ сжимающих операторов, порождаемую оператором $\mathfrak{A}=\alpha(T-\mathbf{1})$. Так как это эндоморфизм, то $v=\mathfrak{A}u$ определено для всех $u\in\mathscr{L}$ и

$$\frac{\mathfrak{Q}(h)-\mathbf{1}}{h}u\to v,\qquad h\to 0+. \tag{10.1}$$

Было бы хорошо, если бы это свойство имело место для всех полугрупп, однако этого невозможно ожидать. Возьмем, например, полугруппу, связанную с броуновским движением. Уравнение диффузии (4.1) показывает, что для дважды непрерывно дифференцируемой u левая часть (10.1) сходится к $1/2u''$. Однако не существует никакого предела, если u не дифференцируема. Тем не менее уравнение диффузии однозначно описывает процесс, так как полугруппа однозначно определяется своим действием на дважды дифференцируемые функции. Поэтому мы не можем ожидать, что (10.1) будет выполняться для *всех* функций u, но для всех практических целей хватает требования, чтобы оно выполнялось для «достаточно многих» функций. Имея это в виду, мы примем

Определение. *Если для некоторых* u *и* v *из* $\mathscr{L}$ *выполняется соотношение* (10.1) (*в смысле равномерной сходимости*),

---

[1]) Это отметил Чжун К. Л. (см. гл. VII, 5).

[2]) О производящих операторах произвольных подчиненных процессов см. пример гл. XIII, (9, б).

то мы полагаем $v = \mathfrak{A}u$. *Так определенный оператор называется производящим оператором*[1]) *полугруппы* $\{\mathfrak{Q}(t)\}$.

Умножая (10.1) слева на $\mathfrak{Q}(t)$, получаем

$$\frac{\mathfrak{Q}(t+h)-\mathfrak{Q}(t)}{h}u \to \mathfrak{Q}(t)v. \tag{10.2}$$

Таким образом, если $\mathfrak{A}u$ существует, то все функции $\mathfrak{Q}(t)u$ лежат в области определения $\mathfrak{A}$ и

$$\frac{\mathfrak{Q}(t+h)-\mathfrak{Q}(t)}{h}u \to \mathfrak{Q}(t)\mathfrak{A}u = \mathfrak{A}\mathfrak{Q}(t)u. \tag{10.3}$$

Это соотношение по существу то же самое, что обратное уравнение для марковских процессов. В самом деле, в обозначениях § 4 мы можем положить

$$u(t, x) = \mathfrak{Q}(t)u_0(x),$$

где u_0 — начальная функция. Тогда (10.3) превращается в

$$\frac{\partial u(t, x)}{\partial t} = \mathfrak{A}u(t, x). \tag{10.4}$$

Это — знакомое нам обратное уравнение, но оно нуждается в правильной интерпретации. Вероятности перехода диффузионных процессов § 4 так гладки, что обратное уравнение верно для *всех* непрерывных начальных функций u_0. В общем случае этого может не быть.

Примеры. *а) Сдвиги.* Пусть $\mathscr{L}$ состоит из всех непрерывных функций на прямой, обращающихся в нуль на бесконечности. Положим $\mathfrak{Q}(t)u(x) = u(x+t)$. Очевидно, что (10.1) выполняется тогда и только тогда, когда u имеет непрерывную производную u', обращающуюся в нуль на бесконечности. В этом случае $\mathfrak{A}u = u'$.

Формально обратное уравнение (10.4) сводится к

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x}. \tag{10.5}$$

Формальное решение, совпадающее при $t=0$ с данной начальной функцией u_0, было бы $u(t, x) = u_0(t+x)$. Но u в самом деле будет решением только в том случае, когда функция u_0 дифференцируема.

---

[1]) При анализе полугрупп со сверткой в предыдущей главе мы ограничивались бесконечно дифференцируемыми функциями, благодаря чему все производящие операторы имели одну и ту же область определения. Столь удобного средства нет в общей теории полугрупп.

б) Как и в § 2, рассмотрим псевдопуассоновский процесс $\mathbf{X}(t)$ и другой процесс $\mathbf{X}^{\#}(t) = \mathbf{X}(t) - ct$. Соответствующие полугруппы связаны очевидным образом: значение $\mathfrak{T}^{\#}(t)u$ в точке x равно значению $\mathfrak{T}(t)u$ в точке $x+ct$. Для производящих операторов имеем связь

$$\mathfrak{A}^{\#} = \mathfrak{A} - c\frac{d}{dx}, \tag{10.6}$$

так что область определения $\mathfrak{A}^{\#}$ ограничена дифференцируемыми функциями. Обратное уравнение выполняется только, если начальная функция u_0 имеет непрерывную производную, но не для произвольных функций. В частности, сами переходные вероятности не обязаны удовлетворять обратному уравнению. Этим объясняются как трудности устаревших теорий [обсуждавшихся в связи с гл. IX, (2.14)], так и необходимость принятия неестественных условий регулярности при выводе прямого уравнения (2.1). ▶

Польза понятия производящего оператора проистекает из следующего факта: *для любой непрерывной*[1]) *полугруппы сжатий область определения* $\mathfrak{A}$ *плотна в* $\mathscr{L}$, *так что производящий оператор однозначно описывает полугруппу.* Простое доказательство этой теоремы будет дано в гл. XIII, 9.

Эта теорема дает нам возможность обращаться с обратными уравнениями, не прибегая к излишним ограничениям, и значительно упрощает вывод этих уравнений. Так, наиболее общая форма диффузионных операторов, упомянутая в сноске 3 на стр. 395 § 4, не могла бы быть получена без априорной уверенности в том, что производящий оператор на самом деле существует.

[1]) В смысле определения § 8.

ГЛАВА XI

ТЕОРИЯ ВОССТАНОВЛЕНИЯ

Процессы восстановления были введены в гл. VI, 6 и были иллюстрированы примерами в гл. VI, 7. Теперь мы начнем с общей теории так называемого уравнения восстановления, с которым часто приходится сталкиваться в различных вопросах. Ярким примером применения общей теоремы восстановления служит предельная теорема § 8. Параграфы 6 и 7 содержат улучшенные и обобщенные варианты некоторых асимптотических оценок, первоначальный вывод которых был трудоемким и требовал применения глубоких аналитических методов. Здесь еще раз иллюстрируется та экономия в работе мысли и количестве аналитических средств, которая достигается за счет общетеоретического подхода к трудным индивидуальным проблемам. Анализ задач теории восстановления с помощью преобразований Лапласа см. в гл. XIV, 1—3.

Немало работ было написано и затрачено много сил с целью освободить теорему восстановления от условия положительности случайных величин. Ввиду солидной истории вопроса мы включим в § 9 новое и сильно упрощенное доказательство общей теоремы.

Общий обзор литературы, применений и обобщений для случая положительных случайных величин имеется в статье Смита: Smith W. L., *J. Roy. Statist. Soc.* (3), **20** (1958), 243—302[1]). О новых результатах см. статью Чжоу и Роббинса (Chow Y. S., Robbins H. E, *Ann. Math. Statist*, **34** (1963), 390—401).

§ 1. Теорема восстановления

Пусть F — распределение, сосредоточенное[2]) на $\overline{0, \infty}$. Иными словами, мы предполагаем, что $F(0)=0$. Мы не требуем существования математического ожидания. Однако в силу условия

---

[1]) Имеется русский перевод: Смит В., Теория восстановления, сб. переводов, «Математика», 5 : 3 (1961), 95—150. — *Прим. перев.*

[2]) Как показано в задаче 1, при $F(0)>0$ (в предположении $F(x)=0$ при $x<0$) существенных изменений не требуется.

положительности мы можем принять [1])

$$\mu = \int_0^\infty yF\{dy\} = \int_0^\infty [1 - F(y)]\,dy, \tag{1.1}$$

где $\mu \leqslant \infty$. При $\mu = \infty$ мы условимся интерпретировать символ μ^{-1} как 0.

В гл. VI, 6 мы определили чистый процесс восстановления, порожденный F. Доминирующую роль при этом играла функция

$$U = \sum_{n=0}^{\infty} F^{n*}, \tag{1.2}$$

которая, как было показано, конечна. Значение $U(x)$ равно математическому ожиданию моментов восстановления на интервале $\overline{|0, x|}$ (нуль считается моментом восстановления). Сейчас мы отвлечемся от вероятностного смысла U и будем изучать чисто аналитическими средствами асимптотическое поведение функции U [как она определена в (1.2)] на бесконечности.

Лучше всего представлять себе U как меру, сосредоточенную на $\overline{|0, \infty|}$. Интервалу $I = \overline{|a, b|}$ соответствует мера $U\{I\} = U(b) - U(a)$. Как обычно, мы обозначаем через $I+t$ интервал $\overline{|a+t, b+t|}$, получаемый из I сдвигом на t. Таким образом, $U\{I+t\} = U(b+t) - U(a+t)$. Отметим, что нуль является атомом этой меры. Соответствующий вес равен 1 и вносится членом F^{0*} ряда (1.2).

Допустим сначала, что все точки роста F принадлежат последовательности $\lambda, 2\lambda, 3\lambda, \ldots$. Согласно определению 3 из гл. V, 2, подобное распределение называется *арифметическим*, а наибольшее λ, обладающее указанным свойством, называется *шагом* F. Теория восстановления, развитая в **1**, гл. XIII, 4, показывает, что в этом случае мера U чисто *аналитическая* и для веса u_n точки $n\lambda$ выполняется соотношение $u_n \to \mu^{-1}\lambda$ при $n \to \infty$. Приводимая ниже теорема обобщает этот результат на случай произвольных распределений [2]), сосредоточенных на $\overline{0, \infty}$. Случай арифметического распределения F ради полноты формулировки упомянут снова.

Теорема 1. (*Теорема восстановления.*) *Если распределение F не является арифметическим, то при любом фиксирован-*

[1]) По поводу равенства двух интегралов в (1.1) см. гл. V, (6.3).

[2]) Случай арифметических распределений был впервые рассмотрен П. Эрдёшем, В. Феллером, Г. Поллардом (1949). Их доказательство было распространено Д. Блекуэллом на произвольные положительные величины. В тексте дано новое доказательство.

ном $h>0$ *и* $t\to\infty$

$$U(t)-U(t-h)\to\frac{h}{\mu}. \tag{1.3}$$

Если же F *— арифметическое распределение, то (1.3) верно для* h*, кратных шагу* λ.

Прежде чем доказывать теорему [1]), рассмотрим некоторые из ее следствий и вариантов. Пусть z — ограниченная функция, равная нулю на $\overline{-\infty, 0}$. Если $U(x)<\infty$ при всех x, то *свертка* $Z=U*z$ равна по определению интегралу [2])

$$Z(x)=\int_0^x z(x-y)\,U\{dy\}, \tag{1.4}$$

который имеет смысл, даже если мера U не ограничена. Для $x<0$ имеем $Z(x)=0$. Как было показано в гл. VI, 6—7, важность функции U объясняется главным образом тем обстоятельством, что (1.4) представляет собой единственное решение *уравнения восстановления*

$$Z=z+F*Z, \tag{1.5}$$

равное нулю при $x<0$. Это уравнение в подробной записи имеет вид

$$Z(x)=z(x)+\int_0^x Z(x-y)\,F\{dy\}, \qquad x>0. \tag{1.6}$$

Если z — непрерывная функция, равная нулю вне некоторого конечного интервала, то, как легко вывести из теоремы 1,

$$Z(t)\to\frac{1}{\mu}\int_0^\infty z(y)\,dy, \qquad t\to\infty \tag{1.7}$$

для неарифметических F, и

$$Z(x+n\lambda)\to\frac{\lambda}{\mu}\sum_k z(x+k\lambda), \qquad n\to\infty \tag{1.8}$$

для арифметического распределения F с шагом λ. Это утверждение не выполняется для произвольных непрерывных функций z, так как эти функции могут сильно осциллировать на бесконечности [пример (г)]. К счастью, осложнений не возникает,

[1]) При первом чтении рекомендуется перейти прямо к § 3.

[2]) Интервал интегрирования считается *замкнутым* и указывается только для ясности; так как функции z и U равны нулю на $-\infty$, 0, то интеграл (1.4) равен интегралу от $-\infty$ до $+\infty$. Это замечание относится ко всем встречающимся в тексте интегралам типа (1.4).

если ограничиться функциями z, для которых интеграл в (1.7) является пределом верхних и нижних сумм в следующем смысле. При фиксированном $h>0$ обозначим через $\underline{m}_n$ и $\overline{m}_n$ минимум и максимум z на интервале $(n-1)h\leqslant x\leqslant nh$ соответственно. Предположим, что ряды

$$\underline{\sigma}=h\Sigma\underline{m}_n, \qquad \overline{\sigma}=h\Sigma\overline{m}_n \tag{1.9}$$

сходятся абсолютно. Их суммы мы назовем нижней и верхней римановыми суммами для данного разбиения. Условимся говорить, что функция z, определенная на $\overline{0,\infty}$, является *непосредственно интегрируемой по Риману*[1]), если эти верхние и нижние суммы существует, и при достаточно малых h $\overline{\sigma}-\underline{\sigma}<\varepsilon$.

Примеры. а) Пусть $z\geqslant 0$ ограниченная и непрерывная функция. Обозначим через μ_n максимум z на интервале $n-1\leqslant x\leqslant n$. Легко видеть, что z непосредственно интегрируема тогда и только тогда, когда $\Sigma\mu_n<\infty$.

б) Пусть z невозрастающая функция и $z(\infty)=0$. Тогда $\Sigma(\overline{m}_n-\underline{m}_n)\leqslant z(0)$, так что или оба ряда в (1.9) сходятся, или оба расходятся. Отсюда следует, что монотонная функция непосредственно интегрируема тогда и только тогда, когда она интегрируема в обычном смысле.

в) Теперь мы дадим пример интегрируемой непрерывной функции, которая *не* является непосредственно интегрируемой. Пусть α — иррациональное число, $0<\alpha<1$. Рассмотрим треугольники единичной высоты с основаниями, лежащими на оси x, причем середины оснований расположены в точках $k+n\alpha$ и столь малы, что треугольники не пересекаются (здесь k и n — натуральные числа). Выберем основания столь малыми, чтобы сумма площадей всех треугольников была конечна. Определим z как непрерывную функцию, график которой состоит из боковых сторон этих треугольников и подходящего множества точек оси x. Так как полная «площадь под графиком» конечна, то функция z интегрируема в обычном смысле. С другой стороны, существует бесконечно много интервалов $(n-1)h\leqslant x\leqslant nh$, содержащих целиком основание какого-либо из наших треуголь-

[1]) Этот термин не является общепринятым, но само понятие известно со времени работ Н. Винера по тауберовым теоремам. Для функций, обращающихся в нуль вне конечного интервала $\overline{0,a}$, «непосредственная» интегрируемость совпадает с обычной интегрируемостью по Риману, однако интеграл Римана на $\overline{0,\infty}$ определяется обычно с помощью предельного перехода при $a\to\infty$.

ников. Для таких интервалов $\bar{m}_n=1$, и потому верхняя сумма $\bar{\sigma}$ не может быть конечна.

г) Мы используем предыдущий пример для того, чтобы показать, что (1.7) может не выполняться, если z не является непосредственно интегрируемой. Пусть F сосредоточено в точках 1 и $1-\alpha$. Тогда U сосредоточено в точках вида $k-n\alpha$. Если $x=r>0$ — целое число, то подинтегральное выражение $z(x-y)$ в $U*r$ равно единице во всех атомах U. Следовательно, $Z(r)=U(r)-1$. Мы видим, что функция Z *не будет даже ограниченной.* ▶

Теорема 2. (*Другая форма теоремы восстановления.*)[1]) *Пусть z непосредственно интегрируема по Риману и $z(x)=0$ при $x<0$. Тогда для неарифметических распределений выполняется* (1.7), *а для арифметических с шагом λ* — (1.8).

Доказательство *эквивалентности обеих теорем.* Предположим, что F — неарифметическое распределение. Чтобы получить (1.3) из (1.7), достаточно положить $z(x)=1$ при $0\leqslant x<h$ и $z(x)=0$ при других x. Для доказательства обратного утверждения рассмотрим сначала частный случай ступенчатых функций со ступеньками равной величины.

Пусть при фиксированных $h>0$ и $n\geqslant 1$ $z_n(x)=1$ для $(n-1)h\leqslant x<nh$ и $z_n(x)=0$ в других точках. Полагая $Z_n=U*z_n$, мы видим из (1.3), что при фиксированном n и $t\to\infty$

$$Z_n(t)=U(t-(n-1)h)-U(t-nh)\to\frac{h}{\mu} \tag{1.10}$$

и что существует верхняя граница M_h для Z_n: $Z_n(t)\leqslant M_h$ при всех n и t.

Пусть $a_k\geqslant 0$ и $\Sigma a_k<\infty$. Тогда ступенчатая функция $z=\Sigma a_k z_h$ непосредственно интегрируема и для $Z=U*z$ выполняются неравенства

$$\sum_{k=1}^{n}a_kZ_k(t)\leqslant Z(t)\leqslant\sum_{k=1}^{n}a_kZ_k(t)+M_h\cdot\sum_{n+1}^{\infty}a_k. \tag{1.11}$$

Так как последнее верно при всех n, то при $t\to\infty$ получаем

$$Z(t)=\frac{h}{\mu}\sum_{1}^{\infty}a_k=\frac{1}{\mu}\int_{0}^{\infty}z(x)\,dx. \tag{1.12}$$

[1]) Теорема 2, несмотря на ее простоту, осталась неотмеченной в литературе. Можно найти лишь некоторые более слабые замены этой теоремы. При этом наблюдается к тому же существенная путаница в вопросах взаимозависимости этих теорем. Большие затруднения вызваны искусственными ограничениями на распределение F.

Таким образом, (1.7) верно для любой ступенчатой функции с одинаковыми ступеньками. Пусть z — произвольная непосредственно интегрируемая функция. Применяя установленный выше результат к двум ступенчатым функциям, получаемым при $a_k=\bar{m}_k$ и $a_k=m_k$ соответственно, мы видим, что все предельные точки для $Z(t)$ лежат между $\mu^{-1}\underline{\sigma}$ и $\mu^{-1}\bar{\sigma}$. Поэтому (1.7) верно и в общем случае.

Эти же рассуждения после очевидных изменений применимы и к случаю арифметических распределений. Доказательство закончено. ▶

Доказательство *теоремы* 1. При

$$z(x)=1-F(x), \quad x>0 \tag{1.13}$$

решение (1.4) уравнения восстановления удовлетворяет соотношению $Z(x)=1$ для $x>0$. Так как функция z монотонна, то из (1.4) следует, что $z(\alpha)[U(x)-U(x-\alpha)]<1$. Поэтому разность $U(x)-U(x-\alpha)$ ограничена, если только α выбрано достаточно малым. Так как каждый интервал может быть разбит на интервалы длины не больше α, то функция $U\{I+t\}$ будет ограниченной при любом фиксированном конечном интервале I. По теореме о выборе (теорема 2 из гл. VIII, 6) существует последовательность чисел $t_k\to\infty$ и мера V, такие, что $U\{t_k+dy\}\to V\{dy\}$ (следует заметить, что V не сосредоточено на положительной полуоси).

Пусть теперь z — непрерывная функция, равная нулю вне некоторого конечного интервала $\overline{0, a}$, и пусть Z снова определена по (1.4). Тогда

$$Z(t_k+x)=\int_{-\infty}^{\infty} z(x-s)\,U\{t_k+ds\}\to\int_{-\infty}^{\infty} z(x-s)\,V\{ds\}, \tag{1.14}$$

где интеграл в действительности берется по интервалу длины не больше a. Правая часть представляет собой ограниченную и непрерывную функцию x, которую мы обозначим через ζ. Тогда $Z(t_k+x)\to\zeta(x)$, и так как Z удовлетворяет уравнению восстановления (1.5), то

$$\zeta(x)=\int_{0}^{\infty}\zeta(x-y)\,F\{dy\}. \tag{1.15}$$

Допустим теперь, что распределение F — неарифметическое. В следующам параграфе будет показано, что в этом случае ζ необходимо будет константой. Это означает, что предел в (1.14) не зависит от x, т. е. мера V должна быть пропорциональна мере

Лебега. Другими словами, мы доказали, что

$$U(t_k) - U(t_k - h) \to \alpha h, \tag{1.16}$$

где α не зависит от h. Это соотношение отличается от (1.3) только тем, что t пробегает последовательность $\{t_k\}$. Предыдущие рассуждения можно применить и для доказательства того, что

$$Z(t_k) \to \alpha \int_0^\infty z(x)\,dx \tag{1.17}$$

для любой непосредственно интегрируемой функции z.

Для функции z, определенной по (1.13), левая часть равна единице. Эта функция является непосредственно интегрируемой, если $\mu < \infty$. Тогда $\alpha = \mu^{-1}$. Если $\mu = \infty$, то интеграл, стоящий справа, расходится и ограниченность $Z(t_k)$ влечет равенство $\alpha = 0$. Следовательно, константа α не зависит от выбора последовательности $\{t_k\}$. Но каждая последовательность чисел $t_k' \to \infty$ содержит подпоследовательность $\{t_k\}$, удовлетворяющую (1.16). Таким образом, теорема верна для всех неарифметических F.

С тривиальными видоизменениями приведенные рассуждения применимы и к арифметическим распределениям (для них, впрочем, теорема восстановления была доказана ранее в 1 гл. XIII). ▶

§ 2*). Уравнение $\zeta = F * \zeta$

В предыдущем доказательстве использовался тот факт, что для неарифметического распределения F ограниченными решениями (1.15) могут быть только константы. Мы выделяем доказательство этого факта ввиду его формального аналитического характера, и в то же самое время мы установим его и для распределений, не сосредоточенных на полупрямой. Этот общий результат будет использован в теореме восстановления в § 9.

Лемма[1]). *Пусть F — распределение вероятностей, не сосредоточенное в нуле, и пусть ζ — ограниченное и непрерывное решение «уравнения свертки»*

$$\zeta(x) = \int_{-\infty}^{+\infty} \zeta(x-y)\,F\{dy\}. \tag{2.1}$$

---

*) Используется только в теоремах восстановления в § 1 и § 9.

[1]) Эта лемма верна для распределений на любых группах, см. Choquet G., Deny J., *C. R. Acad. Sci. Paris*, **250** (1960), 799—801.

Тогда если F — неарифметическое распределение, то $\zeta(x) = \text{const}$. Если F имеет шаг λ, то ζ периодична с периодом λ.

Доказательство. Предположим сначала, что F — неарифметическое распределение. Примем на время без доказательства следующее *утверждение*:

«Пусть ζ — равномерно непрерывное ограниченное решение (2.1), и пусть $m>0$ — точная верхняя граница значений ζ. Тогда существуют интервалы $\overline{t, t+h}$ сколь угодно большой длины, такие, что $\zeta(x) > \frac{1}{2}m$ для всех $t<x<t+h$».

Рассмотрим теперь любое ограниченное непрерывное решение уравнения (2.1). Пусть η_ε обозначает его свертку с нормальным распределением, имеющим нулевое математическое ожидание и дисперсию, равную ε. Функция η_ε и ее производная η_ε' снова будут решениями уравнения (2.1) и потому к ним применимо сделанное выше утверждение. Положим $q = \sup \eta_\varepsilon'(x)$. Если $q>0$, то существует интервал $t<x<t+h$ произвольной длины h, на котором $\eta_\varepsilon'(x) > {}^1/_2 q$. Интегрируя по x, получаем

$$\frac{1}{2}qh < \eta_\varepsilon(t+h) - \eta_\varepsilon(t) \leqslant 2 \cdot \sup|\zeta(x)|. \tag{2.2}$$

Так как h произвольно, то отсюда следует, что $q=0$ и $\eta_\varepsilon' \leqslant 0$. Но те же самые рассуждения применимы к $-\eta_\varepsilon$, так что η_ε равно константе. Полагая $\varepsilon \to 0$, видим, что ζ действительно равно константе.

Для того чтобы закончить доказательство, нам следует установить справедливость утверждения, выделенного выше кавычками. Мы произведем это в два этапа.

(i) Предположим, что ζ достигает максимума m в некоторой точке, например в нуле. Уравнение (2.1) представляет величину $\zeta(0)=m$ в форме взвешенного среднего значений $\zeta(-y) \leqslant m$, и потому в каждой точке роста F должно быть $\zeta(-y)=m$. Но ζ удовлетворяет и уравнению свертки $\zeta = F^{r*} * \zeta$ при $r=1, 2, \ldots$, так что $\zeta(-y)=m$ в каждой точке y, являющейся при каком-либо r точкой роста F^{r*}. Другими словами, обозначая буквой Σ множество точек роста $F, F^{2*}, F^{3*}, \ldots$, имеем $\zeta(-y)=m$ во всех точках $y \in \Sigma$. В соответствии с леммой 2 из гл. V, 4, множество Σ всюду плотно, за исключением того случая, когда F сосредоточено на полуоси. Из плотности Σ очевидным образом вытекает, что $\zeta = \text{const}$. В исключительном случае мы можем допустить, что F сосредоточено на $\overline{0, \infty}$ (но не в нуле). По той же лемме Σ асимптотически плотно на бесконечности, так что $\zeta(-y) \to m$ при $y \to \infty$. Для больших r масса F^{r*} сосредоточи-

вается «вблизи бесконечности». Полагая $r \to \infty$ в соотношении $\zeta = P^{r*} * \zeta$, мы снова видим, что $\zeta(x) = m$ при всех x.

(ii) Таким образом, решение ζ, отличное от постоянной, не может достигать максимума. Поэтому существует последовательность $t_1, t_2, \ldots, \to \pm\infty$, такая, что $\zeta(t_n) \to m$. Определим функции ζ_n равенствами $\zeta_n(x) = \zeta_n(t_n + x)$. По предположению функция ζ равномерно непрерывна. Следовательно, последовательность $\{\zeta_n\}$ является равностепенно непрерывной и содержит подпоследовательность, сходящуюся равномерно на каждом конечном интервале (теорема 3 гл. VIII, 6). Чтобы избежать двойных индексов, предположим, что $\zeta_n \to \eta$. Очевидно, что $\eta = F * \eta$ и функция η равномерно непрерывна. По определению верхней грани $\eta(x) \leqslant m$. Но $\eta(0) = \lim \zeta(t_n) = m$ и по приведенному выше результату $\eta(x) = m$ при всех x. Таким образом, $\zeta_n(x) \to m$ равномерно на $0, h$, откуда следует, что при всех достаточно больших n $\zeta(s) > {}^1/_2 m$ для $t_n < s < t_n + h$. Тем самым лемма доказана для неарифметических распределений.

Для арифметических F утверждается, что при фиксированном a значения $\zeta(a + n\lambda)$ не зависят от n. Доказательство такое же, как и прежде, с той разницей, что m следует определить как точную верхнюю границу указанных значений, а производные следует заменить разностями. ▶

Лемма может быть интерпретирована и доказана вероятностным путем. Именно, пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — независимые случайные величины с одним и тем же распределением F, и $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. Уравнение свертки (2.1) утверждает, что при фиксированном a случайные величины $\zeta(a - \mathbf{S}_n)$ образуют *мартингал* (см. гл. VI, 12). Теорема о сходимости мартингалов из гл. VII, 8 гарантирует существование такой случайной величины $\mathbf{U}_a$, что $\zeta(a - \mathbf{S}_n) \to \mathbf{U}_a$ с вероятностью единица. Лемма утверждает, что $\mathbf{U}_a$ постоянна всюду, за исключением множества меры нуль. Это легко вывести из второго закона «нуля или единицы» (см. гл. IV, 6) [1].

§ 3. Устойчивые (возвратные) процессы восстановления

Мы используем теперь теорему восстановления для вывода различных предельных теорем для процессов восстановления, описанных в гл. VI, 6. Мы рассмотрим последовательность взаимно независимых случайных величин $\mathbf{T}_1, \mathbf{T}_2, \ldots$ с одним и тем же распределением F, которые назовем *промежутками между восстановлениями*. В этом параграфе мы предположим, что F — собственное распределение и $F(0) = 0$. В дополнение к $\mathbf{T}_k$ может

[1]) Doob J. L., Snell J. L., Williamson R. E., in Contributions to Probability and Statistics (сборник в честь Г. Хотеллинга), Стэнфорд, 1960.

быть задана неотрицательная случайная величина $\mathbf{S}_0$, имеющая собственное распределение F_0. Положим

$$\mathbf{S}_n = \mathbf{S}_0 + \mathbf{T}_1 + \ldots + \mathbf{T}_n. \tag{3.1}$$

Случайные величины $\mathbf{S}_n$ называются *моментами восстановления*. Процесс восстановления $\{\mathbf{S}_n\}$ называется *чистым*, если $\mathbf{S}_0 = 0$, и называется процессом *с запаздыванием* в других случаях.

Математическое ожидание числа восстановлений на $\overline{|0, t|}$ равно

$$V(t) = \sum_{n=0}^{\infty} \mathbf{P}\{\mathbf{S}_n \leqslant t\}. \tag{3.2}$$

Воспользуемся обозначением $U = \Sigma F^{n*}$, введенным в (1.2). Тогда

$$V = F_0 * U$$

(и $V = U$ в чистом процессе восстановления). Таким образом [1]), при $h > 0$

$$V(t+h) - V(t) = \int_0^{t+h} [U(t+h-y) - U(t-y)]\, F_0\{dy\}. \tag{3.3}$$

Если F — неарифметическое распределение, то при $t \to \infty$ подинтегральное выражение стремится к $\mu^{-1}h$. Поэтому основная теорема может быть распространена на случай процессов с запаздыванием: если распределение F — неарифметическое, то *математическое ожидание числа восстановлений в* $\overline{|t, t+h|}$ *стремится к* $\mu^{-1}h$. В этом утверждении содержатся две теоремы о равномерной распределенности: во-первых, скорость восстановления стремится к константе и, во-вторых, эта константа не зависит от начального распределения. В этом смысле мы имеем здесь аналог эргодических теорем для цепей Маркова [см. **1**, гл. XV].

Если $\mu < \infty$, то $V(t) \sim \mu^{-1}t$ при $t \to \infty$. Естественно поставить вопрос о том, возможно ли выбрать F_0 так, чтобы получить $V(t) = \mu^{-1}t$, т. е. постоянную скорость восстановления. Но V удовлетворяет уравнению восстановления

$$V = F_0 + F * V, \tag{3.4}$$

и потому $V(t) = \mu^{-1}t$ тогда и только тогда, когда

$$F_0(t) = \frac{t}{\mu} - \frac{1}{\mu}\int_0^t (t-y)\, F\{dy\}. \tag{3.5}$$

[1]) См. сноску 2 на стр. 425.

Интегрирование по частям показывает, что это эквивалентно равенству

$$F_0(t)=\frac{1}{\mu}\int_0^t [1-F(y)]\,dy. \tag{3.6}$$

Функция F_0 является функцией распределения, так что ответ утвердителен: *при начальном распределении* (3.6) *скорость восстановления постоянна*, $V(t)=\mu^{-1}t$.

Распределение (3.6) появляется и как предельное распределение времени ожидания восстановления (или как предельное распределение величины перескока через некоторый уровень). Каждому $t>0$ соответствует случайный индекс $\mathbf{N}_t$, такой, что

$$\mathbf{S}_{\mathbf{N}_t}\leqslant t<\mathbf{S}_{\mathbf{N}_t+1}. \tag{3.7}$$

По терминологии гл. VI, 7 случайная величина $\mathbf{S}_{\mathbf{N}_t+1}-t$ называется *остающимся временем ожидания*, соответствующим моменту t. Обозначим через $H(t,\xi)$ вероятность того, что это время $\leqslant\xi$. Другими словами, $H(t,\xi)$ — это вероятность того, что *первый* после t момент восстановления лежит в интервале $t, t+\xi$ или что перескок через уровень t не превосходит ξ. Это событие происходит тогда и только тогда, когда некоторый момент восстановления $\mathbf{S}_n$ равен $x\leqslant t$, а время до следующего момента восстановления лежит между $t-x$ и $t-x+\xi$. В случае чистого процесса восстановления получаем, суммируя по всем x и n,

$$H(t,\ \xi)=\int_0^t U\{dx\}\,[F(t-x+\xi)-F(t-x)]. \tag{3.8}$$

Этот интеграл содержит ξ как свободный параметр и имеет стандартный вид $U*z$, где $z(t)=F(t+\xi)-F(t)$. Заметим, что последняя функция является непосредственно интегрируемой [1]). Предположим, что распределение F — неарифметическое. Из равенства

$$\int_0^\infty z(t)\,dt=\int_0^\infty ([1-F(t)]-[1-F(t+\xi)]\,dt=$$

$$=\int_0^\xi (1-F(s))\,ds \tag{3.9}$$

[1]) Максимум функции z в $\overline{n,n+1}$ не превосходит $F(n+1+\xi)-F(n)$, и ряд из этих членов сходится.

получаем *предельную теорему*

$$\lim_{t\to\infty} H(t,\ \xi)=\frac{1}{\mu}\int_0^{\xi}[1-F(s)]\,ds \tag{3.10}$$

(легко проверить, что это соотношение верно и для процессов с запаздыванием при любом начальном распределении F_0). Эта предельная теорема замечательна в нескольких отношениях.

Если $\mu<\infty$, то правая часть совпадает с распределением (3.6). Таким образом, при $\mu<\infty$ *остающееся время ожидания имеет собственное предельное распределение*, которое совпадает с распределением, приводящим к постоянной скорости восстановления. В этом примере мы сталкиваемся еще раз со сходимостью к стационарному режиму.

Предельное распределение в (3.10) имеет конечное математическое ожидание в том и только том случае, когда F имеет дисперсию. Это указывает, грубо говоря, на то, что *«вероятности вхождения»* [1]) *ведут себя хуже, чем* F. В случае $\mu=\infty$ имеем при всех ξ

$$H(t, \xi)\to 0, \tag{3.11}$$

т. е. вероятность того, что *перескок через уровень* t *превзойдет сколь угодно большое число* ξ, стремится к единице (для случая правильно меняющихся хвостов распределения более точная информация дается в гл. XIV, 3).

Примеры. а) *Суперпозиция процессов восстановления.* Пусть даны n процессов восстановления. Соединяя все моменты восстановления вместе, мы получим некоторый новый процесс. Вообще говоря, новый процесс *не* будет процессом восстановления, однако для него легко подсчитать время ожидания W первого восстановления после нуля. Мы покажем, что при весьма общих условиях *распределение* W *приближается к показательному закону*, а составной процесс — к процессу Пуассона. Этот результат объясняет, почему многие процессы (такие, как поток вызовов, поступающих на телефонную станцию) похожи на процесс Пуассона.

Рассмотрим n взаимно независимых процессов восстановления, у которых промежутки между восстановлениями имеют распределения $F_1, \ldots, F_n$ с математическими ожиданиями $\mu_1, \ldots, \mu_n$ соответственно. Положим

$$\frac{1}{\mu_1}+\ldots+\frac{1}{\mu_n}=\frac{1}{\alpha}. \tag{3.12}$$

[1]) См. гл. VI, 7.

Мы потребуем, грубо говоря, чтобы моменты восстановления в каждом из объединяемых процессов были расположены крайне редко, так что суммарный эффект складывается из многих малых частей. Чтобы удовлетворить это требование, мы допустим, что при фиксированных k и y вероятности $F_k(y)$ малы, а μ_k велики. Это предположение оказывается осмысленным в форме предельной теоремы.

Рассмотрим «стационарную» ситуацию, предполагая, что рассматриваемые процессы длятся уже длительное время. Тогда для времени ожидания $\mathbf{W}_k$ ближайшего момента восстановления в k-м процессе имеем

$$\mathbf{P}\{\mathbf{W}_k \leqslant t\} \approx \frac{1}{\mu_k}\int_0^t (1-F_k(y))\,dy \approx \frac{t}{\mu_k}. \tag{3.13}$$

[Последняя аппроксимация основана на малости $F_k(y)$.] Время ожидания $\mathbf{W}$ в объединенном процессе равно наименьшему из $\mathbf{W}_k$ и потому

$$\mathbf{P}\{\mathbf{W} > t\} \approx \left(1-\frac{t}{\mu_1}\right)\ldots\left(1-\frac{t}{\mu_n}\right) \approx e^{-t\alpha}. \tag{3.14}$$

Эти рассуждения легко сделать строгими, и при указанных выше условиях показательное распределение будет предельным при $n\to\infty$.

б) *Вероятности достижения областей в случайных блужданиях.* Рассмотрим последовательность независимых случайных величин $\mathbf{X}_1, \mathbf{X}_2, \ldots$ и их частные суммы

$$\mathbf{Y}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n.$$

Для положительных $\mathbf{X}_k$ случайное блуждание $\{\mathbf{Y}_n\}$ представляет собой процесс восстановления, однако сейчас мы рассмотрим произвольные $\mathbf{X}_k$. Допустим, что случайное блуждание устойчиво, так что для любого $t>0$ с вероятностью единица найдется такое n, что $\mathbf{Y}_n > t$. Если $\mathbf{N}$ — наименьший индекс, для которого это верно, то $\mathbf{Y}_\mathbf{N}$ называется точкой *первого вхождения* в $\overline{t, \infty}$. Случайная величина $\mathbf{Y}_\mathbf{N} - t$ равна перескоку через уровень t при первом вхождении и соответствует остающемуся времени ожидания в процессах восстановления. Положим снова $\mathbf{P}\{\mathbf{Y}_\mathbf{N} \leqslant t+\xi\} = H(t, \xi)$ и покажем, каким образом предельная теорема для остающегося времени ожидания может быть применена к этому распределению.

Определим $\mathbf{S}_1$ как точку первого вхождения в $\overline{0, \infty}$ и по индукции определим $\mathbf{S}_{n+1}$ как точку первого вхождения в $\overline{\mathbf{S}_n, \infty}$. Последовательность $\mathbf{S}_1, \mathbf{S}_2 \ldots$ совпадает с *последовательностью лестничных высот*, введенной в гл. VI, 8, и образует процесс

восстановления, так как разности $\mathbf{S}_{n+1} - \mathbf{S}_n$ очевидным образом независимы и имеют такое же распределение, как $\mathbf{S}_1$. Разность $\mathbf{Y}_N - t$ есть остающееся время ожидания в процессе восстановления $\{\mathbf{S}_n\}$, так что можно применить (3.10). ▷

Методом, использованным при доказательстве (3.10), можно показать, что *прошедшее время ожидания* $t - \mathbf{S}_{N_t}$ *имеет то же самое предельное распределение. Для длины промежутка* $\mathbf{L}_t = \mathbf{S}_{N_t+1} - \mathbf{S}_{N_t}$ *между восстановлениями, содержащего* t, *имеем*

$$\mathbf{P}\{\mathbf{L}_t < \xi\} = \int_{t-\xi}^{t} U\{dx\}\,[F(\xi) - F(t-x)]. \tag{3.15}$$

Следовательно,

$$\lim_{t\to\infty} \mathbf{P}\{\mathbf{L}_t < \xi\} = \frac{1}{\mu}\int_0^{\xi} [F(\xi) - F(y)]\,dy = \frac{1}{\mu}\int_0^{\xi} xF\{dx\}. \tag{3.16}$$

Любопытные следствия этой формулы обсуждались в связи с «парадоксом контроля» в гл. VI, 7 и «парадоксом времени ожидания» в гл. I, 4.

Легко видеть, что три семейства случайных величин $t - \mathbf{S}_{N_t}$, $\mathbf{S}_{N_t+1} - t$, $\mathbf{L}_t$ суть однородные марковские процессы. Доказанные нами три предельные теоремы следуют из эргодических теорем для марковских процессов (см. также гл. XIV, 3).

§ 4. Уточнения

В этом параграфе мы покажем, как отражаются на поведении решений уравнения восстановления

$$Z(t) = z(t) + \int_0^t Z(t-y)\,F\{dy\} \tag{4.1}$$

предположения о регулярности распределения F промежутков между восстановлениями. Сами по себе результаты не так уж восхитительны, но они важны для многих применений, а их вывод методологически интересен. Чтобы избежать тривиальностей (и повторений сказанного в 1, гл. XIII), мы предположим, что распределение F — *нерифметическое.*

При $\mu < \infty$ из теоремы восстановления вытекает, что $U(t) \sim t/\mu$. Наша первая цель — оценить разность

$$Z(t) = U(t) - \frac{t}{\mu} \tag{4.2}$$

в случае, когда F имеет дисперсию σ^2. Эта функция, как можно понять из доказательства (3.6) или проверить непосредственно, удовлетворяет (4.1) с

$$z(t)=\frac{1}{\mu}\int_t^\infty [1-F(y)]\,dy. \tag{4.3}$$

Интегрируя по частям, получаем

$$\int_0^\infty z(t)\,dt=\frac{1}{2\mu}\int_0^\infty y^2F\{dy\}=\frac{\sigma^2+\mu^2}{2\mu}. \tag{4.4}$$

Так как функция z монотонна, то по второй теореме восстановления заключаем, что *для неарифметического распределения F с дисперсией σ^2 имеем*

$$0\leqslant U(t)-\frac{t}{\mu}\to\frac{\sigma^2+\mu^2}{2\mu^2}. \tag{4.5}$$

Этот результат значительно сильнее самой теоремы восстановления. Если F имеет моменты более высокого порядка, то можно дать дальнейшие уточнения. Если $\mu<\infty$, но дисперсия не существует, то (4.5) выполняется с заменой правой части на ∞.

Пример. Пусть F — равномерное распределение на $\overline{0,1}$. Из формулы свертки для равномерного распределения легко получить, что

$$U(t)=\sum_{k=0}^{n}(-1)^k e^{t-k}\frac{(t-k)^k}{k!}\quad\text{при}\quad n\leqslant t\leqslant n+1. \tag{4.6}$$

Эту формулу часто открывают заново в теории очередей, но она мало говорит о природе U. Асимптотическая формула $0\leqslant U(t)-2t\to{}^2/_3$ значительно интереснее. Она кажется тривиальной в настоящем контексте, но вне него это не так; [STAM Rev, vol. 5 (1963), 283—287].

Перейдем теперь от асимптотических свойств к свойствам гладкости. Допустим, что F имеет плотность f, и рассмотрим уравнение восстановления

$$u(x)=f(x)+\int_0^x u(x-y)f(y)\,dy. \tag{4.7}$$

Мы знаем, что его решение единственно. Интеграл Z от функции u удовлетворяет (4.1) с $z=F$, следовательно,

$$U(t)=1+\int_0^t u(y)\,dy, \qquad t>0. \tag{4.8}$$

Таким образом, если исключить атом в нуле, то U имеет плотность u, к которой применима теорема восстановления (в предположении, что f достаточно правильна, например, если f, начиная с некоторого места, монотонна). При такого рода условиях *U имеет плотность u, такую, что $u(t)\to\mu^{-1}$.*

Плотности, не являющиеся непосредственно интегрируемыми, вряд ли могут встретиться в практических задачах. Однако и относительно них мы можем сделать некоторые заключения. В самом деле, рассмотрим плотность

$$f_2(t)=\int_0^t f(t-y)\,f(y)\,dy \tag{4.9}$$

распределения $F*F$. Вообще говоря, f_2 «много лучше», чем f. Например, если $f<M$, то из (4.9) и соображений симметрии получаем

$$f_2(t)<2M\left[1-F\left(\frac{1}{2}t\right)\right]. \tag{4.10}$$

Правая часть монотонна и интегрируема при $\mu<\infty$. Разность $u-f$ есть решение уравнения восстановления (4.1), соответствующее $r=f_2$. Таким образом,

$$u(t)-f(t)\to\frac{1}{\mu}. \tag{4.11}$$

Мы получили это предельное соотношение при всего лишь двух условиях: $\mu<\infty$ *и* f *ограничена.* Из сказанного следует, что если колебания f велики, то u изменяется так, что компенсирует их. Рассуждения, которые мы использовали, часто применимы и приводят к сильным результатам.

§ 5. Центральная предельная теорема

Рассмотрим для простоты чистый процесс восстановления и обозначим через $\mathbf{N}_t$ число моментов восстановления на $\overline{0,t}$ (исключая нуль). Событие $\{\mathbf{N}_t\geqslant r\}$ равносильно тому, что r-й момент восстановления находится на $\overline{0,t}$. Поэтому

$$\mathbf{P}\{\mathbf{N}_t\geqslant r\}=F^{r*}(t). \tag{5.1}$$

Следовательно, знание F^{r*} позволяет нам в принципе вычислить распределение $\mathbf{N}_t$. Точные вычисления затруднительны, но, зная асимптотическое поведение F^{r*}, можно получить асимптотиче-

ские формулы для распределения $\mathbf{N}_t$. Для случая когда F имеет конечное математическое ожидание μ и конечную дисперсию σ^2, это было проделано в 1, гл. XIII, 6. Рассуждения не связаны с арифметическим характером распределения F, так что справедлив следующий общий результат: если F имеет математическое ожидание μ и дисперсию σ^2, то при больших t *число* $\mathbf{N}_t$ *моментов восстановления распределено асимптотически нормально с математическим ожиданием* $t\mu^{-1}$ *и дисперсией* $t\sigma^2\mu^{-3}$.

Пример. а) *Счетчики первого типа.* Поступающие частицы образуют пуассоновский поток. Частица, попадающая в счетчик в момент, когда он работает, регистрируется, но «закрывает» счетчик на *фиксированное* время ξ. Частицы, достигающие счетчика в то время, когда он «закрыт», не регистрируются. Для простоты рассмотрим процесс с того момента, когда новая частица попадает в работающий счетчик. Мы имеем *два* процесса восстановления. Первый процесс — входной поток. Это процесс Пуассона, т. е. промежутки между появлениями частиц имеют показательное распределение $1-e^{-ct}$ с математическим ожиданием c^{-1} и дисперсией c^{-2}. Последовательные регистрации образуют вторичный процесс восстановления. Промежутки между последовательными регистрациями представляют собой суммы ξ и показательно распределенных случайных величин. Поэтому их математическое ожидание равно $\xi+c^{-1}$, а дисперсия — c^{-2}. *Таким образом, число регистраций на временно́м интервале* $0, t$ *распределено приблизительно нормально с математическим ожиданием* $tc(1+c\xi)^{-1}$ *и дисперсией* $tc(1+c\xi)^{-3}$.

Из того что эти величины различны, следует, что число регистраций не подчиняется распределению Пуассона. В свое время не было отчетливого понимания того факта, что процесс регистраций существенно отличается от первоначального процесса. Результаты наблюдений привели некоторых физиков к ошибочному заключению, что ливни космических частиц не имеют пуассоновского характера «чистой случайности».

Предельное распределение для $\mathbf{N}_t$ существует тогда и только тогда, когда F принадлежит области притяжения некоторого закона. Области притяжения описаны в гл. IX, 8 и XVII, 5, откуда можно вывести, что $\mathbf{N}_t$ *имеет собственное предельное распределение тогда и только тогда, когда*

$$1-F(x)\sim x^{-\alpha}L(x), \qquad x\to\infty, \tag{5.2}$$

где L — медленно меняющаяся функция и $0<\alpha<2$. Предельное распределение для $\mathbf{N}_t$ легко вычислить. При этом обнаруживаются парадоксальные свойства флуктуаций. Их поведение существенно различно при $\alpha<1$ и $\alpha>1$.

Рассмотрим случай $0<\alpha<1$. Если a_r выбраны так, что

$$r[1-F(a_r)]\to\frac{2-\alpha}{\alpha}, \tag{5.3}$$

то $F^{r*}(a_r x) \to G_\alpha(x)$, где G_α — одностороннее устойчивое распределение с $p=1$, $q=0$, описанное в гл. IX, (8.18) [см. также гл. XVII, (5.20) и гл. XIII, 6]. Пусть r и t возрастают таким образом, что $t \sim a_r x$. Тогда из (5.2) и (5.3) с учетом того, что функция L медленно меняющаяся, получаем

$$r \sim \frac{2-\alpha}{\alpha} \frac{x^{-\alpha}}{1-F(t)}, \tag{5.4}$$

откуда по (5.1)

$$\mathbf{P}\left\{[1-F(t)]\,\mathbf{N}_t \geqslant \frac{2-\alpha}{\alpha}\, x^{-\alpha}\right\} \to G_\alpha(x). \tag{5.5}$$

Это соотношение — аналог центральной предельной теоремы. Частный случай $\alpha = \frac{1}{2}$ содержится в теореме 2 из 1, гл. XIII, 6. Неожиданные черты поведения $\mathbf{N}_t$ связаны с нормирующим множителем $1-F(t)$ в (5.5). Грубо говоря, $1-F(t)$ имеет порядок $t^{-\alpha}$, так что величина $\mathbf{N}_t$ имеет порядок t^{α}; плотность моментов восстановления должна неограниченно убывать (что согласуется с асимптотическим поведением вероятностей достижения барьеров).

При $1<\alpha<2$ распределение F имеет математическое ожидание $\mu<\infty$ и вычисления, аналогичные проделанным выше, показывают, что

$$\mathbf{P}\left\{\mathbf{N}_t \geqslant \frac{t-\lambda(t)\,x}{\mu}\right\} \to G_\alpha(x),$$

где $\lambda(t)$ удовлетворяет соотношению

$$t\,[1-F(\lambda(t))] \to \frac{2-\alpha}{\alpha}\,\mu.$$

В этом случае среднее число моментов восстановления растет линейно, однако наличие нормирующего множителя $\lambda(t)$ показывает, что случайные отклонения от среднего могут быть весьма велики.

§ 6. Обрывающиеся (невозвратные) процессы

Общая теория процессов восстановления с несобственным распределением F почти тривиальна. Однако соответствующее уравнение восстановления часто встречается в различных «масках», когда случайные черты затемняют общее основание. Ясное понимание этого основного факта позволит избежать громоздких рассуждений в отдельных применениях. В частности, асимптотическая формула теоремы 2 влечет результаты, для доказательства которых ранее использовался в специально преобразованном виде аппарат уравнений Винера — Хопфа.

Чтобы избежать путаницы в обозначениях, мы будем использовать для исходного распределения букву L (вместо F). Иными словами, в этом разделе *L обозначает несобственное распределение вероятностей с* $L(0)=0$ *и* $L(\infty)=L_\infty<1$. Оно играет роль распределения (несобственных) промежутков $\mathbf{T}_k$ между последовательными восстановлениями, причем «дефект» $1-L_\infty$ равен вероятности обрыва процесса. Начало координат на оси времени считается моментом восстановления с нулевым номе-

ром, а $\mathbf{S}_n = \mathbf{T}_1 + \ldots + \mathbf{T}_n$ представляет собой n-й момент восстановления; это несобственная случайная величина с распределением L^{n*}, которому соответствует полная масса $L^{n*}(\infty) = L_\infty^n$. «Дефект» $1 - L_\infty^n$ равен вероятности обрыва процесса *до* n-го момента восстановления. Положим снова

$$U = \sum_{n=0}^{\infty} L^{n*}. \tag{6.1}$$

Как и в случае возвратных процессов, $U(t)$ равно *математическому ожиданию числа* моментов восстановления в интервале $0, t$. На этот раз, однако, математическое ожидание полного числа восстановлений является конечным и равно

$$U(\infty) = \frac{1}{1 - L_\infty}. \tag{6.2}$$

Вероятность того, что n-й момент восстановления $\mathbf{S}_n$ будет последним и при этом $\leqslant x$, равна $(1 - L_\infty) L^{n*}(x)$. Таким образом, верна

Теорема 1. *Обрывающийся процесс восстановления, начинающийся в начале координат, оканчивается с вероятностью единица. Момент обрыва* $\mathbf{M}$ (*т. е. максимум последовательности* $0, \mathbf{S}_1, \mathbf{S}_2, \ldots$) *имеет собственное распределение*

$$\mathbf{P}\{\mathbf{M} \leqslant x\} = (1 - L_\infty) U(x). \tag{6.3}$$

Вероятность того, что n-й момент восстановления будет последним, равна $(1 - L_\infty) L_\infty^n$, т. е. число моментов восстановления имеет геометрическое распределение.

Эти результаты возможно изложить и в терминах несобственного *уравнения восстановления*

$$Z(t) = z(t) + \int_0^t Z(t - y)\, L\{dy\}, \tag{6.4}$$

но при несобственном распределении L соответствующая теория тривиальна. В предположении, что $z(x) = 0$ при $x \leqslant 0$, единственное решение задается формулой

$$Z(t) = \int_0^t z(t - y)\, U\{dy\} \tag{6.5}$$

и, как очевидно,

$$Z(t) \to \frac{z(\infty)}{1 - L_\infty}, \tag{6.6}$$

если только $z(t) \to z(\infty)$ при $t \to \infty$.

Примеры. а) Событие $\{\mathbf{M}\leqslant t\}$ происходит, если процесс заканчивается на $\mathbf{S}_0$ или если $\mathbf{T}_1$ принимает некоторое положительное значение $y\leqslant t$ и остающийся процесс «доживает» до момента $\leqslant t-y$. Таким образом, $Z(t)=\mathbf{P}\{\mathbf{M}\leqslant t\}$ *удовлетворяет уравнению*

$$Z(t)=1-L_\infty+\int_0^t Z(t-y)\,L\{dy\}. \tag{6.7}$$

Это уравнение эквивалентно (6.3).

б) *Вычисление моментов.* В последнем соотношении собственное распределение Z представляется в виде суммы двух несобственных распределений: одно определено, как свертка, а другое имеет единственный атом в нуле. Чтобы вычислить математическое ожидание Z, положим

$$E_L=\int_0^\infty xL\{dx\}. \tag{6.8}$$

Аналогичное обозначение примем и для других распределений (как собственных, так и несобственных). Так как L — несобственное распределение, то свертка в (6.7) имеет математическое ожидание $L_\infty\cdot E_Z+E_L$. Таким образом, $E_Z=E_L/(1-L_\infty)$. Для более общего уравнения (6.4) подобным же путем получаем

$$E_Z=\frac{E_z+E_L}{1-L_\infty}. \tag{6.9}$$

Моменты более высокого порядка могут быть подсчитаны этим же методом. ▶

Для применений важно установить *асимптотические оценки* для $Z(\infty)-Z(t)$. Это легко сделать, если несобственное распределение L обладает тем свойством, что для некоторого $\varkappa$

$$\int_0^\infty e^{\varkappa y}L\{dy\}=1. \tag{6.10}$$

Если такое число существует, то оно, очевидно, единственно и $\varkappa>0$. Введем теперь *собственное* распределение вероятностей $L^\#$ равенством

$$L^\#\{dy\}=e^{\varkappa y}L\{dy\} \tag{6.11}$$

и сопоставим каждой функции f новую функцию $f^\#$, определенную соотношением

$$f^\#(x)=e^{\varkappa x}f(x).$$

Достаточно взглянуть на (6.4), чтобы понять, что выполняется уравнение восстановления

$$Z^{\#}(t)=z^{\#}(t)+\int_0^t Z^{\#}(t-y)\,L^{\#}\{dy\} \tag{6.12}$$

и что к нему применима теорема восстановления из § 1. Если $Z^{\#}(t)\to a>0$, то $Z(t)\sim ae^{-\varkappa t}$. Если Z имеет положительный предел, то мы можем применить те же самые рассуждения к разности $Z_1(t)=Z(\infty)-Z(t)$, которая удовлетворяет (6.4) с заменой z на

$$z_1(t)=z(\infty)-z(t)+z(\infty)\frac{L_\infty-L(t)}{1-L_\infty}. \tag{6.13}$$

Легко проверить, что интеграл от $z_1^{\#}(t)=z_1(t)\,e^{\varkappa t}$ равен правой части (6.14). Поэтому верна

Теорема 2. *Если выполняется* (6.10), *то решение уравнения восстановления удовлетворяет условию*

$$\mu^{\#}e^{\varkappa t}[Z(\infty)-Z(t)]\to\frac{z(\infty)}{\varkappa}+\int_0^\infty e^{\varkappa x}[z(\infty)-z(x)]\,dx \tag{6.14}$$

с

$$\mu^{\#}=\int_0^\infty e^{\varkappa y}yL\{dy\} \tag{6.15}$$

при дополнительном предположении, что $\mu^{\#}<\infty$ *и что подинтегральное выражение*[1]) (6.14) *имеет порядок* $O(x^{-1-\varepsilon})$ *при некотором* $\varepsilon>0$.

Для частного случая (6.7) получаем

$$\mathbf{P}\{\mathbf{M}>t\}\sim\frac{1-L_\infty}{\varkappa\mu^{\#}}\,e^{-\varkappa t}. \tag{6.16}$$

В следующем параграфе мы покажем, как поразительно сильна эта оценка в применениях.

Случай $L_\infty>1$. Если $L_\infty>1$, то существует константа $\varkappa<0$, такая, что (6.10) верно, и тогда описанное выше преобразование сводит интегральное уравнение (6.4) к (6.12). Теорема восстановления приводит, таким образом, к точным оценкам асимптотического поведения $Z(t)e^{\varkappa t}$. [Дискретный случай охватывается теоремой 1 из **1**, гл. XIII, 4. О применениях в демографии см. пример **1**, гл. XII, (5, в). Формулировку в терминах преобразований Лапласа см. гл. XIII, (6, в).]

[1]) Это условие нужно лишь для того, чтобы обеспечить выполнение требования «непосредственной интегрируемости» во второй теореме восстановления. С очевидными видоизменениями теорема 2 верна и в случае $\mu^{\#}=\infty$.

§ 7. Применения

Как уже указывалось, теория предыдущего параграфа может быть применена к задачам, которые не связаны непосредственно с процессами восстановления. В этом параграфе мы дадим два не зависящих друг от друга примера.

а) *Крамеровская оценка для риска.* В гл. VI, 5 было показано, что в решении задачи о разорении в обобщенном пуассоновском процессе, а также в решении проблем, связанных с размером хранилищ, назначением пациентов и т. п., используется *распределение R, сосредоточенное на $\overline{0, \infty}$ и удовлетворяющее интегро-дифференциальному уравнению*

$$R'(z) = \frac{\alpha}{c} R(z) - \frac{\alpha}{c} \int_0^z R(z-x) F\{dx\}, \tag{7.1}$$

где F — собственное распределение[1]). Интегрируя (7.1) на $\overline{0, t}$ и производя очевидное интегрирование по частям, получаем

$$R(t) - R(0) = \frac{\alpha}{c} \int_0^t R(t-x)[1 - F(x)]\, dx. \tag{7.2}$$

Здесь $R(0)$ — неизвестная постоянная, но в остальном (7.2) представляет собой уравнение восстановления с *несобственным* распределением L, имеющим плотность $(\alpha/c)[1-F(x)]$. Обозначим математическое ожидание F через μ. Тогда полная масса L равна $L_\infty = \alpha\mu/c$. [Рассматриваемый процесс имеет смысл только при $L_\infty < 1$, так как в противном случае $R(t) = 0$ при всех t.] Заметим, что (7.2) является частным случаем (6.4) и что $R(\infty) = 1$. Вспоминая (6.6), получаем

$$R(0) = 1 - \frac{\alpha\mu}{c}. \tag{7.3}$$

Подставляя это значение в уравнение (7.2), мы приведем его к уравнению вида (6.7) для распределения времени жизни $\mathbf{M}$ обрывающегося процесса, в котором промежуток между соседними восстановлениями имеет распределение L. Из (6.16) следует, что *если существует константа $\varkappa$, такая, что*

$$\frac{\alpha}{c} \int_0^\infty e^{\varkappa x}[1 - F(x)]\, dx = 1 \tag{7.4}$$

---

[1]) Это специальный случай гл. VI, (5.4) с распределением F, сосредоточенным на $\overline{0, \infty}$. Он будет рассмотрен с помощью преобразований Лапласа в гл. XIV, (2, б). Общую ситуацию см. в гл. XII, (5, г).

и

$$\mu^{\#} = \frac{\alpha}{c} \int_0^\infty e^{\varkappa x} x [1 - F(x)]\, dx < \infty, \tag{7.5}$$

то при $t \to \infty$

$$1 - R(t) \sim \frac{1}{\varkappa \mu^{\#}} \left(1 - \frac{\alpha\mu}{c}\right) e^{-\varkappa t}. \tag{7.6}$$

Это известная оценка, принадлежащая Крамеру. Моменты R могут быть вычислены, как указано в примере (6.б).

б) *Пробелы в процессе Пуассона.* В гл. VI, 7 мы вывели уравнение восстановления для распределения V времени ожидания первого пробела длины $\geqslant \xi$ в процессе восстановления. Когда этот процесс является процессом Пуассона, то промежутки между последовательными восстановлениями распределены показательно, и уравнение восстановления гл. VI, (7.2) имеет стандартную форму $V = z + V * L$ с

$$L(x) = \begin{cases} 1 - e^{-cx} & \text{при} \quad x < \xi, \\ 1 - e^{-c\xi} & \text{при} \quad x \geqslant \xi \end{cases} \tag{7.7}$$

и

$$z(x) = \begin{cases} 0 & \text{при} \quad x < \xi, \\ e^{-c\xi} & \text{при} \quad x \geqslant \xi. \end{cases} \tag{7.8}$$

Так как $z(\infty) = 1 - L_\infty$, то решение V будет собственным распределением вероятностей, что соответствует смыслу задачи.

Моменты рассматриваемого нами времени ожидания $\mathbf{W}$ легко вычислить методом, описанным в примере (6, б). Мы получаем

$$\begin{aligned} \mathbf{E}(\mathbf{W}) &= \frac{e^{c\xi} - 1}{c}, \\ \operatorname{Var}(\mathbf{W}) &= \frac{e^{2c\xi} - 1 - 2c\xi e^{c\xi}}{c^2}. \end{aligned} \tag{7.9}$$

Условимся интерпретировать $\mathbf{W}$ как *время ожидания пешеходом возможности пересечь поток машин.* Тогда указанные формулы показывают, как влияет на ответ возрастание плотности потока. Математическое ожидание числа машин, проходящих за время, необходимое для пересечения потока, равно $c\xi$. Полагая $c\xi = 1, 2$, видим, что $\mathbf{E}(\mathbf{W}) \approx 1{,}72\xi$ и $\mathbf{E}(\mathbf{W}) \approx 3{,}2\xi$ соответственно. Дисперсия возрастает от (примерно) ξ^2 до $6\xi^2$. [Явное решение и связь с теоремой о покрытии указаны в примере гл. XIV, (2, а).] Можно применить асимптотическую оценку, даваемую теоремой 6.2. При $c\xi > 1$ основное уравнение (6.10)

сводится к

$$ce^{(\varkappa-c)\xi} = \varkappa, \qquad 0 < \varkappa < c, \tag{7.10}$$

и стандартными вычислениями мы выводим из (6.14), что

$$1 - V(t) \sim \frac{1-\varkappa/c}{1-\varkappa\xi}\, e^{-\varkappa t}. \quad \blacktriangleright$$

§ 8. Существование пределов в случайных процессах

Сила теоремы восстановления наиболее ярко демонстрируется, пожалуй, тем, что она позволяет без всяких усилий доказать существование «стационарного режима» в обширном классе случайных процессов. От самого процесса мы потребуем лишь, чтобы рассматриваемые вероятности были определены. В противном случае теорема имеет чисто аналитический характер[1]).

Рассмотрим случайный процесс со счетным множеством состояний $E_0, E_1, \ldots$ и обозначим $P_k(t)$ вероятность состояния E_k в момент $t>0$. Нижеследующая теорема опирается на существование «рекуррентных событий», т. е. моментов, в которые процесс как бы начинается заново. Более точно мы предполагаем, что существует момент $\mathbf{S}_1$, такой, что течение процесса после $\mathbf{S}_1$ является точной вероятностной копией всего процесса, начинающегося в момент 0. Отсюда вытекает существование целой последовательности моментов $\mathbf{S}_2, \mathbf{S}_3, \ldots$ с тем же самым свойством. Последовательность $\{\mathbf{S}_n\}$ представляет собой возвратный процесс восстановления. Мы допустим, что среднее время между восстановлениями $\mu = \mathbf{E}(\mathbf{S}_1)$ конечно. Обозначим $P_k(t)$ условную вероятность E_k в момент $t+s$ при условии, что $\mathbf{S}_1 = s$. Предполагается, что эти вероятности не зависят от s. При этих условиях мы докажем следующее важное утверждение.

Теорема. *Предел*

$$\lim_{t\to\infty} P_k(t) = p_k \tag{8.1}$$

существует, причем $p_k \geqslant 0$ *и* $\sum p_k = 1$.

Доказательство. Пусть $q_k(t)$ — вероятность объединения событий: $\mathbf{S}_1 > t$ и «в момент t система находится в состоянии E_k». Тогда

$$\sum_{k=0}^{\infty} q_k(t) = 1 - F(t), \tag{8.2}$$

[1]) Более сложные результаты см. в статье Beneš V. E., A «renewal» limit theorem for general stochastic processes, *Ann. Math. Statist.*, **33** (1962), 98—113 или в книге того же автора (1963).

где F — распределение «времен восстановления» $\mathbf{S}_{n+1} - \mathbf{S}_n$. По предположению

$$P_k(t) = q_k(t) + \int_0^t P_k(t-y)F\{dy\}. \tag{8.3}$$

Так как функция $1 - F$ монотонна и имеет конечный интеграл μ, то по второй теореме восстановления

$$\lim_{t\to\infty} P_k(t) = \frac{1}{\mu}\int_0^\infty q_k(t)\,dt. \tag{8.4}$$

Интегрируя (8.2), видим, что сумма этих пределов равна единице. Теорема доказана. ▶

Примечательно, что *существование* пределов (8.1) установлено безотносительно к способу их вычисления.

Замечание. Если F_0 — собственное распределение, то (8.1) влечет

$$\int_0^t P_k(t-y)F_0\{dy\} \to p_k, \quad t\to\infty. \tag{8.5}$$

Таким образом, теорема охватывает и случай процесса восстановления *с запаздыванием*, в котором $\mathbf{S}_0$ имеет распределение F_0.

Примеры. а) *Теория очередей.* Рассмотрим систему (типа телефонной станции, почтового отделения или части вычислительной машины), состоящую из одного или более «обслуживающих устройств», и пусть E_k обозначает состояние: в системе имеется k ожидающих «клиентов». В большинстве моделей процесс начинается заново, когда прибывающие «клиенты» застают систему в состоянии E_0. В этом случае наша теорема применима тогда и только тогда, когда с вероятностью единица наступает такой момент, и когда математическое ожидание времени между последовательными наступлениями конечно [1]).

б) *Двустепенный процесс восстановления.* Допустим, что имеются два возможных состояния E_1 и E_2. Первоначально система находится в E_1. Последовательные промежутки пребывания в E_1 суть случайные величины $\mathbf{X}_j$ с одним и тем же распределением F_1. Они перемежаются с промежутками пребывания в E_2, имеющими одно и то же распределение F_2. Предполагая,

[1]) Как типичный пример см. задачу об очереди к нескольким обслуживающим устройствам, разобранную Кифером и Волфовицем (цит. в гл. VI, 9).

как обычно, что все рассматриваемые случайные величины независимы, мы получаем «вложенный» процесс восстановления с распределением промежутков между восстановлениями, равным $F=F_1*F_2$. Наша теорема применима, если $\mathbf{E}(\mathbf{X}_j)=\mu_1<\infty$ и $\mathbf{E}(\mathbf{Y}_j)=\mu_2<\infty$. Очевидно,

$$q_1(t)=1-F_1(t),$$

и потому при $t\to\infty$ вероятности состояний E_k стремятся к пределам

$$P_1(t)\to\frac{\mu_1}{\mu_1+\mu_2},\quad P_2(t)\to\frac{\mu_2}{\mu_1+\mu_2}. \tag{8.6}$$

Эти рассуждения легко переносятся на случай многостепенных систем.

в) Дифференциальные уравнения из **1**, гл. XVII соответствуют случайным процессам, в которых последовательные возвращения в *любое* состояние образуют процесс восстановления требуемого типа. Поэтому наша теорема гарантирует существование предельных вероятностей в этом случае. Их явный вид можно легко найти из уравнений, которым они должны удовлетворять. [См., например, **1**, гл. XVII, (7.8). Мы вернемся к более систематическому изложению этого вопроса в гл. XIV, 9. Точно такие же рассуждения применимы к полумарковским процессам, описанным в задаче 12 гл. XIV, 10.]

§9 *). Теория восстановления на всей прямой

В этом параграфе теория восстановления будет распространена на распределения, не сосредоточенные на полупрямой. Чтобы избежать тривиальностей, мы предположим, что $F\{-\infty, 0\}>0$, $F\{0, \infty\}>0$ и что F — *неарифметическое* распределение. Случай арифметических распределений требует видоизменений, которые производятся очевидным образом по аналогии с § 1.

Мы напомним (см. гл. VI, 10), что распределение F *невозвратно* тогда и только тогда, когда

$$U\{I\}=\sum_{n=0}^{\infty}F^{n*}\{I\} \tag{9.1}$$

принимает конечные значения для всех конечных интервалов [1]). Для таких распределений естественно возникает вопрос: переносятся ли на них теоремы восстановления § 1? Эта задача ин-

*) Материал не используется в дальнейшем.

[1]) В противном случае $U\{I\}=\infty$ для *каждого* интервала и F называется возвратным.

тересовала многих математиков, и, пожалуй, не из-за ее важности, а в основном по причине связанных с ней неожиданных трудностей. Так, теорема восстановления переносилась шаг за шагом на различные специальные классы невозвратных распределений Блекуэллом, Чжуном, Чжуном и Поллардом, Чжуном и Волфовицем, Карлином и Смитом. Однако общая теорема была доказана только в 1961 г. Феллером и Ореем с использованием как вероятностных приемов, так и анализа Фурье. Приводимое ниже доказательство значительно проще и более элементарно. Заметим, что если F имеет конечное математическое ожидание, то применимо без изменений доказательство § 1.

Пусть z — непрерывная функция, равная нулю вне некоторого конечного интервала $-h<x<h$. Тогда свертка $Z=U*z$ вполне определена равенством

$$Z(x)=\int_{-\infty}^{+\infty} z(x-y)\,U\{dy\}, \tag{9.2}$$

так как в действительности область интегрирования конечна. По теореме 2 гл. VI, 10 функция Z непрерывна и удовлетворяет *уравнению восстановления*

$$Z=z+F*Z. \tag{9.3}$$

По лемме § 2 решение (9.3) определено с точностью до аддитивной константы.

Теорема 1. *Если F имеет математическое ожидание $\mu>0$, то для всякого конечного интервала I длины $h>0$*

$$U\{I+t\}\to\frac{h}{\mu}, \qquad t\to\infty, \tag{9.4}$$

$$U\{I+t\}\to 0, \qquad t\to-\infty. \tag{9.5}$$

Эта теорема может быть легко переформулирована по очевидной аналогии с теоремой 2 § 1. Асимптотические оценки см. в задаче 13.

Доказательство. Доказательство теоремы 1 в § 1 начиналось со вспомогательного утверждения об ограниченности $U\{I+t\}$. Это делалось только ради полноты изложения, так как сам факт был известен из теоремы 2 гл. VI, 10. За исключением этого пункта, можно дословно повторить рассуждение § 1 с очевидной интерпретацией (1.13) как

$$z(x)=F^{0*}(x)-F(x). \tag{9.6}$$

Подстановка в (9.2) дает $Z(x)=F^{0*}(x)$ (т. е. 1 или 0, в зависимости от того, что $x\geqslant 0$ или $x<0$). ▶

Теорема 2. *Если F невозвратно и не имеет математического ожидания, то при $t\to\pm\infty$ $U\{I+t\}\to 0$ для каждого конечного интервала I.*

Доказательство этой теоремы более тонко, и мы предварим его изучением асимптотического поведения функции Z, определенной формулой (9.2).

Лемма. *Пусть функция $z\geqslant 0$ непрерывна и равна нулю для $|x|>h$. Тогда*[1]) *при фиксированном a*

$$Z(t+a)-Z(t)\to 0, \qquad t\to\pm\infty, \tag{9.7}$$

$$Z(t)Z(-t)\to 0, \qquad t\to\pm\infty. \tag{9.8}$$

Доказательство. Мы знаем из теоремы 2 гл. VI, 10, что функция Z равномерно непрерывна. Следовательно, по теореме о выборе (теорема 3 гл. VIII, 6) каждая последовательность чисел $t_k'\to\pm\infty$ содержит подпоследовательность $\{t_k\}$, такую, что $Z(t_k+x)\to\zeta(x)$, где ζ — ограниченная непрерывная функция. Из уравнения восстановления (9.3) вытекает, что $\zeta=F*\zeta$ и потому $\zeta=\text{const}$ (как показано в § 2). Отсюда следует (9.7), так как иначе мы могли бы выбрать $\{t_k\}$ так, что $\zeta(a)\neq\zeta(0)$.

Для доказательства (9.8) напомним, что $\liminf Z(x)=0$ (см. гл. VI, 10). По данному $\varepsilon>0$ выберем a так, что $Z(a)<\varepsilon$, и рассмотрим семейство функций V_τ, определенных равенством

$$V_\tau(x)=Z(x+a+\tau)-\gamma Z(x)Z(\tau), \tag{9.9}$$

где γ — такая постоянная, что $\gamma'=1-\gamma\cdot\|Z\|>0$. Ввиду равномерной непрерывности Z и ввиду (9.7) можно найти такое τ_0, что

$$Z(x+a+\tau)-Z(\tau)>-\varepsilon\gamma' \quad \text{при} \quad \tau>\tau_0, \quad |x|<h. \tag{9.10}$$

Функция V_τ удовлетворяет уравнению восстановления (9.3) с заменой $z(x)$ на $v(x)=z(x+a+\tau)-\gamma Z(\tau)z(x)$. Так как $v(x)\geqslant 0$ при $|x|>h$, то из теоремы 2[2]) гл. VI, 10 можно сделать следующий вывод: если функция V_τ принимает отрицательные

[1]) Легко видеть, что (9.8) эквивалентно соотношению $U\{I+t\}U\{I-t\}\to 0$. Обозначим $\rho(I)$ вероятность попадания в I при случайном блуждании $\{S_n\}$, управляемом функцией F. Тогда (9.8) эквивалентно также соотношению $\rho(I+t)\rho(I-t)\to 0$. Если бы это не имело места, то вероятность попадания в окрестность нуля *после захода* в $I+t$ не стремилась бы к нулю и распределение F не было бы невозвратным

[2]) Для упрощения формулировки эта теорема была дана для случая неотрицательных функций. Однако все рассуждения пригодны и для случая отрицательного минимума.

значения, то она достигает своего минимума в некоторой точке с $|x|\leqslant h$. В силу (9.10) при $\tau>\tau_0$

$$V_\tau(x) > -\varepsilon\gamma' + \gamma' \cdot Z(\tau). \tag{9.11}$$

Таким образом, при $\tau>\tau_0$ *или* $Z(\tau)<\varepsilon$, *или же* $V_\tau(x)\geqslant 0$ *при всех* x. В последнем случае для $x=-\tau$ получаем $\gamma Z(-\tau)Z(\tau)\leqslant Z(a)<\varepsilon$. В любом случае левая часть (9.8) будет $<\varepsilon/\gamma$. ▶

Доказательство *теоремы* 2. Из (9.8) следует

$$\limsup_{x\to\pm\infty} Z(x) = \limsup_{x\to\infty} |Z(x)+Z(-x)| = \eta. \tag{9.12}$$

Допустим, что $\eta>0$, поскольку в противном случае нечего доказывать. Рассмотрим свертки функций Z и z с равномерным распределением на $\overline{0,\ t}$, именно

$$W_t(x)=\frac{1}{t}\int_{x-t}^{x} Z(y)\,dy, \quad w_t(x)=\frac{1}{t}\int_{x-t}^{x} z(y)\,dy. \tag{9.13}$$

Наша ближайшая цель — показать, что при $t\to\infty$ должно иметь место одно из соотношений

$$W_t(t)=\frac{1}{t}\int_{0}^{t} Z(y)\,dy\to\eta \quad \text{или} \quad W_t(0)=\frac{1}{t}\int_{-t}^{0} Z(y)\,dy\to\eta. \tag{9.14}$$

В силу (9.7) верхние границы для $Z(x)$ и $W_t(x)$ (с фиксированным t) — одни и те же. Поэтому максимум W_t будет $\geqslant\eta$. С другой стороны, W_t удовлетворяет уравнению восстановления (9.3) с z, замененным на w_t. Как отмечалось раньше, отсюда вытекает, что максимум W_t достигается в точке, где w_t положительно, т. е. между $-h$ и $t+h$. Далее для $\frac{1}{2}t\leqslant x<t$

$$W_t(x)=\frac{1}{t}\int_{t-x}^{x} Z(y)\,dy+\frac{1}{t}\int_{0}^{t-x}[Z(y)+Z(-y)]\,dy. \tag{8.15}$$

Суммарная длина обоих интервалов интегрирования равна x. Из (9.12) поэтому следует, что при достаточно больших t $W_t(x)<\eta\frac{x}{t}+\varepsilon$. Таким образом, если $W_t(x)\geqslant\eta$, то точка x должна быть близка к t, и $W_t(t)$ должно быть близко к η. Если максимум W_t достигается в точке $x\leqslant\frac{1}{2}t$, то, как показывают аналогичные рассуждения, x должно быть близко к нулю, а $W_t(0)$ — к η.

Мы доказали, что при больших t или $W_t(t)$, или $W_t(0)$ должно быть близко к η. Но из (9.14) и (9.7) следует

$$\limsup [W_t(t) + W_t(0)] \leqslant \eta. \tag{9.16}$$

В силу непрерывности обеих функций должно быть или $W_t(t) \to \eta$ и $W_t(0) \to 0$, или же $W_t(t) \to 0$ и $W_t(0) \to \eta$.

По причине симметрии мы можем допустить, что $W_t(t) \to \eta > 0$. Вернемся к доказательству леммы и примем, что $\varepsilon > 0$ и $\tau_0 > 0$ — заданные числа. При достаточно больших t можно выбрать $\tau > \tau_0$ так, что и $Z(\tau) > \eta - \varepsilon$ и $Z(t-\tau) > \eta - \varepsilon$ [в противном случае было бы $W_t(t) < \eta - \varepsilon$]. По замечанию, следующему за формулой (9.11), функция V_τ положительна. Подставляя $x = t - \tau$ в (9.9), мы видим, что $Z(t+a) > \gamma(\eta - \varepsilon)^2$. Таким образом, $Z(t) > 0$ и $Z(-t) \to 0$ при $t \to \infty$. Для каждого ограниченного интервала I теперь получим

$$U\{I - t\} \to 0, \qquad t \to +\infty. \tag{9.17}$$

Для завершения доказательства мы воспользуемся результатом примера (3, б), касающимся вероятностей достижения состояний в случайном блуждании, управляемом функцией F. Обозначим $H(t, \xi)$ вероятность того, что первое попадание в $\overline{t, \infty}$ произойдет между t и $t+\xi$. Интервал $I+t$ по отношению к $t+x$ занимает такое же положение, как интервал $I - x$ по отношению к t. Поэтому

$$U\{I + t\} = \int_0^\infty H(t, d\xi)\, U\{I - \xi\}. \tag{9.18}$$

Возможны два случая. Если координаты точки первого попадания в $\overline{0, \infty}$ (т. е. первая верхняя лестничная высота) имеют бесконечное математическое ожидание, то $H(t, \xi) \to 0$ при $t \to \infty$ и каждом ξ. В этом случае в интеграле (9.18) главную роль играют только большие значения ξ и в силу (9.17) $U\{I+t\} \to 0$. В противном случае H сходится к некоторому распределению вероятностей и $U\{I+t\} \to \eta' > 0$. Но это невозможно, если F не имеет математического ожидания. В самом деле, рассматривая первый шаг в случайном блуждании, получим

$$1 - H(0, \xi) \geqslant 1 - F(\xi).$$

Следовательно, функция $1 - F(\xi)$ интегрируема на $\overline{0, \infty}$. Таким образом, F имеет математическое ожидание, равное $-\infty$ (в очевидном смысле слова). Поэтому случайное блуждание «уходит» в $-\infty$. Аналог (9.18) для отрицательной полуоси приводит теперь к очевидному противоречию, так как подинтегральное выражение стремится к η', в то время как левая часть стремится к нулю. ▶

§ 10. Задачи

(См. также задачи 8—15 в гл. VI, 13.)

1. Отбрасывая предположение $F(0)=0$, мы заменим F распределением $F^{\#}=pH_0+qF$, где H_0 сосредоточено в нуле и $p+q=1$. Тогда U заменяется на $U^{\#}=U/q$. Покажите, что с вероятностной точки зрения это очевидно. Проверьте утверждение, формально (а) вычисляя свертки (б), из уравнения восстановления.

2. Выведите из (3.4), что $Z(t)=V(t)-V(t-h)$ удовлетворяет стандартному уравнению восстановления с $z(t)=F_0(t)-F_0(t-h)$. Выведите непосредственно из теоремы 1.2, что $V(t)-V(t-h)\to h/\mu$.

3. *Совместное распределение для прошедшего и остающегося времени ожидания.* В обозначениях (3.7) докажите, что при $t\to\infty$

$$\mathbf{P}\left\{t-\mathbf{S}_{\mathbf{N}_t}>x,\ \mathbf{S}_{\mathbf{N}_t+1}-t>y\right\}\to\frac{1}{\mu}\int_{x+y}^{\infty}[1-F(s)]\,ds.$$

(*Указание.* Получите уравнение восстановления для левой части.)

4. *Стационарные свойства.* Рассмотрим процесс восстановления с задержкой, с начальным распределением F_0, задаваемым (3.10). Вероятность того, что n-й момент восстановления $\mathbf{S}_n$ лежит между t и $t+\xi$, равна $H^{\#}(t,\xi)=F_0(t+\xi)-F_0(t)+F_0*H(t,\xi)$. Докажите, что $H^{\#}(t,\xi)=F_0(\xi)$ тождественно. *Указание*: Проверьте дифференцированием, что $H^{\#}$ не зависит от t.

5. *Максимальное наблюденное время жизни.* Пусть $V(t,\xi)$ — вероятность того, что максимальный промежуток между восстановлениями, наблюденный до момента t в стандартном возвратном процессе восстановления, имеет длительность $>\xi$. Покажите, что

$$V(t,\xi)=1-F(\xi)+M\int_0^{\xi}V(t-y,\xi)\,F\{dy\}.$$

Изучите характер решения.

6. Пусть F — собственное распределение с математическим ожиданием $\mu<\infty$, и $Z=z+F*Z$. Если $z(t)\sim t^{\alpha}$ при $t\to\infty$ (с $\alpha>0$), то $Z(t)\sim t^{\alpha+1}/\mu(\alpha+1)$.

7. Если F — собственное распределение и a — постоянная, то интегро-дифференциальное уравнение $Z'=aZ-aZ*F$ сводится к

$$Z(t)=Z(0)+a\int_0^t Z(t-x)\,[1-F(x)]\,dx.$$

8. *Обобщенные счетчики второго типа.* Поступающие частицы образуют пуассоновский поток. Очередная n-я частица «закрывает» счетчик на время $\mathbf{T}_j$ и *аннулирует последействие* (если оно было) предыдущих частиц Предполагается, что случайные величины $\mathbf{T}_j$ независимы между собой, не зависят от поступающего потока и имеют одно и то же распределение G. Пусть $\mathbf{Y}$ —

длина промежутка времени, в течение которого счетчик закрыт, и $Z(t) = \mathbf{P}\{\mathbf{Y} > t\}$. Покажите, что $\mathbf{Y}$ — собственная случайная величина и

$$Z(t) = [1 - G(t)]\, e^{\alpha - t} + \int_0^t Z(t-x)\,[1 - G(x)]\,\alpha e^{-\alpha x}\,dx.$$

Покажите, что этот процесс восстановления будет обрывающимся в том и только том случае, когда G имеет математическое ожидание $\mu < \alpha^{-1}$. Обсудите применимость асимптотических оценок § 6.

9. *Эффект «островков безопасности».* [Пример (7, б).] Два независимых пуассоновских потока автомашин с одинаковыми плотностями движутся по двум сторонам магистрали. Время, необходимое для пересечения магистрали, равно 2ξ. Формулы (7.9) применимы и в этом случае. Наличие «островка безопасности» приводит к тому, что время, необходимое для пересечения магистрали, становится суммой двух независимых случайных величин с математическими ожиданиями и дисперсиями, определяемыми (7.9). Обсудите практические следствия этого факта.

10. Частицы, попадающие на счетчик, образуют возвратный процесс восстановления с распределением F. После каждой регистрации счетчик закрывается на фиксированное время ξ, в течение которого поступающие частицы не оказывают никакого воздействия на счетчик. Покажите, что распределение для промежутка времени между концом «периода нечувствительности» и моментом появления очередной частицы задается формулой

$$\int_0^{\xi} [F(\xi + t - y) - F(\xi - y)]\, U\{dy\}.$$

Если F — показательное распределение, то и это распределение будет показательным.

11. *Нелинейное восстановление.* Частица имеет показательное время жизни, в конце которого она с вероятностью p_k производит k точно таких же частиц, существующих независимо друг от друга ($k = 0, 1, \ldots$). Вероятность $F(t)$ того, что весь процесс оборвется до момента t, удовлетворяет уравнению

$$F(t) = p_0(1 - e^{-\alpha t}) + \sum_{n=1}^{\infty} p_k \int_0^t \alpha e^{-\alpha(t-x)} F^k(x)\,dx.$$

(Неизвестно никакого общего метода обращения с такими уравнениями.)

12. *Теорема восстановления в* $\mathscr{R}^2$. Пусть распределение пары $(\mathbf{X}, \mathbf{Y})$ сосредоточено в первом квадранте. Пусть I — интервал $0 \leqslant x, y \leqslant 1$. Обозначим $I + \mathbf{a}$ (где $\mathbf{a}$ — произвольный вектор) интервал, получаемый сдвигом I на $\mathbf{a}$. Тогда лемма 1.3 может быть обобщена следующим образом. Для любых фиксированных векторов $\mathbf{a}$ и $\mathbf{b}$

$$U\{I + \mathbf{a} + t\mathbf{b}\} - U\{I + t\mathbf{b}\} \to 0 \qquad (*)$$

при $t \to \infty$.

а) Считая это соотношение выполненным, покажите, что из теоремы восстановления для маргинальных распределений вытекает $U\{I + t\mathbf{b}\} \to 0$.

б) Покажите, что доказательство леммы 1.3 тривиальным образом применимо [1]).

13. Пусть F — произвольное распределение в $\mathscr{R}^1$ с математическим ожиданием $\mu>0$ и конечным вторым моментом m_2. Покажите, что

$$\sum_{n=0}^{\infty} F^{n*}(x) - \frac{x_+}{\mu} \to \frac{m_2}{2\mu^2},$$

где, как обычно, x_+ обозначает положительную часть x.

[*Указание.* Обозначим $Z(t)$ левую часть Тогда Z удовлетворяет стандартному уравнению восстановления с

$$z(x) = \begin{cases} \dfrac{1}{\mu}\displaystyle\int_{-\infty}^{t} F(x)\,dx, & t<0, \\ \dfrac{1}{\mu}\displaystyle\int_{t}^{\infty} (1-F(x))\,dx, & t>0. \end{cases}$$

Это замечание можно использовать для уточнения оценок, если существуют моменты высших порядков.]

---

[1]) Более подходящая формулировка задач теории восстановления для плоскости была недавно предложена в работе Bickel P. J., Yahav, Renewal theory in the plane, *Ann. Math. Statist.*, **36** (1965), 946—955. Эти авторы рассматривали математическое ожидание числа попаданий в область, заключенную между окружностями радиусов r и $r+a$ при $r\to\infty$.

ГЛАВА XII

СЛУЧАЙНЫЕ БЛУЖДАНИЯ в $\mathscr{R}$[1])

В этой главе рассматриваются задачи теории случайных блужданий с упором на комбинаторные методы и систематическое использование «лестничных величин». Некоторые из результатов будут выведены заново (и дополнены) в гл. XVIII методом Фурье. (В другом аспекте случайные блуждания изучались в гл. VI, 10.) Мы ограничимся в основном двумя центральными вопросами. Во-первых, будет показано, что полученные в **1**, гл. III любопытные результаты, относящиеся к бросанию монеты, переносятся на очень общий случай и что по существу при этом применимы те же самые методы. Второй вопрос связан со временем первого прохождения и задачами о разорении. Стало модным связывать подобные темы с известной теорией Винера—Хопфа. Однако эти связи не так тесны, как их обычно изображают. Они будут обсуждены в § 3а и гл. XVIII, 4.

Открытие Спарре-Андерсеном в 1949 г. мощных комбинаторных методов в теории флуктуаций представило всю теорию случайных блужданий в новом свете. С тех пор отмечается очень быстрый прогресс, одним из стимулов которого является неожиданное открытие близкой связи между случайными блужданиями и теорией очередей[1]).

Литература по этой теме обширна и порой способна запутать читателя. Теория, излагаемая на следующих страницах, так элементарна и проста, что непосвященный читатель никогда не заподозрил бы, какими трудными были эти задачи до того, как стала понята их истинная природа. Например, элементарные асимптотические оценки § 5 охватывают множество практических интересных результатов, которые ранее были получены глубокими методами и подчас с большой изобретательностью.

Параграфы 6—8 почти не зависят от первой части главы. Вряд ли нужно подчеркивать, что наше изложение является

---

[1]) Впервые такая связь была отмечена, по-видимому, Линдли (D. V. Lindley) в 1952 г. Он вывел интегральное уравнение, которое теперь отнесли бы к уравнениям типа Винера — Хопфа.

односторонним и оставляет в стороне ряд интересных аспектов теории случайных блужданий (таких, как связи с теорией потенциала или теорией групп) [1]).

§ 1. Обозначения и соглашения

Всюду в этой главе $\mathbf{X}_1, \mathbf{X}_2, \ldots$ суть независимые случайные величины с одним и тем же распределением F, таким, что $0<F(0)<1$ [случайные блуждания с $F(0)=0$ охватываются теорией восстановления]. Указанные величины порождают «случайное блуждание», начинающееся в нуле, т. е. последовательность случайных величин

$$\mathbf{S}_0=0, \quad \mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n. \tag{1.1}$$

Иногда мы будем рассматривать отрезок $(\mathbf{X}_{j+1}, \ldots, \mathbf{X}_k)$ данной последовательности $\{\mathbf{X}_j\}$. Ее частичные суммы $0, \mathbf{S}_{j+1}-\mathbf{S}_j, \ldots, \mathbf{S}_k-\mathbf{S}_j$ будут называться *отрезками случайного блуждания*. Индексы будут обычно интерпретироваться как значения временно́го параметра. Таким образом, момент n делит все случайное блуждание на *прошедший отрезок* и *остающийся отрезок*.

Лестничные величины были введены в гл. VI, 8. Однако теперь они будут определены заново. Определение основано на неравенствах и потому существуют четыре типа лестничных величин, соответствующих четырем возможностям, именно $<$, $\leqslant$, $>$, $\geqslant$. Это приводит к двойной классификации, описываемой само собой понятными терминами: *возрастающие* и *убывающие* [2]), *строгие* и *слабые*. Возрастающие и убывающие величины связаны столь же симметрично, как плюс и минус, максимум и минимум. Однако различие между строгими и слабыми величинами создает трудности в определениях и обозначениях. Простейший выход состоит в том, чтобы рассматривать только непрерывные распределения F, так как в этом случае строгие и слабые величины равны друг другу с вероятностью единица. Начинающим можно рекомендовать именно этот путь и не делать различия между строгими и слабыми лестничными величинами, хотя такое различие неизбежно в общей теории, с одной стороны, в таких примерах, как игра с бросанием монеты — с другой.

---

[1]) Другие аспекты представлены в книге Спицера (1964), хотя там рассматриваются только арифметические распределения. О комбинаторных методах, пригодных в многомерном случае, см.: Hobby C., Pyne R., Combinatorial results in multidimensional fluctuation theory, *Ann. Math. Statist.*, **34** (1963), 402–404.

[2]) «Ascending» и «descending». В том случае, когда речь будет идти о *точках*, мы будем применять термины «верхние» и «нижние». — *Прим. перев.*

Рассмотрим последовательность точек $(n, \mathbf{S}_n)$ для $n=1, 2, \ldots$ (нуль исключается). *Первая строго верхняя*[1] *лестничная точка* $(\mathcal{T}_1, \mathcal{H}_1)$ — *это первый член указанной последовательности, для которого* $\mathbf{S}_n>0$. Иными словами, $\mathcal{T}_1$ — это момент первого входа на положительную (открытую) полуось, определяемый соотношениями

$$\{\mathcal{T}_1=n\}=\{\mathbf{S}_1\leqslant 0, \ldots, \mathbf{S}_{n-1}\leqslant 0, \mathbf{S}_n>0\}, \tag{1.2}$$

а $\mathcal{H}_1=\mathbf{S}_{\mathcal{T}_1}$. Величина $\mathcal{T}_1$ называется первым *лестничным моментом*, а $\mathcal{H}_1$ — первой *лестничной высотой*. Эти случайные величины остаются *неопределенными*, если ни одно из событий (1.2) не наступает. Поэтому обе они могут быть не собственными[2]).

Для совместного распределения $(\mathcal{T}_1, \mathcal{H}_1)$ мы примем обозначение

$$\mathbf{P}\{\mathcal{T}_1=n, \mathcal{H}_1\leqslant x\}=H_n(x). \tag{1.3}$$

Тогда маргинальные распределения задаются равенствами

$$\mathbf{P}\{\mathcal{T}_1=n\}=H_n(\infty), \quad n=1, 2, \ldots \tag{1.4}$$

$$\mathbf{P}\{\mathcal{H}_1\leqslant x\}=\sum_{n=1}^{\infty} H_n(x)=H(x). \tag{1.5}$$

Обе случайные величины имеют один и тот же дефект, а именно $1-H(\infty)\geqslant 0$.

Отрезок случайного блуждания, следующий за первым лестничным моментом, является точной вероятностной копией всего случайного блуждания. Его первая лестничная точка является *второй* лестничной точкой всего случайного блуждания и обладает тем свойством, что

$$\mathbf{S}_n>\mathbf{S}_0, \ldots, \mathbf{S}_n>\mathbf{S}_{n-1}. \tag{1.6}$$

Так мы ее и назовем — второй лестничной точкой случайного блуждания. Она имеет вид $(\mathcal{T}_1+\mathcal{T}_2, \mathcal{H}_1+\mathcal{H}_2)$, где пары $(\mathcal{T}_1, \mathcal{H}_1)$ и $(\mathcal{T}_2, \mathcal{H}_2)$ независимы и одинаково распределены. Поступая аналогично, мы можем определить третью, четвертую и т. д. лестничные точки нашего случайного блуждания. Таким образом, *точка* $(n, \mathbf{S}_n)$ *является верхней лестничной точкой, если она удовлетворяет соотношению* (1.6). Лестничная точка с номером r (если она существует) имеет вид $(\mathcal{T}_1+\ldots+\mathcal{T}_r, \mathcal{H}_1+\ldots+\mathcal{H}_r)$, где пары $(\mathcal{T}_k, \mathcal{H}_k)$ взаимно независимы и имеют одно и то же распределение (1.3) [см. рис. 1 в гл. VI, 8].

[1]) См. предыдущую сноску.

[2]) Задачи 3—5 могут служить иллюстрацией, понятной без общей теории.

Для экономии обозначений мы не будем вводить никаких новых букв для сумм $\mathcal{T}_1+\ldots+\mathcal{T}_r$ и $\mathcal{H}_r+\ldots+\mathcal{H}_r$. Они образуют два (возможно, обрывающихся) *процесса восстановления с промежутками между восстановлениями* $\mathcal{T}_k$ *и* $\mathcal{H}_k$. В случайном блуждании, конечно, только $\mathcal{T}_k$ играет роль времени. Отметим, что сами лестничные точки образуют двумерный процесс восстановления.

Обозначим через ψ_0 *атомическое распределение с единичной массой* в нуле и через

$$\psi=\sum_{n=0}^{\infty}H^{n*} \tag{1.7}$$

меру, определяющую среднее число лестничных высот (здесь $H^{0*}=\psi_0$). Соответствующая несобственная функция распределения, задаваемая равенством $\psi(x)=\psi\{-\infty, x\}$, равна нулю при $x<0$, а для положительных x $\psi(x)$ *равна единице плюс среднее число лестничных точек в полосе* $\overline{0, x}$ (время не ограничивается). Мы знаем, что $\psi(x)<\infty$ при всех x и что в случае несобственных лестничных величин

$$\psi(\infty)=\sum_{n=0}^{\infty}H^n(\infty)=\frac{1}{1-H(\infty)}. \tag{1.8}$$

Верхние слабые лестничные точки определяются по (1.6) *с заменой знака* $>$ *на* $\geqslant$. Соответствующие величины будут помечены дополнительной верхней чертой. Так, $\overline{\mathcal{T}}_1$ — это наименьший индекс, такой что $\mathbf{S}_1<0, \ldots, \mathbf{S}_{n-1}<0$, но $\mathbf{S}_n\geqslant 0$. Как отмечалось раньше, утомительная необходимость различать строгие и слабые величины отпадает, если распределение F непрерывно. В общем случае легко выразить распределение $\overline{H}$ для слабых лестничных высот в терминах распределения H. Это дает нам возможность ограничиться единственным распределением, определенным в (1.5). Событие $\{\overline{\mathcal{H}}_1=0\}$ наступает тогда и только тогда, когда случайное блуждание возвращается в нуль без захода в $\overline{0, \infty}$, вероятность чего равна

$$\zeta=\sum_{n=1}^{\infty}\mathbf{P}\{\mathbf{S}_1<0, \ldots, \mathbf{S}_{n-1}<0, \mathbf{S}_n=0\}. \tag{1.9}$$

[Это событие не может произойти, если $\mathbf{X}_1>0$. Следовательно, $0\leqslant\zeta\leqslant F(0)<1$.] С вероятностью $1-\zeta$ первая строгая лестничная точка совпадает с первой слабой лестничной точкой. Следовательно,

$$\overline{H}=\zeta\psi_0+(1-\zeta)H. \tag{1.10}$$

Иначе говоря, распределение первой слабой лестничной точки является смесью распределения H и атомического распределения, сосредоточенного в нуле.

Вероятность того, что случайное блуждание вернется в нуль k раз до первого попадания в $\overline{0, \infty}$, равна $\zeta^k(1-\zeta)$. Математическое ожидание числа таких возвращений равно $1/(1-\zeta)$. Эта величина дает нам ожидаемую кратность каждой лестничной высоты (т. е. число ее появлений до следующей лестничной высоты). Следовательно,

$$\bar{\psi} = \frac{1}{1-\zeta}\psi \tag{1.11}$$

(см. задачу 7). Простота этих соотношений позволяет нам избежать явного использования распределения $\overline{H}$.

В определении возрастающих лестничных величин было выбрано положительное направление. *Строгие и слабые убывающие лестничные величины определяются по симметрии*, т. е. с помощью замены знака $>$ на $<$. В тех редких случаях, когда потребуется специальное обозначение, мы будем помечать убывающие величины верхним индексом «минус». Так, первая строгая нижняя лестничная точка будет $(\bar{\mathcal{T}}_1, \bar{\mathscr{H}}_1)$ и т. д.

Начиная с этого места, мы будем, как правило, рассматривать строгие возрастающие лестничные величины и всюду, где не будет опасности путаницы, мы будем отбрасывать прилагательные «строгие» и «возрастающие».

Пример. *Простое случайное блуждание.* В случайных блужданиях, изучавшихся в первом томе, распределение F было атомическим и приписывало веса p и q точкам 1 и -1 соответственно. Лестничные моменты $\mathcal{T}_k$ были введены в примере **1**, гл. XIII, (1. г), однако лестничные высоты не отмечались ввиду их тривиального характера: ордината r-й лестничной точки (если она существует) обязательно равна r. Единственной неизвестной величиной, связанной с распределением H лестничных высот, является вес, который оно приписывает единственному атому в точке 1. Но x равно вероятности того, что при некотором n $\mathbf{S}_n = 1$. Эта вероятность была вычислена в **1**, гл. XIII и XIV. Нижеследующее вычисление проведено в духе выкладок этой главы.

Обозначим через x (условную) вероятность возвращения в нуль, если первый шаг приводит в -1. В момент когда такое возвращение произошло, условная вероятность попасть в 1 снова будет равна x. Следовательно, x должно удовлетворять квадратному уравнению

$$x = p + qx^2 \tag{1.12}$$

и, как легко видеть, нам нужен *меньший* [1]) корень этого уравнения. Таким образом, $x=1$, если $p\geqslant q$, и $x=p/q$, если $p\leqslant q$. Величина $1-x$ равна дефекту строгих лестничных величин.

Перейдем к *слабой лестничной* высоте $\overline{\mathscr{H}}_1$. Ясно, что вероятность возвращения в нуль без предварительного попадания в $\overline{0,\ \infty}$ равна $\zeta=qx$. Таким образом, $\mathbf{P}\{\overline{\mathscr{H}}_1=0\}=q$ при $p\geqslant q$ и $\mathbf{P}\{\overline{\mathscr{H}}_1\}=p$ при $p\leqslant q$. Событие $\overline{\mathscr{H}}=1$ происходит тогда и только тогда, когда первый шаг приводит в точку 1, так что $\mathbf{P}\{\overline{\mathscr{H}}_1=1\}=p$ при всех обстоятельствах. ▶

§ 2. Двойственность

В 1, гл. III мы установили любопытные свойства флуктуаций, возникающих при бросании монеты. При этом были использованы простые комбинаторные рассуждения, основанные на рассмотрении величин $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, взятых в обратном порядке. Эти же приемы приведут нас теперь к важным и очень общим результатам.

При фиксированном n вы вводим n новых случайных величин равенствами $\mathbf{X}_1^*=\mathbf{X}_n, \ldots, \mathbf{X}_n^*=\mathbf{X}_1$. Соответствующие частные суммы равны $\mathbf{S}_k^*=\mathbf{S}_n-\mathbf{S}_{n-k}$, где $k=0, \ldots, n$. Совместные распределения величин $(\mathbf{S}_0, \ldots, \mathbf{S}_n)$ и $(\mathbf{S}_0^*, \ldots, \mathbf{S}_n^*)$ совпадают. Поэтому отображение $\mathbf{X}_k\to\mathbf{X}_k^*$ отображает любое событие A, связанное с $(\mathbf{S}_0, \ldots, \mathbf{S}_n)$, в событие A^* равной вероятности. Это отображение легко представить себе наглядно, поскольку графики последовательностей $(0, \mathbf{S}_1, \ldots, \mathbf{S}_n)$ и $(0, \mathbf{S}_1^*, \ldots, \mathbf{S}_n^*)$ получаются один из другого поворотом на 180°.

Пример. а) Двойственным к событию $\{\mathbf{S}_1<0, \ldots, \mathbf{S}_{n-1}<0, \mathbf{S}_n=0\}$ будет событие $\{\mathbf{S}_1>0, \ldots, \mathbf{S}_{n-1}>0, \mathbf{S}_n=0\}$. Отсюда следует, что возвращение в нуль без захода в $\overline{-\infty, 0}$ имеет ту же самую вероятность, что и возвращение в нуль без захода в $\overline{0, \infty}$. Иными словами, вероятность ζ, определенная в (1.9), не изменяется при изменении всех знаков неравенств на противоположные. Таким образом, в соотношения, двойственные к (1.10) и (1.11) для убывающих лестничных величин, входит то же самое число ζ. ▶

---

[1]) Если x_k — вероятность того, что $\mathbf{S}_n=1$ при некотором $n\leqslant k$, то $x_1=p$ и $x_{k+1}\leqslant p+qx_k^2$. По индукции заключаем, что x_k не превосходит меньшего корня уравнения (1.12). Подобное рассуждение типично для случаев, где нет единственности.

В этом параграфе мы изучим следствия применения указанной выше процедуры «обращения» к событию (1.6), определяющему (строгую верхнюю) лестничную точку. Двойственное событие определяется неравенствами $\mathbf{S}_n^* > \mathbf{S}_{n-k}^*$, $k = 1, \ldots, n$, т. е. совпадает с событием $\mathbf{S}_k > 0$. Таким образом, для всякого интервала $I \subset \overline{0, \infty}$

$$\mathbf{P}\{\mathbf{S}_n > \mathbf{S}_j \text{ при } j = 0, \ldots, n-1 \text{ и } \mathbf{S}_n \in I\} = \\ = \mathbf{P}\{\mathbf{S}_j > 0 \text{ при } j = 1, \ldots, n \text{ и } \mathbf{S}_n \in I\}. \quad (2.1)$$

Левая часть равна вероятности того, что найдется лестничная точка с абсциссой n и ординатой, принадлежащей I. Правая часть равна вероятности того, что в момент n произойдет попадание в I и при этом до момента n не будет заходов на замкнутую полупрямую $\overline{-\infty, 0}$. Определим случайную величину $\mathbf{Y}_n$ по правилу: $\mathbf{Y}_n = 1$, если это событие наступает, и $\mathbf{Y}_n = 0$ в противоположном случае. Тогда $\mathbf{Y} = \Sigma \mathbf{Y}_n \leqslant \infty$ представляет собой число попаданий в I до первого попадания в $\overline{-\infty, 0}$. То, что эта величина конечна, никоим образом не очевидно. Однако, суммируя равенство (2.1) по всем n, мы получим справа $\mathbf{E}(\mathbf{Y})$ и слева $\psi\{I\}$ [где функция $\psi\{I\}$ определяется в соответствии с (1.7)]. Но $\psi\{I\}$ конечна для ограниченных I. Нами установлена тем самым основная

Лемма о двойственности. *Мера* ψ, *задаваемая равенством* (1.7), *допускает две интерпретации. Пусть* $I \subset \overline{0, \infty}$, *тогда:*

а) $\psi\{I\}$ *есть математическое ожидание числа лестничных точек в* I.

б) $\psi\{I\}$ *есть математическое ожидание числа попаданий в* I *до первого захода в* $\overline{-\infty, 0}$.

Пример. б) *Простое случайное блуждание.* В случайном блуждании, описанном в примере § 1, лестничная точка с ординатой k существует тогда и только тогда, когда при некотором n наступает событие $\{\mathbf{S}_n = k\}$, вероятность чего равна 1 при $p \geqslant q$ и $(p/q)^k$ при $p \leqslant q$. По лемме о двойственности это означает, что *в симметричном случайном блуждании математическое ожидание числа попаданий в состояние* $k \geqslant 1$ *до первого возвращения в нуль равно единице при всех* k. Фантастичность этого результата можно лучше уяснить в терминологии игры с бросанием монеты. Наше утверждение состоит в том, что *до первого возвращения на нулевой уровень средний накопленный выигрыш Петра однажды принимает любое значение* k. Это утверждение обычно порождает недоверие, однако его можно проверить не-

посредственным вычислением (см. задачу 1). (Наш прежний вывод о бесконечности математического ожидания времени до первого возвращения в 0 получается теперь суммированием по k.) ▶

Тождество (2.1) выполняется и при замене строгих неравенств слабыми. Для $I=\overline{0, \infty}$ получаем

$$\mathbf{P}\{\mathbf{S}_n \geqslant \mathbf{S}_0, \ldots, \mathbf{S}_n \geqslant \mathbf{S}_{n-1}\} = \mathbf{P}\{\mathbf{S}_1 \geqslant 0, \ldots, \mathbf{S}_n \geqslant 0\}. \quad (2.2)$$

Левая часть есть просто вероятность того, что $(n, \mathbf{S}_n)$ будет слабой верхней лестничной точкой. Правая часть равна вероятности того, что первое попадание в $\overline{-\infty, 0}$ не происходит до момента n, т. е. равна $1-\mathbf{P}\{\mathcal{T}_1^- \leqslant n\}$. Если случайная величина $\mathcal{T}_1^-$ — собственная, то сумма этих вероятностей равна математическому ожиданию $\mathbf{E}(\mathcal{T}_1^-)$, которое может быть как конечным, так и бесконечным [см. теорему 2 из 1, гл. XI, 1 или гл. V, (6.3)]. Сумма вероятностей, стоящих в левой части (2.2), равна математическому ожиданию числа слабых верхних точек на $\overline{0, \infty}$, т. е. равна $\bar{\varphi}(\infty)$. Отсюда в силу (1.8) и (1.11) для случая собственной величины $\mathcal{T}_1^-$ получаем

$$\mathbf{E}(\mathcal{T}_1^-) = \frac{1}{1-\zeta}\psi(\infty) = \frac{1}{(1-\zeta)(1-H(\infty))}. \quad (2.3)$$

Интерпретация (6.3) в случае $H(\infty)=1$ очевидна.

Эта формула приводит к некоторым важным заключениям. Вспомним, что $0 \leqslant \zeta < 1$ (и что $\zeta=0$ для непрерывной F). Мы видим, что $\mathbf{E}(\mathcal{T}_1^-) < \infty$ в том и только том случае, когда $H(\infty)<1$, т. е. в случае, когда возрастающая лестничная величина $\mathcal{T}_1$ является несобственной. Таким образом, или одна из этих величин будет несобственной, или

$$\mathbf{E}(\mathcal{T}_1) = \infty, \quad \mathbf{E}(\mathcal{T}_1^-) = \infty. \quad (2.4)$$

Если $\mathcal{T}_1^-$ — собственная величина и $\mathbf{E}(\mathcal{T}_1^-) < \infty$, то возрастающий процесс восстановления будет обрывающимся. С вероятностью единица существует последняя лестничная точка и максимум

$$\mathbf{M} = \max\{\mathbf{S}_0, \mathbf{S}_1, \ldots\} \quad (2.5)$$

конечен. При условии, что n-я лестничная точка существует, вероятность ей быть последней равна $1-H(\infty)$. Поэтому [см. гл. XI, (6.3)]

$$\mathbf{P}\{\mathbf{M} \leqslant x\} = [1-H(\infty)] \sum_{n=0}^{\infty} H^{n*}(x) = [1-H(\infty)]\psi(x). \quad (2.6)$$

Рассуждения, предшествующие (2.3), показывают, что в случае несобственной величины $\mathcal{J}_1^-$ сумма вероятностей (2.2) равна бесконечности и, следовательно, $H(\infty)=1$. Невозможно, стало быть, чтобы и возрастающий, и убывающий лестничные процессы были обрывающимися. Таким образом, нами установлена важная

Теорема 1. *Существуют только два типа случайных блужданий.*

(i) *Осциллирующий тип: возрастающий и убывающий процессы восстановления возвратны;* $\mathbf{S}_n$ *колеблется с вероятностью* 1 *между* $-\infty$ *и* ∞*; имеет место* (2.4).

(ii) *Уходящий тип (скажем, в* $-\infty$*). Возрастающий процесс восстановления обрывается; убывающий процесс является собственным; с вероятностью единица* $\mathbf{S}_n$ *стремится к* $-\infty$ *и достигает конечного максимума* $\mathbf{M}\geqslant 0$*; выполняются соотношения* (2.3) *и* (2.6).

[Блуждания типа (ii) очевидным образом невозвратны, но в числе блужданий типа (i) имеются как возвратные, так и невозвратные; см. конец гл. VI, 10.]

Рассмотрим теперь математическое ожидание лестничной высоты $\mathcal{H}_1$. Начнем с простого замечания, а именно *если величина* $\mathcal{H}_1$ *— собственная и* $\mathbf{E}(\mathcal{H}_1)<\infty$, то $\mathbf{X}_1$ *имеет математическое ожидание* $\mathbf{E}(\mathbf{X}_1)\geqslant 0$. В самом деле, возьмем обычное разложение $\mathbf{X}_n=\mathbf{X}_n^+-\mathbf{X}_n^-$ случайной величины $\mathbf{X}_n$ на положительную и отрицательную части. Ограничиваясь первым шагом случайного блуждания, видим, что $\mathbf{P}\{\mathcal{H}_1>x\}\geqslant\mathbf{P}\{\mathbf{X}_1>x\}$ и потому $\mathbf{E}(\mathbf{X}_1^+)\leqslant\mathbf{E}(\mathcal{H}_1)<\infty$ [см. гл. V, (6.3)]. Из того что величина $\mathcal{H}_1$ собственная, вытекает, что с вероятностью единица при бесконечно многих n суммы $\mathbf{S}_n$ будут положительными. Следовательно, для бесконечно многих n

$$n^{-1}(\mathbf{X}_1^-+\ldots+\mathbf{X}_n^-)<n^{-1}(\mathbf{X}_1^++\ldots+\mathbf{X}_n^+).$$

В силу усиленного закона больших чисел [1]) правая часть остается ограниченной с вероятностью единица, в то время как левая часть сходится к $\mathbf{E}(\mathbf{X}_1^-)$. То есть это математическое ожидание не превосходит $\mathbf{E}(\mathbf{X}_1^+)$, так что $\mathbf{X}_1$ в самом деле имеет конечное неотрицательное математическое ожидание.

[1]) См. теорему 2 гл. VII, 7. Этот § 7 помечен звездочкой, и приводимое здесь доказательство может быть опущено, поскольку мы дадим в свое время и аналитическое доказательство нижеследующей теоремы. Отметим попутно, что к положительным случайным величинам усиленный закон больших чисел применим и при бесконечном математическом ожидании, что можно показать с помощью обычного урезания.

Переходя к обратному утверждению, допустим сначала, что $0<\mathbf{E}(\mathbf{X}_1)<\infty$. Мы покажем, что в этом случае $\mathbf{E}(\mathscr{H}_1)<\infty$ и

$$\mathbf{E}(\mathscr{H}_1)=\mathbf{E}(\mathscr{T}_1)\,\mathbf{E}(\mathbf{X}_1). \tag{2.7}$$

Из (2.3), меняя ролями положительные и отрицательные значения, получаем $\mathbf{E}(\mathscr{T}_1)<\infty$. Рассмотрим теперь подпоследовательность последовательности $\{\mathbf{S}_n/n\}$, образованную теми n, которые соответствуют лестничным моментам. Член этой подпоследовательности с номером k равен

$$\frac{\mathscr{H}_1+\cdots+\mathscr{H}_k}{\mathscr{T}_1+\cdots+\mathscr{T}_k}, \tag{2.8}$$

так что при $k\to\infty$ это отношение с вероятностью единица сходится к $\mathbf{E}(\mathbf{X}_1)$. Деля числитель и знаменатель на k, видим, что отношение (2.8) сходится и к $\mathbf{E}(\mathscr{H}_1)/\mathbf{E}(\mathscr{T}_1)$. Поэтому если $\mathbf{E}(\mathbf{X}_1)$ существует и положительно, то (2.7) верно.

Такие же рассуждения показывают, что из $\mathbf{E}(\mathbf{X}_1)=0$ вытекает $\mathbf{E}(\mathscr{T}_1)=\infty$ (однако $\mathscr{H}_1$ может иметь при этом как конечное, так и бесконечное математическое ожидание; см. задачу 11). По теореме 1 случайное блуждание будет осциллирующим [1]). Таким образом, доказана

Теорема 2. *Если* $\mathbf{E}(\mathbf{X}_1)$ *конечно и положительно, то* $\mathscr{H}_1$ *и* $\mathscr{T}_1$ *— собственные величины с конечными математическими ожиданиями. При этом верно* (2.7).

Если $\mathbf{E}(\mathbf{X}_1)=0$, *то* $\mathscr{H}_1$ *и* $\mathscr{T}_1$ *— собственные величины и* $\mathbf{E}(\mathscr{T}_1)=\infty$.

В других случаях или блуждание уходит в $-\infty$ *(тогда* $\mathscr{T}_1$ *и* $\mathscr{H}_1$ *— несобственные величины) или* $\mathbf{E}(\mathscr{T}_1)=\infty$ *и* $\mathbf{E}(\mathscr{H}_1)=\infty$.

Справедливость (2.7) была впервые доказана А. Вальдом в значительно более общей обстановке (в связи с последовательным анализом). Об этом будет идти речь в гл. XVIII, 2. Мы увидим несколько аналитических доказательств [2]) (2.7) (см. задачи 10—11 и 20 и гл. XVIII, 2 и 4).

---

[1]) Теорема 4 гл. VI. 10 содержит более сильный результат, а именно что при $\mathbf{E}(\mathbf{X}_1)=0$ случайное блуждание возвратно.

[2]) Следующее простое доказательство основано на привлечении *мартингалов*. Пусть $\mathbf{E}(\mathbf{X}_1)=\mu>0$. Случайные величины $\mathbf{S}_n-n\mu$ образуют мартингалы. Свойство мартингалов сохраняется и при «правиле остановки», в соответствии с которым процесс останавливается при первом входе случайного блуждания в $\overline{0,\infty}$ (см. следствие в гл. VI, 12). Самая последняя случайная величина равна при этом $\mathscr{H}_1-\mathscr{T}_1\mu$ и имеет, как можно показать, нулевое математическое ожидание. Из теоремы 1 мы знаем, что $\mathbf{E}(\mathscr{T}_1)<\infty$, так что (2.7) верно

§ 3. Распределение лестничных высот. Факторизация Винера — Хопфа

Вычисление распределений лестничных высот, т. е. H и $\overline{H}$, на первый взгляд кажется устрашающей проблемой, и именно так ее первоначально воспринимали. Однако лемма о двойственности приводит к простому решению. Идея состоит в том, что первый вход, скажем, в $\overline{-\infty, 0}$ следует рассматривать вместе с предшествующим ему отрезком случайного блуждания. Мы приходим к изучению видоизмененного случайного блуждания $\{\mathbf{S}_n\}$, обрывающегося в момент первого входа в $\overline{-\infty, 0}$. Обозначим через ψ_n несобственное распределение вероятностей для положения в момент n в этом ограниченном случайном блуждании. Иными словами, для любого интервала I и $n=1, 2, \ldots$ положим

$$\psi_n\{I\} = \mathbf{P}\{\mathbf{S}_1 > 0, \ldots, \mathbf{S}_n > 0, \mathbf{S}_n \in I\} \tag{3.1}$$

(заметим, что отсюда следует равенство $\psi_n\{\overline{-\infty, 0}\}=0$). Как и прежде, ψ_0 обозначает распределение вероятностей, сосредоточенное в нуле. Формула (2.1) показывает, что $\psi_n\{I\}$ равна вероятности того, что $(n, \mathbf{S}_n)$ есть лестничная точка с $\mathbf{S}_n \in I$. Суммируя по всем n, получаем

$$\psi\{I\} = \sum_{n=0}^{\infty} \psi_n\{I\}, \tag{3.2}$$

где ψ — «функция восстановления», введенная в (1.7). Другими словами, для любого интервала I, расположенного на открытой положительной полуоси, $\psi\{I\}$ равно математическому ожиданию числа (строго верхних) лестничных точек, ординаты которых принадлежат I. Положим для интервалов I отрицательной полуоси $\psi\{I\}=0$. Заметим, что ряд (3.2) сходится для любого *ограниченного* интервала I (но может расходиться для $I=\overline{0, \infty}$). Этот неожиданный результат [1]) делает всю последующую теорию невероятно простой.

Изучение первого входа в $\overline{-\infty, 0}$ — это изучение слабо убывающего лестничного процесса и (в обозначениях § 1) точки первого входа $\mathscr{H}_1^-$ и ее распределения $\overline{H}^-$. Для облегчения типографского набора заменим $\overline{H}^-$ на ρ. *Вероятность того, что первый вход в* $\overline{-\infty, 0}$ *происходит в момент* n и внутри интервала I,

[1]) В вероятностных терминах это означает, что в *любом* случайном блуждании для *любого ограниченного* интервала I математическое ожидание числа попаданий в I до первого входа в $\overline{-\infty, 0}$ является конечным.

обозначим через $\rho_n\{I\}$, т. е. при $n=1, 2, \ldots$

$$\rho_n\{I\} = \mathbf{P}\{\mathbf{S}_1 > 0, \ldots, \mathbf{S}_{n-1} > 0, \mathbf{S}_n \leqslant 0, \mathbf{S}_n \in I\} \tag{3.3}$$

(отсюда вытекает, что $\rho_n\{\overline{0, \infty}\}=0$; ρ_0 не определено). На этот раз очевидно, что ряд

$$\rho\{I\} = \sum_{n=1}^{\infty} \rho_n\{I\} \tag{3.4}$$

сходится и определяет (возможно, несобственное) распределение точки первого входа (т. е. $\overline{\mathscr{H}_1^-}$).

Легко вывести рекуррентные соотношения для ψ_n и ρ_n. В самом деле, условная вероятность события $\mathbf{S}_{n+1} \in I$ при условии, что $\mathbf{S}_n = y$, равна $F\{I-y\}$, где $I-y$ — интервал, получаемый сдвигом I на $-y$. Таким образом,

$$\rho_{n+1}\{I\} = \int_{0-}^{\infty} \psi_n\{dy\}\, F\{I-y\}, \quad \text{если} \quad I \subset \overline{-\infty, 0}, \tag{3.5а}$$

$$\psi_{n+1}\{I\} = \int_{0-}^{\infty} \psi_n\{dy\}\, F\{I-y\}, \quad \text{если} \quad I \subset \overline{0, \infty} \tag{3.5б}$$

(нуль дает вклад в интеграл только при $n=0$). Для ограниченных интервалов I сходимость $\sum \psi_n\{I\}$ обеспечивается леммой о двойственности, а $\sum \rho_n\{I\}$ всегда сходится к числу $\leqslant 1$. Суммы этих рядов равны ψ и ρ, и для них мы получаем

$$\rho\{I\} = \int_{0-}^{\infty} \psi\{dy\}\, F\{I-y\}, \quad \text{если} \quad I \subset \overline{-\infty, 0}, \tag{3.6а}$$

$$\psi\{I\} = \int_{0-}^{\infty} \psi\{dy\}\, F\{I-y\}, \quad \text{если} \quad I \subset \overline{0, \infty} \tag{3.6б}$$

с тем соглашением, что в (3.6б) рассматриваются только ограниченные интервалы I. Мы увидим, что соотношения (3.6) практически более полезны, чем представления ρ и ψ рядами. Иногда удобно заменить функции интервала ρ и ψ соответствующими функциями точки

$$\rho(x) = \rho\{\overline{-\infty, x}\} \quad \text{и} \quad \psi(x) = \psi\{\overline{-\infty, x}\}.$$

Очевидно, что (3.6а) равносильно соотношению

$$\rho(x) = \int_{0-}^{\infty} \psi\{dy\}\, F(x-y), \quad x \leqslant 0. \tag{3.7а}$$

Из (3.6б) при $x>0$ получаем

$$\psi(x)=1+\psi\{\overline{0,\ x}\}=1+\int_{0-}^{\infty}\psi\{dy\}\,[F(x-y)-F(-y)].$$

Принимая во внимание (3.7а), мы видим, что (3.6б) эквивалентно

$$\psi(x)=1-\rho(0)+\int_{0-}^{\infty}\psi\{dy\}\,F(x-y),\quad x\geqslant 0. \tag{3.7б}$$

Для упрощения обозначений введем свертку

$$\psi*F=\sum_{n=0}^{\infty}\psi_n*F. \tag{3.8}$$

Так как ψ сосредоточено на $\overline{0,\infty}$, то значение $\psi*F\{I\}$ равно сумме двух интегралов (3.6) и потому конечно. Вспоминая, что ψ имеет единичный атом в нуле, соединим два соотношения (3.6) в одно уравнение свертки

$$\rho+\psi=\psi_0+\psi*F. \tag{3.9}$$

Ввиду того что ρ и $\psi-\psi_0$ сосредоточены на $\overline{-\infty,0}$ и $\overline{0,\infty}$ соответственно, соотношение (3.9) эквивалентно паре соотношений (3.6).

Мы будем рассматривать (3.9) как интегральное уравнение, определяющее неизвестные меры ρ и ψ. Большое число важных теоретических выводов можно получить непосредственно из (3.9). Мы отметим самую замечательную из подобных теорем в форме примера, для того чтобы подчеркнуть, что она не используется в дальнейшем и представляет собой отклонение от основного изложения.

Примеры. а) *Факторизация типа Винера — Хопфа.* Из определения (1.7) функции ψ следует, что она удовлетворяет уравнению восстановления

$$\psi=\psi_0+\psi*H. \tag{3.10}$$

Возьмем свертку (3.9) с ψ_0-H. Член $\psi*H*F$ можно упростить, используя (3.10). Тогда получим

$$F=H+\rho-H*\rho. \tag{3.11}$$

Это тождество равносильно (3.9). Однако оно замечательно тем, что дает представление произвольной функции F в терминах двух (возможно, несобственных) распределений вероятностей,

сосредоточенных на $\overline{-\infty, 0}$ и $\overline{0, \infty}$ соответственно. Уравнение обладает небольшой асимметрией, так как H есть распределение точки первого входа в открытый интервал $\overline{0, \infty}$, в то время как ρ есть аналогичное распределение для замкнутого интервала $\overline{-\infty, 0}$. Но мы знаем из примера (2.а) и из (1.10), что $\rho = \zeta\psi_0 + (1-\zeta)H^-$, где ζ определяется по (1.9). Подставляя это выражение в (3.11), после тривиальных преобразований получаем

$$\psi_0 - F = (1-\zeta)[\psi_0 - H] * [\psi_0 - H^-]. \tag{3.12}$$

Различные варианты этой формулы были открыты независимо один от другого различными методами, и все они вызвали восхищение. Одно видоизменение см. в задаче 19, другое в терминах преобразований Фурье дано в гл. XVIII, 3. Связь с методами Винера — Хопфа обсуждается в § 3а.

б) *Тождество Вальда* (2.7), доказанное в теореме 2.2, является простым следствием (3.11). См. задачи 10 и 11 и гл. XVIII, 2. ▶

Обратимся теперь к (3.9) как к интегральному уравнению относительно неизвестных мер ρ и ψ. Мы покажем, что его решение единственно, если принять во внимание свойства, которые ρ и ψ обязаны иметь в этом контексте. Условимся говорить для краткости, что пара (ρ, ψ) имеет *вероятностный смысл*, если ρ — распределение вероятностей (возможно, несобственное), сосредоточенное на $\overline{0, \infty}$, а $\psi - \psi_0$ — мера, сосредоточенная на $\overline{0, \infty}$ и такая, что для каждого интервала I значения $\psi\{I+t\}$ ограничены (последнее условие вытекает из теорем восстановления, так как $\psi = \sum H^{n*}$).

Теорема 1. *Уравнение свертки* (3.9) [*или равносильная ему пара уравнений* (3.6)] *обладает только одним решением* (ρ, ψ), *имеющим вероятностный смысл* [1]).

Доказательство. Пусть $\rho^{\#}$ и $\psi^{\#}$ — две неотрицательные меры, удовлетворяющие (3.6), и $\psi^{\#} \geqslant \psi_0$. Из (3.6б) по индукции получаем $\psi^{\#} \geqslant \psi_0 + \ldots + \psi_n$ при любом n. Поэтому наше решение ψ *минимально* в том смысле, что для любого другого решения $\psi^{\#}$ с единичным атомом в нуле $\psi^{\#}\{I\} \geqslant \psi\{I\}$ для всех интервалов. Другими словами, $\delta = \psi^{\#} - \psi$ есть *мера*. Из

[1]) Отсюда вытекает, что ρ есть распределение точки первого входа в $\overline{-\infty, 0}$ и $\psi = \Sigma H^{n*}$, где H — распределение точки первого входа в $\overline{0, \infty}$.

(3.6а) видно, что это же верно и относительно $\gamma=\rho^{\sharp}-\rho$. Так как и (ρ, ψ), и $(\rho^{\sharp}, \varphi^{\sharp})$ удовлетворяют (3.9), мы имеем

$$\delta+\gamma=\delta * F. \tag{3.13}$$

Теперь возможны два случая. Если ρ — собственное распределение, то из его минимальности вытекает, что $\rho^{\sharp}=\rho$ и, следовательно, $\gamma=0$. Положим $z(t)=\delta(I+t)$, где I — фиксированный конечный интервал. Тогда функция z будет ограниченным решением уравнения свертки $z=F*z$ и по лемме гл. XI, 2 должно быть $z=\text{const}$. Но при $t\to-\infty$ функция z должна стремиться к нулю. Поэтому $z(t)=0$ при всех t. Если ρ — несобственное распределение, то мы можем утверждать лишь, что $z\leqslant F*z$. В этом случае по индукции получаем

$$z(t)\leqslant \int_{-\infty}^{\infty} z(t-y)\,F^{n*}\{dy\} \tag{3.14}$$

при всех n. Так как распределение ρ — несобственное, то случайное блуждание уходит к ∞ и потому вся масса F^{n*} сосредоточивается вблизи $+\infty$. Но $z(t-y)\to 0$ при $y\to\infty$, следовательно, снова $z(t)=0$. Доказательство закончено. ▶

3а. Интегральное уравнение Винера — Хопфа

Связь между интегральным уравнением (3.9) и обычным уравнением Винера — Хопфа можно лучше всего объяснить, начав с вероятностной задачи, в которой встречается последнее уравнение.

Пример. с) *Распределение максимума.* Предположим для простоты, что распределение F имеет плотность и отрицательное математическое ожидание. Случайное блуждание $\{\mathbf{S}_n\}$ уходит в $-\infty$ и с вероятностью 1 определена конечнозначная случайная величина

$$\mathbf{M}=\max[0,\ \mathbf{S}_1,\ \mathbf{S}_2,\ \ldots]. \tag{3.15}$$

Найдем ее распределение вероятностей $M(x)=\mathbf{P}\{\mathbf{M}\leqslant x\}$, которое по определению сосредоточено на $\overline{0,\ \infty}$. Событие $\{\mathbf{M}\leqslant x\}$ происходит тогда и только тогда, когда $\mathbf{X}_1=y\leqslant x$ и $\max[0,\ \mathbf{X}_1,\ \mathbf{X}_2+\mathbf{X}_3,\ \ldots]\leqslant y-x$. Суммируя по всем возможным значениям y, получаем

$$M(x)=\int_{-\infty}^{x} M(x-y)\,f(y)\,dy,\quad x>0, \tag{3.16}$$

или, что то же самое,

$$M(x)=\int_0^\infty M(s)f(x-s)\,ds, \quad x>0. \tag{3.17}$$

С другой стороны, мы знаем из (2.6), что $M(x)=[1-H(\infty)]\psi(x)$. Мы видели, что ψ удовлетворяет интегральному уравнению (3.7б), в котором теперь следует принять $\rho(0)=1$. Простое интегрирование по частям показывает теперь, что (3.7б) и (3.17) идентичны. ▶

Стандартной формой уравнения Винера — Хопфа является (3.17). Наш пример иллюстрирует путь, на котором это уравнение может возникать в теории вероятностей. Однако ссылки на общую теорию Винера — Хопфа могут запутать положение, так как мы ограничиваемся положительными функциями и мерами, что меняет (и упрощает) характер проблемы.

Искусный метод [1]), использованный Винером и Хопфом, привлек усиленное внимание и был приспособлен для решения различных вероятностных задач, например, Крамером для вывода асимптотических оценок вероятностей разорения. Легкость, с которой подобные оценки получаются при этом подходе, почти обескураживает. Но следует понять более глубокие причины этой легкости. Уравнение (3.17) представляет собой в лучшем случае одно из двух уравнений (3.7) при $\rho(0)<1$ и того меньше. Взятое само по себе, уравнение (3.17) значительно более трудно для изучения, чем пара уравнений (3.7). Например, теорема единственности *неверна* для (3.17) даже в классе распределений вероятностей. Основная идея метода Винера — Хопфа состоит во введении вспомогательной функции, которая в общей теории не имеет никакого особого смысла. Этот остроумный прием на самом деле заменяет отдельное уравнение (3.17) парой уравнений, равносильной (3.7), но единственность при этом теряется. Мы двигались в противоположном направлении, начиная с очевидных рекурсивных соотношений для вероятностей, связанных с двумя неотделимыми друг от друга задачами: изучение первого вхождения в $-\infty, 0$ и случайного блуждания на полупрямой $x>0$ до момента этого вхождения. Таким путем мы получили интегральное уравнение (3.9) из известных соотношений. Единственность *вероятностного* решения было при этом легко доказать. Доказательство сходимости, свойства решений,

[1]) Около 1931 года. Обширная литература появилась после первого монографического изложения: Hopf E., Mathematical problems of radiative equilibrium, Cambridge tracts, 31, 1934.

так же как и связь между распределением M максимума и «мерой восстановления» ψ, основаны на лемме о двойственности.

Возможность подхода к уравнению Винера — Хопфа (3.17), опирающегося на принцип двойственности, была отмечена Ф. Спицером [1]. В обычном способе соединять теорию Винера—Хопфа с вероятностными задачами начальным шагом служат формулы, связанные с (9.3) (в форме, содержащей преобразования Фурье; мы вернемся к этому в гл. XVIII). В настоящее время имеется обширная литература, посвященная применению метода Винера — Хопфа к вероятностным задачам и расширению возможностей комбинаторных методов. Основным орудием служит при этом анализ Фурье [2].

§ 4. Примеры

Явные формулы для распределений первого вхождения в общем случае получить трудно. К счастью, имеется одно примечательное исключение из этого правила, обсуждаемое в примере (а). На первый взгляд распределение F из этого примера представляется искусственным, однако этот тип распределений часто встречается в связи с процессами Пуассона, теорией очередей, задачами о разорении и т. д. Крайняя простота наших общих результатов делает непостижимым тот факт, что на анализ частных случаев было затрачено (часто с повторениями) столько изобретательности и аналитического мастерства.

Пример (в) показывает шаг за шагом процесс полного решения для арифметического распределения F с рациональной производящей функцией. Эти вычисления приведены потому, что точно такой же метод пригоден для рациональных преобразований Лапласа или Фурье. Другой пример указан в задачах 3—6. Пример (б) посвящен некоторым общим соотношениям, имеющим независимый интерес.

Мы придерживаемся обозначений предыдущего параграфа. Таким образом, H и ρ суть распределения точки первого вхождения в $\overline{0, \infty}$ и $\overline{-\infty, 0}$ соответственно. (Другими словами, H и ρ являются распределениями первой строгой верхней и сла-

---

[1]) Spitzer F., The Wiener — Hopf equation whose kernel is a probability density, *Duke Math. J.*, 24 (1957), 327—343.

[2]) Краткий обзор литературы — по причине ее полной неупорядоченности. Методология многих статей находится под влиянием случайностей исторического развития. Значительная часть работ цитируется в книге Кемпермана [Kemperman J. H. B. (1961)]. Обобщения, лежащие за пределами теории вероятностей, проиллюстрированы Бакстером [Baxter G., An operator identity. *Pacific J. Math.*, 4 (1958), 649—663].

бой нижней лестничных высот.) Наконец, $\psi = \sum H^{n*}$ — это функция восстановления, соответствующая H. Нашим основным орудием будет уравнение (3.7а), утверждающее, что при $x<0$ распределение точки первого вхождения в $-\infty, 0$ определяется формулой

$$\rho(x) = \int_{0-}^{\infty} \psi\{dy\} F(x-y). \tag{4.1}$$

а) *Показательный правый хвост распределения*[1]). Допустим для начала, что *левый* хвост распределения F — показательный, т. е. $F(x) = qe^{\beta x}$ при $x<0$. Каково бы ни было ψ, (4.1) показывает, что $\rho(x) = Ce^{\beta x}$ при $x<0$, где C — некоторая постоянная. Произведя это открытие, поменяем полуоси ролями (частью для облегчения ссылок на наши формулы, частью имея в виду наиболее важные применения к теории очередей). Предположим, следовательно, что

$$F(x) = 1 - pe^{-\alpha x} \quad \text{при } x \geqslant 0, \tag{4.2}$$

не делая никаких ограничений при $x<0$. Во избежание ненужных усложнений допустим, что F *имеет конечное математическое ожидание* μ и что F непрерывна. Из предварительного замечания вытекает, что распределение лестничной высоты H имеет плотность, пропорциональную $e^{-\alpha x}$. Рассмотрим теперь отдельно два случая.

(i) Если $\mu \geqslant 0$, то H — собственное распределение и при $x>0$

$$H(x) = 1 - e^{-\alpha x}, \quad \psi(x) = 1 + \alpha x. \tag{4.3}$$

[Последнее равенство тривиально следует как из того, что $\varphi = \sum H^{n*}$, так и из уравнения восстановления (3.10).] Из (4.1) получаем

$$\rho(x) = F(x) + \alpha \int_{-\infty}^{x} F(s)\, ds, \quad x < 0. \tag{4.4}$$

Таким образом, мы имеем явные выражения для всех нужных нам вероятностей. Легкий подсчет показывает, что

$$\rho(0) = 1 - \alpha\mu. \tag{4.5}$$

Это частный случай (2.7), так как в силу (2.3) $(1-\rho(0))^{-1} = \mathbf{E}(\mathcal{T}_1)$.

---

[1]) В случайном блуждании примера гл. VI, (8.б) и в соответствующем процессе ожидания гл. VI, (9. д) *оба* хвоста показательны.

(ii) Если $\mu<0$, то соотношения (4.3) и (4.4) по-прежнему дают решение интегрального уравнения (3.9), однако (4.5) показывает, что это решение не может иметь вероятностного смысла при $\mu<0$. Для отыскания правильного решения заметим, что H имеет плотность $R(x)=(\alpha-\varkappa)e^{-\alpha x}$, где $0<\varkappa<\alpha$, так как распределение H — несобственное. Легко подсчитать, что $\psi'(x)=(\alpha-\varkappa)e^{-\varkappa x}$ при $x>0$. Неизвестная константа $\varkappa$ определяется из условия $\rho(0)=1$. Стандартные вычисления показывают, что $\varkappa$ совпадает с единственным положительным корнем уравнения (4.6). Найдя корень этого трансцендентного уравнения, мы снова получим явные выражения для H, ρ и ψ.

Читатель может легко проверить, что эта же теория применима и тогда, когда величины $\mathbf{X}_1$, $\mathbf{X}_2$, ..., порождающие случайное блуждание, целочисленны и распределение F имеет *геометрически убывающий правый хвост*, то есть если F приписывает целому числу $k>0$ вес $q\beta^k$.

б) *Сопряженные случайные блуждания.* Допустим, что F имеет математическое ожидание $\mu\neq 0$ и что существует число $\varkappa\neq 0$, для которого

$$\int_{-\infty}^{+\infty} e^{\varkappa y}F\{dy\}=1. \tag{4.6}$$

В этом случае интеграл от e^{ty} по отношению к F существует для всех t между 0 и $\varkappa$ и является выпуклой функцией t, производная которой в нуле равна μ. Ясно, что $\varkappa$ является единственным корнем и что $\varkappa$ и μ имеют различные знаки. В последующем $\varkappa$ обозначает именно этот корень.

Пусть γ — произвольная мера на прямой. Свяжем с ней новую меру ${}^{\alpha}\gamma$, полагая

$${}^{\alpha}\gamma\{dy\}=e^{\varkappa y}\gamma\{dy\}. \tag{4.7}$$

Мера ${}^{\alpha}F$, «сопряженная» с F, представляет собой собственное распределение вероятностей. Мы скажем, что случайные блуждания, порождаемые ${}^{\alpha}F$ и F, *сопряжены друг с другом* [1]). Легко видеть, что n-кратная свертка ${}^{\alpha}F$ с собой сопряжена с F^{n*}, так что обозначение ${}^{\alpha}F^{n*}$ не вызывает сомнений. Рекурсивные формулы (3.5) показывают, что преобразования ${}^{\alpha}\rho_n$ и ${}^{\alpha}\psi_n$ имеют тот же самый вероятностный смысл в новом случайном блуж-

[1]) Это понятие применяли Хинчин, Вальд и др., но оно никогда не использовалось в полной мере. Преобразование (4.7) встречалось в теории восстановления (гл. XI, 6) и в форме, замаскированной применением производящих функций, в 1, гл. XIII, 4. Оно появится снова в теории преобразований Лапласа, см. гл. XIII, (1,6). Уравнение (4.6) применяют и в теории Винера — Хопфа

дании, как ρ_n и ψ_n в старом. Вообще преобразования ${}^a\rho$, ${}^aH$, ${}^a\psi$ и т. д. *имеют очевидное значение для случайного блуждания, порождаемого* ${}^aF$. [Это можно увидеть и непосредственно из интегрального уравнения (3.9).]

Мы наметили, таким образом, широко применимый метод преобразования результатов, касающихся случайного блуждания с $\mu<0$, в результаты для случайных блужданий с положительным математическим ожиданием и обратно.

Если $\mu<0$, то лестничная высота имеет несобственное распределение, но ${}^aH$ имеет собственное распределение. Это означает, что

$$\int_0^\infty e^{xy} H\{dy\} = 1. \tag{4.8}$$

Вся сила метода сопряженных случайных блужданий проистекает главным образом из этого замечания. В самом деле, в гл. XI, 6 мы установили, что, зная корень уравнения (4.8), можно дать прекрасные асимптотические оценки для верхнего лестничного процесса. Эти оценки были бы иллюзорны, если бы они требовали знания H, но мы только что видели, что *корни уравнений* (4.6) *и* (4.8) *совпадают*.

в) *Ограниченные арифметические распределения.* Пусть a и b — положительные числа и F — арифметическое распределение со скачками f_k в точках $k=-b, \ldots a$. Меры ψ и ρ также сосредоточены в целых точках. Их скачки в точке k мы обозначим ψ_k и ρ_k соответственно. Первое вхождение в $\overline{-\infty, 0}$ осуществляется в целой точке $\geqslant -b$, так что $\rho_k=0$ при $k<-b$.

Введем производящие функции

$$\Phi(s) = \sum_{k=-b}^{a} f_k s^k, \quad \Psi(s) = \sum_{k=0}^{\infty} \psi_k s^k, \quad R(s) = \sum_{k=-b}^{0} \rho_k s^k. \tag{4.9}$$

Они отличаются от производящих функций из **1**, гл. XI тем, что Φ и R содержат также и отрицательные степени s, однако ясно, что основные свойства и правила действий при этом сохраняются. В частности, $\mu=\Phi'(1)$ есть математическое ожидание F. Свертке распределений соответствует перемножение производящих функций, поэтому основное интегральное уравнение (3.9) равносильно уравнению $\Psi+R=1+\Psi\Phi$ или

$$\Psi(s) = \frac{s^b(R(s)-1)}{s^b(\Phi(s)-1)}. \tag{4.10}$$

Числитель и знаменатель здесь — полиномы степени a и $a+b$ соответственно. Степенной ряд слева сходится при $|s|<1$, так что все корни знаменателя, лежащие внутри единичного круга,

должны быть в то же время и корнями числителя. Мы покажем, что это условие позволяет однозначно восстановить R и Ψ.

Предположим для определенности, что $\mu=0$ (см. задачу 12), и докажем, что знаменатель имеет ровно $b-1$ комплексных корней внутри единичного круга, $a-1$ корней вне его и двойной корень в точке 1. В самом деле, на вещественной оси $\Phi''\geqslant 0$, так что Φ — выпуклая функция с минимумом $\Phi(1)=1$.

Для комплексных s с $|s|\leqslant 1$ имеем $|s^b\Phi(s)|\leqslant 1$ и по теореме Руше[1]) число корней уравнения

$$s^b[\Phi(s)-(1+\varepsilon)]=0$$

в единичном круге равно числу корней уравнения $s^b(1+\varepsilon)=0$, т. е. равно b. Один из этих корней вещественный и он сходится к 1 при $\varepsilon\to 0$, так что знаменатель имеет ровно $b-1$ корней $s_1,\ldots,s_{b-1}$ с $|s_j|<1$. Пусть $\sigma_1,\ldots,\sigma_{a-1}$ — корни с $|\sigma_j|>1$. Тогда знаменатель имеет вид

$$(\Phi(s)-1)=C(s-1)^2(s-s_1)\ldots(s-s_{b-1})(s-\sigma_1)\ldots(s-\sigma_{a-1}). \tag{4.11}$$

Корни $s_1,\ldots,s_{b-1}$ должны совпадать с корнями числителя, и так как коэффициенты ψ_n ограничены, то же самое должно быть верно и для *одного* корня, равного 1. Этим Ψ определяется с точностью до постоянного множителя. Но по определению, $\Psi(0)=1$, и мы получаем *искомую явную формулу*

$$\Psi(s)=\frac{1}{(1-s)\left(1-\frac{s}{\sigma_1}\right)\ldots\left(1-\frac{1}{\sigma_{a-1}}\right)}. \tag{4.12}$$

Стандартное разложение на простейшие дроби приводит к явному выражению для ψ_n. Большое преимущество этого метода в том, что знание доминирующего корня приводит к хорошим асимптотическим оценкам (см. 1, гл. XI, 4).

Для производящей функции R вероятностей первого вхождения ρ_k из (4.10) и (4.12) получаем

$$R(s)=1+C(-1)^a\sigma_1\ldots\sigma_{a-1}\left(1-\frac{1}{s}\right)\left(1-\frac{s_1}{s}\right)\ldots\left(1-\frac{s_{b-1}}{s}\right). \tag{4.13}$$

[Коэффициент C определяется по (4.11) и зависит только от данного распределения $\{f_k\}$.] Снова разложение на простейшие дроби приводит к асимптотическим оценкам (продолжение см. в задачах 12—15).

---

[1]) См., например, Hille E., Analytic function theory, vol. I, section 9.2 (Ginn and Co., 1959) или Маркушевич А. И., Теория аналитических функций, Гостехиздат, М., 1950.

§ 5. Применения

В гл. VI, 9 было показано, что основная задача теории очередей сводится к отысканию распределения M для

$$\mathbf{M}=\max[0,\ \mathbf{S}_1,\ \ldots] \tag{5.1}$$

в случайном блуждании, порожденном величинами $\mathbf{X}_k$ с $\mu=\mathbf{E}(\mathbf{X}_k)<0$. Примеры гл. VI, (9а)—(9в) показывают, что это же распределение возникает и в других обстоятельствах, например в связи с задачей о разорении для обобщенного процесса Пуассона. В последнем случае, как и в теории очередей, основное распределение имеет вид

$$F=A*B, \tag{5.2}$$

где A сосредоточено на $\overline{0,\ \infty}$, а B на $\overline{-\infty,\ 0}$. Мы предположим, что A и B имеют конечные математические ожидания a и $-b$, так что F имеет математическое ожидание $\mu=a-b$. Мы предположим также, что распределение F *непрерывно*. Это позволит избежать утомительной необходимости различать строгие и слабые лестничные величины.

Мы обозначим, так же как в двух предыдущих параграфах, распределения для верхних и нижних лестничных высот через H и ρ соответственно. Иными словами, H и ρ суть распределения для первых вхождений в $\overline{0,\infty}$ и $\overline{-\infty,0}$ (а также в соответствующие замкнутые интервалы). В примере (3в) и в (2.6) было показано, что при $\mu<0$

$$M(x)=\frac{\psi(x)}{\psi(\infty)}=[1-H(\infty)]\sum_0^\infty H^{n*}(x). \tag{5.3}$$

Пример (4а) содержит *явную формулу* [1]) *для этого распределения, справедливую, если один из хвостов F — показательный,* т. е.

$$F(x)=1-pe^{-\alpha x}\quad \textit{при}\quad x>0 \tag{5.4}$$

или $F(x)=qe^{\alpha x}$ при $x<0$.

По весьма счастливому совпадению *условие* (5.4) *выполняется, если F имеет вид* (5.2) с

$$A(x)=1-e^{-\alpha x}\quad \textit{при}\quad x>0. \tag{5.5}$$

---

[1]) Другая явная формула указана в примере (4.в) для случая арифметических распределений F с конечным числом атомов. Эта явная формула слишком сложна для практического использования, но разложение на простейшие дроби приводит к хорошим *асимптотическим оценкам, если известен доминирующий корень знаменателя.* Такой же метод применим в случае *рациональных характеристических функций.* Это замечание охватывает множество частных случаев, разобранных в литературе.

Тогда

$$p = \int_{-\infty}^{0} e^{\alpha y} B\{dy\}. \tag{5.6}$$

Наши простые результаты применимы поэтому к теории очередей, если *или* поступающий поток является пуассоновским, *или* время обслуживания распределено показательно.

Упомянутое условие выполнено и в задаче о разорении для обобщенного процесса Пуассона. Существует необъятная прикладная литература, разбирающая отдельные задачи при тех или иных предположениях относительно распределения B, иногда, как в задаче о разорении, в замаскированной форме. Оказывается, большая общность и значительно большая простота достигаются при использовании только одного условия (5.4) вместо комбинации (5.5) и (5.2). Мы имеем здесь прекрасный пример экономии мысли, присущей общей теории, когда понимание не затемняют случайные черты отдельных примеров.

Примеры. а) *Формула Хинчина — Полячека.* Допустим, что F имеет вид (5.2), где A удовлетворяет (5.5) и $\mu = 1/\alpha - b > 0$. Случайное блуждание уходит в ∞, и мы должны заменить в (5.1) максимум на минимум. Это равносильно замене H в (5.3) распределением ρ из (4.4). Простое интегрирование по частям показывает, что при $x < 0$

$$\rho(x) = \alpha \int_{-\infty}^{x} B(y)\,dy. \tag{5.7}$$

Здесь $\rho(0) = \alpha b$, так что при $x < 0$

$$\mathbf{P}\{\min(\mathbf{S}_0, \mathbf{S}_1, \ldots) \leqslant x\} = (1 - \alpha b) \sum_{0}^{\infty} \rho^{n*}(x). \tag{5.8}$$

Это известная формула Хинчина — Полячека, которую неоднократно открывали заново в частных случаях, неизменно используя метод преобразований Лапласа [неприменимый к более общим распределениям типа (5.4)]. Мы вернемся к этой формуле в примере гл. XVIII, (3в).

б) *Двойственный случай.* Рассмотрим такие же распределения, как и в предыдущем примере, но предположим, что $\mu < 0$. Как было показано во второй части примера (4а), в этом случае

$$\mathbf{P}\{\max(\mathbf{S}_0, \mathbf{S}_1, \ldots) \leqslant x\} = \frac{\varkappa}{\alpha}\psi(x) = 1 - \left(1 - \frac{\varkappa}{\alpha}\right) e^{-\varkappa x}, \tag{5.9}$$

где $\varkappa$ — единственный положительный корень «характеристического уравнения» (4.6). (Этот результат может быть получен

также и методом сопряженных случайных блужданий; достаточно вспомнить, что при $\mu \geqslant 0$ $\psi(x) = 1 + \alpha x$ для $x > 0$.) В применении к теории очередей (5.9) означает, что если *время обслуживания распределено показательно, то распределение времен ожидания стремится к показательному пределу.*

в) *Асимптотические оценки.* Самый вид формулы (5.3) напоминает нам тот факт, что максимум частных сумм $\mathbf{S}_n$ совпадает с максимальной из ординат лестничных точек. Асимптотическая оценка для распределения этого максимума была выведена в гл. XI, (6.16). Она может показаться неприменимой в настоящей ситуации, так как в нее входит положительный корень характеристического уравнения (4.8) для распределения H, а последнее, вообще говоря, неизвестно. Однако мы видели в примере (4.б), что этот корень совпадает с корнем (4.6). Для распределения M максимума частных сумм это означает следующее: *если F имеет отрицательное математическое ожидание и существует корень $\varkappa > 0$ уравнения* (4.6), *то при* $x \to \infty$ $1 - M(x) \sim Ce^{-\varkappa x}$. От специальных свойств распределения F зависит только константа C. Общность этого результата замечательна, в частности, потому, что он следует из простейшей теоремы восстановления и простого преобразования сопряженного случайного блуждания.

г) *Оценка Крамера для вероятностей разорения.* Покажем, как можно практически вычислить константу C Вернемся к общей задаче о разорении (см. гл. VI, 5), которая была переформулирована как задача об очередях в примере гл. VI, (9.в). В настоящих обозначениях мы имеем дело с распределением F вида (5.2), где A — показательное распределение (5.5) и математическое ожидание $\mu = 1/\alpha - b > 0$. Мы ищем распределение (отрицательного) минимума из $\mathbf{S}_0$, $\mathbf{S}_1$, . . . , указанное в (5.8). Именно эта задача [1]) изучалась в примере гл. XI, (7.а). Характеристическое уравнение для ρ равносильно (4.6). Если существует корень $\varkappa < 0$, то (в настоящих обозначениях) из формул (7.5)—(7.6) следует, что при $x \to -\infty$

$$\mathbf{P}\{\min(\mathbf{S}_0, \mathbf{S}_1, \ldots) \leqslant x\} \sim \frac{1 - \alpha b}{|\varkappa| \beta} e^{|\varkappa| x}, \tag{5.10}$$

где

$$\beta = \alpha \int_{-\infty}^{0} e^{-|\varkappa| y} |y| B(y)\, dy. \tag{5.11}$$

[1]) Нашему несобственному распределению ρ, сосредоточенному на $\overline{-\infty, 0}$, в гл. VI, 7 соответствует сосредоточенное на $\overline{0, \infty}$ несобственное распределение L с *плотностью* $(\alpha/c)(1 - F(x))$.

Оценку (5.10) неоднократно выводили в специальных предположениях и весьма трудоемкими методами. Она эквивалентна часто используемой оценке, которая в математической теории страхования была предложена Крамером [1]).

§ 6. Одна комбинаторная лемма

Распределение лестничных моментов связано с одной простой комбинаторной леммой. Вероятностная часть рассуждений будет более ясной, если мы выделим эту лемму особо.

Пусть $x_1, \ldots, x_n$ — n чисел. Рассмотрим их частные суммы

$$s_0=0, \ \ldots, \ s_n=x_1+\ldots+x_n.$$

Мы скажем, что $\nu>0$ есть *лестничный индекс*, если $s_\nu>s_0, \ldots, s_\nu>s_{\nu-1}$, т. е. если s_ν превосходит все предыдущие частные суммы. Ясно, что имеется n лестничных индексов, если все x_ν положительны, и нет ни одного, если все x_ν отрицательны.

Рассмотрим теперь n циклических перестановок

$$(x_1, \ldots, x_n), \ (x_2, \ldots, x_n, x_1), \ \ldots (x_n, x_1, \ldots, x_{n-1})$$

и занумеруем их числами $0, \ldots, n-1$. Частные суммы $s_k^{(\nu)}$ в перестановке с номером ν равны

$$s_k^{(\nu)}=\begin{cases} s_{\nu+k}-s_\nu & \text{при } k=1, \ldots, n-\nu, \\ s_n-s_\nu+s_{k-n+\nu} & \text{при } k=n-\nu+1, \ldots, n. \end{cases} \tag{6.1}$$

Лемма 1. *Пусть $s_n>0$. Обозначим через r число циклических перестановок, в которых n является лестничным индексом. Тогда $r\geqslant 1$ и в каждой из этих перестановок будет ровно r лестничных индексов.*

Примеры. Для $(-1, -1, -1, 0, 1, 10)$ мы имеем $r=1$. Данный порядок — единственный, в котором последняя сумма максимальна. Для $(-1, 4, 7, 1)$ мы имеем $r=3$; перестановки с номерами 0, 2 и 3 содержат по три лестничных индекса. ▶

Доказательство. Выберем ν так, что s_ν максимально. Если таких индексов несколько, то возьмем наименьший из них. Другими словами,

$$s_\nu>s_1, \ \ldots, \ s_\nu>s_{\nu-1}, \quad s_\nu\geqslant s_{\nu+1}, \ \ldots, \ s_\nu\geqslant s_n. \tag{6.2}$$

---

[1]) Более новый вывод методом Винера — Хопфа в комплексной плоскости см в статье Крамера, цитированной в гл. VI. 5. Наша оценка (5.10) совпадает с формулой (57) у Крамера.

Из (6.1) видно, что при этом в ν-й перестановке последняя частная сумма строго максимальна, так что n есть лестничный индекс. Таким образом, $r \geqslant 1$. Не ограничивая общности, допустим, что n есть лестничный индекс в исходной перестановке, т. е. $s_n > s_j$ при всех j. Тогда все величины в первой строке (6.1) будут $< s_n$. Вторая строка показывает, что n будет лестничным индексом для ν перестановок тогда и только тогда, когда $s_\nu > s_1, \ldots, s_\nu > s_{\nu-1}$, т. е. когда в исходной перестановке ν будет лестничным индексом. Таким образом, число перестановок, в которых n есть лестничный индекс, равно числу лестничных индексов. Лемма доказана. ▶

Слабые лестничные индексы определяются аналогично, с той разницей, что строгие неравенства $>$ заменяются на $\geqslant$. В применении к ним предыдущие рассуждения показывают, что верна

Лемма 2. *Если* $s_n \geqslant 0$, *то лемма* 1 *применима и к слабым лестничным индексам.*

§ 7. Распределение лестничных моментов

В предыдущем параграфе наше внимание было сосредоточено на лестничных высотах. Теперь мы перейдем к лестничным моментам. Рассмотрим вероятность того, что n является моментом первого вхождения в $\overline{0, \infty}$, т. е.

$$\tau_n = \mathbf{P}\{\mathbf{S}_1 \leqslant 0, \ldots, \mathbf{S}_{n-1} \leqslant 0, \mathbf{S}_n > 0\}. \tag{7.1}$$

Другими словами, $\{\tau_n\}$ есть (возможно, несобственное) распределение первого лестничного момента $\mathcal{T}_1$. Введем производящую функцию

$$\tau(s) = \sum_{n=1}^{\infty} \tau_n s^n, \quad 0 \leqslant s \leqslant 1. \tag{7.2}$$

Следующая ниже важная теорема была найдена Спарре-Андерсеном. Затем был дан ряд новых ее доказательств. [Улучшенный вариант, принадлежащий Бакстеру, содержится в (9.3).]

Теорема 1.

$$\log \frac{1}{1-\tau(s)} = \sum_{n=1}^{\infty} \frac{s^n}{n} \mathbf{P}\{\mathbf{S}_n > 0\}. \tag{7.3}$$

Доказательство. Фиксируем n и рассмотрим наряду с выборочной точкой $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ также $(n-1)$ точек, получаемых из нее циклическими перестановками и соответствующие

частные суммы $\mathbf{S}_0^{(\nu)}, \ldots, \mathbf{S}_n^{(\nu)}$ [определенные, как в (6.1)]. Фиксируем целое r и определим n случайных величин $\mathbf{Y}^{(\nu)}$ следующим образом: $\mathbf{Y}^{(\nu)}=1$, если n есть r-й лестничный индекс для $(\mathbf{S}_1^{(\nu)}, \ldots, \mathbf{S}_n^{(\nu)})$ и $\mathbf{Y}^{(\nu)}=0$ в других случаях. Случаю $\nu=0$ соответствует последовательность $\{\mathbf{S}_1, \ldots, \mathbf{S}_n\}$, так что $\mathbf{Y}^{(0)}=1$ тогда и только тогда, когда n совпадает с r-м лестничным индексом в нашем случайном блуждании. Но r-й лестничный момент есть сумма r независимых случайных величин, распределенных как $\mathcal{T}_1$, так что

$$\mathbf{P}\{\mathbf{Y}^{(0)}=1\}=\tau_n^{(r)}, \tag{7.4}$$

где $\tau_n^{(r)}$ — это коэффициент при s^n в $\tau^r(s)$. По причине симметрии все n случайных величин $\mathbf{Y}^{(\nu)}$ имеют одно и то же распределение. Поэтому

$$\tau_n^{(r)}=\mathbf{E}(\mathbf{Y}^{(0)})=\frac{1}{n}\mathbf{E}(\mathbf{Y}^{(0)}+\ldots+\mathbf{Y}^{(n-1)}). \tag{7.5}$$

Последняя сумма тривиально равна нулю в области $\{\mathbf{S}_n>0\}$. По лемме (6.1) она может принимать только значение 0 и r, откуда

$$\frac{1}{r}\tau_n^{(r)}=\frac{1}{n}\mathbf{P}\{\mathbf{Y}^{(0)}+\ldots+\mathbf{Y}^{(n-1)}=r\}. \tag{7.6}$$

Суммируя по всем r, получаем

$$\sum_{r=1}^{\infty}\frac{1}{r}\tau_n^{(r)}=\frac{1}{n}\mathbf{P}\{\mathbf{S}_n>0\}, \tag{7.7}$$

так как в каждой точке множества $\{\mathbf{S}_n>0\}$ сумма (7.6) равна целому r. Умножая (7.7) на s^n и суммируя по n, находим

$$\sum_{r=1}^{\infty}\frac{1}{r}\tau^r(s)=\sum_{n=1}^{\infty}\frac{s^n}{n}\mathbf{P}\{\mathbf{S}_n>0\}, \tag{7.8}$$

что совпадает с утверждением (7.3). ▶

Удивительное следствие этой теоремы состоит в том, что распределение лестничных моментов однозначно определяется последовательностью чисел $F^{n*}(0)$. Например, если распределение F симметрично и непрерывно, то при всех n $\mathbf{P}\{\mathbf{S}_n>0\}=\frac{1}{2}$ и левая часть (7.3) равна $\log(1/\sqrt{1-s})$. Таким образом, получаем

Следствие 1. *Если F симметрично и непрерывно, то*

$$\tau(s)=1-\sqrt{1-s}. \tag{7.9}$$

Очевидно, что *теорема* 1 *остается справедливой, если в* (7.1) *и* (7.3) *заменить знак* $>$ *на* $\geqslant$ *и* $\leqslant$ *на* $<$, т. е. если $\mathcal{T}_1$ заменить на слабый лестничный момент $\overline{\mathcal{T}}_1$ и события $\{S_n>0\}$ на $\{S_n\geqslant 0\}$.

Важное следствие доказанной нами теоремы содержит

Теорема 2. *Случайное блуждание уходит в* $-\infty$ *тогда и только тогда, когда*

$$\sum_{n=1}^{\infty}\frac{1}{n}\mathbf{P}\{S_n>0\}<\infty. \tag{7.10}$$

Этот критерий остается справедливым[1]) *при замене событий* $\{S_n>0\}$ *на* $\{S_n\geqslant 0\}$.

Доказательство. Уход в $-\infty$ происходит тогда и только тогда, когда верхний лестничный процесс обрывается, т. е. когда $\mathcal{T}_1$ имеет несобственное распределение. Это равносильно неравенству $\tau(1)<1$. В этом случае обе части (7.3) остаются ограниченными при $s\to 1$. Отсюда видно, что условие (7.10) является необходимым и достаточным. Это же рассуждение применимо к слабым лестничным моментам, что обосновывает заключительное утверждение теоремы. ▶

Мы знаем, что уход в $-\infty$ имеет место, если F имеет математическое ожидание $\mu<0$. Однако не очевидно, что неравенство $\mu<0$ влечет (7.10). Проверка этого факта представляет собой превосходное техническое упражнение, имеющее и методологический интерес (см. задачу 16).

Теорема 3. *Лестничный момент* $\mathcal{T}_1$ *имеет математическое ожидание* (*и собственное распределение*) *тогда и только тогда, когда случайное блуждание уходит в* ∞. *При выполнении последнего условия*

$$\log \mathbf{E}(\mathcal{T}_1)=\log\sum k\tau_k=\sum_{n=1}^{\infty}\frac{1}{n}\mathbf{P}\{S_n\leqslant 0\}. \tag{7.11}$$

(Во всех остальных случаях ряд расходится.)

Доказательство. Для $0<s<1$

$$\log\frac{1-\tau(s)}{1-s}=\sum_{n=1}^{\infty}\frac{s^n}{n}[1-\mathbf{P}\{S_n>0\}]. \tag{7.12}$$

Полагая $s\to 1$, в случае сходимости ряда мы получаем (7.11). В случае его расходимости или распределение $\mathcal{T}_1$ несобственное, или $\mathbf{E}(\mathcal{T}_1)=\infty$. По теореме 2 сходимость имеет место тогда и только тогда, когда случайное блуждание уходит в ∞. ▶

[1]) Мы увидим, что всегда $\Sigma n^{-1}\mathbf{P}\{S_n=0\}<\infty$ [см. (9в)].

В заключение мы обратимся к *альтернативной интерпретации* теоремы 1, которая приводит к закону арксинуса. Положим

$$p(s)=\frac{1}{1-\tau(s)}. \tag{7.13}$$

Из теории рекуррентных событий ясно, что (7.13) есть производящая функция для вероятностей p_n того, что n будет лестничным индексом:

$$p_n=\mathbf{P}\{\mathbf{S}_n>\mathbf{S}_0,\ \ldots,\ \mathbf{S}_n>\mathbf{S}_{n-1}\}. \tag{7.14}$$

Меняя порядок случайных величин $\mathbf{X}_j$, получаем двойственное соотношение

$$p_n=\mathbf{P}\{\mathbf{S}_1>0,\ \mathbf{S}_2>0,\ \ldots,\ \mathbf{S}_n>0\}. \tag{7.15}$$

Для удобства ссылок мы придадим этому результату форму теоремы.

Теорема 4. *Производящая функция* (7.13) *вероятностей* (7.15) *определяется по* (7.3).

В силу симметрии вероятности

$$q_n=\mathbf{P}\{\mathbf{S}_1\leqslant 0,\ \ldots,\ \mathbf{S}_n\leqslant 0\} \tag{7.16}$$

имеют производящую функцию q, определяемую из равенства

$$\log q(s)=\sum_{n=1}^{\infty}\frac{s^n}{n}\mathbf{P}\{\mathbf{S}_n\leqslant 0\}. \tag{7.17}$$

§ 8. Закон арксинуса

Одна из удивительных черт случайных колебаний при бросании монеты находит свое выражение в двух законах арксинуса (**1**, гл. III, 5 и 8). Один из них влечет тот факт, что число положительных членов в последовательности $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_n$ с с большой вероятностью будет ближе к 0 или n, чем к $n/2$ ($n/2$ соответствовало бы наивным ожиданиям). Второй закон приводит к такому же выводу относительно положения максимального члена. Мы покажем теперь, что эти законы верны для всех симметричных и многих других распределений.

В последующих формулировках нам будет мешать то, что максимум может достигаться несколько раз и что частные суммы могут обращаться в нуль. Эти возможности можно не рассматривать, если распределение F непрерывно, так как в этом случае вероятность равенства двух любых частных сумм равна нулю. (Мы советуем читателю ограничиться только этим

случаем.) В общей теории мы условимся рассматривать номер *первой* максимальной суммы, т. е. индекс k, для которого

$$\mathbf{S}_k > \mathbf{S}_0, \ \ldots, \ \mathbf{S}_k > \mathbf{S}_{k-1}, \ \mathbf{S}_k \geqslant \mathbf{S}_{k+1}, \ \ldots, \ \mathbf{S}_k \geqslant \mathbf{S}_n. \tag{8.1}$$

Здесь n фиксировано и k пробегает значения 0, 1, ..., n. При некотором $k \leqslant n$ событие (8.1) должно произойти, так как мы можем определить (собственную) случайную величину $\mathbf{K}_n$ *как номер первого максимума*, т. е. номер, при котором верно (8.1). (Здесь $\mathbf{S}_0 = 0$.) Для наступления (8.1) необходимо и достаточно одновременное наступление событий

$$\{\mathbf{S}_k > \mathbf{S}_0, \ \ldots, \ \mathbf{S}_k > \mathbf{S}_{k-1}\} \text{ и } \{\mathbf{S}_{k+1} - \mathbf{S}_k \leqslant 0, \ \ldots, \ \mathbf{S}_n - \mathbf{S}_k \leqslant 0\}.$$

Первое зависит только от случайных величин $\mathbf{X}_1, \ \ldots, \ \mathbf{X}_k$, а второе — от величин $\mathbf{X}_{k+1}, \ \ldots, \ \mathbf{X}_n$, поэтому они независимы. По самому определению, их вероятности равны p_k и q_{n-k} [см. (7.14) и (7.16)], следовательно, верна

Лемма 1. *Для всех k и n*

$$\mathbf{P}\{\mathbf{K}_n = k\} = p_k q_{n-k}. \tag{8.2}$$

Допустим теперь, что $\mathbf{P}\{\mathbf{S}_n > 0\} = \mathbf{P}\{\mathbf{S}_n \leqslant 0\} = \frac{1}{2}$ при всех n. Производящие функции p и q в этом случае равны и определяются по (7.17). Справа мы узнаем разложение в степенной ряд для логарифма, так что $p(s) = q(s) = 1/\sqrt{1-s}$. Таким образом,

$$p_k q_{n-k} = \binom{-\frac{1}{2}}{k} \binom{-\frac{1}{2}}{n-k} (-1)^n, \tag{8.3}$$

что можно переписать в более удобной форме

$$p_k q_{n-k} = \binom{2k}{k} \binom{2n-2k}{n-k} \frac{1}{2^{2n}}. \tag{8.4}$$

Эта формула совпадает с формулой **1**, гл. III, (5.1) для распределения числа гербов при бросаниях монеты. Ее предельная форма была получена в **1**, гл. III, (5.7). Мы можем теперь утверждать, что верна

Теорема 1. *Если распределение F симметрично и непрерывно, то распределение вероятностей $\mathbf{K}_n$ (номера первого максимума среди $\mathbf{S}_0, \mathbf{S}_1, \ \ldots, \ \mathbf{S}_n$) задается формулами* (8.3) *или* (8.4). *При $n \to \infty$ и фиксированном α, $0 < \alpha < 1$,*

$$\mathbf{P}\{\mathbf{K}_n < n\alpha\} \to 2\frac{1}{\pi}\arcsin\sqrt{\alpha}. \tag{8.5}$$

Предельное распределение имеет плотность $1/(\pi\sqrt{\alpha(1-\alpha)})$, которая неограничена вблизи точек 0 и 1 и имеет минимум в средней точке 1/2. Это показывает, что значения нормированного максимума $\mathbf{K}_n/n$ располагаются вблизи точек 0 и 1 с вероятностью заметно большей, чем вблизи точки 1/2. Более подробное обсуждение см. 1, гл. III, 5 и 8.

Можно было бы ожидать, что формула (8.4) остается верной (по крайней мере асимптотически), если медианы сумм $\mathbf{S}_n$ достаточно быстро стремятся к нулю. В самом деле, по теореме 7.4 всегда

$$p(s)\sqrt{1-s} = \exp\sum_{n=1}^{\infty}\frac{s^n}{n}[\mathbf{P}\{\mathbf{S}_n > 0\} - 1/2]. \tag{8.6}$$

Далее можно рассуждать примерно следующим образом. Если медианы $\mathbf{S}_n$ близки к нулю, то коэффициенты в правой части малы и потому правая часть близка к единице. Следовательно, $p(s)$ близко к $1/\sqrt{1-s}$. Правдоподобно, что соответствующие коэффициенты обоих рядов будут близки друг к другу. Это рассуждение можно сделать строгим, если ряд

$$\sum_{n=1}^{\infty}\frac{1}{n}[\mathbf{P}\{\mathbf{S}_n > 0\} - 1/2] = c \tag{8.7}$$

сходится. При $s \to 1$ правая часть (8.6) стремится к e^c, так что $p(s) \sim e^c/\sqrt{1-s}$. В силу (7.15) p_n монотонны и по уточненной тауберовой теореме 5 из гл. XIII, 5 последнее соотношение влечет

$$p_n \sim e^c\binom{-\frac{1}{2}}{n}(-1)^n, \qquad n \to \infty. \tag{8.8}$$

Для q_n мы получаем точно такое же соотношение с заменой c на $-c$. Поэтому при $n \to \infty$ и $n - k \to \infty$

$$\mathbf{P}\{\mathbf{K}_n = k\} \sim \binom{2k}{k}\binom{2n-2k}{n-k}\frac{1}{2^{2n}}. \tag{8.9}$$

Таким образом, верна

Теорема 1а[1]). *Если ряд* (8.7) *сходится, то выполняются соотношения* (8.9)*и* (8.5).

В гл. XVIII 5 будет показано, что ряд (8.7) сходится, если F имеет нулевое математическое ожидание и конечную дисперсию. К этим распределениям применим, следовательно, закон арксинуса.

В 1, гл. III нам пришлось доказывать оба закона арксинуса по отдельности, однако следующая теорема показывает, что они эквивалентны. Теорема 2 (для непрерывных распределений) была исходной точкой исследований Е. Спарре-Андерсена,

[1]) Эта теорема была доказана трудоемкими вычислениями Спарре-Андерсеном, а также Спицером. Наблюдение, что тауберова теорема устраняет все сложности, принадлежит Спицеру. Обобщения см. в 9. г.

нашедшего новый подход к теории флуктуаций. Первоначальное доказательство было крайне сложным. Сейчас существует несколько доказательств, но приводимое здесь кажется простейшим.

Теорема 2. *Число* Π_n *строго положительных членов последовательности* $S_1, \ldots, S_n$ *имеет то же самое распределение* (8.2), *что и номер* K_n *первого максимального числа.*

Аналогично, распределение числа неотрицательных членов точно такое же, как и распределение номера *последнего* максимума.

Мы начнем с чисто комбинаторной леммы, в которой слово «вероятность» используется только для удобства выражений. Пусть $x_1, \ldots, x_n$ — произвольные числа (не обязательно различные). Рассмотрим $n!$ перестановок $(i_1, \ldots, i_n)$ индексов и соответствующие им частные суммы

$$s_0 = 0, \ \ldots, \ s_n = x_{i_1} + \ \ldots \ + x_{i_n}. \tag{8.10}$$

Все $n!$ перестановок мы рассмотрим как точки выборочного пространства, в котором каждой из них приписывается вероятность $1/n!$ Суммы s_j становятся случайными величинами. Введем четыре новые случайные величины:

$\mathbf{K}$ = номер первого максимума в $(s_0, \ldots, s_n)$;

$\mathbf{K}^*$ = номер последнего максимума в $(s_0, \ldots, s_n)$;

$\mathbf{\Pi}$ = число положительных членов в $(s_1, \ldots, s_n)$;

$\mathbf{\Pi}^*$ = число неотрицательных членов в $(s_1, \ldots, s_n)$.

Все эти случайные величины принимают значения $0, 1, \ldots, n$.

Пример. Ниже мы перечисляем в таблице 12 различимых перестановок четырех чисел 1, 1, —1, —2 и соответствующие значения наших четырех случайных величин.

$(s_0,$	$s_1,$	$s_2,$	$s_3,$	$s_4)$	$\mathbf{K}$	$\mathbf{K}^*$	$\mathbf{\Pi}$	$\mathbf{\Pi}^*$
0	1	2	1	—1	2	2	3	3
0	1	2	0	—1	2	2	2	3
0	1	0	1	—1	1	3	2	3
0	1	0	—2	—1	1	1	1	2
0	1	—1	0	—1	1	1	1	2
0	1	—1	—2	—1	1	1	1	1
0	—1	0	1	—1	3	3	1	2
0	—1	0	—2	—1	0	2	0	1
0	—1	—3	—2	—1	0	0	0	0
0	—2	—1	0	—1	0	3	0	1
0	—2	—1	—2	—1	0	0	0	0
0	—2	—3	—2	—1	0	0	0	0

$\mathbf{K}$ и $\mathbf{\Pi}$ принимают значения 0, 1, 2, 3 с вероятностями $\frac{5}{12}$, $\frac{4}{12}$, $\frac{2}{12}$, $\frac{1}{12}$, а $\mathbf{K}^*$ и $\mathbf{\Pi}^*$ распределены равномерно. ▶

Основную роль играет комбинаторная

Лемма 2. *Случайные величины* $\mathbf{K}$ *и* $\mathbf{\Pi}$ *имеют одно и то же распределение. Это же верно для пары* $\mathbf{K}^*$, $\mathbf{\Pi}^*$

Доказательство. Мы применим индукцию. При $n=1$ лемма верна, как можно установить, рассматривая три случая $x_1>0$, $x_1<0$ и $x_1=0$. Допустим теперь, что лемма верна для $n-1$. Выделим три случая.

(а) Пусть $x_1+\ldots+x_n<0$. Тогда $s_n<0$ для всех перестановок и так как $s_0=0$, максимум никогда не может быть на последнем месте. Все четыре случайные величины зависят поэтому от первых $n-1$ координат, так что утверждение леммы верно в силу предположения индукции.

В качестве подготовки к следующему шагу мы выведем одно уточнение этого результата. Как мы видели в § 2, изменение порядка величин $x_{i_1}, \ldots, x_{i_n}$ равносильно повороту на 180° графика последовательности $(s_1, \ldots, s_n)$, при котором первый максимум становится последним минимумом. Поэтому распределение номера последнего минимума такое же, как распределение $n-\mathbf{K}$, а это есть число неположительных частных сумм. Таким образом, номер последнего (первого) минимума имеет такое же распределение, как число неположительных (отрицательных) членов.

(б) Пусть $x_1+\ldots+x_n>0$. Тогда последнее замечание, касающееся минимумов, можно применить к перестановке $(-x_1, \ldots, x_n)$. Результат тривиально эквивалентен утверждению леммы.

(в) Пусть $x_1+\ldots+x_n=0$. Так же как и в пункте (а), устанавливается, что $\mathbf{K}$ и $\mathbf{\Pi}$ имеют одно и то же распределение. Для $\mathbf{K}^*$ и $\mathbf{\Pi}^*$ нужно повторить рассуждения пункта (б). Доказательство закончено. ▶

Доказательство *теоремы* 2. Мы поступим, как в доказательстве теоремы 7.1. Рассмотрим при фиксированном r $n!$ перестановок $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ и определим $n!$ случайных величин $\mathbf{Y}^{(\nu)}$, полагая $\mathbf{Y}^{(\nu)}=1$, если перестановке номер ν соответствуют r положительных частных сумм. В других случаях $\mathbf{Y}^{(\nu)}=0$. По причинам симметрии все величины $\mathbf{Y}^{(\nu)}$ имеют одно и то же распределение. Если считать естественный порядок $(1, \ldots, n)$ пере-

становкой номер нуль, то

$$\mathbf{P}\{\mathbf{\Pi}_n = r\} = \mathbf{E}(\mathbf{Y}^{(0)}) = \frac{1}{n!}\mathbf{E}\left(\sum \mathbf{Y}^{(\nu)}\right). \tag{8.11}$$

В каждой выборочной точке величина $\frac{1}{n!}\sum \mathbf{Y}^{(\nu)}$ равна значению случайной величины $\mathbf{\Pi}$ из леммы 2. Стало быть, утверждение $\mathbf{P}\{\mathbf{\Pi}_n = r\} = \mathbf{P}\{\mathbf{K}_n = r\}$ немедленно следует из этой леммы. ▶

Уместно заметить, что доказательство не предполагает независимости величин $\mathbf{X}_j$. Нужно лишь, чтобы все $n!$ перестановок $(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_n})$ были одинаково распределены. Другими словами, *теорема 2 остается справедливой для каждой системы n симметрично зависимых случайных величин* (см. гл. VII, 4). Конечно, общее распределение $\mathbf{K}_n$ и $\mathbf{\Pi}_n$ будет зависеть от совместного распределения $\mathbf{X}_j$. Интересен следующий пример. Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ независимы и имеют одно и то же распределение F. Положим $\mathbf{Y}_k = \mathbf{X}_k - \mathbf{S}_n/n$ (где $k = 1, \ldots, n$). Величины $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ симметрично зависимы и их частные суммы равны

$$\mathbf{\Sigma}_k = \mathbf{S}_k - \frac{k}{n}\mathbf{S}_n, \qquad k = 1, \ldots, n-1. \tag{8.12}$$

На графике $(\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_n)$ величина Σ_k равна вертикальному отклонению вершины $\mathbf{S}_k$ от хорды, соединяющей начало координат и концевую точку $(n, \mathbf{S}_n)$.

Допустим теперь, что распределение F непрерывно (с тем, чтобы избежать необходимости различать первый и последний максимумы). С вероятностью единица последовательность $0, \mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_{n-1}$ имеет единственный максимум. Циклической перестановке $(\mathbf{Y}_2, \ldots, \mathbf{Y}_n, \mathbf{Y}_1)$ соответствуют частные суммы $0, \mathbf{\Sigma}_2 - \mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_{n-1} - \mathbf{\Sigma}_1, -\mathbf{\Sigma}_1$, и ясно, что положение максимума смещается при этом на одно место вперед в циклическом порядке (если первоначальный максимум был на нулевом месте, то $\mathbf{\Sigma}_k < 0$ при $k = 1, \ldots, n$ и новый максимум будет на $(n-1)$-м месте). В n циклических перестановках максимум будет появляться ровно один раз на каждом месте и его положение равномерно распределено на множестве $\{0, 1, \ldots, n-1\}$. Мы получаем, таким образом, следующую теорему, принадлежащую Спарре-Андерсену и связанную с теоремой 3 из 1, гл. III, 2, касающейся бросаний монеты.

Теорема 3. *В любом случайном блуждании, порождаемом непрерывным распределением F, и при любом n число вершин среди* $\mathbf{S}_1, \ldots, \mathbf{S}_{n-1}$, *которые лежат выше хорды, соединяющей* $(0, 0)$ *и* $(n, \mathbf{S}_n)$, *с одной и той же вероятностью принимает каждое из значений* $0, 1, \ldots, n-1$.

(Это же верно и для номера максимально удаленной от хорды вершины.)

§ 9. Различные дополнения

(а) Совместные распределения

Для того чтобы получить совместное распределение лестничных величин, достаточно только изменить обозначения в рассуждениях, приводящих к теореме 7.1. Приспосабливая обозначения § 1, обозначим через I интервал на $\overline{0, \infty}$ и через $H_n^{(r)}(I)$

вероятность того, что n совпадает с r-м лестничным моментом и $\mathbf{S}_n \in I$. Положим

$$H\{I, s\} = \sum_{n=1}^{\infty} s^n H_n\{I\}, \quad 0 \leqslant s \leqslant 1. \tag{9.1}$$

По индукции можно установить, что при фиксированном s

$$H^{r*}\{I, s\} = \sum_{n=1}^{\infty} s^n H_n^{(r)}\{I\}. \tag{9.2}$$

Применяя доказательство теоремы 7.1, легко получаем следующий результат Г. Бакстера.

Теорема. *При $I \subset \overline{0, \infty}$ и $0 \leqslant s \leqslant 1$*

$$\sum_{r=1}^{\infty} \frac{1}{r} H^{r*}\{I, s\} = \sum_{n=1}^{\infty} \frac{s^n}{n} \mathbf{P}\{\mathbf{S}_n \in I\}. \tag{9.3}$$

При $I = \overline{0, \infty}$ это утверждение сводится к теореме 7.1. Более простая и доступная для изучения форма будет выведена в гл. XVIII, 3.

(б) Интерпретация производящих функций с помощью «исчезновения» [1])

Указываемая интерпретация может помочь интуиции и упростить формальные вычисления. Рассмотрим при фиксированном s, $0<s<1$, *несобственное* случайное блуждание, которое может оборваться на каждом шаге с вероятностью $1-s$, а в противном случае определяется распределением sF. Тогда $s^n F^{n*}\{I\}$ равно вероятности попасть в I в момент n. Дефект $1-s^n$ равен вероятности «исчезновения» процесса до момента n. Здесь проходят все прежние рассуждения с той разницей, что все распределения становятся несобственными. В частности, (9.1) *представляет собой распределение первой лестничной высоты в нашем случайном блуждании с исчезновением.* Функция (9.2) аналогична L^{r*} (см. § 2—3). Производящая функция $\tau(s)$ равна вероятности появления лестничного индекса.

(в) Рекуррентное событие

$$\{\mathbf{S}_1 \leqslant 0, \ \ldots, \ \mathbf{S}_{n-1} \leqslant 0, \ \mathbf{S}_n = 0\} \tag{9.4}$$

означает возвращение в нуль без предварительного захода на правую полуось. Оно появлялось в § 1 в определении слабых лестничных величин. Обозначим ω_n вероятность *первого* наступления события (9.4) в момент n:

$$\omega_n = \mathbf{P}\{\mathbf{S}_1 < 0, \ \ldots, \ \mathbf{S}_{n-1} < 0, \quad \mathbf{S}_n = 0\}. \tag{9.5}$$

[1]) В оригинале «mortality» — смертность. — *Прим. перев.*

Если $\omega(s)=\sum \omega_r s^r$, то ω^r будет производящей функцией для r-го наступления нашего события, а $1/[1-\omega(s)]$ будет производящей функцией для вероятностей (9.4). Упрощенный вариант доказательства теоремы 7.1 приводит к основному тождеству

$$\log \frac{1}{1-\omega(s)} = \sum_{n=1}^{\infty} \frac{s^n}{n} \mathbf{P}\{\mathbf{S}_n = 0\}. \tag{9.6}$$

Сравнивая его с (7.3), (7.11), (7.17) и т. д., видим, как легко можно перейти от слабых лестничных величин к строгим и обратно. Формула (9.6) подтверждает замечание § 1, что вероятности (9.4) не меняются, если все неравенства заменить противоположными.

(г) Обобщение на произвольные интервалы. Теория § 3 с тривиальными изменениями в обозначениях применима в ситуации, когда полупрямая $\overline{0, \infty}$ заменяется произвольным интервалом A, а полупрямая $-\infty, 0$ — дополнением A' к A. Сохраняется, в частности, без изменений уравнение Винера — Хопфа. Читателю предлагается проследить за деталями; полностью это сделано в гл. XVIII, 1 (см. также задачу 15).

§ 10. Задачи

1. Рассмотрим биномиальное случайное блуждание (пример (2.б)). Пусть e_k — математическое ожидание числа номеров $n \geqslant 0$, таких, что $\mathbf{S}_n = k$, $\mathbf{S}_1 \geqslant 0, \ldots, \mathbf{S}_{n-1} \geqslant 0$ (это событие означает попадание в k без предварительного захода на отрицательную полуось). Пусть $q \geqslant p$.

(а) Рассматривая случай $\mathbf{S}_1 = 1$ и момент первого возвращения в 1, покажите, что $e_k = p[e_{k-1} + e_k]$ при $k \geqslant 1$ и $e_0 = 1 + pe_0$. Выведите, что $e_k = (p/q)^k q^{-1}$ при $k = 0, 1, \ldots$.

(б) Пусть при $k \geqslant 1$ ψ_k обозначает математическое ожидание числа попаданий в k до первого возвращения в 0. Покажите, что (тривиально) $e_k = \psi_k e_0$. Это дает новое доказательство неожиданного результата примера (2.б).

2. *Продолжение.* При $q < p$ рекурсивная формула имеет вид $e_k = p[e_{k-1} + \frac{q}{p} e_k]$. Следовательно, $e_k = p^{-1}$ при $k = 0, 1, 2, \ldots$.

Примечание. Нижеследующие задачи 3—6 могут служить введением к настоящей главе и *могут быть решены до ее изучения*. Они содержат также примеры явных решений основных интегральных уравнений. Сверх того, они показывают силу и красоту метода производящих функций [попробуйте решить уравнение (1) непосредственно!].

3. Случайные величины $\mathbf{X}_k$, порождающие случайное блуждание, имеют одно и то же арифметическое распределение, приписывающее вероятности $f_1, f_2, \ldots$ целым числам 1, 2, ..., и вероятность q числу -1 ($q + f_1 + f_2 + \ldots = 1$). Обозначим через λ_r ($r = 1, 2, \ldots$) вероятность того, что первый положительный член последовательности $\mathbf{S}_1, \mathbf{S}_2, \ldots$ имеет величину r (иными словами, $\{\lambda_r\}$ есть распределение первой лестничной высоты). Покажите, что

(а) Числа λ_r удовлетворяют рекуррентным соотношениям

$$\lambda_r = f_r + q(\lambda_{r+1} + \lambda_1 \lambda_r). \tag{1}$$

(б) Производящие функции связаны равенством

$$\lambda(s) = 1 - \frac{f(s) + qs^{-1} - 1}{\lambda_1 q + qs^{-1} - 1}, \qquad 0 < s < 1. \tag{2}$$

(в) Если $\mathbf{E}(\mathbf{X}_k) = \mu = f'(1) - q > 0$, то существует единственный корень $0 < \sigma < 1$ уравнения

$$f(s) + \frac{q}{s} = 1. \tag{3}$$

Из того, что λ — монотонная функция и $\lambda < 1$ на $\overline{0,1}$, выводится, что

$$\lambda(s) = s\,\frac{\sigma}{q}\,\frac{f(s) - f(\sigma)}{s - \sigma}. \tag{4}$$

Это эквивалентно

$$\lambda_r = \frac{[f_r\sigma + f_{r+1}\sigma^2 + \ldots]}{q}. \tag{5}$$

(г) Если $\mathbf{E}(\mathbf{X}_k) < 0$, то соответствующее решение получаем, полагая в (2) $\lambda_1 = (1-q)/q$. При этом (4) и (5) выполняются с $\sigma = 1$.

4. Решите предыдущую задачу (изменив ее там, где нужно) для слабых лестничных высот. Другими словами, рассмотрите вместо λ_r вероятность того, что первый из неотрицательных членов последовательности $\mathbf{S}_1, \mathbf{S}_2, \ldots$ принимает значение r ($r = 0, 1, \ldots$).

Покажите, что (1) и (4) следует заменить на

$$\gamma_r = f_r + \frac{q}{1 - \gamma_0}\,\gamma_{r+1} \tag{1a}$$

и

$$\gamma(s) = 1 - \frac{q}{\sigma} + s\,\frac{f(s) - f(\sigma)}{s - \sigma}. \tag{4a}$$

5. Пусть в случайном блуждании задачи 3 x обозначает вероятность того, что $\mathbf{S}_n < 0$ при некотором n. Покажите (независимо от этой задачи), что x удовлетворяет уравнению (3) и что, следовательно, $x = \sigma$.

6. *Продолжение.* Покажите, что вероятность выполнения при каком-нибудь $n > 0$ неравенства $\mathbf{S}_n \leqslant 0$ равна $q + f(\sigma) = 1 - q(\sigma^{-1} - 1)$. Проверьте, что $\lambda'(1) = \mu\sigma[q(1-\sigma)]^{-1}$. Это — частный случай (2.7) (или тождество Вальда).

7. Выведите (1.11) из (1.10) непосредственным вычислением.

8. *Вероятности достижения.* Пусть $t \geqslant 0$ и $\xi > 0$. Обозначим $S(t, \xi)$ вероятность того, что первая сумма $\mathbf{S}_n$, превосходящая t, не будет превышать $t + \xi$. Докажите, что G удовлетворяет интегральному уравнению

$$G(t, \xi) = F(t + \xi) - F(t) + \int_{-\infty}^{t+} G(t - y, \xi)\, F\{dy\}.$$

В случае неединственности G является минимальным решением. Распределение H лестничной высоты однозначно определяется формулой $H(\xi) = G(0, \xi)$.

9. Пусть h тот член последовательности $\mathbf{S}_1, \mathbf{S}_2, \ldots$, который следует за первой лестничной высотой $\mathscr{H}_1$ и меньше ее. Распределение h равно $H * H^-$. Рассмотрите процесс восстановления, порождаемый h, и выведите из закона больших чисел (как в теореме 2.2), что верна

Теорема. *Если случайные величины* $\mathscr{H}_1$ и $\mathscr{H}_1^-$ *собственные и имеют математические ожидания, то*

$$\mathbf{E}(\mathscr{H}_1)+\mathbf{E}(\mathscr{H}_1^-)=\mathbf{E}(\mathbf{X}_1)=0.$$

10. *Аналитическое доказательство тождества Вальда* (2.7). Выведите из (3.11), что

$$1-F(x)=[1-\rho(0)][1-H(x)]+\int_{-\infty}^{0+}\rho\{dy\}[H(x-y)-H(x)],$$

$$F(x)=\int_{-\infty}^{x}\rho\{dy\}[1-H(x-y)]$$

при $x>0$ и $x<0$ соответственно.

Установите, что F имеет положительное математическое ожидание μ тогда и только тогда, когда H имеет конечное математическое ожидание ν и $\rho(0)<1$. Выведите интегрированием на $-\infty, \infty$, что $\mu=[1-\rho(0)]\nu$ (что эквивалентно (2.7)).

11. Покажите, исходя из (3.11), что если все три распределения имеют дисперсии [1]), то $\mathbf{E}(\mathbf{X}_1)=0$ и $\mathrm{Var}(\mathbf{X}_1)=-\mathbf{E}(\mathscr{H}_1)\mathbf{E}(\mathscr{H}_1^-)$.

12. *К примеру* (4. в). Если $\mu>0$, то знаменатель имеет положительный корень $s_0<1$, ровно $b-1$ комплексных корней в круге $|s|<s_0$ и $a-1$ комплексных корней в области $|s|>1$.

Случай $\mu<0$ можно описать заменой s на $1/s$.

13. Производящая функция для верхней лестничной высоты в примере (4. в) равна

$$\chi(s)=1-(1-s)\left(1-\frac{s}{\sigma_1}\right)\ldots\left(1-\frac{s}{\sigma_{a-1}}\right).$$

Для нижних лестничных высот результат получается заменой s/σ_k на σ_k/s.

14. *К примеру* (4. в). Допустим, что $\mathbf{X}_j$ принимают значения -2, -1, 0, 1, 2, с вероятностью 1/5 каждое. Покажите, что верхняя лестничная высота имеет распределение, для которого

$$\lambda_1=\frac{1+\sqrt{5}}{3+\sqrt{5}},\qquad \lambda_2=\frac{2}{3+\sqrt{5}}.$$

Для *слабых* высот $\overline{\lambda_0}=\frac{1}{10}(7-\sqrt{5})$, $\overline{\lambda_1}=\frac{1}{10}(1+\sqrt{5})$, $\overline{\lambda_2}=\frac{1}{5}$.

15. Обозначим в примере (4. в) через $\psi_k^{(n)}$ вероятность того, что первые n шагов не выводят из интервала $\overline{|-B, A|}$, а n-й шаг приводит в состояние k (таким образом, $\psi_k^{(n)}=0$ при $k>A$ и $k<-B$; как обычно, $\psi_k^{(0)}$ равно 1 при

---

[1]) Более точные результаты см. в **1**, гл. XVIII, 5, где $\mathbf{S}_N$ и $\mathbf{S}_{\hat{N}}$ играют роль $\mathscr{H}_1$ и $\mathscr{H}_1^-$.

$k=0$ и равно 0 в других случаях). Пусть $\psi_k = \sum \psi_k^{(n)}$ — математическое ожидание числа попаданий в k до первого выхода из $\overline{-B, A}$. Покажите, что

$$\psi_k = \sum_{\nu=-B}^{A} \psi_\nu f_{k-\nu} + \psi_k^{(0)}, \qquad -B \leqslant k \leqslant A,$$

и что при $k > A$ и $k < -B$ величина

$$\rho_k = \sum_{\nu=-B}^{A} \psi_\nu f_{k-\nu}$$

равна вероятности того, что *первый выход* из интервала $\overline{-B, A}$ приводит в точку k. [Эта задача важна для последовательного анализа. Она иллюстрирует ситуацию, описанную в (9. г).]

16. Из теоремы 7.2 вытекает, что при $\mu<0$

$$\sum n^{-1}\mathbf{P}\{\mathbf{S}_n > 0\} < \infty.$$

Проведите следующее *прямое доказательство*. Достаточно показать (вспомните неравенство Чебышева), что

$$\sum \int_{|y|>n} F\{dy\} < \infty, \qquad \sum \frac{1}{n^2} \int_{-n}^{n} y^2 F\{dy\} < \infty.$$

Первое утверждение очевидно. Для доказательства второго разобьем интеграл в сумму интегралов по областям $k-1<|y| \leqslant k$, $k=1, \ldots, n-1$. Меняя порядок суммирования, видим, что весь ряд $<2\mathbf{E}(|\mathbf{X}|)$.

17. Покажите, что при бросании симметричной монеты

$$\sum \frac{s^n}{n} \mathbf{P}\{\mathbf{S}_n = 0\} = \log \frac{2}{1+\sqrt{1-s^2}}.$$

Указание: представьте левую часть как интеграл от 0 до s от функции $[(1-x^2)^{-1/2} - 1]x^{-1}$.

18. Предположим, что случайное блуждание невозвратно, т. е что

$$U\{I\} = \sum_0^\infty F^{n*}\{I\} < \infty$$

для любого конечного интервала. Положим в обозначениях § 3 $\Phi = \sum_0^\infty \rho^{n*}$.

Докажите справедливость уравнения восстановления

$$U = \Phi + U * H.$$

Если ψ^- — аналог ψ для отрицательной полуоси, то $\psi^- = (1-\zeta)\Phi$, как и в (1.11).

19. Выведите, что

$$U = \frac{1}{1-\zeta} \psi * \psi^-,$$

и покажите, что это эквивалентно разложению Винера — Хопфа (3.12).

20. Выведите тождество Вальда $\mathbf{E}(\mathscr{H}_1) = \mathbf{E}(\mathscr{T}_1)\mathbf{E}(\mathbf{X}_1)$ непосредственно из уравнения восстановления задачи 18.

ГЛАВА XIII

ПРЕОБРАЗОВАНИЕ ЛАПЛАСА. ТАУБЕРОВЫ ТЕОРЕМЫ. РЕЗОЛЬВЕНТЫ

Преобразование Лапласа служит мощным практическим орудием. В то же самое время соответствующая теория интересна и сама по себе и по связи ее с другими разделами, например с теорией полугрупп. Теорема о вполне монотонных функциях и основная тауберова теорема с полным правом рассматривались как жемчужины математического анализа. (Хотя приводимые доказательства просты и элементарны, первоначальные исследования в этом направлении требовали оригинальности и силы.) Резольвенты (§ 9—10) являются основой теории полугрупп.

Материал этой главы предназначен для использования в различных целях. Поэтому были приложены значительные усилия, во-первых, чтобы сделать части настолько независимыми друг от друга, насколько это позволяет тема, и, во-вторых, чтобы дать читателю возможность опускать детали. Для дополнительного чтения может служить гл. XIV. Там же можно найти примеры. Остальная часть книги совершенно не зависит от настоящей главы.

Несмотря на то что нам часто будут встречаться правильно меняющиеся функции, мы будем пользоваться только совсем элементарной теоремой 1 из гл. VIII, 8.

§ 1. Определения. Теорема непрерывности

Определение 1. *Пусть F — собственное или несобственное распределение вероятностей, сосредоточенное на $\overline{0, \infty}$. Преобразованием Лапласа φ распределения F называют функцию, определенную для $\lambda \geqslant 0$ равенством*

$$\varphi(\lambda) = \int_{0}^{\infty} e^{-\lambda x} F\{dx\}. \tag{1.1}$$

Здесь и далее подразумевается, что *интервал интегрирования замкнут* (и может быть заменен на $\overline{-\infty, \infty}$). Когда мы говорим о преобразовании Лапласа распределения F, мы неявно

предполагаем, что F сосредоточено на $[0, \infty)$. Как обычно, мы расширяем смысл терминов и говорим о «преобразовании Лапласа случайной величины $\mathbf{X}$», понимая под этим преобразование Лапласа соответствующего распределения. Применяя привычное обозначение для математического ожидания, имеем

$$\varphi(\lambda) = \mathbf{E}(e^{-\lambda \mathbf{X}}). \tag{1.2}$$

Пример. а) Пусть $\mathbf{X}$ принимает значения 0, 1, ... с вероятностями $p_0, p_1, \ldots$. Тогда $\varphi(\lambda) = \Sigma p_n e^{-n\lambda}$. Так как производящая функция равна $P(s) = \Sigma p_n s^n$, то $\varphi(\lambda) = P(e^{-\lambda})$, и преобразование Лапласа отличается от производящей функции только заменой переменных $s = e^{-\lambda}$. Этим объясняется большое сходство свойств преобразований Лапласа и производящих функций. ▶

Польза преобразований Лапласа была бы ограниченной, если бы распределение не восстанавливалось по соответствующему преобразованию. Формулы обращения (§ 4) показывают, как вычислить распределение F, если известно его преобразование; здесь же мы приведем предварительный результат.

Теорема 1[1]). *Различным распределениям вероятностей соответствуют различные преобразования Лапласа.*

Доказательство. Положим $y = e^{-x}$. При изменении x от 0 до ∞ y пробегает интервал $[0, 1]$. Полагая $G(y) = 1 - F(x)$ в точках непрерывности, получим распределение G, сосредоточенное на $[0, 1]$. Теперь $\varphi(\lambda)$ есть предел римановых сумм $\Sigma e^{-\lambda x_k}[F(x_{k+1}) - F(x_k)]$, а последние совпадают с римановыми суммами $\Sigma y_k^{\lambda}[G(y_k) - G(y_{k+1})]$ для математического ожидания y^{λ} по отношению к G. Следовательно, $\varphi(k)$ есть k-й момент G. Поэтому значение $\varphi(1)$, $\varphi(2)$, ... определяют G, а вместе с тем и F. Этот результат сильнее, чем утверждение теоремы (дальнейшее обобщение см. в задаче 11). ▶

Следующий ниже важный результат просто выводится из теоремы 1.

Теорема 2. *(Теорема непрерывности.) Пусть F_n, $n = 1, 2, \ldots$ распределение вероятностей с преобразованием φ_n.*

Если $F_n \to F$, где F — (возможно, несобственное) распределение с преобразованием φ, то $\varphi_n(\lambda) \to \varphi(\lambda)$ при $\lambda > 0$.

---

[1]) Эта теорема вытекает из гл. VII, (6.4); последнее соотношение выводится здесь заново (см. (4.4)).

Обратно, если последовательность $\{\varphi_n(\lambda)\}$ *сходится при каждом* $\lambda>0$ *к пределу* $\varphi(\lambda)$, *то* φ — *преобразование (возможно, несобственного) распределения* F *и* $F_n \to F$.

Предел F *будет собственным в том и только том случае, когда* $\varphi(\lambda) \to 1$ *при* $\lambda \to 0$.

Доказательство. Первая часть содержится в основной теореме о сходимости гл. VIII, 1. Для доказательства второй части привлечем теорему о выборе (теорема 1 из гл. VIII, 6). Пусть $\{F_{nk}\}$ — подпоследовательность, сходящаяся к (возможно, несобственному) распределению F. Если $\varphi_n(\lambda) \to \varphi(\lambda)$, то F является тем единственным распределением, которое имеет преобразование Лапласа φ. Поэтому все сходящиеся подпоследовательности сходятся к одному и тому же пределу F. Отсюда вытекает, что F_n сходится к F. Последнее утверждение ясно из (1.1). ▶

Для ясности изложения мы сохраним всюду, где это возможно, букву F для обозначения распределения вероятностей. Вместо (1.1) мы могли бы взять более общий интеграл вида

$$\omega(\lambda) = \int_0^\infty e^{-\lambda x} U\{dx\}, \tag{1.3}$$

где U — некоторая мера. Важный частный случай представляют интегралы вида

$$\omega(\lambda) = \int_0^\infty e^{-\lambda x} u(x)\,dx, \tag{1.4}$$

где $u \geqslant 0$. Этот интеграл сходится при всех $\lambda \geqslant 0$, если u интегрируема на $0, \infty$. Однако при $\lambda>0$ он может сходиться и для неинтегрируемых u.

Примеры. б) Если $u(x)=x^a$ с $a \geqslant 0$, то $\omega(\lambda)=\Gamma(a+1)/\lambda^{a+1}$ для всех $\lambda>0$.

в) Если $u(x)=e^{ax}$, то $\omega(\lambda)=1/(\lambda-a)$ для $\lambda>a>0$. Интеграл (1.4) расходится для $\lambda \leqslant a$.

г) Если $u(x)=e^{x^2}$, то интеграл (1.4) расходится всюду. ▶

Мы будем интересоваться главным образом мерами U, полученными простыми преобразованиями из распределений вероятностей, и интеграл в (1.3) будет, как правило, сходящимся при всех $\lambda>0$. Однако, исключая меры, для которых сходимость имеет место только при *некотором* λ, мы не получаем никакого выигрыша. Далее из $\omega(a)<\infty$ вытекает $\omega(\lambda)<\infty$ для всех

$\lambda > a$, так что значения λ, для которых интеграл (1.3) сходится, заполняют промежуток a, ∞.

Определение 2. *Пусть U — мера, сосредоточенная на* $0, \infty$. *Если интеграл (1.3) сходится при $\lambda > a$, то определенная при $\lambda > a$ функция ω называется преобразованием Лапласа меры U.*

Если U имеет плотность u, то преобразование Лапласа (1.4) *меры U называется также обычным преобразованием Лапласа функции u.*

Последнее соглашение принимается единственно ради удобства. Чтобы быть систематичным, следовало бы рассмотреть интеграл более общего типа

$$\int_0^\infty e^{-\lambda x} v(x)\, U\{dx\} \tag{1.5}$$

и назвать его «преобразованием Лапласа функции v по отношению к мере U». Тогда преобразование (1.4) получило бы название: «преобразование u по отношению к мере Лебега». Этот подход мог бы иметь теоретическое преимущество, если бы позволял рассматривать знакопеременные функции u и v. Однако для целей настоящей книги самый простой и самый ясный способ действий состоит в том, чтобы связывать понятие преобразования Лапласа только с мерами. Это мы и будем делать [1]).

Если U такая мера, что интеграл (1.3) сходится при $\lambda = a$, то функция

$$\omega(\lambda + a) = \int_0^\infty e^{-\lambda x} e^{-ax} U\{dx\} \tag{1.6}$$

определена при всех $\lambda > 0$ и представляет собой преобразование Лапласа ограниченной меры $U^{\#}\{dx\} = e^{-ax} U\{dx\}$, а $\omega(\lambda + a)/\omega(a)$ является преобразованием Лапласа некоторого распределения *вероятностей*. Поэтому каждая теорема, касающаяся преобразований распределений вероятностей, автоматически распространяется на более широкий класс мер. Поскольку $\omega(\lambda + a)$ получается «сдвигом» ω, мы будем называть

[1]) Терминология еще не вполне сложилась, и в литературе термин «преобразование Лапласа для F» может обозначать как (1.1), так и (2.7). Мы бы назвали (2.7) «обычным преобразованием Лапласа функции распределения F», но в учебниках, которые, как правило, рассматривают только такие преобразования, прилагательное «обычный» бывает опущено. В подобных случаях во избежание путаницы преобразование (1.1) называют преобразованием *Лапласа — Стильтьеса*.

указанный полезный метод «*принципом сдвига*». Заметим, что U однозначно определяется по $U^{\#}$, а $U^{\#}$ — по $\omega(\lambda+a)$ при $\lambda>0$. Поэтому мы можем перефразировать теорему 1 следующим образом.

Теорема 1а. *Мера однозначно определяется значениями соответствующего преобразования Лапласа на некотором интервале* a, ∞.

Теорема 2а. *(Обобщенная теорема непрерывности.) Пусть* U_n $(n=1, 2, \ldots)$ *меры с преобразованиями Лапласа* ω_n. *Если* $\omega_n(\lambda) \to \omega(\lambda)$ *при* $\lambda>a$, *то* ω *является преобразованием Лапласа некоторой меры* U *и*[1]) $U_n \to U$.

Обратно, если $U_n \to U$ *и последовательность* $\{\omega_n(a)\}$ *ограничена, то* $\omega_n(\lambda) \to \omega(\lambda)$ *при* $\lambda>a$.

Доказательство. При фиксированном $\lambda_0>a$ функция $\omega_n(\lambda+\lambda_0)/\omega_n(\lambda_0)$ представляет собой преобразование Лапласа распределения вероятностей

$$U_n^{\#}\{dx\} = \frac{1}{\omega_n(\lambda_0)} e^{-\lambda_0 x} U_n\{dx\}.$$

Применяя теорему 2 к последовательности $\{U_n^{\#}\}$, получаем теорему 2а. ▶

Следующий пример пояснит, зачем нужно условие ограниченности $\{\omega_n(a)\}$.

Пример. д) Пусть U_n приписывает вес e^{n^2} точке n и нуль — дополнению этой точки. Так как $U_n\{0, n\}=0$, мы имеем $U_n \to 0$. Однако $\omega_n(\lambda)=e^{n(r-\lambda)} \to \infty$ при всех $\lambda>0$. ▶

Иногда говорят о *двусторонней преобразовании* Лапласа для распределения F, не сосредоточенного на положительной полупрямой, а именно

$$\varphi(\lambda) = \int_{-\infty}^{+\infty} e^{-\lambda x} F\{dx\}. \tag{1.7}$$

Однако этот интеграл не обязан существовать при $\lambda \neq 0$. Если он существует для $\lambda \neq 0$, то $\varphi(-\lambda)$ называют обычно *производящей функцией моментов*, хотя на самом деле она служит производящей для последовательности $\{\mu_n/n!\}$, где μ_n — n-й момент.

---

[1]) Напомним, что, согласно гл. VIII, 1, $U_n \to U$ тогда и только тогда, когда $U_n\{I\} \to U\{I\}$ для каждого *ограниченного* интервала непрерывности U.

§ 2. Элементарные свойства

В этом параграфе мы перечислим наиболее часто используемые свойства преобразований Лапласа. Аналогии с производящими функциями очевидны.

(i) *Свертки.* Пусть F и G — распределения вероятностей и U — их свертка, т. е.

$$U(x)=\int_0^x G(x-y)\,F\{dy\}. \tag{2.1}$$

Соответствующие преобразования Лапласа подчиняются *правилу умножения*

$$\omega=\varphi\gamma. \tag{2.2}$$

Последнее равносильно утверждению, что для независимых случайных величин $\mathbf{E}(e^{-\lambda(X+Y)})=\mathbf{E}(e^{-\lambda X})\,\mathbf{E}(e^{-\lambda Y})$, что есть специальный случай правила умножения математических ожиданий [1]).

Рассмотрим более общий случай двух произвольных мер F и G, сосредоточенных на $0,\infty$ и таких, что их преобразования Лапласа φ и γ существуют при $\lambda\geqslant a$. Свертка (2.1) определена, и мы покажем, что *правило умножения* (2.2) *остается в силе.* Для этого введем *ограниченную* меру $F^{\#}\{dx\}=e^{-ax}F\{dx\}$ с преобразованием Лапласа $\varphi(\lambda+a)$. Аналогично определим $G^{\#}$ и $U^{\#}$. Умножая (2.1) на e^{-ax}, видим, что свертка $F^{\#}$ и $G^{\#}$ равна $U^{\#}$, так что $\omega(\lambda+a)=\varphi(\lambda+a)\gamma(\lambda+a)$. Последнее равносильно (2.2).

В терминах плотностей (2.1) имеет вид

$$u(x)=\int_0^x g(x-y)\,f(y)\,dy. \tag{2.3}$$

Поэтому правило умножения (2.2) справедливо и для обычных преобразований неотрицательных функций (см. задачу 1).

Примеры. а) *Гамма-распределения.* Плотность $f_\alpha(x)=\dfrac{x^{\alpha-1}}{\Gamma(\alpha)}e^{-x}$ имеет преобразование $\varphi_\alpha(\lambda)=(1+\lambda)^{-\alpha}$. Правило свертки $f_\alpha * f_\beta=f_{\alpha+\beta}$ отражается в очевидном соотношении $\varphi_\alpha\varphi_\beta=\varphi_{\alpha+\beta}$.

[1]) Обратное неверно: две случайные величины могут быть зависимыми, но распределение их суммы определяется по формуле свертки [см. гл. II (4д) и задачу 1 в гл. III, 9].

б) *Степени.* Положим $u_\alpha(x)=x^{\alpha-1}$. Обычное преобразование Лапласа для u_α равно $\omega_\alpha(\lambda)=\lambda^{-\alpha}\Gamma(\alpha)$. Отсюда следует, что свертка (2.3) функций u_α и u_β равна

$$u(x)=\frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}\,x^{\alpha+\beta-1}. \tag{2.4}$$

Это, впрочем, можно вывести из примера (а) с помощью принципа сдвига.

в) *Если $a>0$, то $e^{-a\lambda}\omega(\lambda)$ есть преобразование Лапласа для меры с функцией распределения $U(x-a)$*, т. е. для меры, приписывающей интервалу I вес $U\{I-a\}$. Это очевидным образом следует из определения, но может рассматриваться и как специальный случай теоремы о свертке, так как $e^{-a\lambda}$ является преобразованием распределения, сосредоточенного в точке a. ▶

(ii) *Производные и моменты.* Если F — распределение вероятностей и φ — его преобразование Лапласа (1.1), то φ *обладает производными всех порядков, причем*

$$(-1)^n\varphi^{(n)}(\lambda)=\int_0^\infty e^{-\lambda x}x^n F\{dx\} \tag{2.5}$$

(как всегда, $\lambda>0$). Дифференцирование под знаком интеграла законно, так как формальное дифференцирование приводит к ограниченному и непрерывному подинтегральному выражению.

Из (2.5) вытекает, в частности, что F *имеет конечный n-й момент в том и только том случае, когда существует конечный предел* $\varphi^n(0)$. Поэтому для случайной величины $\mathbf{X}$ мы можем написать

$$\mathbf{E}(\mathbf{X})=-\varphi'(0),\quad \mathbf{E}(\mathbf{X}^2)=\varphi''(0) \tag{2.6}$$

с очевидным соглашением о смысле этих равенств в случае расходимости. Правило дифференцирования (2.5) верно и для произвольных мер F.

(iii) *Интегрирование по частям* приводит от (1.1) к

$$\int_0^\infty e^{-\lambda x}F(x)\,dx=\frac{\varphi(\lambda)}{\lambda},\quad \lambda>0. \tag{2.7}$$

Для распределений вероятностей (2.7) иногда предпочтительнее записывать в форме, содержащей «хвост» распределения,

$$\int_0^\infty e^{-\lambda x}[1-F(x)]\,dx=\frac{1-\varphi(\lambda)}{\lambda}. \tag{2.8}$$

Это соответствует формуле 1, гл. XI, (1.6) для производящих функций.

(iv) *Изменение масштаба.* Из (1.2) выводим, что при любом фиксированном $a>0$ $\mathbf{E}(e^{-a\lambda \mathbf{X}})=\varphi(a\lambda)$, т. е. $\varphi(a\lambda)$ *является преобразованием для распределения* $F\{dx/a\}$ [с функцией распределения $F(x/a)$]. Это соотношение часто используется.

Пример. г) *Закон больших чисел.* Пусть $\mathbf{X}_1$, $\mathbf{X}_2$, ... — независимые случайные величины с одним и тем же преобразованием Лапласа φ. Положим $\mathbf{E}(\mathbf{X}_j)=\mu$. Преобразование Лапласа для суммы $\mathbf{X}_1+\ldots+\mathbf{X}_n$ равно φ^n. Поэтому преобразование для среднего арифметического $[\mathbf{X}_1+\ldots+\mathbf{X}_n]/n$ равно $\varphi^n(\lambda/n)$. Вблизи нуля $\varphi(\lambda)=1-\mu\lambda+o(\lambda)$ [см. (2.6)], так что при $n\to\infty$

$$\lim \varphi^n\left(\frac{\lambda}{n}\right)=\lim\left(1-\frac{\mu\lambda}{n}\right)^n=e^{-\mu\lambda}. \tag{2.9}$$

Но $e^{-\mu\lambda}$ есть преобразование Лапласа для распределения, сосредоточенного в точке μ, следовательно, распределение $[\mathbf{X}_1+\ldots+\mathbf{X}_n]/n$ сходится к этому пределу. Мы получили (слабый) закон больших чисел в форме Хинчина, не требующей существования дисперсии. Доказательство, правда, применимо только к неотрицательным случайным величинам, но оно иллюстрирует изящество метода, основанного на преобразованиях Лапласа. ▶

§ 3. Примеры

а) *Равномерное распределение.* Пусть F обозначает равномерное на $\overline{0,\ 1}$ распределение. Его преобразование Лапласа равно $\varphi(\lambda)=(1-e^{-\lambda})/\lambda$. Поэтому n-кратная свертка F^{n*} имеет преобразование

$$\varphi^n(\lambda)=\sum_{k=0}^{n}(-1)^k\binom{n}{k}e^{-\lambda k}\lambda^{-n}. \tag{3.1}$$

Функция λ^{-n} служит преобразованием для $U(x)=x^n/n!$. Пример (2, в) показывает, что функция $e^{-k\lambda}\lambda^{-n}$ соответствует $(x-k)_+^n/n!$, где x_+ обозначает функцию, равную 0 для $x\leqslant 0$ и x для $x\geqslant 0$. Таким образом,

$$F^{n*}(x)=\frac{1}{n!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}(x-k)_+^n. \tag{3.2}$$

Эта формула была получена прямым вычислением в гл. I, (9.5) и предельным переходом в задаче 20. из 1, гл. XI.

б) *Устойчивое распределение с показателем* $1/2$. Функция распределения

$$G(x)=2[1-\mathfrak{N}(1/\sqrt{x})], \quad x>0 \tag{3.3}$$

(где $\mathfrak{N}$ — стандартное нормальное распределение) встретилась нам в предельной теореме 2 из **1**, гл. III, 6. Мы используем этот результат для вычисления преобразования Лапласа γ для G. Рассмотрим простое симметричное случайное блуждание (бросание монеты) и обозначим через $\mathbf{T}$ момент первого возвращения в нуль. Цитированная теорема утверждает, что G является предельным распределением для нормированной суммы $\mathbf{S}_n = (\mathbf{T}_1+\ldots+\mathbf{T}_n)/n^2$, где $\mathbf{T}_1, \mathbf{T}_2, \ldots$ — независимые случайные величины, распределенные, как $\mathbf{T}$. В соответствии с **1**, гл. XI, (3.11) производящая функция для $\mathbf{T}$ равна $f(s)=1-\sqrt{1-s^2}$. Следовательно, $\mathbf{S}_n$ имеет преобразование Лапласа $\omega_n(\lambda)=f^n(e^{-\lambda/n^2})$. Очевидно, что при $s\to 1$ $\log f(s)\sim -\sqrt{2(1-s)}$. Поэтому при $n\to\infty$ $\log\omega_n(\lambda)\to-\sqrt{2\lambda}$ и *преобразование Лапласа для* G *равно* $\gamma(\lambda)=e^{-\sqrt{2\lambda}}$. Этот результат можно проверить элементарными, но громоздкими вычислениями.

Мы несколько раз отмечали, что *распределение* G *является устойчивым*, но опять-таки прямая проверка требует трудоемких вычислений. В то же время очевидно, что $\gamma^n(\lambda)=\gamma(n^2\lambda)$, т. е. $G^{n*}(x)=G(n^{-2}x)$, чем без всякого труда доказывается устойчивость G.

в) *Степенные ряды и смеси.* Пусть F — распределение вероятностей с преобразованием Лапласа $\varphi(\lambda)$. Мы постоянно сталкивались с распределениями вида

$$G=\sum_{k=0}^{\infty} p_k F^{n*}, \tag{3.4}$$

где $\{p_k\}$ — распределение вероятностей. Если $P(s)=\sum p_k s^k$ — производящая функция для $\{p_k\}$, то преобразование Лапласа для G равно, очевидно,

$$\gamma(\lambda)=P(\varphi(\lambda)). \tag{3.5}$$

Это соотношение распространяется на все степенные ряды с положительными коэффициентами. Перейдем к примерам.

г) *Бесселевы плотности.* Мы видели в примере гл. II, (7, c), что при $r=1, 2, \ldots$ плотность [1])

$$v_r(x)=e^{-x}\frac{r}{x}I_r(x) \tag{3.6}$$

[1]) I_r — фракция Бесселя, определенная в гл. II, (7.1).

соответствует распределению вида (3.4), где F имеет показательное распределение с $\varphi(\lambda)=1/(\lambda+1)$ и $\{p_k\}$ — распределение момента первого перехода через точку $r>0$ в обычном симметричном случайном блуждании. Производящая функция последнего распределения равна

$$P(s)=\left(\frac{1-\sqrt{1-s^2}}{s}\right)^r \tag{3.7}$$

[см. 1, гл. XI, (3.4)]. Подставляя сюда $s=(1+\lambda)^{-1}$, мы видим, что *обычное преобразование Лапласа для плотности вероятности* (3.6) *равно*

$$[\lambda+1-\sqrt{(\lambda+1)^2-1}]^r. \tag{3.8}$$

Утверждение о том, что v_r *является плотностью вероятности, преобразование которой равно* (3.8), доказано нами только для $r=1, 2, \ldots$. Однако оно верно [1]) для *всех* $r>0$ и в этом смысле представляет вероятностный интерес. В самом деле, отсюда вытекает формула свертки $v_r * v_s = v_{r+s}$ и, следовательно, *безграничная делимость* v_r (см. § 7).

д) *Другие бесселевы плотности.* Выберем в (3.4) в качестве F показательное распределение с $\varphi(\lambda)=1/(1+\lambda)$, а в качестве $\{p_k\}$ — распределение Пуассона с $P(s)=e^{-t+ts}$. Легко дать явную формулу для G, но, к счастью, эта задача уже была решена в примере (7, а) гл. II. Мы видели там, что плотность

$$w_\rho(x)=e^{-t-x}\sqrt{\left(\frac{x}{t}\right)^\rho}\, I_\rho(2\sqrt{tx}), \tag{3.9}$$

определенная в гл. II, (7.2), является сверткой нашего распределения G с гамма-плотностью $f_{1,\,\rho+1}$. Отсюда следует, что обычное преобразование Лапласа для w_ρ равно произведению γ на преобразование для $f_{1,\,\rho+1}$, т. е. на $(\lambda+1)^{-(\rho+1)}$. Соответственно *плотность вероятности* (3.9) *имеет преобразование Лапласа*

$$\frac{1}{(\lambda+1)^{\rho+1}}\, e^{-t+t/(\lambda+1)}. \tag{3.10}$$

При $t=1$, используя принцип сдвига (1.6), получаем, что $\sqrt{x^\rho}\, I_\rho(2\sqrt{x})$ *имеет обычное преобразование* $\lambda^{-\rho-1}e^{1/\lambda}$.

§ 4. Вполне монотонные функции. Формулы обращения

Как мы видели в гл. XII, 2, функция f, заданная на $\overline{0,\ 1}$, является производящей функцией для положительной последова-

[1]) Результат принадлежит H. Weber. Элементарное доказательство опубликовано в *J. Soc. Industr. Appl. Math.*, vol. 14 (1966).

тельности $\{f_n\}$ тогда и только тогда, когда f абсолютно монотонна (т. е. когда f имеет неположительные производные $f^{(n)}$ всех порядков). Аналогичная теорема имеет место и для преобразований Лапласа. Разница лишь в том, что теперь знаки производных чередуются.

Определение 1. *Заданная на* $\overline{0, \infty}$ *функция* φ *называется вполне монотонной, если она имеет производные* $\varphi^{(n)}$ *всех порядков и*

$$(-1)^n \varphi^{(n)}(\lambda) \geqslant 0, \quad \lambda > 0. \tag{4.1}$$

При $\lambda \to 0$ значения $\varphi^{(n)}(\lambda)$ приближаются к конечным или бесконечным пределам, которые мы обозначим $\varphi^{(n)}(0)$. Типичные примеры — функции $1/\lambda$ и $1/(1+\lambda)$.

С. Н. Бернштейну (1928) принадлежит следующая прекрасная теорема, которая послужила отправным пунктом многих исследований (ее доказательство упрощалось шаг за шагом [1]).

Теорема 1. *Функция* φ *на* $\overline{0, \infty}$ *является преобразованием Лапласа распределения вероятностей* F *тогда и только тогда, когда она вполне монотонна и* $\varphi(0) = 1$.

Теорема может быть сформулирована и в другой, эквивалентной форме.

Теорема 1а. *Функция* φ *на* $\overline{0, \infty}$ *является вполне монотонной тогда и только тогда, когда она имеет вид*

$$\varphi(\lambda) = \int_0^\infty e^{-\lambda x} F\{dx\}, \quad \lambda > 0, \tag{4.2}$$

где F *(может быть, бесконечная) мера на* $\overline{0, \infty}$.

(По нашему первоначальному соглашению промежуток интегрирования всегда *замкнут*; наличие у F атома в нуле приводит к $\varphi(\infty) > 0$.)

Приводимое доказательство непосредственно применимо и к кажущейся более общей теореме 1а. Однако читатель может легко проверить сам, используя принцип сдвига (1.6), что второй вариант на самом деле вытекает из первого.

Доказательство. Необходимость условия устанавливается дифференцированием [ср. (2.5)]. Предположим теперь, что φ вполне монотонна. Рассмотрим $\varphi(a - as)$ при фиксированном $a > 0$ как функцию от s, $0 < s < 1$. Производные этой функции, очевидно, положительны и по теореме 3 из гл. VII, 2

[1]) См. дальнейшие результаты в задачах 10 и 11.

разложение Тейлора

$$\varphi(a-as)=\sum_{n=0}^{\infty}\frac{(-a)^n\varphi^{(n)}(a)}{n!}s^n \tag{4.3}$$

служит производящей функцией арифметической меры, сосредоточенной на множестве точек 0, 1, 2, ... Преобразование Лапласа этой меры получается заменой s на $e^{-\lambda}$. Поэтому $\varphi_a(\lambda)=\varphi(a-ae^{-\lambda/a})$ является преобразованием Лапласа некоторой меры F_a. Так как при $a\to\infty$ $\varphi_a(\lambda)\to\varphi(\lambda)$, то по обобщенной теореме непрерывности предел преобразований Лапласа мер F_a сам будет преобразованием Лапласа некоторой меры F. ▷

Доказательство закончено, но стоит отметить, что попутно мы доказали соотношение $F=\lim F_a$, где F_a — арифметическое распределение, приписывающее вес $(-a)^n\varphi^{(n)}(a)/n!$ точке n/a. Поэтому верна

Теорема 2. *(Формула обращения.) Если* (4.2) *выполняется при* $\lambda>0$, *то во всех точках непрерывности* [1])

$$F(x)=\lim_{a\to\infty}\sum_{n\leqslant ax}\frac{(-a)^n}{n!}\varphi^{(n)}(a). \tag{4.4}$$

Последняя формула представляет значительный теоретический интерес и приводит к различным полезным заключениям. Примером может служить следующий критерий ограниченности, применяемый в теории полугрупп [см. задачу 9].

Следствие. *Для того чтобы функция* φ *представлялась в форме*

$$\varphi(\lambda)=\int_0^{\infty}e^{-\lambda x}f(x)\,dx,\ \text{где}\ 0\leqslant f\leqslant C, \tag{4.5}$$

необходимо и достаточно, чтобы при всех $a>0$

$$0\leqslant\frac{(-a)^n\varphi^{(n)}(a)}{n!}\leqslant\frac{C}{a}. \tag{4.6}$$

Доказательство. Дифференцируя (4.5) под знаком интеграла, получаем (4.6) [ср. (2.5)]. Обратно, из (4.6) видно, что φ вполне монотонна и потому является преобразованием некоторой меры F. Из (4.4) выводим

$$F(x_2)-F(x_1)\leqslant C(x_2-x_1)$$

[1]) Формула обращения (4.4) была получена в гл. VII, (6.4) как прямое следствие закона больших чисел. В гл. VII, (6.6) указана аналогичная *формула для интегралов вида* (4.5) *с непрерывной* (не обязательно положительной) f.

для любых $x_1<x_2$. Иными словами, F имеет ограниченные разностные отношения и, следовательно, представляет собой интеграл от функции $f\leqslant C$ (см. гл. V, 3). ▶

Теорема 1 приводит к простым *критериям* того, будет ли данная функция преобразованием Лапласа некоторого распределения вероятностей. Для иллюстрации стандартных приемов докажем

Критерий 1. *Если* φ *и* ψ *вполне монотонны, то таково же и их произведение* $\varphi\psi$.

Доказательство. Мы покажем по индукции, что знаки производных $\varphi\psi$ чередуются. Допустим, что для *каждой* пары φ и ψ вполне монотонных функций знаки первых n производных $\varphi\psi$ чередуются. Так как $-\varphi'$ и $-\psi'$ вполне монотонны, то предположение индукции применимо к произведениям $-\varphi'\psi$ и $-\varphi\psi'$. Из равенства $-(\varphi\psi)'=-\varphi'\psi-\varphi\psi'$ заключаем, что знаки первых $(n+1)$-х производных $\varphi\psi$ чередуются. Так как при $n=1$ утверждение тривиально верно, критерий доказан. ▶

Точно так же доказывается полезный

Критерий 2. *Если* φ *вполне монотонна и* ψ *— положительная функция с вполне монотонной производной, то* $\varphi(\psi)$ *вполне монотонна (в частности, функция* $e^{-\psi}$ *вполне монотонна).*

Типичные применения даны в § 6, а также в следующем примере, который часто излагается в литературе с ненужными усложнениями.

Пример. а) *Уравнение, встречающееся в теории ветвящихся процессов.* Пусть φ — преобразование Лапласа для распределения вероятностей F с математическим ожиданием μ, $0<\mu\leqslant\infty$. Пусть $c>0$. Мы докажем, что *уравнение*

$$\beta(\lambda)=\varphi(\lambda+c-c\beta(\lambda)) \tag{4.7}$$

имеет единственный корень $\beta(\lambda)\leqslant 1$ *и что* β *есть преобразование Лапласа для некоторого распределения* B. *Оно будет собственным при* $\mu c\leqslant 1$ *и несобственным в других случаях.*

Доказательство. Рассмотрим уравнение

$$\varphi(\lambda+c-cs)-s=0 \tag{4.8}$$

при фиксированном $\lambda>0$ и $0\leqslant s\leqslant 1$. Левая часть его — выпуклая функция, значение которой при $s=1$ отрицательно, а при $s=0$ — положительно. Отсюда следует существование единственного корня.

Докажем, что корень $\beta(\lambda)$ есть преобразование Лапласа. Положим $\beta_0=0$ и $\beta_{n+1}=\varphi(\lambda+c-c\beta_n)$. Тогда $\beta_0\leqslant\beta_1\leqslant 1$, и так как φ убывает, то $\beta_1\leqslant\beta_2\leqslant 1$. По индукции $\beta_n\leqslant\beta_{n+1}\leqslant 1$. Предел ограниченной монотонной последовательности $\{\beta_n\}$ удовлетворяет (4.7) и, следовательно, $\beta=\lim\beta_n$. Функция $\beta_1(\lambda)=\varphi(\lambda+c)$ вполне монотонна, и критерий 2 показывает, что $\beta_2, \beta_3, \ldots$ также вполне монотонны. По теореме непрерывности то же самое верно для предела β, и остается только узнать, будет ли $\beta(0)=1$ или нет. По построению $s=\beta(0)$ является *наименьшим* корнем уравнения (4.8) при $\lambda=0$. Далее, $s=1$ является корнем. Из выпуклости графика $\varphi(c-cs)$ вытекает, что второй корень $s<1$ существует тогда и только тогда, когда угловой коэффициент графика при $s=1$ больше единицы (т. е. если $-c\varphi'(0)>1$). Только в этом случае распределение, соответствующее β, будет несобственным. Доказательство закончено. (Применения и ссылки см. в гл. XIV, 4.) ▶

§ 5. Тауберовы теоремы

Пусть U — мера, сосредоточенная на $\overline{0, \infty}$ и такая, что ее преобразование Лапласа

$$\omega(\lambda)=\int_0^\infty e^{-\lambda x}U\{dx\} \tag{5.1}$$

существует при $\lambda>0$. Удобно описывать меру U в терминах соответствующей «функции распределения», определенной при $x\geqslant 0$, как $U\{\overline{0, x}\}$. Мы увидим, что при весьма общих условиях поведение ω вблизи нуля однозначно определяет асимптотическое поведение $U(x)$ при $x\to\infty$. Любое подобное соотношение между $\omega(\lambda)$ и $U(x)$ называется тауберовой теоремой. Простейший частный случай получаем, замечая, что мера U конечна тогда и только тогда, когда $\omega(\lambda)$ стремится к конечному пределу $\omega(0)$ при $\lambda\to 0$. В этом случае $U(\infty)=\omega(0)$. Даже этот простейший частный случай позволяет сделать некоторые заключения.

Пример. а) Пусть F — распределение вероятностей с преобразованием Лапласа $\varphi(\lambda)$. Мы знаем [см. (2.8)], что $\omega(\lambda)=\frac{1-\varphi(\lambda)}{\lambda}$ является преобразованием Лапласа меры $U\{dx\}=[1-F(x)]dx$. Здесь $\omega(0)=-\varphi'(0)$. Соотношение $\omega(0)=U(\infty)$ показывает, что конечная производная $\varphi'(0)$ существует тогда и только тогда, когда функция $1-F$ интегрируема на $\overline{0, \infty}$, или, что то же самое, когда F имеет конечное математическое

ожидание. Этот результат содержится в (2.6), и единственная цель настоящего рассуждения — показать, что формулы дифференцирования (2.5) связаны с тауберовыми теоремами.

Чтобы избежать непривлекательных формул, содержащих обратные величины, введем две положительные переменные t и τ, связанные соотношением

$$t\tau = 1. \tag{5.2}$$

Тогда $\tau \to 0$ при $t \to \infty$.

Поясним происхождение тауберовых теорем. При фиксированном t замена переменных $x = ty$ в (5.1) показывает, что $\omega(\tau\lambda)$ является преобразованием Лапласа для несобственной функции распределения $U(ty)$. Так как ω убывает, возможно найти последовательность $\tau_1, \tau_2, \ldots \to 0$, такую, что, когда τ пробегает ее,

$$\frac{\omega(\tau\lambda)}{\omega(\tau)} \to \gamma(\lambda), \tag{5.3}$$

где $\gamma(\lambda)$ конечно при всех $\lambda > 1$. По обобщенной теореме непрерывности предел γ есть преобразование Лапласа некоторой меры G и когда t пробегает последовательность $t_k = 1/\tau_k$

$$\frac{U(tx)}{\omega(\tau)} \to G(x) \tag{5.4}$$

во всех точках непрерывности G. Фиксируя x, мы видим, что асимптотическое поведение $U(t)$ при $t \to \infty$ тесно связано с поведением $\omega(t^{-1})$.

В принципе мы могли бы сформулировать это утверждение как некую всеобъемлющую тауберову теорему, но она была бы неудобна для практического использования. Для того чтобы достигнуть разумной простоты, мы рассмотрим только случай, когда (5.3) верно при *любом* способе приближения τ к нулю (т. е. когда ω правильно изменяется в окрестностях нуля). Элементарная лемма [1]) из гл. VIII, 8 утверждает, что предел γ необходимо имеет вид $\gamma(\lambda) = \lambda^{-\rho}$, где $\rho \geqslant 0$ — некоторая постоянная. Тогда $G(x) = \dfrac{x^\rho}{\Gamma(\rho+1)}$ и, как мы показали, из

$$\frac{\omega(\tau\lambda)}{\omega(\tau)} \to \frac{1}{\lambda^\rho}, \quad \tau \to 0, \tag{5.5}$$

вытекает

$$\frac{U(tx)}{\omega(\tau)} \to \frac{x^\rho}{\Gamma(\rho+1)}, \tag{5.6}$$

[1]) Эта лемма нужна *только* для объяснения формы соотношений (5.5) и (5.7) (которая иначе показалась бы неестественной). Теория правильно меняющихся функций *не* используется нигде в этом параграфе [кроме замечания, что (5.17) влечет (5.16)].

откуда в свою очередь следует

$$\frac{U(tx)}{U(t)} \to x^{\rho}, \quad t \to \infty. \tag{5.7}$$

Иными словами, U правильно меняется на бесконечности и соответствующие показатели у ω и U равны по абсолютной величине. Полагая в (5.6) $x=1$, мы видим, что

$$\omega(\tau) \sim U(t)\,\Gamma(\rho+1). \tag{5.8}$$

Это и есть желаемая тауберова теорема. Мы сформулируем ее вместе с обратным утверждением.

Теорема 1. *Для фиксированного* $0 \leqslant \rho < \infty$ *каждое из соотношений* (5.5)—(5.7) *влечет остальные.*

Доказательство. Как мы уже знаем, (5.5) влечет (5.6), а (5.6) влечет (5.7). Чтобы показать, что (5.7) влечет (5.5), мы повторим предыдущие рассуждения. Функции $U(tx)$ соответствует преобразование Лапласа $\omega(\tau\lambda)$, так что (5.7) приводит к соотношению

$$\frac{\omega(\tau\lambda)}{U(t)} \to \frac{\Gamma(\rho+1)}{\lambda^{\rho}} \tag{5.9}$$

(в предположении, что применима обобщенная теорема непрерывности, т. е. *в предположении*, что левая часть ограничена при некотором λ). Из (5.9) тривиально выводится (5.5), поэтому для доказательства теорема достаточно проверить, что отношения $\omega(\tau)/U(t)$ ограничены.

Разбивая область интегрирования точками t, $2t$, $4t$, ..., получаем

$$\omega(\tau) \leqslant \sum_{0}^{\infty} e^{-2^{n-1}} U(2^n t). \tag{5.10}$$

В силу (5.7) существует число t_0, такое, что $U(2t) < 2^{\rho+1} U(t)$ при $t > t_0$. Последовательное применение этого неравенства дает

$$\frac{\omega(\tau)}{U(t)} \leqslant \sum_{0}^{\infty} 2^{n(\rho+1)} e^{-2^{n-1}}, \tag{5.11}$$

так что левая часть действительно ограничена при $t \to \infty$. ▶

Пример. б) $U(x) \sim \log^2 x$ при $x \to \infty$ тогда и только тогда, когда $\omega(\lambda) \sim \log^2 \lambda$ при $\lambda \to 0$. Аналогично $U(x) \sim \sqrt{x}$ тогда и только тогда, когда $\omega(\lambda) \sim \frac{1}{2}\sqrt{\frac{\pi}{\lambda}}$.

Иногда полезно знать, в какой степени справедлива теорема в пределе при $\rho \to \infty$. Мы сформулируем соответствующий результат как

Следствие. *Если при некотором* $a>1$ *и* $t \to \infty$

$$\text{или } \frac{\omega(\tau\omega)}{\omega(\tau)} \to 0, \text{ или } \frac{U(ta)}{U(t)} \to \infty, \tag{5.12}$$

то

$$\frac{U(t)}{\omega(\tau)} \to 0. \tag{5.13}$$

Доказательство. Если выполняется первое из соотношений (5.12), то при $\lambda>a$ $\frac{\omega(\tau\lambda)}{\omega(\tau)} \to 0$, и по обобщенной теореме непрерывности $\frac{U(tx)}{\omega(\tau)} \to 0$ при всех $x>0$. Так как $\omega(\tau) \geqslant e^{-\alpha} U(ta)$, то и второе из соотношений (5.12) влечет (5.13).

В применениях более удобно формулировать теорему 1 в терминах медленно меняющихся функций. Напомним, что определенная на $\overline{0, \infty}$ положительная функция L называется *медленно меняющейся на бесконечности*, если при каждом фиксированном x

$$\frac{L(tx)}{L(t)} \to 1, \quad t \to \infty. \tag{5.14}$$

Говорят, что L медленно меняется в нуле, если это соотношение верно при $t \to 0$ [т. е. если $L(1/x)$ медленно меняется на бесконечности]. Очевидно, что U имеет вид (5.7) тогда и только тогда, когда $U(x)/x^{\rho}$ медленно меняется на бесконечности. Аналогично (5.5) выполняется тогда и только тогда, когда $\lambda^{\rho}\omega(\lambda)$ медленно меняется в нуле. Поэтому теорема 1 может быть сформулирована следующим образом.

Теорема 2. *Если* L *медленно меняется на бесконечности и* $0 \leqslant \rho < \infty$, *то каждое из соотношений*

$$\omega(\tau) \sim \tau^{-\rho} L\left(\frac{1}{\tau}\right), \qquad \tau \to 0 \tag{5.15}$$

и

$$U(t) \sim \frac{1}{\Gamma(\rho+1)} t^{\rho} L(t), \quad t \to \infty \tag{5.16}$$

влечет другое.

Теорема 2 имеет интересную историю. Переход от (5.16) к (5.15) (от меры к преобразованию Лапласа) называют абелевой теоремой; обратный переход — от (5.15) к (5.16) (от преобразования Лапласа к мере) — называют тауберовой теоремой. В обычном изложении эти две теоремы отделены одна от другой, причем доказательство второй из них заметно сложнее.

В знаменитой работе Харди и Литтльвуда с помощью трудных вычислений был исследован случай $\omega(\lambda) \sim \lambda^{-\rho}$. В 1930 г. И. Карамата произвел сенсацию, дав упрощенное доказательство для указанного специального случая (это доказательство и до сих пор приводится в учебниках по теории функции комплексного переменного и преобразованиям Лапласа). Вскоре после этого он ввел класс правильно меняющихся функций и доказал теорему 2. Доказательство было, однако, слишком сложным для того, чтобы включать его в учебники. Понятие медленного изменения было введено Р. Шмидтом около 1925 г. в связи с этими же вопросами. Наше доказательство упрощает теорию и делает ее цельной. Кроме того, оно приводит к мало известному, но полезному следствию.

Большим преимуществом нашего доказательства является то, что оно без изменений применимо к случаю, когда нуль и бесконечность меняются ролями, т. е. когда $\tau \to \infty$, а $t \to 0$. Этим путем мы приходим к двойственной теореме, связывающей поведение ω на бесконечности и поведение U в нуле [эта теорема будет использована в настоящей книге только для вывода (6.2)].

Теорема 3. *Последние две теоремы и следствие остаются верными если нуль и бесконечность меняются ролями, т. е. если* $\tau \to \infty$ *и* $t \to 0$.

Теорема 2 содержит основной результат этого параграфа, однако для полноты изложения мы сделаем два добавления к ней. Прежде всего если U обладает плотностью $U' = u$, то желательно иметь оценки для u. Так, в примере (а) мы имеем дело с «хвостом» распределения $u = 1 - F$, и оценки для U (для интеграла от u) можно рассматривать лишь как весьма слабую замену. В наиболее общей обстановке возникают сложности, связанные с тем, что «хорошая» функция U может иметь «плохую» плотность u. К счастью, теорему 2 легко видоизменить, предполагая *и монотонной*, что типично для вероятностных применений.

Лемма. *Допустим, что* U *имеет монотонную плотность* u. *Если выполняется* (5.16) *и* $0 < \rho < \infty$, *то*

$$u(t) \sim \frac{1}{\Gamma(\rho)} t^{\rho-1} L(t), \quad t \to \infty. \tag{5.17}$$

[Обратно, (5.17) влечет (5.16), даже если u не монотонна. Это утверждение получается из гл. VIII, (9.5) при $Z = u$ и $p = 0$.]

Доказательство. Для $0 < a < b$ имеем

$$\frac{U(tb) - U(ta)}{U(t)} = \int_a^b \frac{u(ty)\, t}{U(t)}\, dy. \tag{5.18}$$

При $t \to \infty$ левая часть сходится к $b^\rho - a^\rho$. В силу монотонности u отсюда следует, что при фиксированном $y>0$ подинтегральное выражение остается ограниченным. По теореме о выборе [гл. VIII, 6] существует такая последовательность $t_1, t_2, \ldots$ $\ldots, \to \infty$, что, когда t пробегает ее,

$$\frac{u(ty)\,t}{U(t)} \to \psi(y) \tag{5.19}$$

во всех точках непрерывности. Как вытекает из предыдущего, интеграл от ψ по $\overline{a, b}$ равен $b^\rho - a^\rho$, так что $\psi(y) = \rho y^{\rho-1}$. Так как этот предел не зависит от выбора последовательности $\{t_k\}$, то (5.19) верно при любом стремлении t к бесконечности. Для $y=1$ это соотношение сводится к (5.17). ▷

Вряд ли необходимо говорить, что соотношения (5.16) и (5.17) сохраняются, если изменить U в некотором конечном интервале; соответственно и лемма остается верной, если предположить, что u «монотонна, начиная с некоторого места», т. е. монотонна на некотором интервале $\overline{a, \infty}$. Комбинируя лемму с теоремой 2, мы видим, что верна

Теорема 4[1]). *Пусть* $0<\rho<\infty$ *Если* U *имеет монотонную, начиная с некоторого места, производную* u, *то при* $\lambda \to 0$ *(соответственно при* $x \to \infty$*)*

$$\omega(\lambda) \sim \frac{1}{\lambda^\rho} L\left(\frac{1}{\lambda}\right) \text{ тогда и только тогда,} \tag{5.20}$$

$$\text{когда} \quad u(x) \sim \frac{1}{\Gamma(\rho)} x^{\rho-1} L(x).$$

Применение этой теоремы иллюстрируется в следующем параграфе. В заключение мы покажем, как теорема 2 позволяет получить *тауберову теорему для степенных рядов* [она используется в гл. XII, (8.8) и гл. XVII, 5].

Теорема 5. *Пусть* $q_n \geqslant 0$, *и пусть*

$$Q(s) = \sum_{n=0}^{\infty} q_n s^n \tag{5.21}$$

сходится при $0 \leqslant s < 1$. *Если* L *медленно меняется на бесконечности и* $0 \leqslant \rho < \infty$, *то каждое из двух соотношений*

$$Q(s) \sim \frac{1}{(1-s)^\rho} L\left(\frac{1}{1-s}\right), \qquad s \to 1- \tag{5.22}$$

[1]) Эта теорема включает известную тауберову теорему Э. Ландау. Наше доказательство дает новый пример того, как теорема о выборе устраняет аналитические сложности.

и

$$q_0+q_1+\ldots+q_n\sim\frac{1}{\Gamma(\rho+1)}n^{\rho}L(n),\qquad n\to\infty \tag{5.23}$$

влечет другое.

Далее, если последовательность $\{q_n\}$ *монотонна и* $0<\rho<\infty$, *то* (5.22) *равносильно соотношению*

$$q_n\sim\frac{1}{\Gamma(\rho)}n^{\rho-1}L(n),\ n\to\infty. \tag{5.24}$$

Доказательство. Пусть U — атомическая мера, приписывающая точке n вес q_n, и пусть ω — ее преобразование Лапласа. Тогда $\omega(\lambda)=Q(e^{-\lambda})$, так что (5.22) равносильно тому, что $\omega(\lambda)\sim\lambda^{-\rho}L(1/\lambda)$ при $\lambda\to 0$. Поэтому эквивалентность (5.22) и (5.23) следует из теоремы 2.

Чтобы доказать (5.24), заменим U мерой V с плотностью $u(x)=q_n$ при $n\leqslant x<n+1$. Ее преобразование Лапласа равно $\omega(\lambda)\dfrac{1-e^{-\lambda}}{\lambda}\sim\omega(\lambda)$ при $\lambda\to 0$. Очевидно, $V(x)\sim U(x)$ при $x\to\infty$. Поэтому (5.24) вытекает из теоремы 4. ▶

§ 6 *. Устойчивые распределения

Чтобы показать пользу тауберовых теорем, мы выведем общую формулу для устойчивых распределений, сосредоточенных на $0,\ \infty$, и дадим полную характеризацию соответствующих областей притяжения. Доказательства проводятся прямым методом и обладают замечательной простотой по сравнению с методами, необходимыми для распределений, не сосредоточенных на $0,\ \infty$.

Теорема 1. *При фиксированном* $0<\alpha<1$ *функция* $\gamma_\alpha(\lambda)=e^{-\lambda^\alpha}$ *является преобразованием Лапласа распределения* G_α, *обладающего следующими свойствами:*

G_α *устойчиво; более точно, если* $\mathbf{X}_1,\ldots,\mathbf{X}_n$ *независимые случайные величины с распределением* G_α, *то* $(\mathbf{X}_1+\ldots+\mathbf{X}_n)/n^{1/\alpha}$ *снова имеет распределение* G_α; *далее*

$$x^\alpha[1-G_\alpha(x)]\to\frac{1}{\Gamma(1-\alpha)},\ x\to\infty, \tag{6.1}$$

$$e^{x^{-\alpha}}G_\alpha(x)\to 0,\ x\to 0. \tag{6.2}$$

*) За исключением (6.2), результаты этого раздела получены независимо от гл. IX и XVII. Устойчивые распределения были введены в гл. VI, 1.

Доказательство. По второму критерию § 4 функция γ_α вполне монотонна, так как $e^{-\lambda}$ вполне монотонна и λ^α имеет вполне монотонную производную. Поскольку $\gamma_\alpha(0)=1$, мера G_α с преобразованием Лапласа γ_α имеет полную массу, равную единице. Свойство устойчивости очевидным образом вытекает из равенства $\gamma_\alpha^n(\lambda)=\gamma_\alpha(n^{1/\alpha}\lambda)$. Напомним, что по (2.8) функция $\omega(\lambda)=[1-\gamma_\alpha(\lambda)]/\lambda$ является преобразованием Лапласа меры U с монотонной плотностью $u=1-G_\alpha$. При $\lambda\to 0$ имеем $\omega(\lambda)\sim$ $\sim\lambda^{\alpha-1}$, так что (6.1) вытекает из теоремы 5.4. Аналогично (6.2) немедленно получается из следствия теоремы 5.1. ▶

Теорема 2. *Пусть F — распределение вероятностей, сосредоточенное на $\overline{0,\infty}$, и такое, что*

$$F^{n*}(a_n x)\to G(x) \tag{6.3}$$

(в точках непрерывности), где G — собственное распределение, не сосредоточенное в одной точке. Тогда

а) *существует функция L, медленно меняющаяся на бесконечности*[1]*, и константа α, $0<\alpha<1$, такие, что*

$$1-F(x)\sim\frac{x^{-\alpha}L(x)}{\Gamma(1-\alpha)},\quad x\to\infty. \tag{6.4}$$

б) *Обратно, если F имеет вид* (6.4), *то возможно выбрать числа a_n так, что*

$$\frac{nL(a_n)}{a_n^\alpha}\to 1. \tag{6.5}$$

В этом случае выполняется (6.3) *с $G=G_\alpha$.*

Из сказанного следует, что возможные предельные распределения G в (6.3) отличаются от какого-нибудь G_α лишь масштабным множителем. В частности, отсюда видно, что *не существует других устойчивых распределений, сосредоточенных на $\overline{0,\infty}$.*

Доказательство. Если φ и γ — преобразования Лапласа F и G соответственно, то (6.3) равносильно соотношению

$$-n\log\varphi\left(\frac{\lambda}{a_n}\right)\to-\log\gamma(\lambda). \tag{6.6}$$

По критерию гл. VIII, 8 отсюда следует, что $-\log\varphi$ правильно меняется в нуле, т. е.

$$-\log\varphi(\lambda)\sim\lambda^\alpha L\left(\frac{1}{\lambda}\right),\qquad \lambda\to 0, \tag{6.7}$$

[1]) То есть L удовлетворяет (5.14). Множитель $\Gamma(1-\alpha)$ в (6.4) введен лишь для удобства и приводит только к изменению в обозначениях.

где L — медленно меняется на бесконечности и $\alpha \geqslant 0$. Кроме того, $-\log \gamma(\lambda) = C\lambda^{\alpha}$. Так как G не сосредоточено в одной точке, то $0 < \alpha < 1$.

Теперь из (6.7) следует

$$\frac{1-\varphi(\lambda)}{\lambda} \sim \lambda^{\alpha-1} L\left(\frac{1}{\lambda}\right), \qquad \lambda \to 0. \tag{6.8}$$

Здесь левая часть — преобразование Лапласа меры U с монотонной плотностью $1-F$, и в силу теоремы (5.4) соотношения (6.4) и (6.8) эквивалентны. Наконец, если (6.4) выполняется, то можно выбрать числа a_n так, чтобы они удовлетворяли (6.5). Тогда из (6.8) видно, что при $n \to \infty$ левая часть (6.6) стремится к λ^{α}. Доказательство закончено. ▶

(См. задачу 20 о влиянии максимального слагаемого.)

§ 7*. Безгранично-делимые распределения

В соответствии с определением гл. VI, 3 распределение вероятностей U с преобразованием Лапласа ω называется безгранично-делимым, если для любого $n = 1, 2, \ldots$ положительный корень n-й степени $\omega_n = \omega^{1/n}$ является преобразованием Лапласа некоторого распределения вероятностей.

Теорема 1. *Функция ω представляет собой преобразование Лапласа безгранично-делимого распределения тогда и только тогда, когда $\omega = e^{-\psi}$, где ψ имеет вполне монотонную производную и $\psi(0) = 0$.*

Доказательство. Используем критерий 2, § 4. Мы видим, что если $\psi(0) = 0$ и ψ' вполне монотонна, то функция $\omega_n = e^{-\psi/n}$ является преобразованием Лапласа некоторого распределения вероятностей. Следовательно, условие теоремы достаточно.

Для доказательства необходимости положим $\omega = e^{-\psi}$. Замечая, что при $z \to 1$ $\log z \sim -(1-z)$, получаем при фиксированном $\lambda > 0$ и $n \to \infty$

$$\psi(\lambda) = -n \log \omega_n(\lambda) \sim \psi_n(\lambda). \tag{7.1}$$

Здесь для сокращения записи мы обозначили

$$\psi_n(\lambda) = n[1 - \omega_n(\lambda)]. \tag{7.2}$$

Из (7.2) видно, что ψ_n имеет вполне монотонную производную $\psi_n'(\lambda)$. По теореме о среднем $\psi_n(\lambda) = \lambda\psi_n'(\theta\lambda) \geqslant \lambda\psi_n'(\lambda)$ и так как

*) Материал этого параграфа не используется в дальнейшем.

$\psi_n \to \psi$, то последовательность $\{\psi_n'(\lambda)\}$ ограничена при каждом фиксированном $\lambda>0$. Следовательно, из нее можно выбрать сходящуюся подпоследовательность, предел которой автоматически будет вполне монотонной функцией (по обобщенной теореме непрерывности из § 1) Таким образом, ψ является интегралом от вполне монотонной функции. Доказательство закончено. ▶

Другой формой этой теоремы служит

Теорема 2. *Функция ω представляет собой преобразование Лапласа безгранично-делимого распределения тогда и только тогда, когда она имеет вид $\omega = e^{-\psi}$, где*

$$\psi(\lambda) = \int_0^\infty \frac{1-e^{-\lambda x}}{x} P\{dx\} \tag{7.3}$$

и P такая мера, что

$$\int_1^\infty \frac{1}{x} P\{dx\} < \infty. \tag{7.4}$$

Доказательство. В силу теоремы о представлении вполне монотонных функций условия теоремы 1 могут быть заменены следующими: должно быть $\psi(0)=0$ и

$$\psi'(\lambda) = \int_0^\infty e^{-\lambda x} P\{dx\}, \tag{7.5}$$

где P — некоторая мера. Заменяя в интеграле верхний предел числом a, легко получим, что при любом $a>0$

$$\psi(\lambda) \geqslant \int_0^a \frac{1-e^{-\lambda x}}{x} P\{dx\}. \tag{7.6}$$

Отсюда вытекает, что правая часть (7.3) имеет смысл, и условие (7.4) выполнено. Формальное дифференцирование показывает, что правая часть (7.3) представляет собой первообразную от (7.5), равную в нуле нулю. ▶

(См. задачи 14—17.)

Примеры. а) Обобщенное распределение Пуассона

$$U = e^{-c} \sum_0^\infty \frac{c^n}{n!} F^{n*} \tag{7.7}$$

имеет преобразование Лапласа, равное $e^{-c+c\psi}$, и (7.3) верно с $P\{dx\}=cxF\{dx\}$.

б) *Гамма-плотность* $x^{\alpha-1}e^{-x}/\Gamma(a)$ имеет преобразование $\omega(\lambda)=1/(\lambda+1)^a$. Здесь

$$\psi(\lambda)=a\int_0^\infty \frac{1-e^{-\lambda x}}{x}e^{-x}\,dx, \tag{7.8}$$

что можно проверить дифференцированием.

в) *Устойчивые распределения.* Для найденного в § 5 преобразования $\omega(\lambda)=e^{-\lambda^\alpha}$ имеем $\psi(\lambda)=\lambda^\alpha$ и

$$\lambda^\alpha=\frac{\alpha}{\Gamma(1-\alpha)}\int_0^\infty \frac{1-e^{-\lambda x}}{x^{\alpha+1}}\,dx, \tag{7.9}$$

что опять можно проверить дифференцированием.

г) *Бесселевы функции.* Рассмотрим плотность v_r из примера (3, г) с преобразованием Лапласа (3.8). Из его вида ясно, что v_r равна n-кратной свертке плотности $v_{r/n}$ с собой, т. е. v_r безгранично-делима. Формальное дифференцирование показывает, что в этом случае $\psi'(\lambda)=r/\sqrt{(\lambda+1)^2-1}$ и, как нетрудно проверить (см. задачу 5), что ψ' представима в форме (7.5) с

$$P\{dx\}=re^{-x}I_0(x)\,dx.$$

д) *Подчиненность.* Если ψ_1 и ψ_2 — положительные функции с вполне монотонными производными, то, применяя критерии § 4, легко видеть, что функция $\psi(\lambda)=\psi_1(\psi_2(\lambda))$ имеет те же самые свойства. Соответствующее безгранично-делимое распределение представляет особый интерес. Чтобы понять это, обозначим $Q_t^{(i)}$ распределение вероятностей с преобразованием Лапласа $e^{-t\psi_i(\lambda)}$ (где $i=1, 2$) и положим

$$U_t(x)=\int_0^\infty Q_s^{(2)}(x)\,Q_t^{(1)}\{ds\}. \tag{7.10}$$

(Результирующее распределение U_t получается рандомизацией параметра s в $Q_s^{(2)}$.) Преобразование Лапласа для U_t равно

$$\omega_t(\lambda)=\int_0^\infty e^{-s\psi_2(\lambda)}Q_t^{(1)}\{ds\}=e^{-t\psi(\lambda)}. \tag{7.11}$$

Читатель, знакомый с гл. X, 7, легко свяжет (7.10) с *подчиненностью процессов:* U_t подчинено $Q_s^{(2)}$ с управляющим процес-

сом $Q_t^{(1)}$. Мы очень легко получили формулу для преобразования Лапласа нового процесса, хотя и при специальном предположении, что $Q_t^{(2)}$ сосредоточено на $\overline{0, \infty}$.

Заслуживает внимания частный случай, когда $\psi_1(\lambda)=\lambda^\alpha$ и $\psi_2(\lambda)=\lambda^\beta$. Здесь $\psi(\lambda)=\lambda^{\alpha\beta}$. Таким образом, *устойчивый α-процесс, управляемый устойчивым β-процессом, дает устойчивый $\alpha\beta$-процесс.* Читатель может проверить, что по существу это утверждение повторяет формулировку задачи 13. Более общее предложение см. в примере (2, ж) гл. VI. ▶

§ 8*. Многомерный случай

Понятие преобразования Лапласа распространяется на многомерный случай очевидным образом. Даже формула (1.1) не требует изменения, если интерпретировать x как вектор-столбец $(x_1, \ldots, x_n)$, λ — как вектор-строку $(\lambda_1, \ldots, \lambda_n)$, а $\lambda x=\lambda_1 x_1+\ldots$ $\ldots+\lambda_n x_n$ — как скалярное произведение λ и x. В теории вероятностей многомерные преобразования используются в ограниченной степени.

Примеры. а) *Резольвентное уравнение.* Пусть f — непрерывная функция *одной* переменной с обычным преобразованием Лапласа $\varphi(\lambda)$. Рассмотрим функцию $f(s+t)$ от двух переменных s и t. Ее двумерное преобразование Лапласа равно

$$\omega(\lambda, \nu)=\int_0^\infty\int_0^\infty e^{-\lambda s-\nu t}f(s+t)\,ds\,dt. \tag{8.1}$$

После замены переменных $s+t=x$, $-s+t=y$ подинтегральное выражение принимает вид $A(x)e^{cy}$ и интегрирование по y на $\overline{-x, x}$ осуществляется тривиально. В результате получим

$$\omega(\lambda, \nu)=-\frac{\varphi(\lambda)-\varphi(\nu)}{\lambda-\nu}. \tag{8.2}$$

Мы встретимся с этим соотношением в более «возвышенной» обстановке, как с основным *резольвентным уравнением* для полугрупп [см. (10.5)].

б) *Функции Миттаг-Леффлера.* Этот пример иллюстрирует пользу многомерных преобразований как технического средства вычисления некоторых простых одномерных преобразований. Мы докажем сейчас следующее *утверждение*:

*) Материал не используется в дальнейшем.

Если F — устойчивое распределение с преобразованием Лапласа $e^{-\lambda^\alpha}$, то (при фиксированном t) распределение

$$G_t(x) = 1 - F\left(\frac{t}{x^{1/\alpha}}\right), \quad x > 0, \tag{8.3}$$

имеет своим преобразованием Лапласа функцию Миттаг-Леффлера[1])

$$\sum_{k=0}^{\infty} \frac{(-\lambda)^k}{\Gamma(1+k\alpha)} t^{k\alpha}. \tag{8.4}$$

Этот результат довольно интересен, так как распределение G постоянно встречается вместе с F [см., например, гл. IX, (5.5)]. Прямая проверка кажется трудной, однако можно поступить следующим образом. Сначала фиксируем x, а t примем в качестве переменной. Обычное преобразование Лапласа $\gamma_x(v)$ (v — переменная) для $G_t(x)$ равно, очевидно, $(1 - e^{-v^\alpha x})/v$. С точностью до множителя $1/v$ это есть функция распределения по x и *ее* преобразование Лапласа равно

$$\frac{v^{\alpha-1}}{\lambda + v^\alpha}. \tag{8.5}$$

Это есть не что иное, как двумерное преобразование для функции (8.3). Но оно могло бы быть вычислено и другим путем — интегрированием сначала по x, а затем по t. Поэтому (8.5) можно рассматривать, как преобразование по t того преобразования, которое мы ищем. Разлагая (8.5) в геометрический ряд, видим, что (8.5) действительно является преобразованием для (8.4). Утверждение доказано.

Функция Миттаг-Леффлера обобщает экспоненту, к которой она сводится при $\alpha = 1$. ▲

§ 9. Преобразования Лапласа для полугрупп

Понятие преобразования Лапласа может быть распространено на абстрактные функции и интегралы[2]). Однако мы рассмотрим только преобразования Лапласа для полугрупп опера-

[1]) Этот факт обнаружен впервые в связи с теорией рекуррентных событий. Первое аналитическое доказательство полной монотонности функции (8.4) было дано Г. Поллардом.

[2]) Плодотворная теория, охватывающая преобразования вида (9.6), была развита С. Бохнером, Completely monotone functions in partially ordered spaces, *Duke Math. J.*, **9** (1942), 519—526. Распространение на произвольные семейства операторов см. у Е. Хилле и Р. Филипса (1957).

торов, связанных с марковскими процессами[1]). Мы вернемся к основным обозначениям и соглашениям гл. X, 8.

Пусть Σ — некоторое пространство (например, прямая, интервал или совокупность всех целых чисел), и пусть $\mathscr{L}$ — банахово пространство ограниченных функций на нем с нормой $\|u\| = \sup|u(x)|$. Мы предполагаем, что из $u \in \mathscr{L}$ вытекает $|u| \in \mathscr{L}$. Пусть $\{\mathfrak{Q}(t), t>0\}$ — непрерывная полугруппа сжатий $\mathscr{L}$. Иными словами, мы предполагаем, что для каждой функции $u \in \mathscr{L}$ существует функция $\mathfrak{Q}(t)u \in \mathscr{L}$ и что $\mathfrak{Q}(t)$ обладает следующими свойствами: из $0 \leqslant u \leqslant 1$ вытекает $0 \leqslant \mathfrak{Q}(t)u \leqslant 1$; $\mathfrak{Q}(t+s) = \mathfrak{Q}(t)\mathfrak{Q}(s)$ и $\mathfrak{Q}(h) \to \mathfrak{Q}(0) = \mathbf{1}$ (единица обозначает тождественное преобразование)[2]).

Мы начнем с определения интегрирования. Пусть F — произвольное распределение вероятностей на $0, \infty$. Мы желаем определить сжимающий оператор E, отображающий $\mathscr{L}$ в $\mathscr{L}$, который будем обозначать

$$E = \int_0^\infty \mathfrak{Q}(s) F\{ds\} \tag{9.1}$$

и который удовлетворяет равенству

$$\mathfrak{Q}(t)E = E\mathfrak{Q}(t) = \int_0^\infty \mathfrak{Q}(t+s) F\{ds\}. \tag{9.2}$$

Для полугруппы, отвечающей марковскому процессу с переходными вероятностями $Q_t(x, \Gamma)$, этот оператор будет порождаться стохастическим или субстохастическим ядром

$$\int_0^\infty Q_s(x, \Gamma) F\{ds\}. \tag{9.3}$$

Естественное (и почти тривиальное) определение оператора E для случая атомической F с помощью простого предельного перехода приводит к желаемому общему определению.

Пусть $p_j \geqslant 0$ и $p_1 + \ldots + p_r = 1$. Линейная комбинация

$$E = p_1\mathfrak{Q}(t_1) + \ldots + p_r\mathfrak{Q}(t_r) \tag{9.4}$$

снова является сжатием и может рассматриваться как математическое ожидание $\mathfrak{Q}(t)$ по отношению к распределению веро-

[1]) Построение минимального решения в гл. XIV, 7 может служить типичным примером излагаемых далее методов.

[2]) Напомним, что сильная сходимость $T_n \to T$ эндоморфизмов означает, что $\|T_n u - Tu\| \to 0$ для всех $u \in \mathscr{L}$. Мы говорим кратко «непрерывность», подразумевая «сильную непрерывность при $t \geqslant 0$».

ятностей, приписывающему вес p_j в точке t_j. Таким путем определяется оператор (9.1) для специального случая дискретных распределений с конечным числом атомов. При этом (9.2) оказывается выполненным. В общем случае оператор (9.1) определяется предельным переходом, подобным применяемому в теории интеграла Римана: интервал $\overline{0, \infty}$ разбивают на интервалы $I_1, \ldots, I_n$; выбирают $t_j \in I_j$ и образуют риманову сумму $\sum \mathfrak{Q}(t_k) F\{I_k\}$, которая определяет некоторое сжатие. Ввиду равномерной непрерывности [см. гл. X, (8.7)] известное доказательство сходимости римановых сумм применимо без изменений. Таким путем (9.1) определяется как частный случай интеграла Бохнера.

Если полугруппа образована операторами перехода, т. е. если $\mathfrak{Q}(t)\,1 = 1$ для всех t, то $E1 = 1$. Обозначение (9.1) будет использоваться для E, а для функции Ew мы будем использовать обычную запись

$$Ew = \int_0^\infty \mathfrak{Q}(s)\, w \cdot F\{ds\} \tag{9.5}$$

(хотя более логично было бы писать w вне знака интеграла). Значение функции $Ew(x)$ в данной точке x — это обычное математическое ожидание по отношению к F числовой функции $\mathfrak{Q}(s)w(x)$.

В специальном случае $F\{ds\} = e^{-\lambda s} ds$ оператор E называют *интегралом Лапласа полугруппы* или *резольвентой*. Этот оператор мы будем обозначать

$$\mathfrak{R}(\lambda) = \int_0^\infty e^{-\lambda s} \mathfrak{Q}(s)\, ds, \quad \lambda > 0. \tag{9.6}$$

Сжатие $\lambda\, \mathfrak{R}(\lambda)$ является оператором перехода тогда и только тогда, когда все $\mathfrak{Q}(s)$ являются операторами перехода.

Преобразование Лапласа (9.6) приводит к простой характеризации *инфинитезимального оператора* $\mathfrak{A}$ рассматриваемой полугруппы. По определению этого оператора [см. гл. X, 10] имеем

$$\frac{\mathfrak{Q}(h) - 1}{h} u \to \mathfrak{A}u, \quad h \to 0+, \tag{9.7}$$

тогда и только тогда, когда $\mathfrak{A}u$ существует (сходимость означает, что норма разности правой и левой части (9.7) стремится к нулю).

Теорема 1. *При фиксированном* $\lambda > 0$

$$u = \mathfrak{R}(\lambda) w \tag{9.8}$$

тогда и только тогда, когда и принадлежит области определения $\mathfrak{A}$ *и*

$$\lambda u - \mathfrak{A} u = w. \tag{9.9}$$

Доказательство. (i) Определим u равенством (9.8). Опираясь на свойства (9.2) математических ожиданий, получаем

$$\frac{\mathfrak{T}(h) - 1}{h} u = \frac{1}{h} \int_0^\infty e^{-\lambda s} \mathfrak{T}(s+h) w \cdot ds - \frac{1}{h} \int_0^\infty e^{-\lambda s} \mathfrak{T}(s) w \cdot ds. \tag{9.10}$$

Замена переменных $s+h=t$ и очевидные преобразования показывают, что правая часть равна

$$\frac{e^{\lambda h} - 1}{h} \int_0^\infty e^{-\lambda t} \mathfrak{T}(t) w \cdot dt - \frac{e^{\lambda h}}{h} \int_0^h e^{-\lambda t} [\mathfrak{T}(t) w - w] dt - \frac{e^{\lambda h} - 1}{\lambda h} w. \tag{9.11}$$

Так как $\| \mathfrak{T}(t) w - w \| \to 0$ при $t \to 0$, то второй член стремится к нулю при $h \to 0$. Первый интеграл не зависит от h и равен $\mathfrak{R}(\lambda) w = u$. Полагая $h \to 0$, мы видим, что (9.11) сходится к $\lambda u - w$, так что (9.9) верно.

(ii) Предположим, обратно, что $\mathfrak{A} u$ существует, т. е. что выполняется (9.7). Так как оператор сжатия $\lambda \mathfrak{R}(\lambda)$ коммутирует со всеми операторами полугруппы, то из (9.7) вытекает

$$\frac{\mathfrak{T}(h) - 1}{h} \mathfrak{R}(\lambda) u \to \mathfrak{R}(\lambda) \mathfrak{A} u. \tag{9.12}$$

Но как мы только что видели, левая часть стремится к $\lambda \mathfrak{R}(\lambda) u - u$. В получающемся тождестве u представляется как преобразование Лапласа функции w из (9.9). ▶

Следствие 1. *Для данного* $w \in \mathscr{L}$ *существует ровно одно решение и уравнения* (9.9).

Следствие 2. *Две различные полугруппы не могут иметь один и тот же инфинитезимальный оператор* $\mathfrak{A}$.

Доказательство. Пусть w — произвольный элемент $\mathscr{L}$. Значение $\mathfrak{A}$ однозначно определяет преобразование Лапласа $\mathfrak{R}(\lambda) w$ от $\mathfrak{T}(t) w$ при всех $\lambda > 0$. Его значение $\mathfrak{R}(\lambda) w(x)$ в фиксированной точке x является обычным преобразованием Лапласа числовой функции аргумента t, равной $\mathfrak{T}(t) w(x)$. Из теоремы однозначности для преобразований Лапласа заключаем, что $\mathfrak{T}(t) w$ определяется по $\mathfrak{A}$ однозначно. ▶

Соблазнительно было бы получить тауберовы теоремы, аналогичные приведенным в § 5. Однако мы ограничимся совсем простым результатом.

Теорема 2. *При* $\lambda \to \infty$

$$\lambda \mathfrak{R}(\lambda) \to \mathbf{1}. \tag{9.13}$$

Доказательство. Для произвольного $w \in \mathscr{L}$ имеем

$$\|\lambda \mathfrak{R}(\lambda) w - w\| \leqslant \int_0^\infty \|\mathfrak{Q}(t) w - w\| \cdot \lambda e^{-\lambda t} dt. \tag{9.14}$$

При $\lambda \to \infty$ распределение вероятностей с плотностью $\lambda e^{-\lambda t}$ сходится к распределению, сосредоточенному в нуле. Подинтегральное выражение ограничено и стремится к нулю при $t \to 0$. Поэтому интеграл (9.14) стремится к нулю и (9.13) верно. ▷

Следствие 3. *Инфинитезимальный оператор* $\mathfrak{A}$ *имеет область определения, плотную в* $\mathscr{L}$.

Доказательство. Из (9.13) следует, что каждое $w \in \mathscr{L}$ может быть представлено как сильный предел последовательности элементов $\lambda \mathfrak{R}(\lambda) w$. По теореме 1 эти элементы принадлежат области определения $\mathfrak{A}$. ▶

Примеры. а) *Безгранично-делимые полугруппы.* Пусть U — безгранично-делимое распределение с преобразованием Лапласа $\omega = e^{-\psi(\lambda)}$, характеризуемым (7.3). Распределение U_t с преобразованием Лапласа $e^{-t\psi}$ снова безгранично-делимо, и соответствующие операторы свертки $\mathfrak{A}(t)$ образуют полугруппу. Чтобы найти ее инфинитезимальный оператор [1]), возьмем какую-нибудь ограниченную непрерывно дифференцируемую функцию v. Тогда, очевидно,

$$\frac{\mathfrak{A}(t) - \mathbf{1}}{t} v(x) = \int_0^\infty \frac{v(x-y) - v(x)}{y} \cdot \frac{1}{t} y U_t\{dy\}. \tag{9.15}$$

Распределение U_t имеет преобразование Лапласа $e^{-t\psi(\lambda)}$, а мера $t^{-1} y U_t\{dy\}$ имеет преобразование $\psi'(\lambda) e^{-t\psi(\lambda)}$, которое стремится к $\psi'(\lambda)$ при $t \to 0$. Но (7.3) определяет ψ' как преобразование меры P, так что наша мера сходится к P. Первая дробь под знаком интеграла (9.15) является (при фиксированном x) огра-

[1]) Этот вывод дан только как иллюстрация. Вид оператора известен из гл. IX и может быть получен переходом к пределу от обобщенных пуассоновских распределений.

ниченной и непрерывной функцией от y. Поэтому мы получаем соотношение

$$\mathfrak{A}v(x) = \int_0^\infty \frac{v(x-y)-v(x)}{y} P\{dy\}. \tag{9.16}$$

Здесь мы имеем новую интерпретацию меры P, входящей в каноническое представление безгранично-делимого распределения.

б) *Подчиненные полугруппы.* Пусть $\{\mathfrak{Q}(t)\}$ обозначает произвольную марковскую полугруппу, и пусть U_t — безгранично-делимое распределение предыдущего примера. Рандомизацией параметра t можно получить, как объяснено в гл. X, 7, новую марковскую полугруппу $\{\mathfrak{Q}^*(t)\}$.

В принятых здесь обозначениях

$$\mathfrak{Q}^*(t) = \int_0^\infty \mathfrak{Q}(s)\, U_t\{ds\}. \tag{9.17}$$

Полагая для краткости

$$V(s, x) = \frac{\mathfrak{Q}(s)-1}{s} v(x), \tag{9.18}$$

имеем

$$\frac{\mathfrak{Q}^*(t)-1}{t} v(x) = \int_0^\infty V(s, x)\cdot \frac{1}{t} sU_t\{ds\}. \tag{9.19}$$

Пусть v принадлежит области определения $\mathfrak{A}$. При фиксированном x функция V непрерывно зависит от s и при $s=0$ принимает значение $\mathfrak{A}v(x)$. Мы видели в предыдущем примере, что при $t\to 0$ $t^{-1}sU_t\{ds\} \longrightarrow P\{ds\}$. Таким образом, правая часть в (9.19) сходится к некоторому пределу. Поэтому $\mathfrak{A}^*v$ существует и задается формулой

$$\mathfrak{A}^*v(x) = \int_0^\infty V(s, x) P\{ds\}. \tag{9.20}$$

Мы приходим к заключению, что *области определения* $\mathfrak{A}$ *и* $\mathfrak{A}^*$ *совпадают и* [1])

$$\mathfrak{A}^* = \int_0^\infty \frac{\mathfrak{Q}(s)-1}{s} P\{ds\} \tag{9.21}$$

в том смысле, что (9.20) выполняется для всех v в области определения $\mathfrak{A}$. ▶

[1]) Аналог (9.20) для более общих полугрупп был получен Р. С. Филипсом и используется в операционном исчислении, например для представления дробных степеней. Основная цель указанного простого вывода — прояснить вероятностный смысл этого соотношения (который иначе остается затемненным).

§ 10. Теорема Хилле — Иосида

Известная и весьма полезная теорема Хилле — Иосида дает характеризацию инфинитезимальных операторов произвольных полугрупп. Мы изложим ее применительно к случаю сжимающих полугрупп. Эта теорема утверждает, что свойства инфинитезимальных операторов, описанные в предыдущем параграфе, присущи только им.

Теорема 1. (*Хилле — Иосида.*) *Оператор* $\mathfrak{A}$ *с областью определения* $\mathscr{L}' \subset \mathscr{L}$ *представляет собой инфинитезимальный оператор некоторой непрерывной сжимающей полугруппы* $\mathfrak{T}(t)$ *на* $\mathscr{L}$ (*с* $\mathfrak{T}(0) = \mathbf{1}$) *тогда и только тогда, когда он обладает следующими свойствами:*

(i) *уравнение*

$$\lambda u - \mathfrak{A}u = w, \qquad \lambda > 0 \tag{10.1}$$

имеет для каждого $w \in \mathscr{L}$ *ровно одно решение* u;

(ii) *если* $0 \leqslant w \leqslant 1$, *то* $0 \leqslant \lambda u \leqslant 1$;

(iii) $\mathscr{L}'$ *— область определения оператора* $\mathfrak{A}$, *плотна в* $\mathscr{L}$.

Мы знаем уже, что каждый инфинитезимальный оператор обладает этими свойствами, т. е. условия теоремы необходимы. Далее обозначим решение (10.1) через $u = \mathfrak{R}(\lambda) w$. Тогда, как мы знаем, оператор $\mathfrak{R}(\lambda)$ совпадает с преобразованием Лапласа (9.6). Условия теоремы мы можем сформулировать теперь следующим образом.

(i′) Оператор $\mathfrak{R}(\lambda)$ удовлетворяет соотношению

$$\lambda \mathfrak{R}(\lambda) - \mathfrak{A}\mathfrak{R}(\lambda) = \mathbf{1}. \tag{10.2}$$

Область определения $\mathfrak{R}(\lambda)$ совпадает с $\mathscr{L}$, а область значений — с $\mathscr{L}'$, т. е. с областью определения оператора $\mathfrak{A}$.

(ii′) Оператор $\lambda \mathfrak{R}(\lambda)$ является сжатием.

(iii′) Область значения $\mathfrak{R}(\lambda)$ плотна в $\mathscr{L}$.

Как известно из теоремы 9.2, оператор $\mathfrak{R}(\lambda)$ должен удовлетворять соотношению

$$\lambda \mathfrak{R}(\lambda) \to \mathbf{1}, \qquad \lambda \to \infty. \tag{10.3}$$

Из него вытекает, что каждая функция u является пределом своих собственных преобразований, следовательно, область $\mathscr{L}'$ значений $\mathfrak{R}(\lambda)$ плотна в $\mathscr{L}$. Отсюда ясно, что (10.3) может заменить (iii′), т. е. *три условия теоремы эквивалентны условиям* (i′), (ii′), (10.3).

Пусть теперь задано семейство операторов $\mathfrak{R}(\lambda)$ с этими свойствами. Мы намерены построить искомую полугруппу как предел семейства псевдопуассоновских полугрупп. При построении используется

Лемма 1. *Если w принадлежит $\mathscr{L}'$—области определения оператора $\mathfrak{A}$, то*

$$\mathfrak{A}\mathfrak{R}(\lambda)w = \mathfrak{R}(\lambda)\mathfrak{A}w. \tag{10.4}$$

Операторы $\mathfrak{R}(\lambda)$ и $\mathfrak{R}(\nu)$ перестановочны и удовлетворяют резольвентному уравнению

$$\mathfrak{R}(\lambda) - \mathfrak{R}(\nu) = (\nu - \lambda)\mathfrak{R}(\lambda)\mathfrak{R}(\nu). \tag{10.5}$$

Доказательство. Применяя $\mathfrak{A}$ к (10.1), видим, что $\mathfrak{A}u = \mathfrak{R}(\lambda)\mathfrak{A}w$, а это не что иное, как (10.4).

Далее определим z как единственное решение уравнения $\nu z - \mathfrak{A}z = w$. Вычитая это равенство из (10.1), после простых преобразований получим

$$\lambda(u - z) - \mathfrak{A}(u - z) = (\nu - \lambda)z,$$

что совпадает с (10.5). Перестановочность операторов вытекает из симметрии предыдущего соотношения. ▶

Теперь мы в состоянии перейти к построению нашей полугруппы. Положим

$$\mathfrak{A}_\lambda = \lambda[\lambda\mathfrak{R}(\lambda) - \mathbf{1}] = \lambda\mathfrak{A}\mathfrak{R}(\lambda). \tag{10.6}$$

Тогда из (10.4) вытекает, что

$$\mathfrak{A}_\lambda u \to \mathfrak{A}u \tag{10.7}$$

для всех u из $\mathscr{L}'$, т. е. из области определения данного оператора $\mathfrak{A}$. Первое представление (10.6) показывает, что $\mathfrak{A}_\lambda$ есть эндоморфизм, порождающий квазипуассоновскую полугруппу $\mathfrak{Q}_\lambda(t) = e^{t\mathfrak{A}_\lambda}$ (теорема 1 из гл. X, 9). Эти операторы попарно перестановочны.

Мы можем забыть теперь о специальном способе построения $\mathfrak{A}_\lambda$ и получить теорему Хилле — Иосида как частный случай более общей предельной теоремы, которая полезна и сама по себе. В этой теореме λ может пробегать и множество целых чисел.

Лемма 2 *(Лемма об аппроксимации). Пусть $\{\mathfrak{Q}_\lambda(t)\}$ — семейство псевдопуассоновских полугрупп, перестановочных одна с другой и порожденных эндоморфизмами $\mathfrak{A}_\lambda$.*

Если (10.7) выполняется на плотном множестве $\mathscr{L}'$, то

$$\mathfrak{Q}_\lambda(t) \to \mathfrak{Q}(t), \quad \lambda \to \infty, \tag{10.8}$$

где $\{\mathfrak{Q}(t)\}$ — полугруппа сжатий, с инфинитезимальным оператором, совпадающим на $\mathscr{L}'$ с $\mathfrak{A}$.

Далее для $u \in \mathscr{L}'$

$$\|\mathfrak{Q}(t)u - \mathfrak{Q}_\lambda(t)u\| \leqslant t\|\mathfrak{A}u - \mathfrak{A}_\lambda u\|. \tag{10.9}$$

Доказательство. Для любых двух перестановочных сжатий верно тождество

$$S^n - T^n = (S^{n-1} + \ldots + T^{n-1})(S - T),$$

и потому

$$\|S^n u - T^n u\| \leqslant n\|Su - Tu\|. \tag{10.10}$$

В применении к операторам $\mathfrak{Q}_\lambda(t/n)$ это неравенство дает (после очевидных преобразований)

$$\|\mathfrak{Q}_\lambda(t)u - \mathfrak{Q}_\nu(t)u\| \leqslant t\left\|\frac{\mathfrak{Q}_\lambda(t/n) - 1}{t/n}u - \frac{\mathfrak{Q}_\nu(t/n) - 1}{t/n}u\right\|. \tag{10.11}$$

Полагая $n \to \infty$, получаем

$$\|\mathfrak{Q}_\lambda(t)u - \mathfrak{Q}_\nu(t)u\| \leqslant t\|\mathfrak{A}_\lambda u - \mathfrak{A}_\nu u\|. \tag{10.12}$$

Это показывает, что для $u \in \mathscr{L}'$ последовательность $\{\mathfrak{Q}_\lambda(t)u\}$ равномерно сходится при $\lambda \to \infty$. Так как $\mathscr{L}'$ плотно в $\mathscr{L}$, то эта равномерная сходимость имеет место для всех u. Обозначая предел через $\mathfrak{Q}(t)u$, мы видим, что $\mathfrak{Q}(t)$ есть сжатие, удовлетворяющее (10.8). Полугрупповое свойство очевидно. Полагая в (10.12) $\nu \to \infty$, приходим к (10.9). Переписывая левую часть как в (10.11), имеем

$$\left\|\frac{\mathfrak{Q}(t) - 1}{t}u - \frac{\mathfrak{Q}_\lambda(t) - 1}{t}u\right\| \leqslant \|\mathfrak{A}u - \mathfrak{A}_\lambda u\|. \tag{10.13}$$

Выберем λ столь большим, чтобы правая часть стала меньше ε. Для достаточно малых t второе разностное отношение в левой части отличается (по норме) от $\mathfrak{A}_\lambda u$ меньше чем на ε, а от $\mathfrak{A}u$ — меньше чем на 3ε. Таким образом, для $u \in \mathscr{L}'$

$$\frac{\mathfrak{Q}(t) - 1}{t}u \to \mathfrak{A}u, \tag{10.14}$$

чем заканчивается доказательство. ▶

Примеры. *Диффузия.* Пусть $\mathscr{L}$ — семейство непрерывных функций на прямой, обращающихся в нуль на $\pm\infty$. Чтобы перейти к привычным обозначениям, заменим λ на h^{-2} и положим $h \to 0$. Определим разностный оператор ∇_h формулой

$$\nabla_h u(x) = \frac{1}{h^2}\left[\frac{u(x+h) + u(x-h)}{2} - u(x)\right]. \tag{10.15}$$

Первое выражение в скобках определяет оператор перехода, и потому ∇_h порождает полугруппу $e^{t\nabla_h}$ операторов перехода

(марковскую полугруппу). Операторы ∇_h перестановочны один с другим, и для функций, имеющих три ограниченные производные, $\nabla_h u \to \frac{1}{2} u''$ равномерно. Лемма 2 показывает существование предельной полугруппы $\{\mathfrak{T}(t)\}$, порождаемой таким оператором $\mathfrak{A}$, что $\mathfrak{A}u = \frac{1}{2} u''$ (по крайней мере для всех достаточно гладких u).

В этом частном случае, как мы знаем, $\{\mathfrak{T}(t)\}$ есть полугруппа сверток с нормальными распределениями с дисперсией t, так что мы не получаем новой информации. Пример тем не менее указывает, как легко можно установить (иногда) существование полугруппы с заданным инфинитезимальным оператором. Проведенное выше рассуждение применимо, скажем, к более общим дифференциальным операторам и граничным условиям (см. задачи 18, 19). ▶

Замечание *по поводу резольвентного уравнения и полной монотонности.* Тождество (10.5) может быть переписано в форме

$$\frac{\mathfrak{R}(\lambda) - \mathfrak{R}(\nu)}{\lambda - \nu} = -\mathfrak{R}(\nu)\,\mathfrak{R}(\lambda), \tag{10.16}$$

это и называют обычно *резольвентным уравнением.* Мы вывели его из условий теоремы Хилле — Иосида. Верно и обратное: *если семейство сжатий* $\lambda\mathfrak{R}(\lambda)$ *удовлетворяет* (10.16) *и область значений* $\mathfrak{R}(\lambda)$ *плотна, то существует такой оператор* $\mathfrak{A}$, *что верно* (10.2).

[Прежде всего из (10.16) следует, что область значений $\mathfrak{R}(\lambda)$ не зависит от λ. В самом деле, если $u = \mathfrak{R}(\lambda)z$, то $u = \mathfrak{R}(\nu)z'$, где $z' = z - (\lambda - \nu)u$. Если определить операторы $\mathfrak{A}_\lambda$ формулой $\mathfrak{A}_\lambda \mathfrak{R}(\lambda) = \lambda\mathfrak{R}(\lambda) - \mathbf{1}$, то $\mathfrak{A}_\lambda u = \lambda u - z$ и $\mathfrak{A}_\nu u = \nu u - z'$. В силу (10.16) $\mathfrak{A}_\lambda u = \mathfrak{A}_\nu u$, т. е. все операторы $\mathfrak{A}_\lambda$ идентичны, что доказывает наше утверждение.]

Если полугруппа $\mathfrak{T}(t)$ соответствует переходным вероятностям $Q_t(x, \Gamma)$ некоторого марковского процесса, то [см. (9.3)] операторы $\mathfrak{R}(\lambda)$ индуцируются субстохастическими ядрами

$$\int_0^\infty e^{-\lambda t} Q_t(x, \Gamma)\, dt \tag{10.17}$$

и при фиксированных x и Γ резольвентное уравнение (10.16) сводится к (8.2). Оно указывает, что свертка Q_s и Q_t дает Q_{s+t} (уравнение Чепмена — Колмогорова).

Из определения $\mathfrak{R}(\lambda)$ ясно, что формальные производные могут быть введены как сильные пределы разностных отношений. Так же как и в случае обычных преобразований, получаем

$$(-1)^n \mathfrak{R}^{(n)}(\lambda) = \int_0^\infty e^{-\lambda s} s^n \mathfrak{T}(s)\, ds. \tag{10.18}$$

Так как эти операторы *положительны*, то семейство $\{\mathfrak{R}(\lambda)\}$ представляет собой абстрактный аналог *вполне монотонных функций.* Это замечание

проливает новый свет на резольвентное уравнение (10.16). Действительно, полагая $\nu \to \lambda$, мы получаем слева производную $\mathfrak{R}'(\lambda)$, так что $-\mathfrak{R}'(\lambda) = \mathfrak{R}^2(\lambda)$. По индукции выводим, что $(-1)^n \mathfrak{R}^{(n)}(\lambda) = n! \mathfrak{R}^{n+1}(\lambda)$, где правая часть — положительный оператор. Таким образом, резольвентное уравнение влечет полную монотонность семейства $\{\mathfrak{R}(\lambda)\}$.

Абстрактная теория лишь повторяет теоремы, касающиеся обычных преобразований Лапласа. Чтобы подчеркнуть это еще раз, мы докажем *формулу обращения.*

Теорема 2. *При фиксированном* $t>0$ *и* $\lambda \to \infty$

$$\frac{(-1)^{n-1}}{(n-1)!} \mathfrak{R}^{(n-1)}\left(\frac{n}{t}\right)\left(\frac{n}{t}\right)^n \to \mathfrak{L}(t). \tag{10.19}$$

Доказательство. Из (10.18) видно, что левая часть является интегралом от $\mathfrak{L}(s)$ по мере с плотностью $\dfrac{e^{-ns/t}(ns/t)^{n-1}}{t(n-1)!}\, n$, которой соответствуют математическое ожидание t и дисперсия t^2/x. При $n \to \infty$ эта мера сходится к распределению, сосредоточенному в t, и ввиду непрерывности $\mathfrak{L}(s)$ отсюда вытекает (10.19) [так же как и в случае обычных функций, формула (10.19) — то же самое, что гл. VII, (1.6)].

§ 11. Задачи

1. *Обобщение теоремы о свертке.* Пусть u_1 и u_2 — две (скажем, непрерывные и ограниченные) функции на $\overline{0, \infty}$. Определим их «свертку по отношению к F» как

$$u(x) = \int_0^x u_1(x-y)\, u_2(y)\, F\{dy\}.$$

Пусть ω, ω_1 и ω_2 — преобразования Лапласа u, u_1 и u_2 по отношению к F, определенные, как в (1.5). Тогда $\omega = \omega_1 \omega_2$.

2. Пусть ω — преобразование (1.3) меры U. Тогда ω интегрируема на $\overline{0, 1}$ и $\overline{1, \infty}$ тогда и только тогда, когда функция $1/x$ интегрируема по отношению к U на $\overline{1, \infty}$ и $\overline{0, 1}$ соответственно.

3. *Равенство Парсеваля.* Если $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины с распределениями F и G и преобразованиями φ и γ, то преобразование для $\mathbf{XY}$ равно

$$\int_0^\infty \varphi(\lambda y)\, G\{dy\} = \int_0^\infty \gamma(\lambda y)\, F\{dy\}.$$

4. Из примера (3, д) выведите с помощью интегрирования, что $e^{1/\lambda} - 1$ *является обычным преобразованием* $I_1(2\sqrt{x})/\sqrt{x}$.

5. Покажите, исходя из определения гл. II, (7.1), что *обычное преобразование Лапласа функции* $I_0(x)$ равно $\omega_0(\lambda) = 1/\sqrt{\lambda^2 - 1}$ при $\lambda > 1$. [Вспомните тождество **1**, гл. II, (12.5) для $\binom{2n}{n}$.]

6. *Продолжение.* Покажите, что $I_0' = I_1$ и что, следовательно, I_1 имеет обычное преобразование Лапласа, равное $\omega_1(\lambda) = \omega_0(\lambda) R(\lambda)$, где $R(\lambda) = \lambda - \sqrt{\lambda^2 - 1}$.

7. *Продолжение.* Покажите, что $2I_n' = I_{n-1} + I_{n+1}$ при $n = 1, 2, \ldots$. Отсюда по индукции выведите, что I_n имеет обычное преобразование $\omega_n(\lambda) = \omega_0(\lambda) R^{(n)}(\lambda)$.

8. *Рациональные преобразования Лапласа.* Пусть $\varphi(\lambda) = U(\lambda)/V(\lambda)$, где U и V — полиномы без общих корней и степень m полинома U меньше, чем степень V. Предположим, что уравнение $V(\lambda) = 0$ имеет m различных (действительных или мнимых) корней $\lambda_1, \ldots, \lambda_m$. Тогда [1, гл. XI, (4.3)]

$$\varphi(\lambda) = \frac{\delta_1}{\lambda - \lambda_1} + \ldots + \frac{\delta_m}{\lambda - \lambda_m}. \qquad (*)$$

Покажите, что φ является обычным преобразованием Лапласа функции

$$f(x) = \delta_1 e^{\lambda_1 x} + \ldots + \delta_m e^{\lambda_m x}. \qquad (**)$$

Покажите, что вклад двух комплексно сопряженных корней $\alpha \pm i\beta$ имеет форму $e^{\alpha x}[a \cos \beta x + b \sin \beta x]$. Проанализируйте асимптотическое поведение f при $x \to \infty$ в терминах действительных частей корней.

Замечание. Многие аналитические функции, принадлежащие классу мероморфных функций, допускают разложение на простейшие дроби типа (*), но с заменой конечной суммы равномерно сходящимся рядом. Если действительные части α_j всех λ_j меньше a, то φ при $\lambda > a$ представляет собой обычное преобразование Лапласа некоторой функции f и асимптотическое поведение f на бесконечности определяется наибольшим из чисел α_j. Все сказанное легко распространяется и на случай кратных корней.

9. Покажите, что следствие теоремы 4.2 остается верным, если заменить f и $\varphi^{(n)}$ их абсолютными значениями.

10. *Интерполяция вполне монотонных функций.* Пусть $a_0, a_1, \ldots$ вполне монотонная последовательность, т. е. $(-1)^k \Delta^k a_n \geqslant 0$ (обозначения для разностей см. гл. VII, 1; мы используем шаг, равный единице). Докажите, что

$$f(\lambda) = \sum_{k=0}^{\infty} \Delta^k a_r \binom{\lambda - r}{k}, \qquad \lambda > 0 \qquad (*)$$

($r = 1, 2, \ldots$) не зависит от r и является вполне монотонной функцией, причем $f(n) = a_n$ для $n = 1, 2, \ldots$.

Указание. Положим

$$f_{r,N}(\lambda) = \sum_{k=0}^{N} \Delta^k a_r \binom{\lambda - r}{k}.$$

При $\lambda < r$ все члены справа положительны. Оцените $f_{r+1,N} - f_{r,N}$ и покажите, что при фиксированном N и $\lambda < r$ последовательность $f_{r,N}(\lambda)$ убывает с ростом r. Следовательно, $f_{r,N}(n) \leqslant a_n$ при $r \geqslant n$, и ряд (*) сходится. Так как $f_{r,N}$ является положительной линейной комбинацией полиномов $(r - \lambda)(r + 1 - \lambda) \ldots (r + k - 1 - \lambda)$, которые, очевидно, вполне монотонны при $\lambda < r$, то f — вполне монотонна. [Приведенная конструкция допускает обобщение на случай $a_n = \varphi(\lambda_n)$, где $\{\lambda_n\}$ — возрастающая последовательность

чисел с $\Sigma 1/\lambda_n = \infty$ (см. Duke Math. J., **5** (1939), 661—674). *Отсюда вытекает, что вполне монотонная функция однозначно определяется своими значениями* $\varphi(\lambda_n)$, *если* $\Sigma 1/\lambda_n = \infty$.]

11. *Продолжение.* Если последовательность $\{a_n\}$ вполне монотонна, то значения a_r, a_{r+1}, ... однозначно определяют значения a_1, a_2, ..., а также максимальное число a с тем свойством, что $a_0 \geqslant a$ и последовательность a, a_1, a_2, ... снова вполне монотонна.

12. Предполагая, что функция $e^{-x}I_n(x)$ монотонна на бесконечности, выведите из задачи 7, что

$$e^{-x}I_n(x) \sim \frac{1}{\sqrt{2\pi x}}, \qquad x \to \infty.$$

13. Пусть $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины с преобразованиями Лапласа φ и $e^{-\lambda^\alpha}$ соответственно. Тогда $\mathbf{Y}\mathbf{X}^{1/\alpha}$ имеет преобразование Лапласа $\varphi(\lambda^\alpha)$.

14. Пусть F — распределение с преобразованием φ. Если $a>0$, то $\varphi(\lambda+a)/\varphi(a)$ является преобразованием распределения $e^{-ax}F\{dx\}/\varphi(a)$. Установите, что при фиксированном $t>0$ функция[1]) $\exp\left[-t\sqrt{2\lambda+a^2}+at\right]$ есть преобразование безгранично-делимого распределения с плотностью

$$\frac{t}{\sqrt{2\pi x^3}} \exp\left[-\frac{1}{2}\left(\frac{t}{\sqrt{x}} - a\sqrt{x}\right)^2\right].$$

15. Каждое безгранично-делимое распределение является пределом обобщенных пуассоновских распределений.

16. Допустим, что в каноническом представлении (7.4) некоторого безгранично-делимого распределения $P(x) \sim x^c L(x)$ при $x \to \infty$, где $0<c<1$. Тогда $1 - F(x) \sim \frac{1}{1-c} x^{c-1} L(x)$. [Продолжение см. в примере (3, к) гл. XVII.]

17. Пусть P — производящая функция некоторой целочисленной безгранично-делимой случайной величины и φ — преобразование Лапласа некоторого распределения вероятностей. Тогда $P(\varphi)$ безгранично-делима.

18. *Диффузия с поглощающим экраном.* В примере § 10 примем $x>0$ и положим в определении ∇_h $u(x-h)=0$ при $x-h \leqslant 0$. Покажите, что доказательство сходимости остается в силе, если $\mathscr{L}$ обозначает пространство непрерывных функций с $u(\infty)=0$, $u(0)=0$, причем последнее условие нельзя отбросить. Предельная полугруппа описана в примере (5, б) гл. X.

19. *Отражающие экраны.* В примере § 10 примем $x>0$ и положим в определении ∇_h $u(x-h)=u(x+h)$ при $x-h<0$ Тогда $\nabla_h u$ сходится к пределу для каждой u с тремя ограниченными производными и с $u'(0)=0$. Область $\mathscr{L}'$ (область определения оператора $\mathfrak{A}$) выделяется этим граничным условием. Предельная полугруппа описана в примере (5, д) гл. X.

20. *Влияние максимального члена при сходимости к устойчивым распределениям.* Пусть $\mathbf{X}_1$, $\mathbf{X}_2$, ..., — независимые случайные величины с одним

[1]) Эта формула встречается в применениях и неоднократно выводилась с помощью длинных вычислений.

и тем же распределением F, удовлетворяющим (6.4), т. е. принадлежащим области притяжения устойчивого закона G_α. Положим $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ и $\mathbf{M}_n = \max[\mathbf{X}_1, \ldots, \mathbf{X}_n]$. Докажите, что отношение $\mathbf{S}_n/\mathbf{M}_n$ имеет преобразование Лапласа $\omega_n(\lambda)$, сходящееся к [1])

$$\omega(\lambda) = \frac{e^{-\lambda}}{1 + \alpha \int_0^1 (1 - e^{-\lambda t})\, t^{-\alpha-1}\, dt}. \qquad (*)$$

Следовательно, $\mathbf{E}(\mathbf{S}_n/\mathbf{M}_n) \to 1/(1-\alpha)$.

Указание. Оценивая интеграл по области $\mathbf{X}_j \leqslant \mathbf{X}_1$, получаем

$$\omega_n(\lambda) = ne^{-\lambda} \int_0^\infty F\{dx\} \left(\int_0^\infty e^{-\lambda y/x} F\{dy\} \right)^{n-1}.$$

Произведем замену $y = tx$ и затем $x = a_n s$, где a_n удовлетворяет (6.5). Внутренний интеграл, как легко видеть, равен

$$1 - \frac{1 - F(a_n s)}{n[1 - F(a_n)]} - \frac{1}{n} \int_0^1 (1 - e^{-\lambda t}) \frac{F\{a_n\, dt\}}{1 - F(a_n)} + o\left(\frac{1}{n}\right) =$$

$$= 1 - \frac{s^{-\alpha}\psi(\lambda)}{n} + o\left(\frac{1}{n}\right),$$

где $\psi(\lambda)$ обозначает знаменатель в (*). Таким образом,

$$\omega_n(\lambda) \to e^{-\lambda} \int_0^\infty e^{-s^{-\alpha}\psi(\lambda)} \cdot \frac{\alpha\, ds}{s^{\alpha+1}} = \omega(\lambda).$$

---

[1]) Этот результат был получен Д. Дарлингом в терминах характеристических функций. См. *Trans. Amer. Math. Soc.*, **73** (1952), 95—107.

ГЛАВА XIV

ПРИМЕНЕНИЕ ПРЕОБРАЗОВАНИЯ ЛАПЛАСА

Эта глава может служить для дополнительного чтения после гл. XIII. Она охватывает несколько независимых друг от друга тем от практических задач (§ 1, 2, 4, 5) до общих теорем существования § 7. Предельные теоремы § 3 иллюстрируют силу методов, развитых в связи с понятием правильно меняющихся функций. В последнем параграфе описываются средства анализа асимптотических свойств марковских процессов и времен первого прохождения для них.

§ 1. Уравнение восстановления: теория

Вероятностные основы читатель может найти в гл. VI, 6—7. Хотя теории восстановления целиком посвящена гл. XI, здесь мы излагаем независимый и значительно менее сложный подход. Сравнение методов и результатов поучительно. При известных основах теории преобразований Лапласа настоящий подход проще и прямо ведет к цели. Однако результат основной теоремы восстановления пока не удается получить с помощью преобразований Лапласа. С другой стороны, преобразования Лапласа позволяют легче получить предельные теоремы § 3 и явные решения типа, рассматриваемого в § 2.

Объектом нашего изучения будет интегральное уравнение

$$V(t) = G(t) + \int_0^t V(t-x)\,F\{dx\}, \tag{1.1}$$

в котором F и G — заданные монотонные непрерывные справа функции, равные нулю при $t<0$. Мы рассмотрим их как функции распределения некоторых мер и предположим, что F не сосредоточено в нуле и что их преобразования Лапласа

$$\varphi(\lambda) = \int_0^\infty e^{-\lambda t} F\{dt\}, \quad \gamma(\lambda) = \int_0^\infty e^{-\lambda t} G\{dt\} \tag{1.2}$$

определены при $\lambda>0$. Так же как и в предыдущей главе, все интервалы интегрирования считаются *замкнутыми*. Мы докажем, что (1.1) имеет ровно одно решение V; оно будет функцией распределения, преобразование Лапласа которой существует при всех $\lambda>0$. Если G имеет плотность g, то V имеет плотность v. Последняя удовлетворяет интегральному уравнению

$$v(t)=g(t)+\int_0^t v(t-x)\,F\{dx\}, \tag{1.3}$$

получаемому из (1.1) дифференцированием.

Вспоминая правило свертки, мы получаем для преобразования Лапласа ψ распределения V (или обычного преобразования его плотности) соотношение $\psi=\gamma+\psi\varphi$, откуда формально следует

$$\psi(\lambda)=\frac{\gamma(\lambda)}{1-\varphi(\lambda)}. \tag{1.4}$$

Чтобы доказать, что это формальное решение является преобразованием Лапласа некоторой меры (или плотности), мы рассмотрим отдельно три случая (из которых только первые два имеют вероятностное значение).

Случай (а). *F есть распределение вероятностей,* не сосредоточенное в нуле. Тогда $\varphi(0)=1$ и $\varphi(\lambda)<1$ при $\lambda>0$. Соответственно ряд

$$\sum_0^\infty \varphi^n=\frac{1}{1-\varphi}=\omega \tag{1.5}$$

сходится при $\lambda>0$. Очевидно, что функция ω вполне монотонна и поэтому является преобразованием Лапласа некоторой меры U (теорема 1 из гл. XIII,4). Далее, $\psi=\omega\gamma$ есть преобразование Лапласа для свертки $V=U*G$, т. е.

$$V(t)=\int_0^t G(t-x)\,U\{dx\}. \tag{1.6}$$

Наконец, если G имеет плотность g, то V имеет плотность $v=U*g$. Таким образом, мы доказали *существование* и *единственность* решения нашего интегрального уравнения. ▶

Асимптотическое поведение V на бесконечности описывается тауберовой теоремой 2 из гл. XIII, 4. Рассмотрим типичный случай, когда $G(\infty)<\infty$ и F имеет конечное математическое ожидание μ. В окрестности нуля $\psi(\lambda)\sim\mu^{-1}G(\infty)\lambda^{-1}$, откуда следует,

что

$$V(t) \sim \mu^{-1} G(\infty) \cdot t, \quad t \to \infty. \tag{1.7}$$

Из теоремы восстановления гл. XI,1 вытекает более точный результат

$$V(t+h) - V(t) \to \mu^{-1} G(\infty) h,$$

но его нельзя получить из тауберовых теорем. [Последние дают лучший результат, если F не имеет математического ожидания; см. § 3.]

Случай (б). *F — несобственное распределение, $F(\infty) < 1$.* Допустим для простоты, что $G(\infty) < \infty$. Можно применить предшествующие рассуждения с заметным упрощением: так как $\varphi(0) = F(\infty) < 1$, то $\omega(0) < \infty$ и мера V *на этот раз ограничена.*

Случай (в). *В этом случае $F(\infty) > 1$.* Для малых λ знаменатель в (1.4) отрицателен и для таких λ $\omega(\lambda)$ не может быть преобразованием Лапласа. К счастью, это обстоятельство не создает неудобств. Во избежание тривиальностей допустим, что F не имеет атома в нуле, так что $\varphi(\lambda) \to 0$ при $\lambda \to \infty$. В этом случае уравнение $\varphi(\varkappa) = 1$ имеет единственный корень $\varkappa > 0$. Рассуждения пункта (а) применимы без изменений при $\lambda > \varkappa$. Иными словами, существует единственное решение V, но его преобразование Лапласа ω сходится только при $\lambda > \varkappa$. Для этих значений ω по-прежнему определяется формулой (1.4) [1].

§ 2 Уравнение типа уравнения восстановления: примеры

а) *Время ожидания пробелов в процессе Пуассона.* Пусть V обозначает распределение времени ожидания до окончания первого пробела длины ξ в процессе Пуассона с параметром c (т. е в процессе восстановления с показательными временами между поступлениями). Распределение V изучалось аналитически в примере гл. XI, (7.б). Соответствующие практические ситуации (задержка пешехода или автомобиля, пытающегося пересечь поток автомашин, время блокировки в счетчиках Гейгера второго типа и т. п.) были описаны в гл. VI, 7. Здесь мы выведем нужное нам уравнение восстановления заново.

Время ожидания, если ожидание началось в момент 0, необходимо превосходит ξ. Оно меньше $t > \xi$, если до момента ξ не будет ни одного поступления (вероятность чего $e^{-c\xi}$), или же если

---

[1]) Решение V имеет вид $V\{dx\} = e^{\varkappa x} V^{\#}\{dx\}$, где $V^{\#}$ есть решение [с преобразованием Лапласа $\psi^{\#}(\lambda) = \psi(\lambda + \varkappa)$] стандартного уравнения восстановления (1.1) с распределением F, замененным на собственное распределение вероятностей $F^{\#}\{dx\} = e^{-\varkappa x} F\{dx\}$, и с G, замененным на $G^{\#}\{dx\} = e^{-\varkappa x} G\{dx\}$.

первое поступление происходит в моменты $x<\xi$ и остающееся время ожидания $\leqslant t-x$. В силу свойства отсутствия последействия вероятность $V(t)$ того, что время ожидания $\leqslant t$, равна

$$V(t)=e^{-c\xi}+\int_0^{\xi} V(t-x)\cdot e^{-cx}c\,dx \tag{2.1}$$

при $t\geqslant\xi$ и $V(t)=0$ при $t<\xi$. Несмотря на непривычный вид, (2.1) есть уравнение восстановления стандартного типа (1.1), в котором F имеет плотность $f(x)=ce^{-cx}$ на $0<x<\xi$, в то время как G сосредоточено в точке ξ. Таким образом,

$$\varphi(\lambda)=\frac{c}{c+\lambda}(1-e^{-(c+\lambda)\xi}),\quad \gamma(\lambda)=e^{-(c+\lambda)\xi} \tag{2.2}$$

и, следовательно, преобразование ψ для V равно

$$\psi(\lambda)=\frac{(c+\lambda)\,e^{-(c+\lambda)\xi}}{\lambda+ce^{-(c+\lambda)\xi}}. \tag{2.3}$$

Выражения гл. XI, (7.9) для математического ожидания и дисперсии получаются отсюда простым дифференцированием [1]). То же самое верно и для моментов высших порядков.

Поучителен вывод из (2.3) *явной формулы* для решения. По причинам, которые станут понятны, мы рассмотрим *хвост* $1-V(t)$ распределения. Его обычное преобразование Лапласа равно $[1-\psi(\lambda)]/\lambda$ [см. XIII, (2.8)]. Последняя функция может быть разложена в геометрический ряд

$$\frac{1-\psi(\lambda)}{\lambda}=\xi\sum_{n=1}^{\infty}c^{n-1}\xi^{n-1}\left\{\frac{1-e^{-(c+\lambda)\xi}}{(c+\lambda)\xi}\right\}^n. \tag{2.4}$$

Выражение в скобках отличается от преобразования Лапласа равномерного распределения только масштабным множителем ξ и заменой λ на $\lambda+c$. Как неоднократно отмечалось, эта замена соответствует умножению плотности на e^{-ct}. Таким образом,

$$1-V(t)=e^{-ct}\sum_{n=1}^{\infty}c^{n-1}\xi^{n-1}f^{n*}\left(\frac{t}{\xi}\right), \tag{2.5}$$

где f^{n*} есть плотность n-кратной свертки равномерного распределения с собою. Используя гл. I, (9.6), окончательно получаем

$$1-V(t)=e^{-ct}\sum_{n=1}^{\infty}\frac{(ct)^{n-1}}{(n-1)!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}\left(1-k\,\frac{\xi}{t}\right)_+^{n-1}. \tag{2.6}$$

[1]) Чтобы избежать громоздкого дифференцирования дробей, знаменатель лучше убрать, перенеся его в левую часть.

Интересна связь этого результата с *теоремами о покрытии*. Как было показано в гл. I, (9.9), при фиксированных t и ξ внутренняя сумма представляет собой вероятность того, что, *выбирая наудачу на* $\overline{0, t}$ $n-1$ *точек, мы разобьем этот интервал на* n *частей, каждая из которых не превышает* ξ. Отметим, что время ожидания превосходит t тогда и только тогда, когда каждый подинтервал $\overline{0, t}$ длины ξ содержит по крайней мере один момент поступления. Таким образом, (2.6) утверждает: если в процессе Пуассона на $\overline{0, t}$ происходит ровно $n-1$ поступлений, то *условное распределение моментов поступления равномерно*. Если исходить из этого факта, то (2.6) будет следствием теоремы о покрытии. С другой стороны, (2.6) дает новое доказательство теоремы о покрытии с помощью *рандомизации*.

б) *Задача о разорении в обобщенном процессе Пуассона.* В качестве второго иллюстративного примера мы возьмем *интегро-дифференциальное уравнение*

$$R'(t) = \frac{\alpha}{c} R(t) - \frac{\alpha}{c} \int_0^t R(t-x) F\{dx\}, \tag{2.7}$$

в котором F — распределение вероятностей с конечным математическим ожиданием μ. Это уравнение было получено в гл. VI, 5, где мы отметили его связь с теорией страхования, задачами хранения запасов и т. п. Решение этого уравнения и его асимптотические свойства были найдены другими методами в примере гл. XI, (7а).

Задача состоит в том, чтобы найти *распределение вероятностей* R, удовлетворяющее (2.7). Это уравнение связано с уравнением восстановления и может быть исследовано теми же приемами. Переходя к обычным преобразованиям Лапласа и отмечая, что

$$\rho(\lambda) = \int_0^\infty e^{-\lambda x} R(x)\, dx = \frac{1}{\lambda} \int_0^\infty e^{-\lambda x} R'(x)\, dx + \frac{1}{\lambda} R(0), \tag{2.8}$$

мы находим

$$\rho(\lambda) = \frac{R(0)}{1 - \frac{\alpha}{c} \frac{1-\varphi(\lambda)}{\lambda}} \cdot \frac{1}{\lambda}, \tag{2.9}$$

где φ — преобразование Лапласа для F. Вспоминая, что $[1-\varphi(\lambda)]/\lambda$ есть преобразование Лапласа, мы видим, что первая дробь в правой части имеет форму (1.4). Следовательно, эта дробь есть преобразование Лапласа — Стильтьеса некоторой меры R. Множитель $1/\lambda$ соответствует интегрированию. Поэтому $\rho(\lambda)$ является *обычным* преобразованием Лапласа некоторой функции распределения $R(x)$ [см. (2.8)]. Так как $R(x) \to 1$ при $x \to \infty$, то по теореме 4 из гл. XIII, 5 $\rho(\lambda) \to 1$ при $\lambda \to 0$. Из (2.9)

мы получаем теперь значение неизвестной постоянной $R(0)$

$$R(0)=1-\frac{\alpha}{c}\mu. \tag{2.10}$$

В соответствии с этим наша задача имеет *единственное решение* при $\alpha\mu<c$ *и не имеет решений при* $\alpha\mu\geqslant c$. Этот результат следовало бы ожидать исходя из вероятностных соображений.

Формула (2.9) встречается и в теории очередей (под названием формулы Хинчина — Полячека, см. пример гл. XII, (5а)). Много статей посвящено выводу явных формул при специальных предположениях. В случае *чисто пуассоновского процесса* F сосредоточено в точке 1 и $\varphi(\lambda)=e^{-\lambda}$. Выражение для ρ теперь почти такое же, как (2.3), и тот же метод легко дает *явную форму решения*

$$R(x)=\left(1-\frac{\alpha}{c}\right)\sum_{k=0}^{\infty}\left(\frac{-\alpha}{c}\right)^{k}\frac{(x-k)_{+}^{k}}{k!}\exp\left(\frac{\alpha}{c}(x-k)_{+}\right). \tag{2.11}$$

Эта формула, хотя она практически бесполезна, интересна наличием экспоненциальных функций с *положительными* показателями, которые должны занятным образом «компенсировать» друг друга. Формула была известна в связи с задачами страхования [1]) с 1934 г., но неоднократно переоткрывалась.

§ 3. Предельные теоремы, включающие распределения арксинуса

Стало привычным называть распределения, сосредоточенные на $\overline{0,1}$ и имеющие плотности

$$q_{\alpha}(x)=\frac{\sin\pi\alpha}{\pi}x^{-\alpha}(1-x)^{\alpha-1}, \qquad 0<\alpha<1, \tag{3.1}$$

«обобщенными распределениями арксинуса», хотя на самом деле это специальные бета-распределения. Частный случай $\alpha=\frac{1}{2}$ соответствует функции распределения $2\pi^{-1}\arcsin\sqrt{x}$, играющей важную роль в теории флуктуаций в случайных блужданиях. Все большее число исследований приводит к предельным распределениям, связанным с q_{α}. Сложность вычисления этих распределений делает появление q_{α} мистическим. Подлинная причина заключается в близкой связи q_{α} с функциями распределения, имеющими правильно меняющиеся хвосты, т. е. функциями распределения вида.

$$1-F(x)=x^{-\alpha}L(x), \qquad 0<\alpha<1, \tag{3.2}$$

[1]) Явное решение в задаче о разорении до момента t дано Пайком (Pyke R., The supremum and infimum of the Poisson process, *Ann. Math. Statist.*, **30** (1959), 568—576).

где $L(tx)/L(t)\to 1$ при $t\to\infty$. Для таких функций теорема восстановления может быть дополнена в том смысле, что для «функции восстановления» $U=\Sigma F^{n*}$ выполняется соотношение

$$U(t)\sim \frac{1}{\Gamma(1-\alpha)\,\Gamma(1+\alpha)}\,\frac{t^{\alpha}}{L(t)}, \qquad t\to\infty. \tag{3.3}$$

Другими словами, если F меняется правильно, то это же верно и для U. Известно (хотя и не очевидно), что константа в (3.3) равна $(\sin\pi\alpha)/\pi\alpha$, так что соотношение (3.3) может быть переписано в виде

$$[1-F(x)]\,U(x)\to\frac{\sin\pi\alpha}{\pi\alpha}, \qquad x\to\infty. \tag{3.4}$$

Лемма. *Если F удовлетворяет* (3.2), *то выполняется* (3.4).

Доказательство. По тауберовой теореме 4 гл. XIII, 5

$$1-\varphi(\lambda)\sim\Gamma(1-\alpha)\,\lambda^{\alpha}L(1/\lambda),\ \lambda\to 0.$$

Преобразование Лапласа для U равно $\sum\varphi^n=1/(1-\varphi)$ и по теореме 2 из гл. XIII, 5, верно (3.3). ▶

Рассмотрим теперь последовательность положительных независимых случайных величин $\mathbf{X}_k$ с одним и тем же распределением F и их частные суммы $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$. При фиксированном t обозначим при $\mathbf{N}_t$ случайный индекс, такой, что

$$\mathbf{S}_{\mathbf{N}_t}\leqslant t\leqslant\mathbf{S}_{\mathbf{N}_t+1}. \tag{3.5}$$

Мы интересуемся длинами двух подинтервалов

$$\mathbf{Y}_t=t-\mathbf{S}_{\mathbf{N}_t}\ \text{ и }\ \mathbf{Z}_t=\mathbf{S}_{\mathbf{N}_t+1}-t.$$

Они были введены в гл. VI, 7 под названием *«прошедшее время ожидания»* и *«остающееся время ожидания»* в момент t. Причины интереса, который могут представлять эти величины, объяснялись ранее по-разному. В гл. XI, 3 было доказано, что при $t\to\infty$ случайные величины $\mathbf{Y}_t$ и $\mathbf{Z}_t$ имеют (одно и то же) собственное предельное распределение тогда и только тогда, когда F имеет конечное математическое ожидание. В противном случае $\mathbf{P}\{\mathbf{Y}_t\leqslant x\}\to 0$ при любом фиксированном x. Аналогичное соотношение верно и для $\mathbf{Z}_t$. Побочным результатом наших исследований является следующая интересная теорема (первоначальное доказательство которой представляло значительные аналитические трудности [1])).

[1]) Дынкин Е. Б., Некоторые предельные теоремы для сумм независимых случайных величин с бесконечными математическими ожиданиями, *Изв. АН СССР*, серия матем., **19** (1955), 247—266.

Теорема. *Если имеет место* (3.2), *то нормированная случайная величина* $\mathbf{Y}_t/t$ *имеет предельную плотность* q_α, *определяемую по* (3.1), *а* $\mathbf{Z}_t/t$ *имеет предельную плотность*[1])

$$p_\alpha(x)=\frac{\sin\pi\alpha}{\pi}\cdot\frac{1}{x^\alpha(1+x)}, \qquad x>0. \tag{3.6}$$

Доказательство. Неравенство $tx_1<\mathbf{Y}_t<tx_2$ осуществляется тогда и только тогда, когда при некоторых n и y, $1-x_2<y<1-x_1$, имеем $\mathbf{S}_n=ty$ и $\mathbf{X}_{n+1}>t(1-y)$. Суммируя по всем возможным n и y, получаем

$$\mathbf{P}\{tx_1<\mathbf{Y}_t<tx_2\}=\int_{1-x_2}^{1-x_1}[1-F(t(1-y))]\,U\{t\,dy\}, \tag{3.7}$$

или, используя (3,4),

$$\mathbf{P}\{tx_1<\mathbf{Y}_t<tx_2\}\sim\frac{\sin\pi\alpha}{\pi\alpha}\int_{1-x_2}^{1-x_1}\frac{1-F(t(1-y))}{1-F(t)}\cdot\frac{U\{t\,dy\}}{U(t)}. \tag{3.8}$$

Теперь $\dfrac{U(ty)}{U(t)}\to y^\alpha$, так что мера $U\{tdy\}/U(t)$ сходится к мере с плотностью $\alpha y^{\alpha-1}$. В то же время первый множитель под интегралом сходится к $(1-y)^{-\alpha}$. В силу монотонности наших функций сходимость равномерна. Поэтому

$$\mathbf{P}\{tx_1<\mathbf{Y}_t<tx_2\}\to\frac{\sin\pi\alpha}{\pi}\int_{1-x_2}^{1-x_1}y^{\alpha-1}(1-y)^{-\alpha}\,dy, \tag{3.9}$$

чем доказано первое утверждение. Для вероятности $\mathbf{P}\{\mathbf{Z}_t>ts\}$ мы получаем такой же интеграл, но в пределах от 0 до $1/(1+s)$. Теперь дифференцированием получаем (3.6). ▶

Примечательно, что плотность q_α становится бесконечной вблизи точек 0 и 1. Следовательно, наиболее вероятные значения величины $\mathbf{Y}_t/t$ лежат вблизи точек 0 и 1.

Легко так усовершенствовать наши рассуждения, чтобы получить утверждения, обратные к лемме и теореме. При этом станет видно, что условие (3.2) *необходимо* для существования предельного распределения величины $\mathbf{Y}_t/t$. С другой стороны, (3.2) характеризует области притяжения устойчивых распределений. Этим и объясняется частое появление q_α в связи с подобными распределениями.

---

[1]) Так как $\mathbf{S}_{\mathbf{N}_t+1}=\mathbf{Z}_t+t$, то распределение $\mathbf{Z}_t/\mathbf{S}_{\mathbf{N}_t+1}$ получается из (3.6) заменой переменных $x=y(1-y)$. Таким образом, мы видим, что величина $\mathbf{Z}_t/\mathbf{S}_{\mathbf{N}_t+1}$ *имеет предельную плотность* q_α.

§ 4. Периоды занятости и соответствующие ветвящиеся процессы

В примере гл. XIII, (4.a) было показано, что если φ есть преобразование Лапласа некоторого распределения вероятностей F с математическим ожиданием μ, то уравнение

$$\beta(\lambda) = \varphi(\lambda + c - c\beta(\lambda)), \quad \lambda > 0, \tag{4.1}$$

имеет единственное решение β. Сверх того, β *является преобразованием Лапласа некоторого распределения* B; *при этом* B *собственно, если* $c\mu \leqslant 1$, *и несобственно в противном случае.* Этот простой и красивый результат находит все больше и больше применений. Стоит поэтому объяснить вероятностный смысл уравнения (4.1) и его применения.

Вывод уравнения (4.1) и аналогичных уравнений оказывается простым для тех, кто привык выражать вероятностные соотношения непосредственно в терминах преобразований Лапласа. Типичная ситуация такова. Рассмотрим случайную сумму $\mathbf{S_N} = \mathbf{X}_1 + \ldots + \mathbf{X_N}$, где $\mathbf{X}_j$ независимы и имеют преобразование Лапласа $\gamma(\lambda)$ и $\mathbf{N}$ — не зависящая от $\mathbf{X}$ случайная величина с производящей функцией $P(s)$. Преобразование Лапласа $\mathbf{S_N}$ равно, очевидно, $P(\gamma(\lambda))$ [см. пример гл. XIII, (3.c)]. Для пуассоновской случайной величины $\mathbf{N}$ это преобразование Лапласа имеет вид $e^{-\alpha[1-\gamma(\lambda)]}$. Как мы неоднократно видели, в применениях параметр α часто считают случайным с некоторым распределением U. Используя ту же терминологию, что и для распределений, мы можем сказать, что $e^{-\alpha[1-\gamma(\lambda)]}$ есть условное преобразование Лапласа для $\mathbf{S_N}$ при данном значении α нашего параметра. Безусловное преобразование Лапласа получается интегрированием по отношению к U. Благодаря специальному виду подинтегрального выражения результат равен, очевидно, $\omega(1-\gamma(\lambda))$, где ω обозначает преобразование Лапласа для U.

Примеры. а) *Периоды занятости*[1]). Клиенты (или вызовы) приходят к обслуживающему прибору (линии связи) в момен-

[1]) То, что (4.1) возникает в связи с периодами занятости, было отмечено Д. Кендаллом (Kendall D. G., Some problems in the theory of queues., *J. Roy. Statist. Soc. (B)*, **13** (1951), 151—185). Красивая редукция к ветвящимся процессам была предложена И. Гудом (Good I. J.). Уравнение (4.1) равносильно уравнению

$$B(t) = \sum \int_0^t e^{-cx} \frac{(cx)^n}{n!} B^{n*}(t-x)\, F\{dx\},$$

которое часто называют *интегральным уравнением Такача.* Внутреннюю простоту теории не всегда понимают.

ты, образующие процесс Пуассона с плотностью c. Последовательные времена обслуживания предполагаются независимыми случайными величинами с одним и тем же распределением F. Допустим, что в момент 0 приходит вызов и линия свободна. Обслуживание начинается мгновенно. Вызовы, поступающие в моменты, когда линия занята, встают в очередь, и обслуживание продолжается без перерывов до тех пор, пока не исчезнет очередь. Под *периодом занятости* понимается интервал от нуля до первого момента, когда линия снова становится свободной. Его продолжительность является случайной величиной. Мы обозначим через B и β ее распределение и преобразование Лапласа соответственно.

В терминологии ветвящихся процессов клиент, начинающий период занятости, есть «предок». Клиенты, прибывающие в моменты, когда все еще обслуживается первый, суть его прямые «потомки» и т. д. При условии, что предок уходит в момент x, число $\mathbf{N}$ его прямых потомков распределено по закону Пуассона с математическим ожиданием cx. Обозначим $\mathbf{X}_j$ *полное время обслуживания j-го прямого потомка и всех его потомков.* Хотя соответствующие промежутки времени не обязаны следовать непосредственно один за другим, их общая сумма имеет, очевидно, такое же распределение, как и период занятости. Полное время обслуживания всех (прямых и непрямых) потомков равно поэтому $\mathbf{S}_N = \mathbf{X}_1 + \ldots + \mathbf{X}_N$, где $\mathbf{X}_j$ имеет преобразование Лапласа β и все случайные величины независимы. Чтобы получить период занятости, мы должны добавить время x обслуживания предка. Соответственно при данной продолжительности x времени обслуживания предка длительность периода занятости $x+\mathbf{S}_N$ имеет (условное) преобразование Лапласа $e^{-x[\lambda+c-c\beta(\lambda)]}$. Параметр x имеет распределение F. Интегрирование по x приводит к (4.1).

Если распределение B несобственно, то его дефект $1-B(\infty)$ равен вероятности того, что период занятости никогда не кончится (переполнение). Условие $c\mu \leqslant 1$ означает, что математическое ожидание полного времени обслуживания всех вызовов, поступающих за единицу времени, не должно превышать единицы. Нетрудно вычислить, исходя из (4.1), математическое ожидание и дисперсию B.

Частный случай. Если $F(t) = 1-e^{-\alpha t}$, то (4.1) становится квадратным уравнением относительно β и B можно выразить через функцию Бесселя. Это проделано в примере (6.б) [указанный результат был использован в примере с очередями гл. VI, (9.е)].

б) *Задержки в уличном движении*[1]). Допустим, что автомобили, проходящие данный участок дороги, образуют процесс Пуассона с параметром c. Пусть движение задерживается на время δ (красным светом или какой-либо другой причиной). Когда движение возобновляется, в очереди будут стоять $\mathbf{K}$ автомобилей; $\mathbf{K}$ имеет распределение Пуассона. Так как r-й автомобиль в очереди не может двинуться ранее $r-1$ автомобилей, находящихся впереди него, то каждый автомобиль создает задержку для всех следующих за ним автомобилей. Естественно предположить, что отдельные задержки суть независимые случайные величины с одним и тем же распределением F. При наличии ожидающих вновь прибывший автомобиль обязан стать в очередь, тем самым увеличивая общую задержку. Положение такое же, как в предыдущем примере, с той лишь разницей, что здесь имеется $\mathbf{K}$ «предков». Полная задержка, создаваемая каждым автомобилем и всеми его прямыми и непрямыми потомками, имеет преобразование Лапласа, удовлетворяющее (4.1), а полный «период занятости» — промежуток от возобновления движения до первого момента, когда нет ожидающих автомобилей, — имеет преобразование Лапласа $e^{-c\delta[1-\beta(\lambda)]}$. Легко вычислить математическое ожидание задержки. Этот результат можно использовать для выяснения эффекта, создаваемого последовательными светофорами, и т. п. (см. задачи 5 и 6).

§ 5. Диффузионные процессы

Как мы знаем, в одномерном броуновском движении переходные вероятности нормальны, а время первого прохождения имеет устойчивое распределение с показателем $\frac{1}{2}$ (см. пример гл. VI, (2.e)). Имея в своем распоряжении эти результаты, мы не в праве ожидать новой информации от использования преобразований Лапласа. Причиной возвращения заново к диффузионным уравнениям служит то, что метод сам по себе поучителен и применим к наиболее общим диффузионным уравнениям (с той разницей, что *явные решения* не удается получить при произвольных коэффициентах). Для упрощения записи мы допустим с самого начала, что переходные вероятности Q_t имеют плотности q_t (хотя метод, который мы укажем, был бы применим без всяких специальных ограничений).

---

[1]) Этот пример возник под влиянием изучения Дж. Литтлом числа задержания автомобилей [Little J. D. C., *Operations Res.*, **9** (1961), 39—52]

Мы начнем с частного случая броуновского движения. Пусть f — данная непрерывная ограниченная функция. Положим

$$u(t, x) = \int_{-\infty}^{+\infty} q_t(x, y) f(y)\, dy. \tag{5.1}$$

Нашим исходным пунктом будет установленный в гл. X, (4а) факт, что (по крайней мере для достаточно гладких f) u удовлетворяет диффузионному уравнению

$$\frac{\partial u(t, x)}{\partial t} = \frac{1}{2} \frac{\partial^2 u(t, x)}{\partial x^2} \tag{5.2}$$

и начальному условию $u(t, x) \to f(x)$ при $t \to 0$. Вводя обычное преобразование Лапласа

$$\omega_\lambda(x) = \int_0^\infty e^{-\lambda t} u(t, x)\, dt, \tag{5.3}$$

мы выводим из (5.2), что [1])

$$\lambda \omega_\lambda - \frac{1}{2} \omega_\lambda'' = f, \tag{5.4}$$

а из (5.1), что

$$\omega_\lambda(x) = \int_{-\infty}^{+\infty} K_\lambda(x, s) f(s)\, ds, \tag{5.5}$$

где $K_\lambda(x, y)$ — обычное преобразование Лапласа функции $q_t(x, y)$. В теории дифференциальных уравнений K_λ называют функцией Грина для уравнения (5.4). В нашем случае

$$K_\lambda(x, y) = \frac{1}{\sqrt{2\lambda}} e^{-\sqrt{2\lambda}\,|x-y|}. \tag{5.6}$$

В самом деле, легко проверить, что при ограниченной и непрерывной f уравнение (5.4) имеет единственное ограниченное решение, что оно определяется по (5.5) и (5.6).

Мы намерены вывести (5.6) из вероятностных соображений, применимых и к более общим уравнениям и приводящих к явным выражениям для времен первого прохождения. Мы примем как данное, что траектория $\mathbf{X}(t)$ непрерывно зависит от t. Пусть $\mathbf{X}(0) = x$. Обозначим через $F(t, x, y)$ вероятность того, что

[1]) Читатели, ознакомившиеся с параграфами, посвященными полугруппам, заметят, что мы имеем дело с марковской полугруппой, порождаемой дифференциальным оператором $A = \frac{1}{2} d^2/dx^2$. Дифференциальное уравнение (5.4) есть частный случай основного уравнения гл. XIII, (10.1), встречающегося в теореме Хилле — Иосиды.

точка y будет достигнута до момента t. Мы назовем F распределением момента первого перехода из x в y и обозначим его преобразование Лапласа через $\varphi_\lambda(x, y)$.

Пусть $x<y<z$. Тогда событие $\mathbf{X}(t)=z$ имеет место тогда и только тогда, когда первый переход в y происходит в некоторый момент $\tau<t$ и когда за ним следует переход из y в z за время $t-\tau$. Таким образом, $q_t(x, z)$ есть свертка $F(t, x, y)$ и $q_t(y, z)$, откуда

$$K_\lambda(x, z)=\varphi_\lambda(x, y)K_\lambda(y, z). \tag{5.7}$$

Выберем теперь в качестве f функцию, равную нулю на $\overline{-\infty, z}$. В этом случае (5.5) и (5.7) показывают, что

$$\omega_\lambda(x)=\varphi_\lambda(x, y)\omega_\lambda(y), \qquad x<y. \tag{5.8}$$

В то же время в силу (5.4) при фиксированном y $\varphi_\lambda(x, y)$ удовлетворяет дифференциальному уравнению

$$\lambda\varphi_\lambda-\frac{1}{2}\frac{\partial^2\varphi_\lambda}{\partial x^2}=0, \qquad x<y. \tag{5.9}$$

Ограниченное на $-\infty$ решение необходимо имеет вид $C_\lambda e^{-\sqrt{2\lambda}\,x}$. Так как по (5.8) $\varphi_\lambda(x, y)\to 1$ при $x\to y$, то $\varphi_\lambda(x, y)=e^{\sqrt{2\lambda}\,(x-y)}$ при $x<y$. По соображениям симметрии выводим отсюда, что *преобразование Лапласа времени первого перехода из x в y равно*

$$\varphi_\lambda(x, y)=e^{-\sqrt{2\lambda}\,|x-y|}. \tag{5.10}$$

В нашем случае K_λ инвариантно относительно сдвигов. Следовательно, $K_\lambda(y, y)$ не зависит от y, и из (5.7) вытекает равенство

$$K_\lambda(x, y)=C_\lambda e^{-\sqrt{2\lambda}\,|x-y|}.$$

Итак, мы определили K_λ с точностью до постоянного множителя C_λ; из того, что функции $f=1$ соответствует решение $\omega_\lambda(x)=1/\lambda$, легко выводим $\sqrt{2\lambda}\,C_\lambda=1$. Формула (5.6) доказана.

Следующие далее примеры показывают, как вычислить вероятность достичь какую-либо точку $y_1>x$ ранее, чем будет достигнута некоторая другая точка $y_2<x$. В то же время эти примеры проиллюстрируют, как учитываются *граничные условия*.

Примеры. а) *Один поглощающий экран.* Броуновское движение на $\overline{0, \infty}$ с поглощающим экраном в нуле получается остановкой обычного броуновского движения с $\mathbf{X}(0)=x>0$, когда оно достигает начала координат. Обозначим соответствующие вероятности перехода через $q_t^{\text{абс}}(x, y)$. Аналогично изменим другие обозначения. В неограниченном броуновском движении плотность вероятности перехода от $x>0$ к $y>0$ с заходом в 0 равна

свертке плотности первого перехода из x в 0 и $q_t(0, y)$. Соответствующее преобразование Лапласа равно $\varphi_\lambda(x, 0)K_\lambda(0, y)$. Следовательно, должно быть

$$K_\lambda^{\text{абс}}(x, y) = K_\lambda(x, y) - \varphi_\lambda(x, 0)K_\lambda(0, y), \tag{5.11}$$

где $x > 0$, $y > 0$. Это эквивалентно

$$K_\lambda^{\text{абс}}(x, y) = \frac{1}{\sqrt{2\lambda}}\left[e^{-\sqrt{2\lambda}\,|x-y|} - e^{-\sqrt{2\lambda}\,(x+y)}\right] \tag{5.12}$$

или

$$q_t^{\text{абс}}(x, y) = q_t(x, y) - q_t(x, -y) \tag{5.13}$$

в соответствии с решением гл. X, (5.5), найденным *методом отражений*.

Рассуждения, приводящие к (5.7), применимы без изменений и к процессу с поглощающим экраном. Поэтому из (5.12) мы выводим, что при $0<x<y$

$$\varphi_\lambda^{\text{абс}}(x, y) = \frac{e^{\sqrt{2\lambda}\,x} - e^{-\sqrt{2\lambda}\,x}}{e^{\sqrt{2\lambda}\,y} - e^{-\sqrt{2\lambda}\,y}}. \tag{5.14}$$

Это есть[1]) преобразование Лапласа *для вероятности того, что в неограниченном броуновском движении с* $\mathbf{X}(0) = x$ *точка* $y > x$ *будет достигнута до момента* t *и без захода в нуль*. При $\lambda \to 0$ приходим к выводу: *вероятность того, что точка* y *будет достигнута ранее точки* 0, *равна* x/y (точно так же как в симметричном случайном блуждании, соответствующем схеме Бернулли (см. задачу о разорении в **1**, гл. XIV, 2).

Отметим, наконец, что $\omega_\lambda^{\text{абс}}$ является единственным решением (5.4), ограниченным на $\overline{0, \infty}$ и удовлетворяющим граничному условию $\omega_\lambda^{\text{абс}}(0) = 0$. [В частности, функции $f = 1$ соответствует решение $\omega_\lambda^{\text{абс}}(x) = (1 - e^{-\sqrt{2\lambda}\,x})/\lambda$. Это обычное преобразование функции $1 - F(t, x, 0)$. Ее значения равны вероятности того, что поглощение не произойдет до момента t; интегрирование по частям показывает, что преобразование Лапласа — Стильтьеса функции $F(t, x, 0)$ равно $e^{-\sqrt{2\lambda}\,x}$, что соответствует (5.10).]

Можно предложить читателю следующее поучительное упражнение: вывести (используя метод, примененный к неограниченному случайному блужданию (5.12) и (5.14)) непосредственно из дифференциального уравнения (5.4) и граничного условия $\omega_\lambda^{\text{абс}}(0) = 0$.

---

[1]) При фиксированном y $\varphi_\lambda^{\text{абс}}$ является решением дифференциального уравнения (5.9), обращающимся в нуль при $x = 0$ и в 1 при $x = y$. В такой форме результат применим к любой тройке точек $a<x<b$ и $a>x>b$ и к более общим дифференциальным уравнениям.

б) *Два поглощающих экрана.* Рассмотрим теперь броуновское движение, начинающееся в точке x интервала $\overline{0, 1}$ и оканчивающееся при достижении точки 0 или точки 1. Легче всего этот процесс получить из предыдущего процесса с одним поглощающим экраном, ставя дополнительный экран в точке 1. При этом рассуждения, приводящие к (5.11), применимы без изменений. Переходные плотности $q_t^{\#}(x, y)$ нового процесса имеют преобразование Лапласа $K_\lambda^{\#}$, определяемое соотношением

$$K_\lambda^{\#}(x, y) = K_\lambda^{\text{абс}}(x, y) - \varphi_\lambda^{\text{абс}}(x, 1) K_\lambda^{\text{абс}}(1, y), \qquad (5.15)$$

где x и y принадлежат интервалу $\overline{0, 1}$. [Отметим, что выполняются граничные условия $K_\lambda^{\#}(0, y) = K_\lambda^{\#}(1, y)$.] Простые вычисления показывают, что

$$K_\lambda^{\#}(x, y) = \frac{e^{-\sqrt{2\lambda}|x-y|} - e^{-\sqrt{2\lambda}(2-|x-y|)} - e^{-\sqrt{2\lambda}(x+y)} - e^{-\sqrt{2\lambda}(2-x-y)}}{\sqrt{2\lambda}\left(1-e^{-2\sqrt{2\lambda}}\right)}. \qquad (5.16)$$

Разлагая дробь $1/\left[1-e^{-2\sqrt{2\lambda}}\right]$ в геометрический ряд, мы приходим к другому представлению

$$K_\lambda^{\#}(x, y) = \frac{1}{\sqrt{2\lambda}} \sum_{n=-\infty}^{+\infty} \left[e^{-\sqrt{2\lambda}|x-y+2n|} - e^{-\sqrt{2\lambda}|x+y+2n|}\right]. \qquad (5.17)$$

Последнее эквивалентно решению гл. X, (5.7), полученному с помощью *принципа отражений.* ▶

Сходные рассуждения применимы и к *более общему диффузионному уравнению*

$$\frac{\partial u(t, x)}{\partial t} = \frac{1}{2} a(x) \frac{\partial^2 u(t, x)}{\partial x^2} + b(x) \frac{\partial u(t, x)}{\partial x}, \quad a > 0, \qquad (5.18)$$

на конечном или бесконечном интервале. Вместо (5.4) мы получаем

$$\lambda\omega_\lambda - \frac{1}{2} a\omega_\lambda'' - b\omega_\lambda' = f, \qquad (5.19)$$

и решение снова имеет форму (5.5) с функцией Грина K_λ вида (5.7). При этом $\varphi_\lambda(x, y)$ есть преобразование плотности первого перехода от x к $y > x$. При фиксированном y эта функция должна удовлетворять дифференциальному уравнению, соответствующему (5.9), а именно

$$\lambda\varphi_\lambda - \frac{1}{2} a\varphi_\lambda'' - b\varphi_\lambda' = 0. \qquad (5.20)$$

Эта функция должна быть ограничена на левом конце. Кроме того, $\varphi_\lambda(y, y)=1$. Эти условия определяют φ_λ однозначно, за исключением случая, когда (5.20) имеет ограниченное решение. В последнем случае (как и в приведенных выше примерах) следует добавить надлежащие граничные условия (см. задачи 7 и 8).

§ 6. Процессы размножения и гибели. Случайные блуждания

В этом параграфе мы исследуем связь между процессами размножения и гибели из 1, гл. XVII, 5 и рандомизированным случайным блужданием из гл. II, 7. Основная цель состоит в том, чтобы проиллюстрировать технику применения преобразований Лапласа и правильное использование граничных условий.

Рассмотрим простое случайное блуждание, начинающееся в нуле, в котором размер отдельного скачка равен 1 или -1 с вероятностями p и q соответственно. Промежутки между последовательными скачками предполагаются независимыми показательно распределенными случайными величинами с математическим ожиданием $1/c$. Вероятность $P_n(t)$ находиться в состоянии n в момент t была найдена в гл. II, (7.7), но мы проведем вычисления заново, с новой точки зрения. Составляя уравнение для $P_n(t)$, мы будем рассуждать следующим образом. Состояние $n\neq 0$ может наблюдаться в момент t только при условии, что до момента t произошел скачок. Пусть первый скачок происходит в момент $x<t$ и приводит в точку 1. Тогда (условная) вероятность состояния n в момент t равна $P_{n-1}(t-x)$. Таким образом, при $n\neq 0$

$$P_n(t)=\int_0^t ce^{-cx}\,[pP_{n-1}(t-x)+qP_{n+1}(t-x)]\,dx. \tag{6.1}$$

При $n=0$ к правой части следует добавить член e^{-ct}, соответствующий отсутствию скачков на $\overline{0, t}$. Мы получаем, таким образом, бесконечную систему *уравнений свертки*, и, как мы увидим, она легко решается. Прежде чем переходить к решению, заметим, что наша система уравнений эквивалентна бесконечной системе *дифференциальных уравнений* [1])

$$P_n'(t)=-cP_n(t)+cpP_{n-1}(t)+cqP_{n-1}(t) \tag{6.2}$$

с начальными условиями $P_0(0)=1$, $P_n(0)=0$ при $n\neq 0$. В самом деле, после замены переменных $y=t-x$ уравнения свертки (6.1) можно продифференцировать, что приведет к (6.2).

[1]) Эта система представляет собой частный случай системы уравнений 1, гл XVII, (5, г) для процессов размножения и гибели и может быть получена таким же путем.

Системы (6.1) и (6.2) эквивалентны, но последняя имеет то формальное преимущество, что специальная роль состояния $n=0$ проявляется только в начальных условиях.

Для применения преобразований Лапласа безразлично, отправляемся ли мы от (6.1) или от (6.2). Умножая на $e^{-\lambda t}$ и интегрируя, получаем

$$\pi_n(\lambda)=\frac{c}{c+\lambda}[p\pi_{n-1}(\lambda)+q\pi_{n+1}(\lambda)], \qquad n\neq 0, \tag{6.3а}$$

$$\pi_0(\lambda)=\frac{1}{c+\lambda}+\frac{c}{c+\lambda}[p\pi_{-1}(\lambda)+q\pi_1(\lambda)], \tag{6.3б}$$

где, конечно,

$$\pi_n(\lambda)=\int_0^\infty e^{-\lambda t}P_n(t)\,dt. \tag{6.4}$$

Система линейных уравнений (6.3) аналогична системе, уже появлявшейся в связи со случайными блужданиями в **1**, гл. XIV. Мы можем решить ее таким же методом. Квадратное уравнение

$$cqs^2-(c+\lambda)s+cp=0 \tag{6.5}$$

имеет один корень, ограниченный при всех $\lambda>0$,

$$s_\lambda=\frac{c+\lambda-\sqrt{(c+\lambda)^2-4c^2pq}}{2cq}, \tag{6.6}$$

и другой корень, σ_λ, стремящийся к бесконечности вместе с λ. Легко проверить, что при произвольных постоянных A_λ и B_λ линейные комбинации $\pi_n(\lambda)=A_\lambda s_\lambda^n+B_\lambda\sigma_\lambda^n$ удовлетворяют (6.3а) при $n=1, 2, \ldots$. Коэффициенты можно выбрать так, чтобы получить для $\pi_0(\lambda)$ и $\pi_1(\lambda)$ правильные значения. Если π_0 и π_1 даны, то, исходя из (6.3а), можно вычислить последовательно $\pi_2, \pi_3, \ldots$. Поэтому при $n\geqslant 0$ *каждое* решение имеет вид $\pi_n(\lambda)=A_\lambda s_\lambda^n+B_\lambda\sigma_\lambda^n$. Так как наше решение обязано оставаться ограниченным, то необходимо $B_\lambda=0$ и

$$\pi_n(\lambda)=\pi_0(\lambda)\,s_\lambda^n, \qquad n=0, 1, 2, \ldots. \tag{6.7}$$

При $n=-1, -2, \ldots$ мы имеем аналогичное выражение, в которое входит другой корень. Однако проще заметить, что $\pi_{-n}(\lambda)$ *можно получить из* $\pi_n(\lambda)$, *меняя* p *и* q *местами. Отсюда и из* (6.3б) *выводим*

$$\pi_0(\lambda)=\frac{1}{\sqrt{(c+\lambda)^2-4c^2pq}}, \tag{6.8}$$

так что все $\pi_n(\lambda)$ определяются единственным образом.

Найдя преобразование Лапласа для решения, мы можем пойти тремя путями. Во-первых, мы можем пытаться извлечь максимум информации из тауберовых теорем, вычисления моментов и т. п. В большинстве практических ситуаций нет другого выбора. Во-вторых, таблицы преобразований Ланласа могут дать явные выражения для неизвестных $P_n(t)$. В рассмотренном сейчас случае мы пришли к преобразованию Лапласа функций Бесселя I_n из задачи 7 гл. XIII, 11. Подбирая очевидным образом параметры расположения, видим, что $P_n(t)=a_n(ct)$, где a_n определяется по гл. II, (7.7). Наконец в исключительно благоприятной ситуации, когда явное решение известно заранее, мы можем рассматривать наш прием как метод вычисления преобразования Лапласа для $P_n(t)$. В этом смысле, мы дали *новый вывод формулы для преобразования Лапласа функции Бесселя* I_n.

Для преобразований Лапласа типично, что ряд заключений можно сделать прямо из формы решения (6.7). Так как перемножение преобразований Лапласа соответствует свертке, то из (6.7) следует, что P_n имеет вид $P_n=F^{n*}*P_0$, где F (возможно, несобственное) — распределение вероятностей с преобразованием, равным s_λ. С вероятностной точки зрения очевидно, что F должно быть распределением момента первого прохождения через точку 1 (и поэтому F^{n*} должно быть распределением времени первого прохождения через точку n). Как было показано в примере гл. XIII, (3г), функция $(\lambda-\sqrt{\lambda^2-1})^r$ представляет собой (при $\lambda>1$) обычное преобразование Лапласа для $(r/x)I_r(x)$. Замена λ на $\lambda/2c\sqrt{pq}$ меняет лишь масштабный параметр, а замена λ на $\lambda+c$ соответствует умножению плотности на e^{-cx}. Отсюда следует, что при $n>0$ s_λ^n есть *обычное преобразование Лапласа для*

$$f_n(t)=\sqrt{\left(\frac{p}{q}\right)^n}\,\frac{n}{t}\,I_n(2c\sqrt{pq}\,t)e^{-ct}, \tag{6.9}$$

последняя функция есть плотность времени первого прохождения через точку n. Это согласуется с гл. II, (7.7), а в симметричном случае $p=q$ также и с гл. XIII, (3.6).

Как мы уже видели в 1, гл. XVII, 7, различные задачи связи и обслуживания приводят к той же самой системе дифференциальных уравнений (6.2) с той лишь разницей, что рассматриваются лишь неотрицательные значения n и что граничному значению $n=0$ соответствует иное уравнение. Как действует наш метод в подобных случаях, покажут два следующих примера.

Примеры. а) *Очередь к одному обслуживающему прибору.* Мы рассмотрим пример **1**, гл. XVII, (7.б) в предположении, что имеется очередь к одной линии. Состояние системы описывается числом лиц в очереди, включая и обслуживаемое лицо. Как промежутки между вызовами, так и длительности обслуживания считаются распределенными показательно. Чтобы согласовать обозначения, мы примем их математические ожидания равными $1/(cp)$ и $1/(cq)$ соответственно (в прежних обозначениях $cp=\lambda$ и $cq=\mu$). При $n\geqslant 1$ инфинитезимальные вероятности перехода точно такие же, как в нашем случайном блуждании. Поэтому выполняются уравнения (6.2). При $n=0$, однако, имеем

$$P_0'(t) = -cpP_0(t) + cqP_1(t). \tag{6.10}$$

При начальном условии $P_0(0)=1$ уравнения для преобразований Лапласа при $n\geqslant 1$ будут совпадать с (6.3а), а (6.3б) заменится на

$$(cp+\lambda)\pi_0(\lambda) = 1 + cq\pi_1(\lambda). \tag{6.11}$$

Как и прежде, при $n\geqslant 1$ имеем $\pi_n(\lambda)=\pi_0(\lambda)\,s_\lambda^n$, но в силу (6.11)

$$\pi_0(\lambda) = \frac{1}{cp+\lambda-cqs_\lambda} = \frac{1-s_\lambda}{\lambda}. \tag{6.12}$$

(Здесь было использовано квадратное уравнение (6.5) для s_λ.) Суммируя геометрическую прогрессию, получаем $\pi_n(\lambda)+\pi_{n+1}(\lambda)+\ldots = s_\lambda^n/\lambda$. Сравнивая с предыдущим результатом, выводим, наконец, что при $n>0$

$$P_n(t)+P_{n+1}(t)+\ldots = F_n(t), \tag{6.13}$$

где F_n — распределение с плотностью (6.9). При $n=0$ левая часть равна, конечно, единице.

б) *Флуктуации в период занятости.* Рассмотрим по-прежнему один обслуживающий прибор, но только в период занятости. Другими словами, предполагается, что в момент 0 поступает вызов, причем на свободную линию, и что мы обрываем процесс, когда линия впервые снова становится свободной. Аналитически это означает, что теперь рассматриваются значения $n\geqslant 1$ и начальное условие имеет вид $P_1(0)=1$. При $n\geqslant 2$ в уравнениях (6.2) ничто не меняется, однако в связи с отсутствием нулевого состояния в первом уравнении отпадает член $cpP_0(t)$. Таким образом, преобразование Лапласа $\pi_n(\lambda)$ удовлетворяет при $n\geqslant 2$ уравнениям (6.3а). Кроме того,

$$(\lambda+c)\pi_1(\lambda) = 1 + cq\pi_2(\lambda). \tag{6.14}$$

Как и прежде, при $n \geqslant 2$ получаем $\pi_n(\lambda) = \pi_1(\lambda)\, s_\lambda^{n-1}$. Однако $\pi_1(\lambda)$ следует определить, исходя из (6.14). Принимая во внимание квадратное уравнение (6.5), легко выводим, что $\pi_n(\lambda) = s_\lambda^n/cp$. Таким образом, мы установили *окончательный результат*: $P_n(t) = f_n(t)/cp$, где f_n определяется по (6.9).

Сумма $P(t) = \sum P_n(t)$ равна вероятности того, что *продолжительность периода занятости превосходит* t. Из дифференциальных уравнений видно, что $P'(t) = -cqP_1(t)$. Таким образом, *продолжительность периода занятости имеет плотность*

$$-P'(t) = c\sqrt{\frac{q}{p}}\,\frac{1}{t} I_1\left(2c\sqrt{pq}\,t\right) e^{-ct}. \tag{6.15}$$

Этот результат был получен другим методом в конце примера (4. а) и был использован при изучении длины очереди в примере гл. VI, (9. д) (см. задачу 11).

§ 7. Дифференциальные уравнения Колмогорова [1])

Мы вернемся теперь к марковским процессам, состояния которых описываются целыми числами 1, 2, ... Дифференциальные уравнения Колмогорова были выведены в 1, гл. XVII, 9 и заново в гл. X, 3. В этом параграфе дано независимое изложение, основанное на преобразованиях Лапласа.

Чтобы сделать это изложение замкнутым, мы приведем новый вывод основных уравнений, на этот раз в форме уравнений свертки.

Основное предположение состоит в том, что если в некоторый момент τ $\mathbf{X}(\tau) = i$, то в течение некоторого промежутка времени $\tau \leqslant t < \tau + \mathbf{T}$, длина которого имеет показательную плотность $c_i e^{-c_i x}$, значение $\mathbf{X}(t)$ остается постоянным. Вероятность того, что $\mathbf{X}(\tau + \mathbf{T}) = j$, равна p_{ij}. Пусть дано, что $\mathbf{X}(0) = i$. Тогда вероятность $P_{ik}(t)$ того, что $\mathbf{X}(t) = k \neq i$, может быть вычислена суммированием по всем возможным значениям момента и размера первого скачка:

$$P_{ik}(t) = \sum_{j=1}^{\infty} \int_0^t c_i e^{-c_i x} p_{ij} P_{jk}(t-x)\, dx \qquad (k \neq i). \tag{7.1a}$$

[1]) Приводимые теоремы и доказательства в равной степени применимы и к скачкообразным процессам, описанным в гл. X, 3. В качестве хорошего упражнения можно предложить провести доказательство в терминах самих распределений вероятностей (а не их преобразований Лапласа). Этот вариант менее элегантен, но применим и в нестационарном случае.

Вероятностное изложение, опирающееся на изучение траекторий, см. в книге Чжуна, 1960.

Обобщение на *полумарковские процессы* см. в задаче 12.

При $k=i$ мы должны добавить член, соответствующий отсутствию скачков

$$P_{ii}(t)=e^{-c_i t}+\sum_{j=1}^{\infty}\int_0^t c_i e^{-c_i x}p_{ij}P_{ji}(t-x)\,dx. \qquad (7.1б)$$

Эти уравнения могут быть объединены, если использовать символ Кронекера δ_{ik}, равный 1 при $k=i$ и 0 при $k\neq i$.

Нашим отправным пунктом будет обратное уравнение (7 1)[1]; при любых данных $c_i>0$ и стохастической матрице $\mathbf{p}=(p_{ik})$ мы ищем стохастические матрицы $P(t)=(P_{ik}(t))$, удовлетворяющие (7 1)

Предполагая, что в любой конечный интервал времени происходит лишь конечное число скачков, мы можем видоизменить наши рассуждения, рассматривая момент x последнего скачка до момента t Вероятность скачка из j в k имеет плотность $\sum P_{ij}(x)c_j p_{jk}$, а вероятность отсутствия скачка между x и t равна $e^{-c_k(t-x)}$. Поэтому вместо (7.1) мы получаем *прямое уравнение*

$$P_{ik}(t)=\delta_{ik}e^{-c_i t}+\int_0^t\sum_{j=1}^{\infty}P_{ij}(x)\,c_j p_{jk}e^{-c_k(t-x)}\,dx. \qquad (7.2)$$

Однако, как мы увидим, существуют процессы с *бесконечным числом скачков*, удовлетворяющие обратному уравнению. Следовательно, прямое уравнение не вытекает из основных предположений о рассматриваемом процессе

В терминах преобразований Лапласа

$$\Pi_{ik}(\lambda)=\int_0^{\infty}e^{-\lambda t}P_{ik}(t)\,dt \qquad (7.3)$$

обратное уравнение (7.1) принимает вид

$$\Pi_{ik}(\lambda)=\frac{\delta_{ik}}{\lambda+c_i}+\frac{c_i}{\lambda+c_i}\sum_{j=1}^{\infty}p_{ij}\Pi_{jk}(\lambda). \qquad (7.4)$$

---

[1]) Замена переменных $y=t-x$ облегчает дифференцирование, производя которое найдем, что уравнения свертки (7 1) эквивалентны системе дифференциальных уравнений

$$P'_{ik}(t)=-c_i P_{ik}(t)+c_i\sum_j p_{ij}P_{jk}(t)$$

при начальных условиях $P_{ii}(0)=1$ и $P_{ik}(0)=0$ при $k\neq i$ Эта система согласуется с I, гл XVII, (9 14) с той разницей, что там коэффициенты c_i и p_{ij} зависели от времени и P_{ik} было функцией двух переменных τ и t, а не только разности $t-\tau$

Мы перейдем теперь к более удобным матричным обозначениям (правила действий над матрицами вполне применимы к бесконечным матрицам с неотрицательными элементами). Мы вводим матрицы $\Pi(\lambda) = (\Pi_{ik}(\lambda))$, $P(t) = (P_{ik}(t))$, $\mathbf{p} = (p_{ik})$ и диагональную матрицу $\mathbf{c}$ с элементами c_i. Через $\mathbf{1}$ мы обозначим вектор-столбец, все компоненты которого равны 1. Суммы по строкам матрицы A образуют вектор-столбец $A\mathbf{1}$. Наконец, I обозначает единичную матрицу.

Из (7.4) видно, что *обратные уравнения* (7.1) превращаются в

$$(\lambda + \mathbf{c})\,\Pi(\lambda) = I + \mathbf{cp}\Pi(\lambda), \tag{7.5}$$

а *прямые уравнения* — в

$$\Pi(\lambda)(\lambda + \mathbf{c}) = I + \Pi(\lambda)\,\mathbf{cp}. \tag{7.6}$$

Чтобы построить *минимальное решение*, полагаем последовательно

$$(\lambda + \mathbf{c})\,\Pi^{(0)}(\lambda) = I, \qquad (\lambda + \mathbf{c})\,\Pi^{(n+1)}(\lambda) = I + \mathbf{cp}\Pi^{(n)}(\lambda). \tag{7.7}$$

Для сумм по строкам матрицы $\lambda\Pi^{(n)}(\lambda)$ мы вводим обозначение

$$\lambda\Pi^{(n)}(\lambda)\,\mathbf{1} = \mathbf{1} - \xi^{(n)}(\lambda). \tag{7.8}$$

Подставляя в (7.7) и вспоминая, что $p\mathbf{1} = \mathbf{1}$, получаем

$$(\lambda + c)\,\xi^{(n+1)}(\lambda) = \mathbf{cp}\xi^{(n)}(\lambda). \tag{7.9}$$

Так как $\xi^{(0)} \geqslant 0$, то $\xi^{(n)}(\lambda) \geqslant 0$ при всех n. Поэтому матрицы $\lambda\Pi^{(n)}(\lambda)$ — субстохастические. Их элементы суть неубывающие функции n. Поэтому существует конечный предел

$$\Pi^{(\infty)}(\lambda) = \lim_{n\to\infty} \Pi^{(n)}(\lambda) \tag{7.10}$$

и матрица $\lambda\Pi^{(\infty)}(\lambda)$ — стохастическая или субстохастическая.

Очевидно, что $\Pi^{(\infty)}(\lambda)$ удовлетворяет обратному уравнению (7.5) и что для любого другого неотрицательного решения тривиально верно неравенство $\Pi(\lambda) \geqslant \Pi^{(0)}(\lambda)$. По индукции $\Pi(\lambda) \geqslant \Pi^{(n)}(\lambda)$ при всех n. Таким образом,

$$\Pi(\lambda) \geqslant \Pi^{(\infty)}(\lambda). \tag{7.11}$$

Менее очевидно, что $\Pi^{(\infty)}(\lambda)$ удовлетворяет и прямому уравнению (7.6). Для доказательства этого мы установим по индукции, что

$$\Pi^{(n)}(\lambda)(\lambda + \mathbf{c}) = I + \Pi^{(n-1)}(\lambda)\,\mathbf{cp}. \tag{7.12}$$

Это верно при $n=1$. Предполагая справедливость (7.12) и производя подстановку в (7.7), получаем

$$(\lambda+\mathbf{c})\,\Pi^{(n+1)}(\lambda)(\lambda+\mathbf{c})=\lambda I+\mathbf{c}+[I+\mathbf{c}\mathbf{p}\Pi^{(n-1)}(\lambda)]\,\mathbf{c}\mathbf{p}. \quad (7.13)$$

Выражение в скобках равно $(\lambda+\mathbf{c})\Pi^{(n)}(\lambda)$. Умножая (7.13) слева на $(\lambda+\mathbf{c})^{-1}$, приходим к (7.12) с заменой n на $n+1$. Следовательно, это соотношение верно при всех n, и поэтому $\Pi^{(\infty)}(\lambda)$ удовлетворяет прямому уравнению. Повторяя предыдущие рассуждения, найдем, что любое неотрицательное решение удовлетворяет (7.11). Нами доказана, таким образом,

Теорема 1. *Существует матрица* $\Pi^{(\infty)}(\lambda)\geqslant 0$, *у которой суммы по строкам не превосходят* λ^{-1} *и которая удовлетворяет уравнениям* (7.5) *и* (7.6). *Сверх того, для любого неотрицательного решения уравнения* (7.5) *или* (7.6) *выполняется неравенство* (7.11).

Мы будем называть $\Pi^{(\infty)}(\lambda)$ *минимальным* решением.

Теорема 2. *Минимальное решение является преобразованием Лапласа семейства субстохастических или стохастических матриц* $P(s)$, *удовлетворяющих уравнению Чепмена — Колмогорова*

$$P(s+t)=P(s)P(t), \quad (7.14)$$

а также прямому и обратному уравнениям (7.1)—(7.2). *При этом или все матрицы* $P(t)$ *и* $\lambda\Pi^{(\infty)}(\lambda)$ $(t>0,\ \lambda>0)$ *строго стохастические, или ни одна из них не обладает этим свойством.*

Доказательство. Отбросим верхний индекс ∞ и будем писать $\Pi(\lambda)$ вместо $\Pi^{(\infty)}(\lambda)$. Из определения (7.7) ясно, что $\Pi_{ik}^{(n)}(\lambda)$ есть преобразование положительной функции $P_{ik}^{(n)}$, представляющее собой свертку конечного числа показательных распределений. В силу (7.8) суммы по строкам элементов $P^{(n)}(t)$ образуют монотонно возрастающую последовательность, ограниченную **1**. Поэтому $\Pi(\lambda)$ есть преобразование матрицы $P(t)$, стохастической или субстохастической. Из (7.5)—(7.6) ясно, что $P(t)$ удовлетворяет первоначальным прямому и обратному уравнениям; отсюда вытекает, что $P(t)$ непрерывно зависит от t. Ясно, что если для $P(t)$ i-я сумма по строкам <1 при некотором t, то i-я сумма по строкам матрицы $\Pi(\lambda)$ будет $<\lambda^{-1}$ при всех λ и обратно.

Выразим уравнение (7.14) в терминах преобразований Лапласа. Для этого умножим его на $e^{-\lambda t-\nu s}$ и произведем интегрирование по t и s. Справа получим произведение матриц $\Pi(\lambda)\,\Pi(\nu)$.

Левая часть легко оценивается с помощью подстановки $x=t+s$, $y=-t+s$. Окончательный результат таков

$$-\frac{\Pi(\nu)-\Pi(\lambda)}{\nu-\lambda}=\Pi(\lambda)\Pi(\nu). \tag{7.15}$$

Обратно, (7.15) влечет (7.14) [это рассуждение повторяется в примере гл. XIII, (8а)].

Для доказательства (7.15) рассмотрим матричное уравнение

$$(\lambda+\mathbf{c})Q=A+\mathbf{cp}Q. \tag{7.16}$$

Если A и Q неотрицательны, то, очевидно, $Q\geqslant(\lambda+\mathbf{c})^{-1}A=\Pi^{(0)}(\lambda)A$ и по индукции $Q\geqslant\Pi^{(n)}(\lambda)A$ при всех n. Таким образом $Q\geqslant\Pi(\lambda)A$. Теперь $\Pi(\nu)$ удовлетворяет (7.16) с $A=I+(\lambda-\nu)\Pi(\nu)$ и потому при $\lambda>\nu$

$$\Pi(\nu)\geqslant\Pi(\lambda)+(\lambda-\nu)\Pi(\lambda)\Pi(\nu). \tag{7.17}$$

С другой стороны, правая часть (7.17) удовлетворяет прямому уравнению (7.6) с λ, замененным на ν. Поэтому она не меньше $\Pi(\nu)$, и в (7.17) имеет место знак равенства. Доказательство закончено [1]). ▶

Чтобы понять, является ли матрица $\lambda\Pi^{(\infty)}(\lambda)$ строго стохастической [2]), мы вернемся к соотношениям (7.8) и (7.9). Так как элементы $\xi_i^{(n)}(\lambda)$ суть невозрастающие функции n, то существует предел $\xi(\lambda)=\lim\xi^{(n)}(\lambda)$, такой, что

$$\lambda\Pi^{(\infty)}(\lambda)\mathbf{1}=\mathbf{1}-\xi(\lambda) \tag{7.18}$$

и

$$(\lambda+\mathbf{c})\xi(\lambda)=\mathbf{cp}\xi(\lambda),\qquad 0\leqslant\xi(\lambda)\leqslant\mathbf{1}. \tag{7.19}$$

С другой стороны, мы имеем

$$(\lambda+\mathbf{c})\xi^{(0)}(\lambda)=\mathbf{c1}=\mathbf{cp1} \tag{7.20}$$

и потому $\xi^{(0)}(\lambda)\geqslant\xi(\lambda)$ для *любого* вектора, удовлетворяющего (7.19). Из (7.9) по индукции выводим $\xi^{(n)}(\lambda)\geqslant\xi(\lambda)$ для всех n, так, что вектор $\xi(\lambda)$ в (7.18) является максимальным вектором, удовлетворяющим (7.19).

---

[1]) Уравнение (7.15) есть резольвентное уравнение для семейства сжатий $\lambda\Pi(\lambda)$ на банаховом пространстве ограниченных векторов-столбцов. Мы видели в гл. XIII, 10, что оно выполняется тогда и только тогда, когда область значений этих преобразований не зависит от λ. Это гарантируется их минимальным характером.

[2]) *Предостережение.* Казалось бы, что формальное умножение прямого уравнения на вектор-столбец $\mathbf{1}$ должно привести к тождеству $\lambda\Pi(\lambda)\mathbf{1}=\mathbf{1}$. Однако получающийся ряд может расходиться. Эта процедура законна, если все c_i ограничены (следствие 1).

Таким образом, верна *

Теорема 3. *Дефекты по строкам минимального решения описываются максимальным вектором* $\xi(\lambda)$, *удовлетворяющим* (7.19).

Следовательно, матрица $\lambda\Pi^{(\infty)}(\lambda)$ будет строго стохастической тогда и только тогда, когда (7.19) влечет $\xi(\lambda)=0$.

Следствие 1. *Если* $c_i \leqslant M < \infty$ *при всех* i, *то минимальное решение будет строго стохастическим (так что ни прямое, ни обратное уравнения не имеют других допустимых решений).*

Доказательство. Дробь $c/(\lambda+c)$ является возрастающей функцией c. Из (7.19) по индукции получаем

$$\xi(\lambda) \leqslant \left(\frac{M}{\lambda+M}\right)^n \cdot \mathbf{1} \tag{7.21}$$

при всех n. Следовательно, $\xi(\lambda)=0$. ▶

Если $A(\lambda)$ — матрица с элементами вида $\xi_i(\lambda)\eta_k(\lambda)$ с произвольными $\eta_k(\lambda)$, то $\Pi(\lambda)+A(\lambda)$ снова будет решением обратного уравнения (7.5). Всегда можно выбрать $A(\lambda)$ так, чтобы получить допустимые матрицы, удовлетворяющие уравнению Чепмена — Колмогорова. Эта процедура проиллюстрирована в следующем параграфе. Соответствующие процессы характеризуются наличием бесконечного числа скачков в конечном промежутке времени. Любопытно, что прямые уравнения могут выполняться, даже когда их интерпретация в терминах последнего скачка недопустима.

Таковы основные результаты. Мы закончим параграф критерием, полезным в применениях и интересным с той точки зрения, что при его доказательстве появляются понятия теории *потенциала:* ядро Г в (7.25) служит типичным примером потенциала.

Допустим, что $c_i>0$. Перепишем (7.19) в форме

$$\xi(\lambda) + \lambda \mathbf{c}^{-1}\xi(\lambda) = \mathbf{p}\xi(\lambda). \tag{7.22}$$

Умножая на $\mathbf{p}^k$ и суммируя по $k=0, \ldots, n-1$, получаем

$$\xi(\lambda) + \lambda \sum_{k=0}^{n-1} \mathbf{p}^k \mathbf{c}^{-1}\xi(\lambda) = \mathbf{p}^n\xi(\lambda). \tag{7.23}$$

Отсюда следует, что $\mathbf{p}^n\xi(\lambda)$ монотонно зависит от n, так что $\mathbf{p}^n\xi(\lambda) \to x$, где x — минимальный вектор-столбец, удовлетворяющий [1])

$$\mathbf{p}x = x, \quad \xi(\lambda) \leqslant x \leqslant \mathbf{1}. \tag{7.24}$$

[1]) Нетрудно видеть, что x не зависит от λ и $\lambda\Pi^{(\infty)}(\lambda)x = x - \xi(\lambda)$ (см. сноску 2 на стр. 557).

Определим теперь матрицы (с возможно бесконечными элементами) равенством

$$\Gamma = \sum_{k=1}^{\infty} \mathbf{p}^k \mathbf{c}^{-1}. \tag{7.25}$$

Полагая в (7.23) $n \to \infty$, получаем

$$\xi(\lambda) + \lambda \Gamma \xi(\lambda) = x. \tag{7.26}$$

Отсюда вытекает, что $\xi_k(\lambda) = 0$ при каждом k, для которого $\Gamma_{kk} = \infty$. Так будет, если k является возвратным состоянием цепи Маркова с матрицей $\mathbf{p}$. Мы имеем

Следствие 2. *Минимальное решение будет строго стохастическим (и потому единственным), если в дискретной цепи Маркова с матрицей* $\mathbf{p}$ *все состояния возвратны.*

§ 8. Пример: чистый процесс размножения

Вместо того чтобы продолжать изложение общей теории, мы подробно рассмотрим процессы, в которых возможны лишь переходы $i \to i+1$. Они дают хорошую иллюстрацию того, какими могут быть процессы в обстановке неединственности. Во избежание тривиальностей допустим, что $c_i > 0$ при всех i. По определению $p_{i,\,i+1} = 1$ и $p_{i,\,k} = 0$ для всех других пар индексов. Обратное и прямое уравнения сводятся теперь к

$$(\lambda + c_i)\Pi_{ik}(\lambda) - c_i \Pi_{i+1,\,k}(\lambda) = \delta_{ik} \tag{8.1}$$

и

$$(\lambda + c_k)\Pi_{ik}(\lambda) - c_{k-1}\Pi_{i,\,k-1}(\lambda) = \delta_{ik}, \tag{8.2}$$

где $\delta_{ik} = 1$ при $i = k$ и 0 в других случаях. Положим для краткости

$$\rho_i = \frac{c_i}{c_i + \lambda}, \quad r_i = \frac{1}{c_i + \lambda}. \tag{8.3}$$

Функция ρ_i есть преобразование Лапласа (показательного) времени пребывания в состоянии i, а r_i есть обычное преобразование Лапласа для вероятности того, что это время превосходит t. В дальнейшем следует иметь в виду зависимость r_j и ρ_j от λ.

а) *Минимальное решение.* Легко проверить, что

$$\Pi_{ik}(\lambda) = \begin{cases} \rho_i \rho_{i+1} \cdots \rho_{k-1} r_k & \text{для } k \geqslant 1, \\ 0 & \text{для } k < 1 \end{cases} \tag{8.4}$$

является минимальным решением как для (8.1), так и для (8.2). В (8.4) отражается тот факт, что переходы из i в $k > i$ невозмож-

ны, а момент прихода в $k>i$ равен сумме $k-i$ независимых времен пребывания в $i, i+1, \ldots, k-1$.

Посмотрим, будут ли суммы по строкам давать единицу. Отметим, что $\lambda r_k = 1 - \rho_k$, откуда

$$\lambda[\Pi_{ii}(\lambda) + \ldots + \Pi_{i,\,i+n}(\lambda)] = 1 - \rho_i \ldots \rho_{i+n}. \tag{8.5}$$

Переходя к логарифмам, видим, что произведение в правой части стремится к нулю тогда и только тогда, когда ряд $\sum \frac{1}{c_n}$ расходится. Таким образом, *переходные вероятности процесса, определяемого* (8.4), *дают в сумме единицу тогда и только тогда, когда* $\sum \frac{1}{c_n} = \infty$. (Этот результат был получен в [1], гл. XVII, 4.) В этом случае нет никаких неожиданностей: величины c_n однозначно определяют процесс размножения, удовлетворяющий нашим основным исходным постулатам.

Предположим теперь, что

$$\sum 1/c_n < \infty. \tag{8.6}$$

Дефект $1 - \sum_k p_{ik}(t)$ равен вероятности того, что к моменту t система прошла через *все* состояния или «достигла границы ∞». Момент достижения есть сумма времен пребывания в $i, i+1, \ldots$. Этот ряд сходится с вероятностью единица, так как сходится в силу (8.6) ряд из соответствующих математических ожиданий. В процессе, начинающемся в состоянии i, *момент достижения* ∞ *имеет преобразование Лапласа*

$$\xi_i = \lim_{n\to\infty} \rho_i \rho_{i+1} \ldots \rho_{i+n} \tag{8.7}$$

и ξ_i удовлетворяют уравнению (7.19), а именно

$$(\lambda + c_i)\xi_i = c_i \xi_{i+1}. \tag{8.8}$$

б) *Процесс с возвращением.* Отправляясь от процесса (8.4), можно определить ряд новых процессов следующим образом. Выберем числа q_i, такие, что $q_i \geqslant 0$, $\sum q_i = 1$. Мы требуем, чтобы после достижения ∞ *система мгновенно приходила в состояние* i *с вероятностью*[1]) q_i.

Первоначальный процесс повторяется заново до второго достижения ∞. Время между двумя достижениями есть случайная величина с преобразованием Лапласа

$$\tau(\lambda) = \sum q_i \xi_i. \tag{8.9}$$

[1]) Вариант процесса с возвращением получается при $\Sigma q_i < 1$; при достижении ∞ процесс с вероятностью $1 - \Sigma q_i$ обрывается.

Марковский характер процесса требует, чтобы после второго достижения процесс вел себя прежним образом. Теперь мы опишем переходные вероятности $P_{ik}^{\text{ret}}(t)$ нового процесса в терминах соответствующих преобразований Лапласа $\Pi_{ik}^{\text{ret}}(A)$. Вероятность перехода из состояния i в момент 0 в состояние k в момент t без промежуточного захода в ∞ имеет преобразование (8.4). Вероятность достичь k после ровно одного захода в ∞ имеет преобразование Лапласа $\xi_i \sum_j q_j \Pi_{jk}(\lambda)$ и момент второго захода в ∞ имеет преобразование $\xi_i \tau(\lambda)$. Рассматривая последующие возвращения, мы приходим к формуле

$$\Pi_{ik}^{\text{ret}}(\lambda) = \Pi_{ik}(\lambda) + \xi_i \frac{1}{1-\tau(\lambda)} \sum_j q_j \Pi_{jk}(\lambda), \tag{8.10}$$

где $[1-\tau(\lambda)]^{-1} = \sum \tau^n(\lambda)$ учитывает многократные прохождения через ∞. Тривиальные вычисления показывают, что суммы по строкам в (8.10) равны $1/\lambda$, поэтому $\Pi_{ik}^{\text{ret}}(\lambda)$ суть преобразования строго стохастических матриц переходных вероятностей $P^{\text{ret}}(t)$.

Легко проверить, что *новый процесс удовлетворяет обратным уравнениям* (8.1), *но не удовлетворяет прямым уравнениям* (8.2). Все происходит так, как и должно быть: постулаты, приводящие к прямому уравнению, нарушены, так как последний скачок не обязан существовать.

в) *Двусторонний процесс размножения.* Чтобы получить процесс, удовлетворяющий и прямому и обратному уравнениям, мы видоизменим процесс размножения, приняв в качестве состояний системы числа 0, ± 1, ± 2, ... Все другие условия остаются прежними: постоянные $c_i > 0$ определены для всех целых i и переходы возможны только из i в $i+1$. Предположим опять, что $\sum 1/c_n < \infty$, где суммирование идет от $-\infty$ до ∞.

По-прежнему *минимальное* решение определяется по (8.4). Предел

$$\eta_k = \lim_{i \to -\infty} \Pi_{ik}(\lambda) = r_k \rho_{k-1} \rho_{k-2} \rho_{k-3} \cdots \tag{8.11}$$

существует и может быть интерпретирован как преобразование «вероятности $P_{-\infty,k}(t)$ перехода из $-\infty$ в момент 0 в k в момент t». При такой начальной точке процесс пробегает все состояния от $-\infty$ до ∞ и «достигает ∞» в момент с преобразованием Лапласа $\xi_{-\infty} = \lim_{n \to -\infty} \xi_n$. Определим теперь новый процесс следующим образом. Он начинается как процесс, соответствующий минимальному решению (8.4), но по достижении ∞ он

перескакивает в $-\infty$ и таким образом продолжается неопределенно долго. Построение примера (б) приводит к следующим преобразованиям переходных вероятностей:

$$\Pi^{\sharp}_{ik}(\lambda) = \Pi_{ik}(\lambda) + \frac{\xi_i \eta_k}{1-\xi_{-\infty}}. \tag{8.12}$$

Легко проверить, *что* $\Pi^{\sharp}_{ik}$ *удовлетворяет и обратному и прямому уравнениям* (8.1) *и* (8.2). В то же время процесс удовлетворяет предположениям, приводящим к обратным уравнениям, но не предположениям, приводящим к прямым уравнениям.

§ 9. Вычисление $P(\infty)$ и времен первого прохождения

Можно ожидать, что поведение при $t \to \infty$ переходных вероятностей $P_{ij}(t)$ в марковском процессе с целочисленными состояниями аналогично поведению переходных вероятностей за большое число шагов в дискретных цепях Маркова (с тем, однако, приятным упрощением, что помехи, создаваемые периодичностью, исчезают). В теореме 1 этот факт устанавливается как простое следствие эргодических теорем **1**, гл. XV. Затем нашей основной заботой будет вычисление пределов для переходных вероятностей процессов § 7. Мы покажем также, как могут быть исследованы времена первого прохождения. Используемые методы применимы в широком круге задач.

Теорема 1. *Пусть* $P(t)$ *— семейство стохастических матриц, для которого*

$$P(s+t) = P(s)P(t) \tag{9.1}$$

и $P(t) \to I$ *при* $t \to 0$. *Если никакой из элементов* P_{ik} *не равен нулю тождественно*[1]), *то при* $t \to \infty$

$$P_{ik}(t) \to u_k, \tag{9.2}$$

где или $u_k = 0$ *при всех* k *или*

$$u_k > 0, \quad \sum_k u_k = 1 \tag{9.3}$$

и

$$\sum_j u_j P_{jk}(t) = u_k. \tag{9.4}$$

---

[1]) Это условие введено только во избежание тривиальных осложнений; этого же можно достигнуть, ограничиваясь надлежаще выбранными множествами состояний.

Нетрудно видеть, что наши условия влекут строгую положительность $P_{ik}(t)$ при всех t.

Вторая возможность осуществляется, если найдется вероятностный вектор $(u_1, u_2, \ldots)$, *удовлетворяющий при некотором* $t>0$ *соотношению* (9.4). *В этом случае* (9.4) *верно при всех* $t>0$ *и вероятностный вектор* u *определяется единственным способом.*

(Как уже объяснялось в **1**, гл. XVII, 6, важная особенность состоит в том, что пределы не зависят от i. Иначе говоря, влияние начальных условий асимптотически пренебрежимо.)

Доказательство. Фиксируем $\delta>0$ и рассмотрим дискретную цепь Маркова с матрицей вероятностей перехода $P(\delta)$. Матрица перехода за n шагов будет $P^n(\delta)=P(n\delta)$. Если при некотором n все элементы $P_{ik}(n\delta)$ положительны, то цепь неприводима и непериодична, и по эргодической теореме **1**, гл. XV, 6, наши утверждения верны для значений t, пробегающих подпоследовательность δ, 2δ, 3δ, ... Так как любые два рациональных числа имеют бесчисленное мпожество общих кратных, то предел при $n\to\infty$ $P_{ik}(n\delta)$ будет одним и тем же для всех рациональных δ. Для завершения доказательства достаточно проверить, что $P_{ik}(t)$ есть равномерно непрерывная функция t, положительная для больших t. По (9.1)

$$P_{ii}(s)P_{ik}(t) \leqslant P_{ik}(s+t) \leqslant P_{ik}(t)+[1-P_{ii}(s)] \tag{9.5}$$

[первое неравенство тривиально, а второе вытекает из того факта, что члены $P_{ij}(s)$ с $j\neq i$ в сумме дают $1-P_{ii}(s)$]. Для достаточно малых s имеем $1-\varepsilon\leqslant P_{ii}(s)\leqslant 1$, так что из (9.5) следует равномерная непрерывность P_{ik}. Из (9.5) видно также, что если $P_{ik}(t)>0$, то $P_{ik}(t+s)>0$ в некотором интервале значений s фиксированной длины. Поэтому или P_{ik} тождественно равно нулю или начиная с некоторого момента становится положительным. ▶

Мы применим теперь этот результат к минимальному решению § 7, предполагая, что оно является строго стохастическим (и стало быть, *единственным*). В матричных обозначениях (9.4) означает $uP(t)=u$, или, в терминах преобразований Лапласа,

$$u\lambda\Pi(\lambda)=u. \tag{9.6}$$

Пусть вектор u удовлетворяет (9.6) при некотором частном значении $\lambda>0$. Тогда резольвентное уравнение (7.15) показывает справедливость (9.6) при всех $\lambda>0$, а вместе с тем и справедливость (9.4) при всех $t>0$. Подставляя (9.6) в прямое уравнение (7.6), получаем

$$u\mathbf{cp}=u\mathbf{c}, \tag{9.7}$$

причем компоненты $u_k c_k$ конечны (хотя возможно не ограничены). С другой стороны, если u — вероятностный вектор, удовлетворяющий (9.7), то из (9.12) по индукции следует, что $u\lambda\Pi^{(n)}(\lambda) \leqslant u$ при всех n и потому $u\lambda\Pi(\lambda) \leqslant u$. Так как матрица $\lambda\Pi(\lambda)$ строго стохастическая, то суммы компонент в правой и левой частях должны быть равны. Поэтому верно (9.6). Нами доказана

Теорема 2. *Если минимальное решение является строго стохастическим* (*и потому единственным*), *то соотношения* (9.2) *выполняются с* $u_k>0$ *тогда и только тогда, когда найдется вероятностный вектор* u, *удовлетворяющий* (9.7).

Отсюда вытекает, в частности, что решение u уравнения (9.7) единственно.

Вероятностная интерпретация. Рассмотрим для определенности простейший случай, когда дискретная цепь с вероятностями перехода p_{ij} эргодична. Иными словами, мы предполагаем, что существует строго положительный вероятностный вектор α $(\alpha_1, \alpha_2, \ldots)$, такой, что $\alpha\mathbf{p}=\alpha$ и $p_{ik}^{(n)} \to \alpha_k$ при $n \to \infty$. Тогда ясно, что если $\sigma = \sum \alpha_k c_k^{-1} < \infty$, *то вероятностный вектор с компонентами* $u_k = \alpha_k c_k^{-1}/\sigma$ *удовлетворяет* (9.7). В то же время при $\sigma=\infty$ решение не существует.

Интуитивно ясно, что *переходы* в нашем процессе такие же, как и в дискретной цепи Маркова с матрицей $\mathbf{p}$, однако моменты их осуществления иные. Рассмотрим, например, какое-либо состояние. Пометим его индексом 0. Последовательные *времена пребывания* в 0 чередуются с *временами отсутствия*, в продолжение которых система находится в состояниях $j>0$. Число переходов в состояние j регулируется матрицей $\mathbf{p}$, а продолжительность пребывания в нем зависит от c_j. В дискретной цепи Маркова при большом числе переходов частоты попадания в j и 0 относятся как $\alpha_j : \alpha_0$. Следовательно, α_j/α_0 равно числу попаданий в j за время отсутствия в 0. Каждое пребывание в j продолжается в среднем время $1/c_j$. Поэтому за большой промежуток времени времена пребывания в состояниях j и 0 (а следовательно и вероятности этих состояний) относятся как $\alpha_j c_j^{-1} : \alpha_0 c_0^{-1}$ или $u_j : u_0$.

Эти рассуждения могут быть сделаны строгими, даже в случае когда $P_{ij}(t) \to 0$. В соответствии с теорией Дермана, отмеченной в **1**, гл. XV, 11, если $\mathbf{p}$ порождает неприводимую и возвратную цепь, то существует вектор α, такой что $\alpha\mathbf{p}=\alpha$. При этом α определяется однозначно с точностью до постоянного множителя; $\alpha_k \geqslant 0$, но ряд $\Sigma\alpha_k$ может расходиться. Даже в этом случае отношения $\alpha_j : \alpha_0$ допускают данную выше частотную интерпретацию, так что рассуждения имеют общий характер. *Если* $\Sigma\alpha_n c_k^{-1}<\infty$, *то* (9.2)—(9.4) *выпол-*

няются с u_k, пропорциональными $\alpha_k c_k^{-1}$. В противном случае $P(t) \to 0$ при $t \to \infty$. Интересно отметить, что пределы u_k могут быть положительны, даже если все состояния дискретной цепи нулевые.

Существование пределов $P_{ik}(\infty)$ можно установить также, привлекая аргументы теории восстановления, тесно связанные с поведением времен возвращения. Покажем, как можно вычислить распределение времен возвращения и первого прохождения. Припишем состояниям номера 0, 1, 2, ... и выделим состояние 0. Рассмотрим новый процесс, который совпадает с нервоначальным до момента первого достижения 0, а затем остающийся в нуле навсегда. Другими словами, новый процесс получается из старого постановкой в нуле поглощающего экрана. Обозначим переходные вероятности видоизмененного процесса через ${}^0P_{ik}(t)$. Тогда ${}^0P_{00}(t) = 1$. В терминах первоначального процесса ${}^0P_{i0}(t)$ равно *вероятности первого перехода из $i \neq 0$ в 0 до момента t*, а ${}^0P_{ik}(t)$ равно вероятности перехода из i в k без промежуточного захода в 0. Из вероятностных соображений ясно, что матрица ${}^0P(t)$ должна удовлетворять тем же самым обратному и прямому уравнениям, что и $P(t)$, с той разницей, что c_0 заменяется на 0. Теперь мы пойдем в обратном направлении: *мы изменим обратное и прямое уравнения, заменяя c_0 на 0, и покажем, что единственное решение этих уравнений имеет требуемые свойства.*

Пусть ξ — вектор, совпадающий с нулевым столбцом $\Pi(\lambda)$. Обратное уравнение показывает, что вектор

$$(\lambda + \mathbf{c} - \mathbf{cp})\,\xi = \eta \qquad (9.8)$$

имеет компоненты 1, 0, 0 ... . Теперь обратное уравнение для ${}^0\Pi(\lambda)$ получается заменой c_0 на 0, так что если ξ обозначает нулевой столбец ${}^0\Pi(\lambda)$, то вектор (9.8) имеет компоненты $\eta_1 = \eta_2 = \ldots = 0$, но $\eta_0 = \rho \neq 0$. Отсюда следует, что вектор с компонентами $\xi_k = \Pi_{k0}(\lambda) - \rho\,{}^0\Pi_{k0}(\lambda)$ удовлетворяет (9.8) с $\eta = 0$. Так как матрица $\lambda\Pi(\lambda)$ строго стохастическая, то $\xi_k = 0$ при всех k (теорема 7.3). В силу ${}^0\Pi_{00}(\lambda) = 1/\lambda$, мы имеем при $k \geqslant 0$

$$\Pi_{k0}(\lambda) = \lambda\,{}^0\Pi_{k0}(\lambda)\,\Pi_{00}(\lambda). \qquad (9.9)$$

Учитывая первое уравнение (9.8), мы видим, что

$$\Pi_{00}(\lambda) = \frac{1}{\lambda + c_0} + \frac{c_0}{\lambda + c_0} \sum_j p_{0j} \lambda\,{}^0\Pi_{j0}(\lambda)\,\Pi_{00}(\lambda). \qquad (9.10)$$

Уравнения (9.9) и (9.10) суть *уравнения восстановления* с очевидным вероятностным содержанием. В самом деле, пусть процесс начинается в точке $k > 0$. Тогда ${}^0\Pi_{k0}$ есть обычное преобразование Лапласа для вероятности ${}^0P_{k0}(t)$ того что первое

попадание в 0 произойдет до момента t. Следовательно, $\lambda^0\Pi_{k0}(\lambda)$ представляет собой преобразование Лапласа распределения F_k *момента первого достижения нуля*. Таким образом, (9.9) утверждает, что $P_{k0}(t)$ есть свертка F_k и P_{00}. Событие $\mathbf{X}(t)=0$ осуществляется тогда и только тогда, когда первое попадание в 0 происходит в некоторый момент $x<t$ и когда $t-x$ единиц времени спустя система снова находится в 0.

Аналогично $\sum p_{0j}\lambda^0\Pi_{j0}$ задает распределение F_0 времени, проведенного вне 0, т. е. времени между двумя последовательными пребываниями в состоянии 0. Следовательно, множитель при $\Pi_{00}(\lambda)$ в правой части (9.10) соответствует *времени ожидания первого возвращения в нуль*, если система первоначально в нем находится. (Это есть полный период=время пребывания+время отсутствия.) Уравнение восстановления (9.10) выражает $P_{00}(t)$ как сумму вероятности того, что время пребывания в нуле превосходит t, и вероятности события $\mathbf{X}(t)=0$ после первого возвращения в момент $x<t$. Если состояние 0 возвратно, то из (9.10) и теоремы восстановления следует, что

$$P_{00}(\infty)=\frac{1}{1+c_0\mu}, \tag{9.11}$$

где μ — математическое ожидание времени отсутствия в 0 и $c_0^{-1}+\mu$ — математическое ожидание длины полного цикла.

§ 10. Задачи

1. Положим в уравнении восстановления (1.3)

$F'(t)=g(t)=e^{-t}t^{p-1}/\Gamma(p)$. Тогда

$$\psi(\lambda)=\frac{(\lambda+1)^p}{(\lambda+1)^p-1}. \tag{10.1}$$

Разложением на простейшие дроби покажите, что при целых [1]) p

$$v(t)=\frac{1}{p}\sum_{k=0}^{p-1}a_k e^{-(1-a_k)t}, \tag{10.2}$$

где $a_k=e^{-i2\pi k/p}$ и $i^2=-1$.

2. *Потерянные вызовы.* На линию поступает пуассоновский поток вызовов. Время обслуживания имеет преобразование Лапласа φ. Дано, что линия свободна в момент 0. Обозначим через $U(t)$ вероятность того, что все прибывающие до момента t вызовы застают линию свободной. Покажите, что

[1]) Корни знаменателя будут одни и те же при $p=n$ и $p=1/n$, но решения совсем различны. Это показывает, что популярное «разложение по корням знаменателя» требует осторожности при иррациональной функции ψ.

обычное преобразование Лапласа для U равно

$$\omega(\lambda) = \frac{1}{\lambda+\alpha} + \frac{\alpha}{\lambda+\alpha} \cdot \frac{1}{\lambda+\alpha-\alpha\varphi(\lambda+\alpha)}.$$

Среднее время ожидания первого потерянного вызова равно

$$\frac{1}{\alpha} \frac{1}{\alpha[1-\varphi(\alpha)]}.$$

3. Если F имеет математическое ожидание μ и дисперсию σ^2 и если $c\mu<1$, то решение уравнения (4.1) для периода занятости имеет дисперсию $(\sigma^2+c\mu^3)/(1-c\mu)$.

4. Если F — показательное распределение, то распределение периода занятости, удовлетворяющее (4.1), задается функцией Бесселя.

5. В примере (4.6) производящая функция общего числа задержанных автомобилей равна $e^{c\delta[\psi(s)-1]}$, где

$$\psi(s) = s\varphi(c - c\psi(s)). \tag{10.3}$$

6. Покажите, что если φ есть преобразование Лапласа собственного распределения вероятностей, то решение уравнения (10.3) есть производящая функция некоторого (возможно, несобственного) распределения. Последнее будет собственным тогда и только тогда, когда F имеет математическое ожидание $\mu \leqslant 1/c$.

7. Исходя из (5.7), покажите, что функция Грина общего диффузионного уравнения (5.19) в любом интервале необходимо имеет вид

$$K_\lambda(x, y) = \begin{cases} \dfrac{\xi_\lambda(x)\,\eta_\lambda(y)}{W(y)} & \text{при} \quad x \leqslant y, \\ \dfrac{\eta_\lambda(x)\,\xi_\lambda(y)}{W(y)} & \text{при} \quad x \geqslant y, \end{cases} \tag{10.4}$$

где ξ_λ и η_λ суть решения

$$\lambda\varphi - \frac{1}{2}\, a\varphi'' - b\varphi' = 0, \tag{*}$$

ограниченные соответственно на левом и правом концах. Если (*) не имеет ограниченных решений, то ξ_λ и η_λ определяются с точностью до мультипликативной константы, которую можно включить в W (в противном случае нужно наложить подходящие граничные условия).

Покажите, что функция ω_λ, определенная по (10.4) и (5.5), удовлетворяет дифференциальному уравнению (5.19) в том и только том случае, когда W есть вронскиан

$$W(y) = \xi'_\lambda(y)\,\eta_\lambda(y) - \xi_\lambda(y)\,\eta'_\lambda(y). \tag{10.5}$$

Решения ξ_λ и η_λ обязательно монотонны и потому $W(y) \neq 0$.

8. *Продолжение.* При $x<y$ момент первого перехода из x в y имеет преобразование Лапласа $\xi_\lambda(x)/\xi_\lambda(y)$. При $x>y$ оно равно $\eta_\lambda(x)/\eta_\lambda(y)$.

9. Покажите, что метод, использованный в § 5 для диффузионных уравнений, применим и к общему *процессу размножения и гибели.*

10. Разберите пример (6. а) для случая $a>1$ линий (явное вычисление a констант затруднительно и не рекомендуется).

11. В условиях примера (6. б) выведите непосредственно из дифференциальных уравнений, что период занятости имеет математическое ожидание $\frac{1}{c(q-p)}$ и дисперсию $\frac{1}{c^2(q-p)^3}$.

12. *Полумарковские процессы.* Полумарковский процесс с состояниями 1, 2, ... отличается от марковского тем, что времена пребывания могут зависеть и от конечного состояния: пусть система *входит* в состояние i в момент τ; тогда условная вероятность того, что пребывание в i оборвется до момента $\tau+t$ скачком в k, равна $F_{ik}(t)$. Тогда $\sum_k F_{ik}(t)$ дает распределение времени пребывания, а $p_{ik}=F_{ik}(\infty)$ равно вероятности скачка в состояние k. Обозначим через $P_{ik}(t)$ вероятность k в момент $t+\tau$ при условии, что *вход* в i произошел в момент τ. Выведите аналог обратного уравнения Колмогорова. В понятных обозначениях для преобразований Лапласа имеем

$$\Pi(\lambda) = \gamma(\lambda) + \Phi(\lambda)\,\Pi(\lambda),$$

где $\gamma(\lambda)$ — диагональная матрица с элементами $\left[1-\sum_k \varphi_{ik}(\lambda)\right]/\lambda$. При $F_{ik}(t)=p_{ik}(1-e^{-c_i t})$ это сводится к обратному уравнению (7.5). Построение минимального решения § 7 осуществимо [1]).

---

[1]) Детали см. *Proc. Nat. Acad. Sci.*, **51** (1964); 653—659. Полумарковские процессы были введены П. Леви и В. Смитом и были исследованы, в частности, Пайком.

ГЛАВА XV

ХАРАКТЕРИСТИЧЕСКИЕ ФУНКЦИИ

Эта глава содержит основы теории характеристических функций и совсем не зависит от гл. VI, VII, IX—XIV. Более глубокое изложение анализа Фурье отложено до гл. XIX.

§ 1. Определение. Основные свойства

Производящей функцией неотрицательной целочисленной случайной величины $\mathbf{X}$ называют функцию $\mathbf{E}(s^{\mathbf{X}})$, определенную для $0 \leqslant s \leqslant 1$, т. е. математическое ожидание величины $s^{\mathbf{X}}$. Как было показано в гл. XIII, замена переменной $s = e^{-\lambda}$ дает возможность изучать любые неотрицательные случайные величины. Пользу от таких характеристик мы получаем в первую очередь из-за мультипликативного свойства $s^{x+y} = s^x \cdot s^y$ и $e^{-\lambda(x+y)} = e^{-\lambda x} e^{-\lambda y}$. Этим свойством обладает и показательная функция от чисто мнимого аргумента, т. е. функция, определенная для действительных x равенством

$$e^{i\zeta x} = \cos \zeta x + i \sin \zeta x, \tag{1.1}$$

где ζ — действительное число и $i^2 = -1$. Так как эта функция ограничена, то ее математическое ожидание всегда существует. Использование $\mathbf{E}(e^{i\zeta \mathbf{X}})$ в качестве замены для производящей функции дает нам мощный универсальный метод, но он достигается ценой введения комплексозначных функций и случайных величин. Заметим, однако, что аргументы остаются точками числовой прямой (или будут позже точками пространства $\mathscr{R}^r$)

Под комплекснозначной функцией $w = u + iv$ понимается пара *действительных* функций u и v, определенных для *действительных* x. Математическое ожидание $\mathbf{E}(w)$ — это не более чем сокращенное обозначение для $\mathbf{E}(u) + i\mathbf{E}(v)$. Как обычно, через $\overline{w} = u - iv$ мы обозначаем функцию, комплексно сопряженную к w, а через $|w|$ абсолютную величину функции w (т. е. $|w|^2 = w\overline{w} = u^2 + v^2$). Основные свойства математических ожиданий сохраняются. Объяснения требует только *теорема о среднем значении: если* $|w| \leqslant a$, *то* $|\mathbf{E}(w)| \leqslant a$. В самом деле, по

неравенству Шварца

$$|\mathbf{E}(w)|^2 = (\mathbf{E}(u))^2 + (\mathbf{E}(v))^2 \leqslant \mathbf{E}(u^2) + \mathbf{E}(v^2) = \mathbf{E}(|w|^2) \leqslant a^2. \quad (1.2)$$

Две комплекснозначные случайные величины $\mathbf{W}_j = \mathbf{U}_j + i\mathbf{V}_j$ называются *независимыми*, если независимы пары $(\mathbf{U}_1, \mathbf{V}_1)$ и $(\mathbf{U}_2, \mathbf{V}_2)$. Обычным разложением на действительную и мнимую части можно показать, что мультипликативное свойство $\mathbf{E}(\mathbf{W}_1\mathbf{W}_2) = \mathbf{E}(\mathbf{W}_1) \cdot \mathbf{E}(\mathbf{W}_2)$ выполняется (эта формула иллюстрирует преимущества записи в комплексной форме). После этих подготовительных замечаний дадим определение аналога производящей функции.

Определение. Пусть $\mathbf{X}$ — *случайная величина с распределением вероятностей* F. *Характеристической функцией распределения* F (*или случайной величины* $\mathbf{X}$) *называется функция* φ, *определенная для действительных* ζ *формулой*

$$\varphi(\zeta) = \int_{-\infty}^{+\infty} e^{i\zeta x} F\{dx\} = u(\zeta) + iv(\zeta), \quad (1.3)$$

где

$$u(\zeta) = \int_{-\infty}^{+\infty} \cos \zeta x \cdot F\{dx\}, \quad v(\zeta) = \int_{-\infty}^{+\infty} \sin \zeta x \cdot F\{dx\}. \quad (1.4)$$

Для распределения F с плотностью f имеем, конечно,

$$\varphi(\zeta) = \int_{-\infty}^{+\infty} e^{i\zeta x} f(x)\, dx. \quad (1.5)$$

Замечание по терминологии. В анализе Фурье функцию φ принято называть *преобразованием Фурье — Стильтьеса распределения* F. Такие преобразования определены для любой конечной меры, и термин «характеристическая функция» подчеркивает, что значение мер на всем пространстве равно единице (для других мер характеристические функции не определены). Но интегралы типа (1.5) встречаются во многих ситуациях, и мы будем говорить, что (1.5) определяет *обычное преобразование Фурье* функции f. Характеристическая функция распределения F является обычным преобразованием Фурье плотности f (когда последняя существует), но термин «преобразование Фурье» применяется также и к другим функциям [1]). ▶

[1]) Более общим образом можно определить «преобразование Фурье функции f по отношению к данной мере»; тогда (1.3) будет преобразованием функции, тождественно равной единице, по отношению к F.

Для удобства ссылок мы перечислим некоторые основные свойства характеристических функций.

Лемма 1. *Пусть* $\varphi = u + iv$ — *характеристическая функция случайной величины* $\mathbf{X}$ *с распределением* F. *Тогда*

а) φ *непрерывна*;

б) $\varphi(0) = 1$ *и* $|\varphi(\zeta)| \leqslant 1$ *при всех* ζ;

в) $a\mathbf{X} + b$ *имеет характеристическую функцию*

$$\mathbf{E}(e^{i\zeta(a\mathbf{X}+b)}) = e^{ib\zeta}\varphi(a\zeta) \tag{1.6}$$

(*в частности,* $\bar{\varphi} = u - iv$ *служит характеристической функцией величины* $-\mathbf{X}$);

г) u — *четная, а* v — *нечетная функции; характеристическая функция действительна тогда и только тогда, когда распределение* F *симметрично;*

д) *для всех* ζ

$$0 \leqslant 1 - u(2\zeta) \leqslant 4(1 - u(\zeta)). \tag{1.7}$$

(Другие варианты леммы см. в задачах 1—3.)

Доказательство. а) Заметим, что

$$|e^{i\zeta(x+h)} - e^{i\zeta x}| = |e^{i\zeta h} - 1|. \tag{1.8}$$

Правая часть не зависит от x и может быть сделана сколь угодно малой при достаточно малом h. Таким образом, φ на самом деле даже *равномерно* непрерывна. Свойство (б) с очевидностью вытекает из теоремы о среднем значении, а свойство (в) не требует пояснений. Для доказательства пункта (г) мы привлечем тот факт, что различным распределениям соответствуют различные характеристические функции. В самом деле, φ действительна тогда и только тогда, когда $\varphi = \bar{\varphi}$, т. е. тогда и только тогда, когда $\mathbf{X}$ и $-\mathbf{X}$ имеют одинаковые характеристические функции. Но это равносильно тому, что $\mathbf{X}$ и $-\mathbf{X}$ имеют одно и то же распределение, т. е. тому, что F симметрично. Наконец, чтобы доказать (д), рассмотрим элементарное тригонометрическое соотношение

$$1 - \cos 2\zeta x = 2(1 - \cos^2 \zeta x) \leqslant 4(1 - \cos \zeta x), \tag{1.9}$$

справедливое, поскольку $0 \leqslant 1 + \cos \zeta x \leqslant 2$. Переходя к математическим ожиданиям, получаем (1.7). ▷

Рассмотрим теперь две случайные величины $\mathbf{X}_1$ и $\mathbf{X}_2$ с распределениями F_1 и F_2 и характеристическими функциями φ_1 и φ_2 соответственно. Если $\mathbf{X}_1$ и $\mathbf{X}_2$ независимы, то мультипликативное свойство показательной функции приводит к равенству

$$\mathbf{E}(e^{i\zeta(\mathbf{X}_1+\mathbf{X}_2)}) = \mathbf{E}(e^{i\zeta\mathbf{X}_1})\,\mathbf{E}(e^{i\zeta\mathbf{X}_2}). \tag{1.10}$$

Этот простой результат часто употребляется, поэтому мы отметим его в форме леммы.

Лемма 2. *Свертка* $F_1 * F_2$ *имеет характеристическую функцию* $\varphi_1\varphi_2$.

Другими словами, сумме $\mathbf{X}_1+\mathbf{X}_2$ *двух независимых случайных величин соответствует произведение* $\varphi_1\varphi_2$ *их характеристических функций* [1]).

Если $\mathbf{X}_2$ имеет такое же распределение, как $-\mathbf{X}_1$, то сумма $\mathbf{X}_1-\mathbf{X}_2$ оказывается симметризованной величиной. Поэтому справедливо

Следствие. *Функция* $|\varphi|^2$ *является характеристической функцией симметризованного распределения* ${}^0F$.

Следующая лемма дает описание *арифметических* распределений.

Лемма 3. *Если* $\lambda\neq 0$, *то три следующих утверждения эквивалентны*:

а) $\varphi(\lambda)=1$;

б) φ *имеет период* λ, *т. е.* $\varphi(\zeta+n\lambda)=\varphi(\zeta)$ *для всех* ζ *и* n;

в) *все точки роста распределения* F *находятся среди точек* $0, \pm h, \pm 2h, \ldots$, где $h=2\pi/\lambda$.

Доказательство. Так как $1-\cos\lambda x\geqslant 0$ для всех x, то математическое ожидание этой функции может обращаться в нуль только тогда, когда $1-\cos\lambda x=0$ в каждой точке роста F и, следовательно, (а) влечет (в). Обратно, если верно (в) и F приписывает точке nh вес p_n, то $\varphi(\zeta)=\sum p_n e^{inh\zeta}$. Эта фупкция имеет, очевидно, период $\lambda=2\pi/h$, так что (б) и (а) выполняются. Наконец из (б) тривиально следует (а). ▶

Доказанная лемма покрывает и крайний случай, когда распределепие F полностью сосредоточено в нуле. Тогда $\varphi(\zeta)=1$ для всех ζ, так что каждое число будет периодом для φ. В общем случае из того, что λ — период для φ, вытекает, что периодами будут все кратные λ: $\pm\lambda, \pm 2\lambda, \ldots,$. Но для отличной от константы периодической функции φ существует наименьший положительный период. Его называют *истинным периодом*. Аналогично для арифметического распределения F существует *наибольшее* положительное h, для которого выпол-

---

[1]) Обратное неверно: в гл II, (4, д) и в задаче 1 гл. III, 9 показано, что в некоторых исключительных случаях сумма двух *зависимых* величин может иметь распределение $F_1 * F_2$ и, следовательно, характеристическую функцию $\varphi_1\varphi_2$.

няется свойство (в). Такое h называют *шагом* распределения F. Из леммы 3 следует, что шаг h и (истинный) период λ связаны равенством $\lambda h = 2\pi$. Итак, за исключением случаев, когда $\varphi(\zeta) \neq 1$ для всех $\zeta \neq 0$ и когда $\varphi(\zeta) = 1$ тождественно, существует наименьшее $\lambda > 0$, такое, что $\varphi(\lambda) = 1$, но $\varphi(\zeta) \neq 1$ для $0 < \zeta < \lambda$.

Все сказанное можно сформулировать в более общем виде. Вместо $\varphi(\lambda) = 1$ предположим, что только $|\varphi(\lambda)| = 1$. Тогда существует такое вещественное число b, что $\varphi(\lambda) = e^{ib\lambda}$, и мы можем применить предыдущий результат к случайной величине $\mathbf{X} - b$ с характеристической функцией $\varphi(\zeta) e^{-ib\zeta}$, равной 1 при $\zeta = \lambda$. Каждый период последней функции автоматически будет периодом для $|\varphi|$. Таким образом, доказана

Лемма 4. *Существуют только следующие три возможности:*

а) $|\varphi(\zeta)| < 1$ *для всех* $\zeta \neq 0$;

б) $|\varphi(\lambda)| = 1$ *и* $|\varphi(\zeta)| < 1$ *для* $1 < \zeta < \lambda$. *В этом случае* $|\varphi|$ *имеет период* λ *и существует такое вещественное число* b, *что распределение* $F(x+b)$ *является арифметическим с шагом* $h = 2\pi/\lambda$;

в) $|\varphi(\zeta)| = 1$ *для всех* ζ. *В этом случае* $\varphi(\zeta) = e^{ib\zeta}$ *и* F *сосредоточено в точке* b.

§ 2. Специальные плотности. Смеси

Для облегчения ссылок мы даем таблицу характеристических функций наиболее употребительных плотностей и указываем метод, каким получена каждая функция.

Замечания. 1) *Нормальная плотность.* Для того, кто не боится интегрирования в комплексной плоскости, результат есть очевидное следствие подстановки $y = x - i\zeta$. Чтобы доказать приведенную формулу, не выходя из области вещественных чисел, можно использовать дифференцирование и интегрирование по частям, что даст $\varphi'(\zeta) = -\zeta\varphi(\zeta)$. Так как $\varphi(0) = 1$, то $\ln \varphi(\zeta) = -{}^1/_2\zeta^2$, что и утверждается.

2) — 3) *Равномерные плотности.* Вычисления здесь очевидны. Распределения (2) и (3) отличаются одно от другого масштабными параметрами. Связь между их характеристическими функциями можно получить из формулы (1.6).

4) *Треугольная плотность.* Легко провести, используя интегрирование по частям, прямое вычисление. Другой способ: заметим, что треугольная плотность является сверткой плотности,

Таблица

№	Название	Формула	Область определения	Характеристическая функция
1	Нормальная плотность	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$	$-\infty < x < \infty$	$e^{-\frac{1}{2}\zeta^2}$
2	Равномерная плотность	$\frac{1}{a}$	$0 < x < a$	$\frac{e^{ia\zeta}-1}{ia\zeta}$
3	Равномерная плотность	$\frac{1}{2a}$	$\lvert x\rvert < a$	$\frac{\sin a\zeta}{a\zeta}$
4	Треугольная плотность	$\frac{1}{a}\left(1-\frac{\lvert x\rvert}{a}\right)$	$\lvert x\rvert < a$	$2\frac{1-\cos a\zeta}{a^2\zeta^2}$
5	———	$\frac{1}{\pi}\frac{1-\cos ax}{ax^2}$	$-\infty < x < \infty$	$\begin{cases} 1-\frac{\lvert\zeta\rvert}{a} \text{ при } \lvert\zeta\rvert \leqslant a \\ 0 \text{ при } \lvert\zeta\rvert > a \end{cases}$
6	Гамма-плотности	$\frac{1}{\Gamma(t)} x^{t-1}e^{-x}$	$x > 0,\ t > 0$	$\frac{1}{(1-i\zeta)^t}$
7	Двустороннее показательное распределение	$\frac{1}{2} e^{-\lvert x\rvert}$	$-\infty < x < \infty$	$\frac{1}{1+\zeta^2}$
8	Распределение Коши	$\frac{1}{\pi}\frac{t}{t^2+x^2}$	$-\infty < x < \infty$ $t > 0$	$e^{-t\lvert\zeta\rvert}$
9	Бесселева плотность	$e^{-x}\frac{t}{x} I_t(x)$	$x > 0,\ t > 0$	$[1-i\zeta-\sqrt{(1-i\zeta)^2-1}]^t$
10	Гиперболический косинус	$\frac{1}{\pi \operatorname{ch} x}$	$-\infty < x < \infty$	$\frac{1}{\operatorname{ch}(\pi\zeta/2)}$

равномерной на $-1/2a<x<1/2a$, с собой. В силу (3) ее характеристическая функция равна $\left(\frac{2}{a\zeta}\cdot\sin\frac{a\zeta}{2}\right)^2$.

5) Эта плотность получается применением формулы обращения (3.5) к треугольной плотности (4). См. также задачу 4. Характеристическая функция этой плотности весьма важна, так как многие доказательства в анализе Фурье опираются на использование характеристических функций, равных нулю вне некоторого конечного интервала.

6) *Гамма-плотности.* Можно использовать подстановку $y=x(1-i\zeta)$ или же разложить $e^{-i\zeta x}$ в степенной ряд и исполь-

зовать тождество

$$(-1)^n \binom{-a}{n} n! = \frac{\Gamma(a+n)}{\Gamma(a)}.$$

В специальном случае $t=1$ (показательное распределение) вычисления можно провести без выхода из действительной области, повторно применяя интегрирование по частям. По индукции то же самое можно проделать при всех целых t.

7) *Двустороннее показательное распределение* получается симметризацией показательного распределения, так что формула для соответствующей характеристической функции получается из (6) при $t=1$. Это легко проверить и непосредственно, с помощью повторного применения интегрирования по частям.

8) *Распределение Коши.* Снова формула вытекает из предшествующей и формулы обращения (3.5). Прямая проверка представляет собой стандартное упражнение на отыскание вычетов [1]).

9) *Бесселева плотность.* Это — преобразование Фурье, аналогичное преобразованию Лапласа, полученному в гл. XIII, (3, г).

10) *Гиперболический косинус.* Соответствующая функция распределения равна $F(x)=1-(2/\pi)\operatorname{arctg} e^{-x}$. Формула (10) не очень важна, но она любопытна тем, что в ней появляются *«взаимные пары»*: плотность и характеристическая функция получаются одна из другой линейным преобразованием аргумента и линейным преобразованием самой функции (нормальная плотность служит первым примером такой связи). Для вычисления характеристической функции можно разложить плотность в ряд

$$\frac{1}{2\pi} \sum (-1)^k e^{-(2k+1)|x|}.$$

Применяя результат (7) к отдельным членам ряда, получаем каноническое разложение характеристической функции на простейшие дроби [2]). ▶

Возвращаясь к общей теории, укажем метод построения новых характеристических функций, исходя из перечисленных. Принцип весьма прост и, как показывает пример (б), с его помощью можно избежать длинных вычислений.

---

[1]) См., например, Hille E., Analytic function theory, Boston, 1959, 1, стр. 247.

[2]) Ср. Маркушевич А. И. Теория аналитических функций, М., ГИТТЛ, 1950, гл. IV, 4. — *Прим. перев.*

Лемма. *Пусть* F_0, F_1, ... — *распределения вероятностей с характеристическими функциями* φ_0, φ_1, ... *Если* $p_k \geqslant 0$ *и* $\sum p_k = 1$, *то смесь*

$$U = \sum p_k F_k \tag{2.1}$$

есть распределение вероятностей с характеристической функцией

$$\omega = \sum p_k \varphi_k. \tag{2.2}$$

Примеры. а) *Суммы случайного числа слагаемых.* Пусть $\mathbf{X}_1$, $\mathbf{X}_2$, ... — независимые случайные величины с одним и тем же распределением F и характеристической функцией φ. Пусть $\mathbf{N}$ — целочисленная случайная величина с производящей функцией $P(s) = \sum p_k s^k$ и не зависящая от всех $\mathbf{X}_j$. Тогда случайная сумма $\mathbf{X}_1 + \ldots + \mathbf{X}_\mathbf{N}$ имеет распределение (2.1) с $F_k = F^{k*}$, и соответствующая характеристическая функция равна

$$\omega(\zeta) = P(\varphi(\zeta)). \tag{2.3}$$

Наибольшего внимания заслуживает частный случай — *обобщенное распределение Пуассона.* Здесь $p_k = e^{-t} t^k / k!$ и

$$\omega(\zeta) = e^{-t + t\varphi(\zeta)} \tag{2.4}$$

б) *Вогнутые полигоны.* Из табл. 1 (п. 5) мы видим, что функция

$$\varphi(\zeta) = \begin{cases} 1 - |\zeta| & \text{для } |\zeta| \leqslant 1, \\ 0 & \text{для } |\zeta| \geqslant 1 \end{cases} \tag{2.5}$$

характеристическая. При любых положительных a_1, ..., a_n смесь

$$\omega(\zeta) = p_1 \varphi\left(\frac{\zeta}{a_1}\right) + \ldots + p_n \varphi\left(\frac{\zeta}{a_n}\right) \tag{2.6}$$

является четной характеристической функцией, график которой на интервале $\overline{0, \infty}$ есть вогнутый полигон (см. рис. 1). Действительно, не ограничивая общности, можно предположить, что

$$a_1 < a_2 < \ldots < a_n.$$

На интервале $0 < \zeta < a_1$ график функции ω имеет вид отрезка прямой с угловым коэффициентом $-(p_1/a_1 + \ldots + p_n/a_n)$. На интервале между a_1 и a_2 в этом выражении нужно отбросить член p_1/a_1, и т. д. Наконец, между a_{n-1} и a_n график совпадает с отрезком прямой с угловым коэффициентом $-p_n/a_n$. Следовательно, на $\overline{0, \infty}$ график составлен из n конечных отрезков

с убывающим наклоном и интервала $\overline{a_n, \infty}$ оси ζ. Как легко видеть, каждый полигон с этими свойствами может быть графиком функции типа (2.6) (n «сторон» пересекают ось ω в точках p_n, p_n+p_{n-1}, ..., $p_n+\ldots+p_1=1$). Отсюда можно заключить, что *любая четная функция* $\omega \geqslant 0$ *с* $\omega(0)=1$, *график которой на* $\overline{0, \infty}$ *есть вогнутый полигон, является характеристической функцией.*

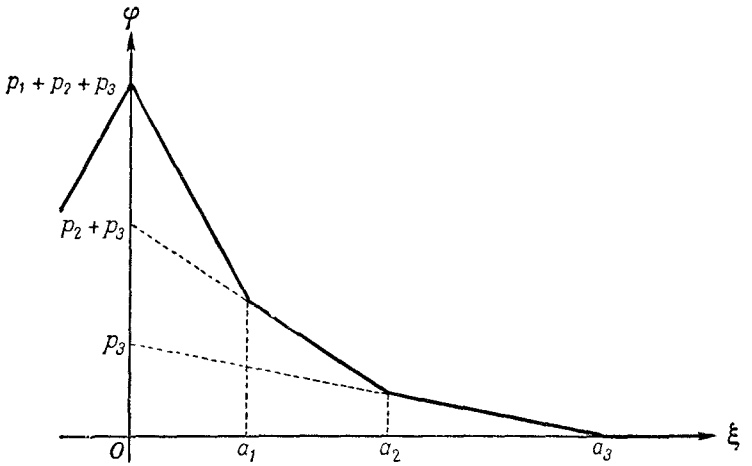


Рис. 1.

Простой переход к пределу приводит к известному критерию Пойа [пример (3, б)] и указывает его естественный источник. Но даже и наш частный критерий приводит к удивительным и интересным результатам. ▶

Неожиданные примеры. (i) *Две различные характеристические функции на конечном интервале* $\overline{-a, a}$ могут быть равны. Это очевидное следствие примера (б) показывает, что свойства регулярности характеристических функций существенно отличаются от свойств регулярности производящих функций и преобразований Лапласа. Мы еще вернемся к этому вопросу в гл. XIX, 5.

(ii) *Соотношение* $F*F_1=F*F_2$ *между тремя распределениями вероятностей не влечет* [1]) *равенства* $F_1=F_2$. Действитель-

[1]) Статистики и астрономы иногда задаются вопросом, содержит ли данное распределение нормальную компоненту. Этот вопрос осмыслен, так как характеристическая функция нормального распределения $\mathfrak{N}_a$ не имеет нулей и, следовательно, из $\mathfrak{N}_a * F_1 = \mathfrak{N}_a * F_2$ вытекает $\varphi_1=\varphi_2$, и по теореме единственности $F_1=F_2$.

но, выберем две различные характеристические функции φ_1 и φ_2, совпадающие при $|\zeta|<1$. Для функции φ, определенной по (2.5), имеем $\varphi\varphi_1=\varphi\varphi_2$.

(iii) Еще более удивительно то, что *можно привести пример двух вещественных характеристических функций* φ_1 *и* φ_2, *таких, что* $|\varphi_1|=|\varphi_2|$ *тождественно* (*следовательно,* $\varphi_1^2=\varphi_2^2$). В качестве φ_1 возьмем периодическую функцию периода 2, определенную при $|\zeta|\leqslant 1$ равенством $\varphi_1(\zeta)=1-|\zeta|$. Пусть $\varphi_2(\zeta)=2[\varphi_1(\zeta/2)-1/2]$. На рис. 2 график функции φ_1 изображен

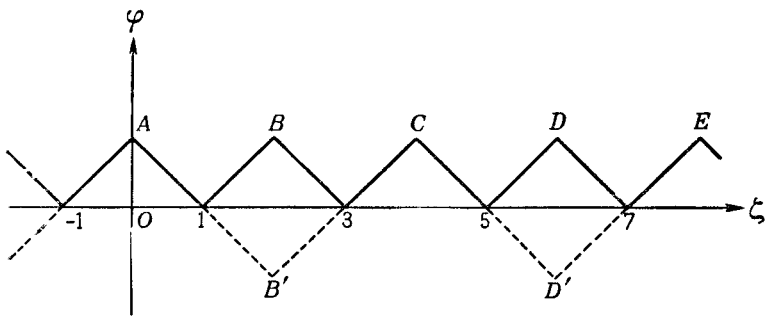


Рис. 2.

сплошной ломаной линией, а график функции φ_2 получается из него «отражением» от оси ζ каждого второго треугольника. Можно показать, что φ_1 — характеристическая функция арифметического распределения, приписывающего вес, равный 1/2, точке 0 [1]). Выбрасывая массу, сосредоточенную в нуле, и увеличивая вдвое массы каждой из оставшихся точек, мы получаем новое распределение с характеристической функцией $\varphi=2[\varphi_1-1/2]$. Так как $\varphi_2(\zeta)=\varphi(\zeta/2)$, то φ_2 также является характеристической функцией.

---

[1]) Прямая проверка требует утомительных вычислений, но, как будет показано в гл. XIX, 5, φ_1 — характеристическая функция просто потому, что она является периодическим продолжением характеристической функции (п. 5 в табл. 1). В силу четности функции φ_1 имеем.

$$\varphi_1(\zeta)=\sum_{-\infty}^{+\infty} p_k\cos k\pi\zeta.$$

Интегрируя по ζ в пределах $-1<\zeta<1$, получаем $p_0=1/2$. Другие атомы суть нечетные числа $\pm(2k+1)$ с весами $2/[(2k+1)\pi]^2$ [см. гл. XIX, (5.6)].

§ 3. Единственность. Формулы обращения

Пусть F и G — два распределения с характеристическими функциями φ и γ соответственно. Тогда

$$e^{-i\zeta t}\varphi(\zeta) = \int_{-\infty}^{+\infty} e^{i\zeta(x-t)} F\{dx\}. \tag{3.1}$$

Интегрируя по $G\{d\zeta\}$, имеем

$$\int_{-\infty}^{+\infty} e^{-i\zeta t}\varphi(\zeta)\, G\{d\zeta\} = \int_{-\infty}^{+\infty} \gamma(x-t) F\{dx\}. \tag{3.2}$$

Это соотношение известно под названием *равенства Парсеваля* (которое, впрочем, может быть записано во многих эквивалентных формах; мы вернемся к этому в гл. XIX).

Из равенства Парсеваля можно сделать удивительно много выводов. Здесь мы используем его для доказательства основной теоремы единственности (которая уже применялась в части (г) леммы 1.1).

Теорема 1. *Различным распределениям вероятностей соответствуют различные характеристические функции.*

Доказательство. Возьмем в качестве G нормальное распределение $\mathfrak{N}_a$ с плотностью $a\nu(ax)$. Его характеристическая функция равна $\gamma(\zeta) = \sqrt{2\pi}\,\nu(\zeta a)$, и поэтому (3.2) принимает вид

$$a\int_{-\infty}^{+\infty} e^{-i\zeta t}\varphi(\zeta)\,\nu(a\zeta)\,d\zeta = \sqrt{2\pi}\int_{-\infty}^{+\infty} \nu\left(\frac{x-t}{a}\right) F\{dx\} \tag{3.3}$$

или

$$\frac{a}{2\pi}\int_{-\infty}^{+\infty} e^{-i\zeta t}\varphi(\zeta)\, e^{-1/2a^2\zeta^2}\,d\zeta = \int_{-\infty}^{+\infty} \nu\left(\frac{x-t}{a}\right) F\{dx\}. \tag{3.4}$$

Обозначим последнее выражение через $f_a(t)$. Правая часть равенства (3.4) показывает, что f_a является плотностью свертки $F_a = \mathfrak{N}_a * F$ распределения F с нормальным распределением с дисперсией a^2. Следовательно, $F_a \to F$ при $a \to 0$. Левая часть равенства (3.4) выражает F_a через φ. Мы не только видим, что F однозначно определяется по φ, но и получили метод вычисления F. ▶

Перед тем как применить последнее замечание, докажем важное следствие теоремы единственности.

Теорема 2. (*Теорема непрерывности.*) *Последовательность* $\{F_n\}$ *распределений вероятностей сходится к распределению вероятностей* F *тогда и только тогда, когда последовательность* $\{\varphi_n\}$ *соответствующих характеристических функций сходится к непрерывной предельной функции* φ.

В таком случае φ *есть характеристическая функция распределения* F *и сходимость* $\varphi_n \to \varphi$ *равномерна на каждом конечном интервале.*

Доказательство. а) *Собственная* сходимость $F_n \to F$ влечет сходимость соответствующих математических ожиданий для любой ограниченной непрерывной функции u. Для $u(x) = e^{i\zeta x}$ (при фиксированном ζ) получаем $\varphi_n(\zeta) \to \varphi(\zeta)$, где φ — характеристическая функция распределения F. Равномерность этой сходимости на любом конечном интервале немедленно вытекает из следствия гл. VIII, 1 [см. пример гл. VIII, (1, ∂)].

б) Пусть $\varphi_n \to \varphi$. По теореме о выборе (гл. VIII, 6) существуют последовательность $\{n_k\}$ и (может быть несобственное) распределение F, такие, что $F_{n_k} \to F$. В применении к паре F_{n_k}, φ_{n_k} формула (3.3) дает

$$a \int_{-\infty}^{+\infty} e^{-i\zeta t}\varphi_{n_k}(\zeta)\, \mathfrak{n}(a\zeta)\, d\zeta = \sqrt{2\pi} \int_{-\infty}^{+\infty} \mathfrak{n}\left(\frac{x-t}{a}\right) F_{n_k}\{dx\}. \quad (3.3a)$$

Переходя к пределу при $k \to \infty$, мы получаем (3.3) (в левой части используется теорема об ограниченной сходимости, в правой — то, что $\mathfrak{n}$ равно нулю на бесконечности; см. гл. VIII, 1). Таким образом, (3.3) выполняется. Так как $\sqrt{2\pi}\,\mathfrak{n} \leqslant 1$, то правая часть не превосходит $F\{-\infty, \infty\}$. С другой стороны, при $a \to 0$ нормальное распределение сосредоточивается вблизи нуля и левая часть стремится к $\varphi(0) = 1$. Поэтому $F\{-\infty, \infty\} = 1$ и сходимость $F_{n_k} \to F$ — собственная. По первой части теоремы φ является характеристической функцией распределения F, и теорема единственности гарантирует, что все сходящиеся подпоследовательности $\{F_{n_k}\}$ имеют один и тот же предел F. Следовательно, $F_n \to F$, чем и заканчивается доказательство. ▶

Следствие. *Непрерывная функция, являющаяся пределом (в смысле точечной сходимости) последовательности характеристических функций, сама является характеристической.*

Примеры. а) Пусть G — непрерывное распределение с характеристической функцией γ. Тогда $|\gamma(\zeta)| < 1$ при $\zeta \neq 0$. Последовательность $\{G^{n*}\}$ *не* будет сходящейся в собственном смысле

слова, а предел $\gamma^n(\zeta)$ существует во всех точках. Таким образом, условие *непрерывности* предельной функции φ существенно как в приведенной теореме, так и в ее следствии (см. задачу 9).

б) *Критерий Пойа. Пусть ω — вещественная непрерывная четная функция, равная единице в нуле и равная нулю на бесконечности. Если ее график на 0, ∞ вогнут, то ω — характеристическая функция.* Действительно, мы видели в примере (2, б), что это утверждение верно для случая, когда график функции ω — вогнутый полигон. Так как полигоны, вписанные в вогнутую линию, сами вогнуты, то общее утверждение вытекает из приведенного выше следствия. Этот критерий (вместе со сложным доказательством) в свое время играл интересную роль. Пойа с его помощью доказал в 1920 г., что $e^{-|\zeta|^\alpha}$ при $0<\alpha\leqslant 1$ является характеристической функцией устойчивого распределения (говорят, что Коши был уверен в этом, но не имел доказательства). В действительности $e^{-|\zeta|^\alpha}$ есть характеристическая функция также и при $1<\alpha\leqslant 2$, но критерий уже неприменим. ▶

Мы отложим до гл. XIX систематическое применение метода, развитого при доказательстве теоремы 1. Мы здесь используем его для вывода важной теоремы, которая уже была использована при составлении таблицы в § 2 (п. 5 и 8). Условимся для краткости писать $\varphi\in L$, в том и только том случае, если функция $|\varphi|$ интегрируема на $-\infty$, ∞.

Теорема 3. (*Обращение интеграла Фурье.*) *Пусть φ — характеристическая функция распределения F и $\varphi\in L$. Тогда F имеет ограниченную непрерывную плотность f, задаваемую формулой*

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\zeta x}\varphi(\zeta)\,d\zeta. \tag{3.5}$$

Доказательство. Напомним, что обе части равенства (3.4) определяют плотность f_a такого распределения F_a, что $F_a\to F$ при $a\to 0$. Из левой части равенства (3.4) мы видим, что $f_a(t)\to f(t)$ ограничено (f — ограниченная непрерывная функция (3.5)). Поэтому для любого ограниченного интервала I

$$F_a\{I\}=\int_I f_a(t)\,dt\to\int_I f(x)\,dx. \tag{3.6}$$

Но если I — интервал непрерывности для F, то $F_a\{I\}$ стремится к $F\{I\}$, так что f действительно является плотностью распределения F.

Следствие. *Если* $\varphi \geqslant 0$, *то* $\varphi \in L$ *тогда и только тогда, когда соответствующее распределение* F *имеет ограниченную плотность.*

Доказательство. По теореме 3 интегрируемость функции φ влечет за собой существование у F непрерывной ограниченной плотности. Обратно, если F имеет плотность $f<M$, то мы получаем из (3.4) при $t=0$

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\varphi(\zeta)\,e^{-1/2a^2\zeta^2}\,d\zeta=\frac{1}{\sqrt{2\pi}\cdot a}\int_{-\infty}^{+\infty}e^{-x^2/(2a^2)}f(x)\,dx<M. \tag{3.7}$$

Подинтегральное выражение слева неотрицательно, и, если бы φ не была интегрируема, левый интеграл стремился бы к ∞ при $a\to 0$. ▶

Примеры. в) *Тождество Планшереля.* Пусть распределение F имеет плотность f и характеристическую функцию φ. Тогда $|\varphi|^2\in L$ тогда и только тогда, когда $f^2\in L$, и в этом случае

$$\int_{-\infty}^{+\infty}f^2(y)\,dy=\frac{1}{2\pi}\int_{-\infty}^{+\infty}|\varphi(\zeta)|^2\,d\zeta. \tag{3.8}$$

Действительно, $|\varphi|^2$ — это характеристическая функция симметризованного распределения ${}^0F$. Если $|\varphi|^2\in L$, то плотность

$${}^0f(x)=\int_{-\infty}^{+\infty}f(y-x)\,f(y)\,dy \tag{3.9}$$

распределения ${}^0F$ ограничена и непрерывна. Применяя (3.5) к ${}^0f$ и полагая $x=0$, получаем (3.8). Обратно, если $f^2\in L$, то применение неравенства Шварца к (3.9) показывает, что ${}^0f$ ограничена, и в силу приведенного выше следствия $|\varphi|^2\in L$. Мы вернемся к соотношению (3.8) в гл. XIX, 7.

г) *Теорема непрерывности для плотностей.* Пусть φ_n и φ — *интегрируемые* характеристические функции, для которых

$$\int_{-\infty}^{+\infty}|\varphi_n(\zeta)-\varphi(\zeta)|\,d\zeta\to 0. \tag{3.10}$$

По следствию теоремы 3 соответствующие распределения F_n и F имеют ограниченные и непрерывные плотности f_n и f. Из (3.10) и формулы обращения ясно, что $f_n \to f$. [При этом сходимость равномерна: разность $|f_n - f|$ меньше умноженного на $1/(2\pi)$ интеграла (3.10).] См. задачу 13.

д) *Формула обращения для функций распределения.* Пусть F — распределение с характеристической функцией φ, и пусть $h>0$ — произвольно, но фиксировано. Тогда

$$\frac{F(x+h)-F(x)}{h} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \varphi(\zeta)\,\frac{1-e^{-i\zeta h}}{i\zeta h}\,e^{-i\zeta h}\,d\zeta, \qquad (3.11)$$

как только интеграл в правой части сходится абсолютно (например, если подинтегральная функция есть $O(1/\zeta^2)$, т. е. если $|\varphi(\zeta)| = O(1/\zeta)$ при $\zeta\to\infty$). В самом деле, левая часть представляет собой *плотность* свертки с распределением, равномерным на $-h$, 0; множитель при $e^{-i\zeta x}$ под интегралом есть характеристическая функция этой свертки. Поэтому (3.11) представляет собой не что иное, как частный случай общей формулы обращения (3.5). ▶

Замечание *о так называемой формуле обращения.* Формула (3.11) применима только в случае, когда функция $|\varphi(\zeta)/\zeta|$ интегрируема на бесконечности. В общем же случае применяются формулы, получаемые ее видоизменением. Например, пусть F_a снова обозначает свертку распределения F с симметричным относительно нуля нормальным распределением с дисперсией a^2. Тогда по (3.11)

$$\frac{F_a(x+h)-F_a(x)}{h} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \varphi(\zeta)\,e^{-\frac{1}{2}a^2\zeta^2}\,\frac{1-e^{-i\zeta h}}{i\zeta h}\,e^{-i\zeta x}\,d\zeta. \qquad (3.12)$$

Если x и $x+h$ — точки непрерывности распределения F, то правая часть стремится к $[F(x+h)-F(x)]/h$ при $a\to 0$. Это — типичный пример «теоремы обращения». Можно предложить бесчисленное множество эквивалентных формулировок. Традиционная форма получается заменой в (3.12) нормального распределения распределением, равномерным на $-t, t$, и переходом к пределу при $t\to\infty$. Формулы обращения все еще остаются популярной темой, хотя они и потеряли в значительной степени свою важность. Заметим, что вывод их из интеграла Дирихле выпадает из логической структуры всей теории.

От распределений с интегрируемыми характеристическими функциями мы перейдем к *решетчатым распределениям.* Пусть F приписывает вес p_k точке $b+kh$, где $p_k \geqslant 0$ и $\sum p_k = 1$. Характеристическая функция φ имеет вид

$$\varphi(\zeta) = \sum_{-\infty}^{+\infty} p_k e^{i(b+kh)\zeta}. \qquad (3.13)$$

Предположим, что $h>0$.

Теорема 4. *Если* φ *— характеристическая функция вида* (3.13), *то*

$$p_r = \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} \varphi(\zeta)\, e^{-i(b+rh)\zeta}\, d\zeta. \tag{3.14}$$

Доказательство. Подинтегральное выражение может быть представлено в виде ряда, где множитель при p_k равен $e^{i(k-r)h\zeta}$. Интеграл от него равен 0, если $k \neq r$ и $2\pi/h$, если $k=r$, так что (3.14) верно. ▶

§ 4. Свойства регулярности

Основной результат этого параграфа может быть грубо резюмирован так: чем меньше «хвосты» распределения F, тем глаже его характеристическая функция φ; обратно, чем глаже F, тем лучше ведет себя φ на бесконечности (леммы 2 и 4). Многие оценки, связанные с характеристическими функциями, опираются на оценку ошибки, с которой функция e^{it} аппроксимируется конечным отрезком своего ряда Тейлора. Следующая ниже лемма показывает, что ошибка мажорируется первым отброшенным членом.

Лемма 1[1]). *Для* $n=1, 2, \ldots$ *и* $t>0$

$$\left| e^{it} - 1 - \frac{it}{1!} - \ldots - \frac{(it)^{n-1}}{(n-1)!} \right| \leqslant \frac{t^n}{n!}. \tag{4.1}$$

Доказательство. Обозначим выражение, стоящее под знаком абсолютной величины, через $\rho_n(t)$. Тогда

$$\rho_1(t) = i \int_0^t e^{ix}\, dx \tag{4.2}$$

и для $n \geqslant 2$

$$\rho_n(t) = i \int_0^t \rho_{n-1}(x)\, dx. \tag{4.3}$$

По теореме о среднем $|\rho_1(t)| < t$, и неравенство (4.1) можно получить по индукции. ▶

---

[1]) То же самое доказательство показывает, что ошибка, получаемая при оставлении в ряде Тейлора для $\sin t$ или $\cos t$ конечного числа членов, *имеет тот же знак, что и первый отброшенный член, и меньше его по абсолютной величине.* Например, $1 - \cos t \leqslant t^2/2$

В дальнейшем F будет обозначать произвольное распределение, а φ — его характеристическую функцию. Моменты и абсолютные моменты распределения F (коль скоро они существуют) обозначим

$$m_n = \int_{-\infty}^{+\infty} x^n F\{dx\}, \quad M_n = \int_{-\infty}^{+\infty} |x|^n F\{dx\}. \tag{4.4}$$

Лемма 2. *Если $M_n<\infty$, то n-я производная от φ существует, непрерывна и задается формулой*

$$\varphi^{(n)}(\zeta) = i^n \int_{-\infty}^{+\infty} e^{i\zeta x} x^n F\{dx\}. \tag{4.5}$$

Доказательство. Разностное отношение для φ равно

$$\frac{\varphi(\zeta+h)-\varphi(\zeta)}{h} = \int_{-\infty}^{+\infty} e^{i\zeta x} \frac{e^{ihx}-1}{h} F\{dx\}. \tag{4.6}$$

В соответствии с последней леммой подинтегральное выражение мажорируется функцией $|x|$, и потому для $n=1$ утверждение (4.5) вытекает из теоремы о мажорированной сходимости. Общий случай доказывается по индукции. ▶

Следствие. *Если $m_2<\infty$, то*

$$\varphi'(0) = im_1, \quad \varphi''(0) = -m_2. \tag{4.7}$$

Утверждение, *обратное* к последнему [1]), также верно: *если $\varphi''(0)$ существует, то $m_2<\infty$.*

Доказательство. Обозначая действительную часть функции φ через u, имеем

$$\frac{1-u(h)}{h^2} = \int_{-\infty}^{+\infty} \frac{1-\cos hx}{h^2x^2} \cdot x^2 F\{dx\}. \tag{4.8}$$

Так как $u''(0)$ существует, то производная u' определена в окрестности нуля и в нуле непрерывна. В частности, $u'(0)=0$, так как u — четная функция. По теореме о среднем значении

[1]) Метод не применим к первой производной. Проблема отыскания условий существования $\varphi'(0)$, привлекавшая долгое время внимание, решена в примере гл. XVII, (1, в).

существует такое число θ, что $0<\theta<1$ и

$$\left|\frac{u(h)-1}{h^2}\right|=\left|\frac{u'(\theta h)}{h}\right|\leqslant\left|\frac{u'(\theta h)}{\theta h}\right|. \tag{4.9}$$

При $h\to 0$ правая часть стремится к $u''(0)$. Но отношение под интегралом в (4.8) стремится к $^1/_2$, так что интеграл стремится к ∞, если $m_2=\infty$. См. задачу 11.

Примеры. а) Функция ψ, отличная от постоянной и такая, что $\psi''(0)=0$, не может быть характеристической, так как соответствующее распределение должно было бы иметь второй момент, равный нулю. Например, функция $e^{-|\zeta|^\alpha}$ при $\alpha>2$ не будет характеристической.

б) *Слабый закон больших чисел.* Пусть $\mathbf{X}_1$, $\mathbf{X}_2$, ... независимые случайные величины с $\mathbf{E}(\mathbf{X}_j)=0$ и одной и той же характеристической функцией φ. Положим $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$. Среднее арифметическое $\mathbf{S}_n/n$ имеет характеристическую функцию $\varphi^n(\zeta/n)$. В окрестности нуля $\varphi(h)=1+o(h)$, и потому $\varphi(\zeta/n)=1+o(1/n)$ при $n\to\infty$. Переходя к логарифмам, мы видим, что $\varphi^n(\zeta/n)\to 1$. По теореме непрерывности (теорема 2 из § 3) отсюда вытекает, что *распределение* величины $\mathbf{S}_n/n$ сходится к распределению, сосредоточенному в нуле. Это и есть слабый закон больших чисел. Такая простота и ясность доказательства вообще типичны для метода характеристических функций. Некоторое видоизменение этого рассуждения приведет нас к доказательству центральной предельной теоремы. ▶

Лемма 3. (*Риман — Лебег.*) *Если g интегрируема и*

$$\gamma(\zeta)=\int_{-\infty}^{+\infty} e^{i\zeta x}g(x)\,dx, \tag{4.10}$$

то $\gamma(\zeta)\to 0$ при $\zeta\to\pm\infty$.

Доказательство. Утверждение проверяется непосредственно для ступенчатых функций с конечным числом ступенек. Для произвольной интегрируемой функции g и любого $\varepsilon>0$ найдется (теорема об аппроксимации в среднем из гл. IV, 2) такая ступенчатая функция g_1 с конечным числом ступенек, что

$$\int_{-\infty}^{+\infty}|g(x)-g_1(x)|\,dx<\varepsilon. \tag{4.11}$$

Соответствующие преобразования (4.10) связаны неравенством $|\gamma-\gamma_1|<\varepsilon$, и потому существует окрестность бесконечности

$\pm\infty$, в которой $|\gamma(\zeta)|<2\varepsilon$. Так как ε произвольно, то $\gamma(\zeta)\to 0$. ▶

В качестве простого следствия получается

Лемма 4. Если F имеет плотность f, то $\varphi(\zeta)\to 0$ при $\zeta\to\pm\infty$. Если f имеет интегрируемую n-ю производную $f^{(n)}$, то $|\varphi(\zeta)|=o(|\zeta|^{-n})$ при $|\zeta|\to\infty$.

Доказательство. Первое утверждение содержится в лемме 3. Если f' интегрируема, то интегрирование по частям показывает, что

$$\varphi(\zeta)=\frac{1}{i\zeta}\int_{-\infty}^{+\infty} e^{i\zeta x}f'(x)\,dx \tag{4.12}$$

и, следовательно, $|\varphi(\zeta)|=o(|\zeta|^{-1})$, и т. д. ▶

Дополнение. Разложение Тейлора для характеристических функций.

Неравенство (4.1) можно переписать в виде

$$\left|e^{i\zeta x}\left(e^{itx}-1-\frac{itx}{1!}-\ldots-\frac{(itx)^{n-1}}{(n-1)!}\right)\right|\leqslant\frac{|tx|^n}{n!}. \tag{4.13}$$

Отсюда, используя (4.5), получаем

$$\left|\varphi(\zeta+t)-\varphi(\zeta)-\frac{t}{1!}\varphi'(\zeta)-\ldots-\frac{t^{n-1}}{(n-1)!}\varphi^{(n-1)}(\zeta)\right|<M_n\frac{|t|^n}{n!}. \tag{4.14}$$

Если $M_n<\infty$, то это неравенство верно при любых ζ и t и дает верхнюю границу для модуля разности между φ и первыми членами разложения Тейлора. В специальном случае, когда F сосредоточено в точке 1, неравенство (4.14) превращается в (4.1).

Предположим теперь, что все моменты существуют и что

$$\overline{\lim_{n\to\infty}}\frac{1}{n}M_n^{1/n}=\lambda<\infty. \tag{4.15}$$

Из формулы Стирлинга для $n!$ тривиально вытекает, что при $|t|<1/(3\lambda)$ правая часть неравенства (4.14) стремится к нулю при $n\to\infty$, так что ряд Тейлора для φ сходится в некотором интервале вокруг ζ. Отсюда следует, что φ аналитична в некоторой полосе, окружающей действительную ось, и, стало быть, однозначно определяется своим разложением в степенной ряд в окрестности нуля. Но $\varphi^{(n)}(0)=(i)^n m_n$, так что φ однозначно определяется моментами m_n распределения F. Соответственно, *если* (4.15) *выполняется, то распределение F однозначно определяется своими моментами и* φ аналитична в некоторой окрестности действительной оси. Этот критерий единственности слабее, чем указанное в гл. VII достаточное условие (3.11), принадлежащее Карлеману: $\sum M_n^{1/n}=\infty$. Но оба критерия не так уж далеки один от другого (пример распределения, не определяемого однозначно своими моментами, приведен в гл. VII, 3).

§ 5. Центральная предельная теорема для одинаково распределенных слагаемых

Исследования, связанные с центральной предельной теоремой, сильно влияли на развитие и совершенствование методов, применяемых теперь повсюду в теории вероятностей, поэтому представляется поучительным сравнение различных доказательств. До последнего времени метод характеристических функций (впервые употребленный П. Леви) был несравненно проще, чем прямой подход, примененный Линдебергом (мы не говорим здесь о других доказательствах). Освобожденный от усложнений современный вариант последнего метода (изложенный в гл. VIII, 4) не труднее метода характеристических функций и, кроме того, имеет некоторые преимущества. С другой стороны, метод характеристических функций приводит к уточнениям, которые в настоящий момент не удается получить прямыми методами. К их числу относятся локальные предельные теоремы этого параграфа, а также оценки ошибок и асимптотические разложения, изучаемые в следующей главе. Мы выделяем случай слагаемых, имеющих одно и то же распределение, частично из-за его важности и частично потому, что желаем на примере простейшего случая показать существо метода.

Всюду в этом параграфе $\mathbf{X}_1$, $\mathbf{X}_2$, ... предполагаются независимыми случайными величинами с одним и тем же распределением F, которому соответствует характеристическая функция φ. Мы предполагаем также, что

$$\mathbf{E}(\mathbf{X}_j)=0, \quad \mathbf{E}(\mathbf{X}_j^2)=1 \tag{5.1}$$

и обозначаем $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$.

Теорема 1 [1]). *Распределение величины $S_n/\sqrt{n}$ сходится к нормальному распределению $\mathfrak{N}$.*

В силу теоремы непрерывности (теорема 2 из § 3) это утверждение эквивалентно следующему: при $n\to\infty$ и всех ζ

$$\varphi^n\left(\frac{\zeta}{\sqrt{n}}\right)\to e^{-1/2\zeta^2}. \tag{5.2}$$

---

[1]) Существование дисперсии вовсе не обязательно для асимптотической нормальности величины $\mathbf{S}_n$. Необходимые и достаточные условия даны в теореме 1а в гл. XVII, 5.

Доказательство. Из леммы 4.2 и формулы (4.7) ясно, что при $n \to \infty$ и фиксированном ζ

$$\varphi\left(\frac{\zeta}{\sqrt{n}}\right) = 1 - \frac{1}{2n}\zeta^2 + o\left(\frac{1}{n}\right). \tag{5.3}$$

Из формулы Тейлора для логарифмов следует, что $\ln(1+z) \sim z$ при $z \to 0$. Поэтому (5.3) дает

$$n \log \varphi\left(\frac{\zeta}{\sqrt{n}}\right) \to -\frac{1}{2}\zeta^2, \tag{5.4}$$

что равносильно (5.2). ▶

Естественно ожидать, что когда F обладает плотностью f, *плотность* величины $S_n/\sqrt{n}$ должна сходиться к нормальной плотности $\mathfrak{n}$. Это не всегда так, но, к счастью, исключения носят «патологический» характер. Сформулированная ниже теорема охватывает много практически интересных случаев.

Теорема 2. *Если функция* $|\varphi|$ *интегрируема, то случайная величина* $S_n/\sqrt{n}$ *имеет плотность, равномерно сходящуюся к нормальной плотности* $\mathfrak{n}$.

Доказательство. Формула обращения интегралов Фурье (3.5) применима как к f_n, так и к $\mathfrak{n}$, и потому

$$|f_n(x) - \mathfrak{n}(x)| \leqslant \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left|\varphi^n\left(\frac{\zeta}{\sqrt{n}}\right) - e^{(-1/2)\zeta^2}\right| d\zeta. \tag{5.5}$$

Мы должны показать, что правая часть стремится к нулю при $n \to \infty$.

Выберем $\delta > 0$ так, чтобы

$$|\varphi(\zeta)| \leqslant e^{(-1/4)\zeta^2} \quad \text{при } |\zeta| < \delta. \tag{5.6}$$

Это возможно, так как в нуле обе функции равны единице, их первые производные равны нулю, а вторая производная левой функции, равная -1, меньше второй производной правой функции (равной $-1/2$). Мы разобьем теперь интеграл (5.5) на три части и покажем, что каждую часть можно сделать меньше ε при достаточно большом n. 1) Как мы видели в последнем доказательстве, внутри фиксированного интервала $-a \leqslant \zeta \leqslant a$ подинтегральное выражение равномерно стремится к нулю, так что интеграл по этому промежутку стремится к нулю. 2) Для $a < |\zeta| < \delta\sqrt{n}$ подинтегральное выражение меньше, чем $2e^{(-1/4)\zeta^2}$, так что величина соответствующего интеграла меньше ε, если только a взято достаточно большим. 3) Мы знаем из леммы 1.4, что $|\varphi(\zeta)| < 1$ для $\zeta \neq 0$, и из леммы 4.3, что $\varphi(\zeta) \to$

$\to 0$ при $|\zeta| \to \infty$. Отсюда следует, что максимум функции $|\varphi(\zeta)|$ для $|\zeta| \geqslant \delta$ равен числу $\eta < 1$. Поэтому часть интеграла (5.5), взятая по области $|\zeta| > \delta\sqrt{n}$, меньше, чем

$$\eta^{n-1} \int_{-\infty}^{+\infty} \left|\varphi\left(\frac{\zeta}{\sqrt{n}}\right)\right| d\zeta + \int_{|\zeta| > \delta\sqrt{n}} e^{-(1/2)\zeta^2} d\zeta. \tag{5.7}$$

Первый интеграл равен интегралу от $\sqrt{n}\,|\varphi(y)|$, так что выражение (5.7) стремится к нулю. ▶

На самом деле из доказательства вытекает [1]) более сильный результат, а именно: если $|\varphi|^r \in L$ *при некотором целом* r, *то* $f_n \to \mathfrak{n}$ *равномерно для всех* x. В то же время следствие теоремы 3.3 показывает, что если $|\varphi|^r$ не интегрируема ни при каком r, то каждая из функций f_n не ограничена. Приведем любопытные примеры, подтверждающие, что такие патологические случаи действительно встречаются.

Примеры. а) Для $x > 0$ и $p \geqslant 1$ положим

$$u_p(x) = \frac{1}{x \log^{2p} x}. \tag{5.8}$$

Пусть g — плотность распределения, сосредоточенного на интервале $\overline{0,1}$, и пусть $g(x) > u_p(x)$ на некотором интервале $\overline{0,h}$. Существует интервал $\overline{0,\delta}$, в котором u_p монотонно убывает, и внутри этого интервала

$$g^{2*}(x) \geqslant \int_0^x u_p(x-y)\, u_p(y)\, dy > x u_p^2(x) = u_{2p}(x). \tag{5.9}$$

По индукции устанавливаем, что при $n = 2^k$ существует интервал $\overline{0,h_n}$, в котором $g^{n*} \geqslant u_{np}$, и, следовательно, $g^{n*}(x) \to \infty$ при $x \to 0+$. Таким образом, каждая свертка g^{n*} не ограничена.

б) Вариант предыдущего примера. Та же самая патологическая особенность проявляется здесь в еще более резкой форме. Пусть v — плотность, получаемая симметризацией функции g. Положим

$$f(x) = \frac{1}{2}\,(v(x+1) + v(x-1)). \tag{5.10}$$

Тогда $f(x)$ будет четной плотностью, сосредоточенной на $\overline{-2,2}$. Мы можем предположить, что соответствующая дисперсия равна 1. Анализируя предыдущий пример, мы видим, что v непрерывна всюду, кроме нуля, где она не ограничена. Это же верно и по отношению ко всем сверткам v^{n*}. Теперь $f^{2n*}(x)$ является линейной комбинацией функций $v^{2n*}(x+2k)$ с $k = 0, \pm 1, \pm 2, \ldots, \pm n$ и, следовательно, не ограничена во всех этих точках. Плотность нормированной суммы $\mathbf{S}_{2n}/\sqrt{2n}$ случайных величин X_j (с плотностью f каждая) определяется формулой $f_{2n}(x) = \sqrt{2n}\, f^{2n*}(x\sqrt{2n})$. Она непрерывна всюду, кроме $2n+1$ точек вида $2k/\sqrt{2n}$ $(k = 0, \pm 1, \ldots, \pm n)$, в которых она

[1]) Единственное изменение состоит в том, что в (5.7) множитель η^{n-1} заменяется на η^{n-r}, а φ на φ^r.

не ограничена. Так как каждой рациональной точке t соответствует бесконечно много пар (k, n) целых чисел, таких, что $2k/\sqrt{2n} = t$, то мы приходим к выводу: *распределение случайной величины* $\mathbf{S}_n/\sqrt{n}$ *стремится к* $\mathfrak{N}$, *но плотности* f_n *не сходятся ни в одной рациональной точке и последовательность* $\{f_n\}$ *не будет ограниченной ни в одном интервале.*

в) *См. задачу* 17. ▶

Чтобы сделать картину полной, перейдем теперь к *решетчатым распределениям*. Предположим, что случайные величины $\mathbf{X}_j$ принимают значения вида b, $b \pm h$, $b \pm 2h$, ... Предположим, что h является *шагом* распределения, т. е. максимальным положительным числом с указанным свойством. Лемма 1.4 утверждает, что $|\varphi|$ имеет период $2\pi/h$ и, следовательно, функция $|\varphi|$ не интегрируема. Однако теореме 2 соответствует аналогичная теорема, касающаяся «весов» возможных значений (атомов) распределения величины $\mathbf{S}_n/\sqrt{n}$. Все эти атомы находятся среди точек вида $x = (nb + kn)/\sqrt{n}$, где $k = 0, \pm 1, \pm 2, \ldots$. Положим для таких x

$$p_n(x) = \mathbf{P}\left\{\frac{\mathbf{S}_n}{\sqrt{n}} = x\right\}. \tag{5.11}$$

Для других x функция $p_n(x)$ не будет определена. Таким образом, значения x в соотношении (5.12) принадлежат наименьшей решетке, содержащей все атомы случайной величины $\mathbf{S}_n/\sqrt{n}$.

Теорема 3[1]). *Если* F — *решетчатое распределение с шагом* h, *то при* $n \to \infty$

$$\frac{\sqrt{n}}{h}\, p_n(x) - \mathfrak{n}(x) \to 0 \tag{5.12}$$

равномерно относительно x.

Доказательство. По (3.14) имеем

$$\frac{\sqrt{n}}{h}\, p_n(x) = \frac{1}{2\pi} \int_{-\sqrt{n}\pi/h}^{\sqrt{n}\pi/h} \varphi^n\left(\frac{\zeta}{\sqrt{n}}\right) e^{-ix\zeta}\, d\zeta. \tag{5.13}$$

Используя снова формулу обращения (3.5) для нормальной плотности $\mathfrak{n}$, получаем, что левая часть в (5.12) не превосходит

$$\frac{1}{2\pi} \int_{-\sqrt{n}\pi/h}^{\sqrt{n}\pi/h} \left|\varphi^n\left(\frac{\zeta}{\sqrt{n}}\right) - e^{(-1/2)\zeta^2}\right| d\zeta + \frac{1}{2\pi} \int_{|\zeta| > \sqrt{n}\pi/h} e^{(-1/2)\zeta^2}\, d\zeta. \tag{5.14}$$

[1]) Предполагается, как и раньше, что выполнены условия (5.1). — *Прим. перев.*

Повторяя в точности доказательство теоремы 2, видим, что первый интеграл стремится к нулю. Ясно, что и второй интеграл стремится к нулю. Этим завершается доказательство. ▶

§ 6. Условие Линдеберга

Рассмотрим теперь последовательность таких независимых случайных величин $\mathbf{X}_k$, что

$$\mathbf{E}(\mathbf{X}_k)=0, \quad \mathbf{E}(\mathbf{X}_k^2)=\sigma_k^2. \tag{6.1}$$

Мы обозначим распределение величины $\mathbf{X}_k$ через F_k, характеристическую функцию через φ_k, а также обозначим, как обычно, $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ и $s_n^2=\mathrm{Var}(\mathbf{S}_n)$. Таким образом,

$$s_n^2=\sigma_1^2+\ldots+\sigma_n^2. \tag{6.2}$$

Мы будем говорить, что выполняется *условие Линдеберга*, если при любом фиксированном $t>0$

$$\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x|>ts_n} x^2F_k\{dx\}\to 0, \quad n\to\infty. \tag{6.3}$$

Грубо говоря, это условие требует, чтобы дисперсии σ_k^2 образовывались в основном за счет масс, сосредоточенных в интервалах длины, малой по сравнению с s_n. Далее ясно, что σ_k^2/s_2^n меньше, чем t^2 плюс левая часть в (6.3). Так как t произвольно, то из (6.3) следует, что при любом $\varepsilon>0$ и достаточно большом n

$$\frac{\sigma_k}{s_n}\leqslant\varepsilon, \quad k=1, 2, \ldots, n. \tag{6.4}$$

Условие Линдеберга было введено в гл. VIII (соотношение (4.15)), и следующая ниже теорема совпадает с теоремой 3 из гл. VIII, 4. Каждое доказательство имеет свои преимущества. Приводимое здесь доказательство позволяет установить, что условие Линдеберга в некотором смысле и необходимо; это доказательство применимо для вывода асимптотических разложений (гл. XVI) и теорем о сходимости плотностей (задача 19).

Теорема 1. (*Линдеберг.*) *Если выполняется условие* (6.3), *то распределения нормированных сумм* $\mathbf{S}_n/s_n$ *сходятся к нормальному распределению* $\mathfrak{N}$.

Доказательство. Учитывая нормировку (6.1), имеем

$$\varphi_k\left(\frac{\zeta}{s_n}\right)-1+\frac{1}{2}\frac{\sigma_k^2}{s_n^2}\zeta^2=\int_{-\infty}^{+\infty}\left(e^{i\zeta x/s_n}-1-\frac{i\zeta x}{s_n}+\frac{1}{2}\frac{\zeta^2x^2}{s_n^2}\right)F_k\{dx\}. \quad (6.5)$$

Основное неравенство (4.1) показывает, что при достаточно больших n

$$\left|\varphi_k\left(\frac{\zeta}{s_n}\right)-1\right|\leqslant\frac{1}{2}\zeta^2\cdot\frac{\sigma_k^2}{s_n^2}<\frac{1}{2}\varepsilon\zeta^2. \quad (6.6)$$

Из формулы Тейлора видно, что

$$\sum_{k=1}^{n}|\ln(1+z_k)-z_k|\leqslant\sum_{k=1}^{n}|z_k|^2<\varepsilon\sum_{k=1}^{n}|z_k| \quad \text{при } |z_k|<\varepsilon<\tfrac{1}{2}. \quad (6.7)$$

Отсюда, полагая $z_k=\varphi_k(\zeta/s_n)-1$, получаем при фиксированном ζ и $n\to\infty$

$$-\sum_{k=1}^{n}\ln\varphi_k\left(\frac{\zeta}{s_n}\right)\sim\sum_{k=1}^{n}\left(\varphi_k\left(\frac{\zeta}{s_n}\right)-1\right). \quad (6.8)$$

Мы должны доказать, что правая часть стремится к $(-1/2)\zeta^2$. С этой целью мы оценим подинтегральное выражение в (6.5) с помощью неравенства (4.1). На отрезке $|x|\leqslant ts_n$ мы возьмем оценку $\frac{|\zeta x|^3}{s_n^3}\leqslant t\frac{|\zeta|^3}{s_n^2}x^2$, а для $|x|>ts_n$ — более грубую оценку $(\zeta^2x^2)/s_n^2$. Тогда

$$\sum_{k=1}^{n}\left|\varphi\left(\frac{\zeta}{s_n}\right)-1+\frac{1}{2}\frac{\sigma_k^2}{s_n^2}\zeta^2\right|\leqslant t|\zeta^3|+\zeta^2\cdot\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x|>ts_n}x^2F_k\{dx\}. \quad (6.9)$$

Множитель при ζ^2 — это как раз то выражение, которое входит в условие Линдеберга (6.3). Так как t можно выбрать сколь угодно малым, то сумма, стоящая слева, стремится к 0. Поэтому правая часть соотношения (6.8) в пределе равна $(-1/2)\zeta^2s_n^{-2}\sum\sigma_k^2=(-1/2)\zeta^2$, чем доказательство и заканчивается. ▶

За примерами читатель отсылается к гл. VIII, 4, к задачам 17—20 в гл. VIII, 10, и к задаче 20 ниже.

Следующая теорема утверждает, что *в предположении*

$$\frac{\sigma_n}{s_n}\to 0, \quad s_n\to\infty \quad (6.10)$$

условие Линдеберга является также необходимым.

Теорема 2. *Допустим, что распределение величины* $\mathbf{S}_n/s_n$ *сходится к* $\mathfrak{N}$ *и что выполняется* (6.10). *Тогда справедливо условие Линдеберга* (6.3).

Доказательство. Покажем сначала, что справедливо неравенство (6.4). В самом деле, найдется такое ν, что $\sigma_k/s_k<\varepsilon$ при $k>\nu$. Тогда при $n\geqslant k>\nu$ имеем $\sigma_k/s_k\leqslant\sigma_n/s_k<\varepsilon$, а ν величин $\sigma_1/s_n, \ldots, \sigma_\nu/s_n$ стремятся к нулю. Мы уже видели, что (6.4) влечет (6.8). Поэтому мнимая часть суммы (6.8) стремится к нулю. Следовательно, при фиксированных $t>0$ и ζ и $n\to\infty$

$$\frac{1}{2}\zeta^2-\sum_{k=1}^{n}\int_{|x|\leqslant ts_n}\left(1-\cos\frac{\zeta x}{s_n}\right)F_k\{dx\}=$$

$$=\sum_{k=1}^{n}\int_{|x|>ts_n}\left(1-\cos\frac{\zeta x}{s_n}\right)\zeta_k\{dx\}+o(1). \quad (6.11)$$

Подинтегральное выражение справа не превосходит $2<2x^2/t^2s_n^2$, а слева — не превосходит $\zeta^2x^2/2s_n^2$. Деля (6.11) на $^1/_2\zeta^2$, получаем

$$1-\frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x|\leqslant ts_n}x^2F_k\{dx\}\leqslant\frac{4}{t^2\zeta^2}+o(1). \quad (6.12)$$

Левая часть не зависит от ζ, а правую часть можно сделать сколь угодно малой, если величину ζ выбрать достаточно большой. Следовательно, левая часть стремится к нулю, что, очевидно, равносильно (6.3). ▶

Отношение σ_n/s_n можно принять за меру того вклада, который вносит слагаемое $\mathbf{X}_n$ в нормированную сумму $\mathbf{S}_n/s_n$. Тогда условие (6.10) можно описать следующим образом: асимптотически $\mathbf{S}_n/s_n$ является суммой *«многих индивидуально пренебрежимых слагаемых»*. В принципе распределение величины $\mathbf{S}_n/s_n$ может сходиться к $\mathfrak{N}$, даже если (6.10) не выполняется. Но такой тип сходимости радикально отличается от того, который интуитивно связывается с центральной предельной теоремой.

Примеры. а) Пусть F_k — нормальное распределение с дисперсией σ_k^2. Тогда $\mathbf{S}_n/s_n$ имеет нормальное распределение $\mathfrak{N}$ независимо от того, выполнено или нет условие (6.10).

б) Пусть G_k — распределение с нулевым математическим ожиданием и единичной дисперсией и такое, что $G_k\to\mathfrak{N}$. Положим $F_k(x)=G_k(x/k!)$. Тогда $s_2^{n-1}=o(s_n^2)$. Следовательно, дис-

персия случайной величины $\mathbf{S}_{n-1}/s_n$ стремится к нулю. Поэтому ее распределение асимптотически то же, что и распределение величины $\mathbf{S}_n/s_n$, которое в свою очередь асимптотически эквивалентно распределению G_n случайной величины $\mathbf{X}_n/\sigma_n$. Таким образом, $\mathbf{S}_n$ распределено асимптотически нормально, но это потому лишь, что $\mathbf{X}_n$ обладает таким же свойством. ▶

Последний пример иллюстрирует типичную ситуацию, возникающую в случае, когда (6.10) не выполняется. Это можно увидеть из следующей теоремы (в которой n может пробегать и подпоследовательность $n_1, n_2, \ldots$).

Предположим, что распределение величины $\mathbf{S}_n/s_n$ *стремится к* $\mathfrak{N}$ *и* $\sigma_n/s_n \to p \neq 0$. *Тогда распределение величины* $\mathbf{X}_n/s_n$ *асимптотически нормально с дисперсией* p^2.

Доказательство. Для простоты предположим сначала, что распределения F_k симметричны. По теореме о выборе можно найти такую последовательность номеров $n_1, n_2, \ldots$, что по этой последовательности распределения величин $\mathbf{S}_{n-1}/s_n$ и $\mathbf{X}_n/s_n$ сходятся к пределам U и V (последние распределения обязательно будут собственными). Тогда $U * V = \mathfrak{N}$, и по теореме Крамера (теорема 1 из § 8) отсюда следует, что U и V нормальны и сумма их дисперсий равна единице. Дисперсии величин $\mathbf{S}_{n-1}/s_n$ и $\mathbf{X}_n/s_n$ сходятся к $1-p^2$ и p^2 соответственно. Дисперсии распределений U и V не могут превзойти эти пределы. Отсюда следует, что V нормально и имеет дисперсию p^2.

В случае асимметричных распределений мы применяем симметризацию. Из сказанного выше вытекает, что симметризованное распределение ${}^0F_n$ асимптотически нормально. Теорема Крамера показывает, что это же верно и для распределения F_n. ▶

Предыдущие теоремы не должны пониматься неправильно: распределение величины $\mathbf{S}_n/s_n$ *может быть асимптотически нормально, даже если* (6.10) *выполняется, а условие Линдеберга* (6.3) *нет.*

Пример. в) Пусть $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ — независимые, одинаково распределенные величины с $\mathbf{E}(\mathbf{Y}_k)=0$ и $\mathrm{Var}(\mathbf{Y}_k)=1$. Пусть $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ не зависят друг от друга и от $\{\mathbf{Y}_k\}$, и пусть $\mathbf{P}\{\mathbf{Z}_n=\pm n\}=(1/2)n^{-2}$ и $\mathbf{P}\{\mathbf{Z}_n=0\}=1-n^{-2}$. Тогда $\mathbf{E}(\mathbf{Z}_n)=0$ и $\mathrm{Var}(\mathbf{Z}_n)=1$. Согласно лемме Бореля — Кантелли, с вероятностью единица лишь конечное число величин $\mathbf{Z}_k$ отлично от нуля. Следовательно $(\mathbf{Z}_1+\ldots+\mathbf{Z}_n)/\sqrt{n}$ стремится по вероятности к нулю (см. гл. VIII, 2). В то же время распределение $(\mathbf{Y}_1+\ldots+\mathbf{Y}_n)/\sqrt{n}$ сходится к $\mathfrak{N}$. Имея это в виду, положим $\mathbf{X}_k=\mathbf{Y}_k+\mathbf{Z}_k$. Тогда ясно, что распределение величины $\mathbf{S}_n/\sqrt{n}$ стремится к $\mathfrak{N}$. Но $\mathrm{Var}(\mathbf{S}_n)=2n$ и распределение *нормированной суммы* $\mathbf{S}_n/s_n$ *стремится к нормальному распределению с дисперсией* $1/2$. ▶

Этот пример показывает, что классическая нормировка сумм $\mathbf{S}_n$, превращающая математическое ожидание в нуль, а дис-

персию в единицу, не *всегда* естественна. Чтобы обеспечить сходимость к $\mathfrak{N}$, порой нужны другие способы нормировки. Мы можем пойти и дальше. Существенной чертой последнего примера было то, что ряд $\sum \mathbf{P}\{\mathbf{Z}_k \neq 0\}$ сходится. Но это свойство может выполняться и при отсутствии у $\mathbf{Z}_k$ математических ожиданий. Следовательно, *возможно построить величины* $\mathbf{X}_k$, *не имеющие математических ожиданий, но такие, что при надлежащем выборе нормирующих констант* a_n *распределение суммы* $\mathbf{S}_n/a_n$ *сходится к* $\mathfrak{N}$. Предыдущие теоремы можно изменить так, чтобы охватить эту общую ситуацию. Мы, однако, не будем входить в детали, так как условия сходимости содержатся в теоремах, касающихся «серий» $\{\mathbf{X}_k/a_n\}$ (гл. XVIII). Но метод, развитый при доказательстве теорем 1 и 2, легко приложим к случаю, когда распределения F_k симметричны. Соответствующие результаты поучительны. Они содержатся в задачах 21 и 22, которые не должны доставить каких-либо затруднений.

§ 7. Характеристические функции многомерных распределений

Теория характеристических функций в пространствах нескольких измерений так близка к теории для случая $\mathscr{R}^1$, что ее систематическое изложение представляется излишним. Для того чтобы описать основные идеи и обозначения, достаточно рассмотреть случай двух измерений. Мы будем через $\mathbf{X}$ обозначать в этом параграфе пару действительных случайных величин $\mathbf{X}_1$ и $\mathbf{X}_2$ с заданным совместным распределением вероятностей F. Мы будем рассматривать $\mathbf{X}$ *как вектор-столбец* с компонентами $\mathbf{X}_1$ и $\mathbf{X}_2$; аналогично аргумент x в $F(x)$ следует понимать как вектор-столбец с компонентами x_1, x_2. С другой стороны, аргумент ζ соответствующей характеристической функции будет пониматься как *вектор-строка* $\zeta=(\zeta_1, \zeta_2)$. Это соглашение удобно тем, что скалярное произведение записывается в виде $\zeta x=\zeta_1 x_1+\zeta_2 x_2$. *Характеристическая функция* φ *для* $\mathbf{X}$ *(или для* F*) определяется равенством*

$$\varphi(\zeta) = \mathbf{E}(e^{i\zeta\mathbf{X}}). \tag{7.1}$$

Это определение по форме такое же, как и в одномерном случае, однако показатель степени имеет новый смысл и интегрирование производится по отношению к двумерному распределению.

Основные свойства двумерных характеристических функций очевидны. Например, при $\zeta_2=0$ скалярное произведение ζx превращается в $\zeta_1 x_1$ и, следовательно, $\varphi(\zeta_1, 0)$ *представляет собой*

характеристическую функцию распределения [1]) $\mathbf{X}_1$. При любых *фиксированных значениях* параметров ζ_1, ζ_2 линейная комбинация $\zeta_1\mathbf{X}_1+\zeta_2\mathbf{X}_2$ *является (одномерной) случайной величиной с характеристической функцией*

$$\mathbf{E}\left(e^{i\lambda(\zeta_1\mathbf{X}_1+\zeta_2\mathbf{X}_2)}\right)=\varphi(\lambda\zeta_1,\ \lambda\zeta_2). \tag{7.2}$$

Здесь ζ_1 и ζ_2 фиксированы, а независимой переменной служит λ. В частности, характеристическая функция суммы $\mathbf{X}_1+\mathbf{X}_2$ равна $\varphi(\lambda, \lambda)$. Таким путем двумерная характеристическая функция порождает одномерные характеристические функции всех линейных комбинаций $\zeta_1\mathbf{X}_1+\zeta_2\mathbf{X}_2$. Обратно, если известны распределения всех таких комбинаций, мы можем вычислить все выражения типа $\varphi(\lambda\zeta_1, \lambda\zeta_2)$, а вместе с ними и двумерную характеристическую функцию [2]). Следующий пример показывает пользу и гибкость этого подхода. В примере используются обозначения, введенные в гл. III, 5.

Примеры. а) *Многомерные нормальные характеристические функции.* Пусть $\mathbf{X}=(\mathbf{X}_1, \mathbf{X}_2)$ (который надо считать вектором-столбцом!) имеет невырожденное нормальное распределение. Предположим для простоты, что $\mathbf{E}(\mathbf{X})=0$. Обозначим *матрицу ковариаций* $\mathbf{E}(\mathbf{X}\mathbf{X}^T)$ буквой C. Ее элементами будут $c_{kk}=\mathrm{Var}\,(\mathbf{X}_k)$ и $c_{12}=c_{21}=\mathrm{Cov}\,(\mathbf{X}_1, \mathbf{X}_2)$. При фиксированных ζ_1 и ζ_2 линейная комбинация $\zeta\mathbf{X}=\zeta_1\mathbf{X}_1+\zeta_2\mathbf{X}_2$ имеет математическое ожидание 0 и дисперсию

$$\sigma^2=\zeta C\zeta^T=c_{11}\zeta_1^2+2c_{12}\zeta_1\zeta_2+c_{22}\zeta_2^2. \tag{7.3}$$

Поэтому характеристическая функция для $\zeta\mathbf{X}$ (ее аргумент обозначим через λ) имеет вид $e^{(-1/2)\sigma^2\lambda^2}$, следовательно, двумерная характеристическая функция для $\mathbf{X}$ равна

$$\varphi(\zeta)=e^{-(1/2)\,\zeta C\zeta^T}. \tag{7.4}$$

Точно такое же рассуждение показывает, что и *r-мерное невырожденное нормальное распределение с нулевым математическим ожиданием и матрицей ковариаций C имеет характеристическую функцию* (7.4). С точностью до множителя (—1/2) показатель есть квадратичная форма от r переменных с матрицей C (заметим, что в выражении для плотности показатель

---

[1]) Иногда говорят — «частного» или «маргинального» распределения $\mathbf{X}_1$. — *Прим. перев.*

[2]) Этим мы доказали попутно, что *распределение вероятностей в $\mathcal{R}^2$ однозначно определяется по вероятностям всех полуплоскостей.* Этот результат (отмеченный Крамером и Вальдом) пока не доказан элементарными методами. Его применение к проблеме моментов указано в задаче 23.

содержит квадратичную форму с матрицей C^{-1}; это прямое обобщение одномерного случая, где показатели имеют вид $-{}^1/_2\sigma^2\zeta^2$ и $-{}^1/_2\sigma^{-2}\zeta^2$ соответственно). ▶

Иногда желательно перейти от переменных $(\mathbf{X}_1, \mathbf{X}_2)$ и (ζ_1, ζ_2) к *полярным координатам*. Положим

$$\mathbf{X}_1 = \mathbf{R}\cos\mathbf{\Phi}, \quad \mathbf{X}_2 = \mathbf{R}\sin\mathbf{\Phi}, \quad \zeta_1 = \rho\cos\theta, \quad \zeta_2 = \rho\sin\theta. \tag{7.5}$$

В этих обозначениях имеем

$$\varphi(\zeta) = \mathbf{E}\,(e^{i\rho\mathbf{R}\cos(\theta-\mathbf{\Phi})}). \tag{7.6}$$

Следует иметь в виду, однако, что это *не есть* характеристическая функция пары случайных величин $\mathbf{R}$, $\mathbf{\Phi}$; последняя равна $\mathbf{E}\,(e^{i(\zeta_1\mathbf{R}+\zeta_2\mathbf{\Phi})})$.

б) *Круговая симметрия в* $\mathscr{R}^2$. Предположим, что пара $(\mathbf{X}_1, \mathbf{X}_2)$ описывает «случайно направленный вектор» (см. гл. I, 10). Совместное распределение для $(\mathbf{R}, \mathbf{\Phi})$ является произведением распределения G для $\mathbf{R}$ и равномерного на $-\pi, \pi$ распределения. Ясно, что математическое ожидание в (7.6) не зависит от θ, так что φ принимает форму

$$\varphi(\zeta_1, \zeta_2) = \int_0^\infty G\,\{dr\} \int_{-\pi}^{\pi} e^{i\rho r\cos t}\,\frac{dt}{2\pi}. \tag{7.7}$$

Внутренний интеграл легко вычисляется разложением экспоненты в степенной ряд. Повторное интегрирование по частям показывает, что интеграл от $\cos^{2k} t$ равен $\binom{2k}{k}2^{-2k}$, в то время как интеграл от $\cos^{2k+1} t$ равен нулю. Полагая

$$J_0(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,k!}\left(\frac{x}{2}\right)^{2k}, \tag{7.8}$$

легко получаем

$$\varphi(\zeta_1, \zeta_2) = \int_0^\infty J_0(r\rho)\,G\,\{dr\}, \tag{7.9}$$

где $\rho^2 = \zeta_1^2 + \zeta_2^2$. Функция J_0 — это функция Бесселя нулевого порядка. При этом $J_0(x) = I_0(ix)$, где I_0 определена в гл. II, (7.1).

В специальном случае (рассмотренном *Рэлеем*) распределение G сосредоточено в точке 1. Таким образом, *характеристическая функция суммы n независимых единичных случайно направленных векторов* (см. гл. I, 10) *равна* $J_0^n\left(\sqrt{\zeta_1^2+\zeta_2^2}\right)$.

в) *Сферическая симметрия в* $\mathscr{R}^3$. Рассуждения предыдущего примера можно повторить при любом числе измерений. При этом изменяется лишь форма внутреннего интеграла в (7.7). В трехмерном случае он сводится к

$$\frac{1}{2}\int_0^{\pi} e^{i\rho r\cos t}\sin t\,dt=\frac{\sin \rho r}{\rho r}. \tag{7.10}$$

В частности, если G сосредоточено в точке 1, мы получаем

$$\varphi(\zeta_1,\zeta_2,\zeta_3)=\frac{\sin\rho}{\rho},\quad \text{где}\quad \rho^2=\zeta_1^2+\zeta_2^2+\zeta_3^2. \tag{7.11}$$

Полагая $\zeta_2=\zeta_3=0$. находим характеристическую функцию компоненты $\mathbf{X}_1$ случайно направленного единичного вектора. Таким образом, мы по-новому доказали, что *эта компонента распределена равномерно на* $-1,\ 1$ (см. гл. I, 10). ➤

Мы предоставляем читателю проверить, что основные теоремы, касающиеся характеристических функций в одномерном случае, обобщаются без существенных изменений. Так, *теорема об обращении интегралов Фурье в* $\mathscr{R}^2$ *утверждает, что, если* φ *(абсолютно) интегрируема на всей плоскости, то* $\mathbf{X}$ *имеет ограниченную и непрерывную плотность, задаваемую формулой*

$$f(x_1,x_2)=\frac{1}{(2\pi)^2}\int_{-\infty}^{+\infty}\int e^{-i(x_1\zeta_1+x_2\zeta_2)}\varphi(\zeta_1,\zeta_2)\,d\zeta_1\,d\zeta_2. \tag{7.12}$$

г) *Двумерное распределение Коши.* Мы намерены показать, что двумерный аналог распределения Коши задается плотностью

$$f(x_1,x_2)=\frac{1}{2\pi\sqrt{(t^2+x^2+y^2)^3}}. \tag{7.13}$$

Характеристическая функция для этой плотности равна

$$\varphi(\zeta_1,\zeta_2)=e^{-t\sqrt{\zeta_1^2+\zeta_2^2}}, \tag{7.14}$$

откуда видно, что распределение (7.13) обладает основными свойствами распределения Коши; в частности, оно *устойчиво*: если $\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(n)}$ — взаимно независимые векторы с плотностью (7.13), то их среднее арифметическое $(\mathbf{X}^{(1)}+\ldots+\mathbf{X}^{(n)})/n$ имеет ту же самую плотность.

Простейший способ проверить, что (7.14) является характеристической функцией плотности (7.13) [а также проверить, что (7.13) есть плотность вероятности], состоит в следующем. Мы покажем, что пара функций f и φ удовлетворяет соотноше-

нию (7.12). В полярных координатах (7.5) интеграл в (7.12) принимает вид [1])

$$-\frac{1}{(2\pi)^2}\frac{d}{dt}\int_{-\pi}^{\pi}d\theta\int_{0}^{\infty}e^{-\rho\,|t+ir\cos(\theta-s)|}\,d\rho=$$

$$=-\frac{1}{(2\pi)^2}\frac{d}{dt}\int_{-\pi}^{\pi}\frac{d\theta}{t+ir\cos(\theta-s)}. \tag{7.15}$$

Так как величина интеграла не зависит от s, мы можем принять $s=0$ и вычислить его вещественную часть

$$2\int_{-1/2\pi}^{1/2\pi}\frac{t}{t^2+r^2\cos^2\theta}\,d\theta=\frac{2\pi}{\sqrt{t^2+r^2}} \tag{7.16}$$

(используется стандартная подстановка $t\cdot\operatorname{tg}\theta=y\sqrt{t^2+r^2}$). Дифференцирование по t показывает теперь, что величина (7.15) равна $f(x_1, x_2)$, как и утверждалось. ▶

§ 8*. Две характеризации нормального распределения

Мы начнем с известной теоремы, высказанной П. Леви и доказанной в 1936 г. Г. Крамером. К сожалению, ее доказательство опирается на теорию аналитических функций и потому несколько выходит за рамки нашей трактовки теории характеристических функций.

Теорема 1. *Пусть* $\mathbf{X}_1$ *и* $\mathbf{X}_2$ *— независимые случайные величины, сумма которых распределена нормально. Тогда* $\mathbf{X}_1$ *и* $\mathbf{X}_2$ *имеют нормальное распределение.*

Другими словами, нормальное распределение может быть разложено только на нормальные компоненты. Доказательство будет опираться на лемму, представляющую самостоятельный интерес.

Лемма. *Пусть* F *— такое распределение вероятностей, что*

$$f(\eta)=\int_{-\infty}^{+\infty}e^{\eta^2x^2}F\{dx\}<\infty \tag{8.1}$$

---

[1]) Дифференцирование по t введено, чтобы устранить множитель ρ, который усложнил бы вычисления. Этот прием часто бывает полезен.

*) В этом параграфе разбирается специальная тема. Результаты в дальнейшем не используются.

при некотором $\eta>0$. *Тогда его характеристическая функция (рассматриваемая как функция комплексного аргумента* ζ*) является целой функцией. Если* $\varphi(\zeta)\neq 0$ *для всех комплексных* ζ, *то* F — *нормальное распределение.*

Доказательство леммы. При любых комплексных ζ и действительных x, η имеем $|x\zeta|\leqslant \eta^2x^2+\eta^{-2}|\zeta|^2$. Поэтому интеграл, определяющий φ, сходится при всех комплексных ζ и

$$|\varphi(\zeta)|\leqslant e^{\eta^{-2}|\zeta|^2}\cdot f(\eta). \tag{8.2}$$

Это значит, что φ есть целая функция порядка не более 2, и если такая функция не имеет нулей, то $\log\varphi(\zeta)$ является многочленом не выше второй степени [1]). Следовательно, $\varphi(\zeta)=e^{-\frac{1}{2}a\zeta^2+ib\zeta}$, где a и b — некоторые, возможно комплексные числа. Но φ — характеристическая функция, и поэтому $-i\varphi'(0)$ равно математическому ожиданию, а $-\varphi''(0)$ — второму моменту соответствующего распределения. Отсюда видим, что b вещественно и $a\geqslant 0$, так что распределение F нормально. ▶

Доказательство теоремы 1. Не ограничивая общности, мы можем предположить, что случайные величины $\mathbf{X}_1$ и $\mathbf{X}_2$ центрированы своими медианами. Тогда

$$\mathbf{P}\{|\mathbf{X}_1+\mathbf{X}_2|>t\}\geqslant\frac{1}{2}\mathbf{P}\{|\mathbf{X}_1|>t\}. \tag{8.3}$$

Теперь с помощью обычного интегрирования по частям [см. гл. V, 6] находим, что

$$f(\eta)\leqslant\eta^2\int_0^\infty x\cdot e^{\eta^2x^2}[1-F(x)+F(-x)]\,dx. \tag{8.4}$$

Поэтому функции f_k, соответствующие $\mathbf{X}_k$, удовлетворяют неравенству $f_k(\eta)\leqslant 2f(\eta)$. Как мы уже видели, отсюда вытекает, что характеристические функции φ_1 и φ_2 определены для всех комплексных ζ. Так как $\varphi_1(\zeta)\varphi_2(\zeta)=e^{(-1/2)a\zeta^2+ib\zeta}$, то ни φ_1, ни φ_2 не могут обращаться в нуль. Поэтому $\mathbf{X}_1$ и $\mathbf{X}_2$ нормальны. ▶

Мы переходим к доказательству другого характеристического свойства нормального распределения, о котором уже шла речь в гл. III, 4.

---

[1]) См., например, Hille E., Analytic Function Theory Boston, 1962, т. II, стр. 199 (теорема Адамара о факторизации). Если не привлекать теорию целых функций, то можно рассуждать и так: $\psi=\log\varphi$ регулярна во всей плоскости, тогда из неравенства (8.2) вытекает, что $|\mathrm{Re}\,\psi(\zeta)|\leqslant C|\zeta|^2$ для больших $|\zeta|$. Отсюда и из усиленной теоремы Лиувилля (там же, стр. 193) выводим требуемое заключение.

Теорема 2. *Пусть* $\mathbf{X}_1$ *и* $\mathbf{X}_2$ — *независимые случайные величины, и пусть*

$$\mathbf{Y}_1 = a_{11}\mathbf{X}_1 + a_{12}\mathbf{X}_2, \quad \mathbf{Y}_2 = a_{21}\mathbf{X}_1 + a_{22}\mathbf{X}_2. \tag{8.5}$$

Если $\mathbf{Y}_1$ *и* $\mathbf{Y}_2$ *взаимно независимы, то или все четыре случайные величины распределены нормально, или же преобразование* (8.5) *тривиально в том смысле, что либо* а) $\mathbf{Y}_1 = a\mathbf{X}_1$, $\mathbf{Y}_2 = b\mathbf{X}_2$, *либо* б) $\mathbf{Y}_1 = a\mathbf{X}_2$, $\mathbf{Y}_2 = b\mathbf{X}_1$.

Доказательство. При специальном предположении, что случайные величины $\mathbf{X}_j$ имеют непрерывные плотности, теорема была доказана в гл. III, 4. Доказательство опиралось на решение функционального уравнения [гл. III, (4.3)]. Мы покажем теперь, что характеристические функции φ_j случайных величин $\mathbf{X}_j$ удовлетворяют функциональному уравнению того же типа. Мы покажем сначала, что достаточно рассматривать действительные характеристические функции. При этом мы продемонстрируем пользу теоремы 1.

Примеры. а) *Редукция к случаю симметричных распределений.* Введем случайные величины $\mathbf{X}_1^-$ и $\mathbf{X}_2^-$, которые не зависят друг от друга и от $\mathbf{X}_j$ и которые распределены как $-\mathbf{X}_1$ и $-\mathbf{X}_2$ соответственно. Линейное преобразование (8.5) переводит симметризованные случайные величины ${}^0\mathbf{X}_j = \mathbf{X}_j + \mathbf{X}_j^-$ в пару $({}^0\mathbf{Y}_1, {}^0\mathbf{Y}_2)$ *симметричных* и независимых случайных величин. Если теорема верна для таких случайных величин, то величины ${}^0\mathbf{X}_j$ нормальны и по теореме 1 $\mathbf{X}_j$ также нормальны.

б) *Функциональные уравнения.* В силу независимости $\mathbf{Y}_1$ и $\mathbf{Y}_2$ совместная характеристическая функция для $(\mathbf{Y}_1, \mathbf{Y}_2)$ должна представляться в форме произведения соответствующих характеристических функций, т. е.

$$\mathbf{E}\left(e^{i(\zeta_1\mathbf{Y}_1+\zeta_2\mathbf{Y}_2)}\right) = \mathbf{E}\left(e^{i\zeta_1\mathbf{Y}_1}\right)\mathbf{E}\left(e^{i\zeta_2\mathbf{Y}_2}\right). \tag{8.6}$$

Подставляя сюда выражение (8.5), мы получаем следующее соотношение между характеристическими функциями для $\mathbf{X}_1$ и $\mathbf{X}_2$:

$$\varphi_1(a_{11}\zeta_1 + a_{21}\zeta_2)\,\varphi_2(a_{12}\zeta_1 + a_{22}\zeta_2) = \\ = \varphi_1(a_{11}\zeta_1)\,\varphi_2(a_{12}\zeta_1)\,\varphi_1(a_{21}\zeta_2)\,\varphi_2(a_{22}\zeta_2). \tag{8.7}$$

Теперь (8.7) совпадает с соотношением (4.3) из гл. III (с тем лишь исключением, что a_{12} и a_{21} меняются местами). Для действительных четных функций лемма из гл. III, 4 утверждает, что в невырожденном случае $\varphi_j(\zeta) = e^{-a_j\zeta^2}$. Отсюда следует, что $\mathbf{X}_j$ нормальны.

§ 9. Задачи

1. Выведите из неравенства (1.7) (без всяких вычислений), что для любой характеристической функции φ

$$|\varphi(\zeta)|^2 \leqslant 1 - \frac{1-|\varphi(2\zeta)|}{4} \leqslant e^{-\frac{1}{4}(1-|\varphi(2\zeta)|)}. \tag{9.1}$$

2. Пусть $\varphi = u + iv$ — характеристическая функция. Покажите, что

$$u^2(\zeta) \leqslant \frac{1}{2}(1+u(2\zeta)). \tag{9.2}$$

Последнее в свою очередь влечет неравенство

$$|\varphi(\zeta)|^2 \leqslant \frac{1}{2}(1+|\varphi(2\zeta)|). \tag{9.3}$$

Указание. Для (9.2) используйте неравенство Шварца, для (9.3) рассмотрите характеристические функции вида $e^{ia\zeta}\varphi(\zeta)$.

3. Докажите (обозначения те же, что и выше), что

$$|\varphi(\zeta_2)| - \varphi(\zeta_1)|^2 \leqslant 2[1-u(\zeta_2-\zeta_1)].$$

При $\zeta_2 = -\zeta_1$ отсюда следует (1.7).

4. Из элементарных формул выведите (без явного использования интегрирования), что характеристическая функция φ плотности $\frac{1}{\pi}\frac{1-\cos x}{x^2}$ отличается лишь постоянным множителем от функции $2|\zeta|-|\zeta+1|-|\zeta-1|$. Отсюда вытекает, что $\varphi(\zeta) = 1-|\zeta|$ для $|\zeta| \leqslant 1$.

5. Из характеристической функции для плотности $ae^{-a|x|}$ получите дифференцированием по a новую характеристическую функцию. Используйте этот результат — покажите, что свертка данного распределения с собой имеет плотность $(1/4)ae^{-a|x|}(1+a|x|)$.

6. Покажите, что $\varphi(\zeta) = e^{t(e^{-|\zeta|}-1)}$ есть характеристическая функция обобщенного распределения Пуассона, порожденного распределением Коши. Найдите его плотность для $x>0$.

7. Исходя из п. 10 таблицы на стр. 574, покажите, что $2\pi^2 x(\operatorname{sh} x)^{-1}$ есть плотность с характеристической функцией $\frac{2}{1+\operatorname{ch}(\pi\zeta)}$. (*Указание.* Используйте задачу 6 из гл. II, 9.)

8 [1]). Пусть $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины с распределениями F и G и характеристическими функциями φ и γ соответственно.

---

[1]) Комбинируя (9.4) с теоремой, указанной в сноске на стр. 197, получаем следующий критерий (принадлежащий Хинчину): функция ω *является характеристической функцией унимодального распределения тогда и только тогда, когда* $\omega(\zeta) = \int_0^1 \varphi(\zeta/x)\,dx$, *где* φ — *характеристическая функция.*

Покажите, что *произведение* **XY** *имеет характеристическую функцию*

$$\int_{-\infty}^{+\infty} \gamma\left(\frac{\zeta}{x}\right) F\{dx\} = \int_{-\infty}^{+\infty} \varphi\left(\frac{\zeta}{x}\right) G\{dx\}. \tag{9.4}$$

9. В теореме непрерывности 3.2 и ее следствии достаточно предполагать, что $\varphi(\zeta) = \lim \varphi_n(\zeta)$ существует при всех ζ и что φ непрерывна *в нуле* (тогда φ автоматически непрерывна всюду). *Эквивалентное* условие состоит в том, что сходимость равномерна в некоторой окрестности нуля.

10. Если $\{\varphi_n\}$ — такая последовательность характеристических функций, что $\varphi_n(\zeta) \to 1$ при $-\delta < \zeta < \delta$. Тогда $\varphi_n(\zeta) \to 1$ при всех ζ.

11. *Обобщение утверждения, обратного к* (4.7). Рассматривая распределения $\frac{1}{m_{2k}} x^{2k} F\{dx\}$ (если они существуют), докажите по индукции, что распределение F имеет конечный момент порядка m_{2r} тогда и только тогда, когда существует $2r$-я производная характеристической функции φ в нуле.

12. Пусть f — плотность вероятности с *положительной и интегрируемой* характеристической *функцией*. Тогда f имеет единственный максимум в нуле. Если вторая производная f'' существует, то

$$f(0) > f(x) > f(0) - \frac{x^2}{2} f''(0);$$

аналогичное разложение верно и для первых $2r$ членов формулы Тейлора [заметьте, что f — четная функция и, следовательно, $f^{(2k+1)}(0) = 0$].

13. Пусть g — четная плотность со строго положительной характеристической функцией γ.

Тогда

$$g_a(x) = \frac{g(x)[1 - \cos ax]}{1 - \gamma(a)}$$

есть плотность вероятности с характеристической функцией

$$\gamma_a(\zeta) = \frac{2\gamma(\zeta) - \gamma(\zeta + a) - \gamma(\zeta - a)}{2[1 - \gamma(a)]}. \tag{9.5}$$

При $a \to \infty$ *имеем* $\gamma_a \to \gamma$, *но* g_a *не стремится к* g. Это показывает, что в теореме непрерывности для плотностей условие (3.10) существенно.

14. Если γ — *произвольная* положительная характеристическая функция, то γ_a (см. 9.5) — также характеристическая функция.

15. Пусть φ — действительная характеристическая функция с непрерывной второй производной φ''. Тогда [исключая случай $\varphi(\zeta) = 1$ для всех ζ]

$$\psi(\zeta) = \frac{1 - \varphi(\zeta)}{\zeta^2} \cdot \frac{2}{|\varphi''(0)|}$$

будет характеристической функцией. Ей соответствует четная плотность f_2, определяемая для $x > 0$ формулой

$$\frac{2}{|\varphi''(0)|} \int_x^\infty [1 - F(t)]\, dt.$$

Распространите это утверждение на моменты высших порядков.

16. Пусть f — четная плотность с характеристической функцией φ. Для $x>0$ положим

$$g(x)=\int_x^\infty \frac{f(s)\,ds}{s}, \quad g(-x)=g(x).$$

Тогда g — снова четная плотность, и ее характеристической функцией будет

$$\gamma(\xi)=\frac{1}{\xi}\int_0^\xi \varphi(s)\,ds.$$

17. Если u — плотность, то и

$$f(x)=\sum c_k 2^k u(2^k x)$$

(при $c_k>0, \Sigma c_k=1$) — также плотность. (Если u равна нулю вне $\overline{-1,1}$, то лишь конечное число членов отлично от нуля и вопрос о сходимости не возникает.) Обозначим через φ характеристическую функцию для f. Покажите, что при надлежащем выборе постоянных c_k

$$\int |\varphi(\xi)|^n\,d\xi=\infty, \quad n=1, 2, \ldots,.$$

Указание. Вспомните, что $\left(\sum c_k p_k\right)^n \geqslant \sum c_k^n p_k^n$ при $p_k \geqslant 0$. Можно взять $c_k \sim 1/k^2$.

18. Докажите центральную предельную теорему (формула 4.4 из гл. VIII)) для сумм случайного числа слагаемых методом характеристических функций.

19. Выведите из теоремы 6.1 центральную предельную теорему для плотностей, аналогичную теореме 5.2.

20. Пусть $\mathbf{X}_k$ — независимые случайные величины, такие, что $\mathbf{X}_k$ принимают значения ± 1 и $\pm\sqrt{k}$ с вероятностями, равными $(1/2)(1-k^{-1})$ и $1/2k^{-1}$ соответственно. Покажите, что ни при каких нормирующих константах распределение величины $\mathbf{S}_n/a_n$ не сходится к $\mathfrak{N}$.

21. *Обобщенная центральная предельная теорема.* Предположим, что распределение F_k симметрично.

а) Слегка изменяя доказательство теоремы 6.1, покажите, что *если при любом* $t>0$

$$\sum_{k=1}^n \int_{|x|>ta_n} F_k\{dx\}\to 0, \quad a_n^2\sum_{k=1}^n \int_{|x|\leqslant ta_n} x^2 F_k\{dx\}\to 1, \tag{9.6}$$

то распределение величины $\mathbf{S}_n/a_n$ *стремится к* $\mathfrak{N}$.

б) Используя доказательство теоремы 6.2, покажите, что *эти условия становятся также необходимыми* в предположении $\mathbf{X}_k/a_n \xrightarrow{p} 0$ для $k=1, \ldots, n$. (Если $a_n\to\infty$ монотонно, то достаточно требовать $\mathbf{X}_n/a_n\xrightarrow{p}0$.)

22. *Продолжение.* Для того чтобы существовали нормирующие константы a_n, удовлетворяющие условиям (9.6), необходимо и достаточно, чтобы

для некоторой последовательности $t_n \to \infty$

$$\sum_{k=1}^{n} \int_{|x|>t_n} F_k\{dx\} \to 0, \quad \frac{1}{t_n^2} \sum_{k=1}^{n} \int_{|x|<t_n} x^2 F_k\{dx\} \to \infty.$$

В последнем случае можно взять

$$a_n^2 = \sum_{k=1}^{n} \int_{|x|<t_n} x^2 F_k\{dx\}.$$

(Обычно этот критерий применяется без затруднений.)

23. *Проблема моментов в* $\mathscr{R}^2$. Пусть $\mathbf{X}_1$ и $\mathbf{X}_2$ — две случайные величины с совместным распределением F. Положим $A_k = \mathbf{E}(|\mathbf{X}_1|^k + \mathbf{E}(|\mathbf{X}_2|^k)$. Покажите, что F однозначно определяется своими моментами, если $\overline{\lim} k^{-1} A_k^{1/k} < \infty$.

Указание: Используйте критерий единственности (4.15) и сноску 2 на стр. 597. Оцените k-й момент линейной комбинации $\mathbf{X}_1$ и $\mathbf{X}_2$ в терминах A_k.

24. *Вырожденное двумерное распределение.* Пусть φ — *одномерная* характеристическая функция и a_1, a_2 — произвольные константы. Покажите, что $\varphi(a_1\zeta_1 + a_2\zeta_2)$ как функция от ζ_1 и ζ_2 является *двумерной* характеристической функцией пары $(\mathbf{X}_1, \mathbf{X}_2)$, для которой тождественно

$$a_2\mathbf{X}_1 = a_1\mathbf{X}_2.$$

Сформулируйте обратное утверждение. Рассмотрите специальный случай $a_2 = 0$.

25. Пусть $\mathbf{X}, \mathbf{Y}, \mathbf{U}$ — взаимно независимые случайные величины с характеристическими функциями φ, γ, ω. Покажите, что произведение $\varphi(\zeta_1)\gamma(\zeta_2)\omega(\zeta_1+\zeta_2)$ представляет собой двумерную характеристическую функцию, и выразите соответствующий случайный вектор через $\mathbf{X}, \mathbf{Y}, \mathbf{U}$.

Указание. Рассмотрите трехмерную характеристическую функцию.
Ответ: $(\mathbf{U}+\mathbf{X}, \mathbf{U}+\mathbf{Y})$.

ГЛАВА XVI[1])

АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ, СВЯЗАННЫЕ С ЦЕНТРАЛЬНОЙ ПРЕДЕЛЬНОЙ ТЕОРЕМОЙ

Вопросы, рассматриваемые в настоящей главе, технически довольно трудны. Они группируются вокруг двух проблем. Одна — вывод оценок для точности центральной предельной теоремы и улучшение этих результатов за счет построения асимптотических разложений. Другая проблема — уточнение центральной предельной теоремы для далеких от математического ожидания значений аргумента, для которых классические формулировки мало содержательны.

Чтобы облегчить подход к важным теоремам и объяснить основные идеи, мы выделяем сначала случай одинаково распределенных слагаемых (§ 1—6). Методы, развитые в § 6 для больших отклонений, не опираются на содержание первых пяти параграфов, в которых излагается теория, существенно использующая два приема: прямую оценку абсолютно сходящихся интегралов Фурье и сглаживание. Мы различаем две главные идеи, рассматривая сначала асимптотическое разложение для плотностей, хотя это и приводит к некоторым повторениям и уменьшает изящество построений.

Кульминационный пункт главы — теорема Берри — Эссеена (§ 5). Метод сглаживания, описанный в § 3, был впервые использован Берри в доказательстве этой теоремы. Сейчас употребляется бесчисленное множество приемов сглаживания. Заметим, к сожалению, что долгая и примечательная история предмета этой главы привела к печальному обстоятельству: случайности исторического развития все еще влияют на подход к отдельным вопросам. Возникшее в результате разнообразие средств и избыток созданных ad hoc методов привели к тому, что эта область прославилась своей хаотичностью.

Систематическое использование метода Берри и современных неравенств позволяет, к счастью, придать всей теории удивительную цельность и простоту [2]).

---

[1]) Эта глава посвящена специальным вопросам и может быть опущена при первом чтении.

[2]) Наилучшим введением в асимптотические разложения может служить книга Г. Крамера (1962). В ней содержатся теоремы § 2 и 4 для одинаково распределенных слагаемых и несколько более сильный вариант теорем 7.2 и 7.3. Материал § 1—5 изложен и в монографии Гнеденко и Колмогорова (1954).

§ 1. Обозначения

Через F мы будем обозначать всюду (кроме последнего параграфа, где рассматривается случай неодинаково распределенных величин) одномерное распределение вероятностей с характеристической функцией φ. Момент порядка k (если он существует) будет обозначаться μ_k;

$$\mu_k = \int_{-\infty}^{+\infty} x^k F\{dx\}. \tag{1.1}$$

Предположим, что $\mu_1 = 0$. Положим, как обычно, $\mu_2 = \sigma^2$. Распределение, полученное после нормировки n-кратной свертки F, обозначим F_n, т. е.

$$F_n(x) = F^{n*}(x\sigma\sqrt{n}). \tag{1.2}$$

Плотность F_n (если она существует) обозначим f_n.

За исключением § 6 (посвященного большим отклонениям), нам придется иметь дело с функциями вида

$$u(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\zeta x} v(\zeta)\, d\zeta, \tag{1.3}$$

к которым мы постоянно будем применять очевидную оценку

$$|u(x)| \leqslant \frac{1}{2\pi}\int_{-\infty}^{+\infty} |v(\zeta)|\, d\zeta. \tag{1.4}$$

Как u, так и v будут интегрируемыми. Если u — плотность, то v будет соответствующей характеристической функцией. Для упрощения формулировок примем следующее

Соглашение. *Функция v в* (1.3) *будет называться преобразованием Фурье для u, а правая часть неравенства* (1.4) — *нормой Фурье для u.*

Как всегда, нормальная плотность обозначается

$$n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \tag{1.5}$$

Ее преобразование Фурье — характеристическая функция $e^{(-1/2)\zeta^2}$. Повторным дифференцированием получаем тождество

$$\frac{d^k}{dx^k} n(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\zeta x}(-i\zeta)^k e^{-\frac{1}{2}\zeta^2} d\zeta, \tag{1.6}$$

верное для $k=1, 2, \ldots$. Очевидно, что левая часть имеет вид

$$\frac{d^k}{dx^k}\,\mathfrak{n}(x)=(-1)^k H_k(x)\,\mathfrak{n}(x), \tag{1.7}$$

где H_k — многочлен степени k. Многочлены H_k называют *многочленами Эрмита* [1]).

В частности,

$$H_1(x)=x, \quad H_2(x)=x^2-1, \quad H_3(x)=x^3-3x. \tag{1.8}$$

Характеристическое свойство многочленов H_k состоит, следовательно, в том, что *преобразование Фурье для* $H_k(x)\,\mathfrak{n}(x)$ *равно* $(i\zeta)^k e^{(-1/2)\zeta^2}$.

§ 2. Асимптотические разложения для плотностей

Центральную предельную теорему для плотностей (теорема 5.2 из гл. XV) можно значительно усилить, если допустить конечность моментов высшего порядка. Важную роль играет предположение [2]) о том, что

$$\int_{-\infty}^{+\infty} |\varphi(\zeta)|^\nu\, d\zeta < \infty \tag{2.1}$$

при некотором $\nu \geqslant 1$. Доказательство, данное в гл. XV, 5, можно суммировать, грубо говоря, следующим образом: разность $u_n=f_n-\mathfrak{n}$ имеет преобразование Фурье

$$v_n(\zeta)=\varphi^n\left(\frac{\zeta}{\sigma\sqrt{n}}\right)-e^{(-1/2)\zeta^2}. \tag{2.2}$$

Интеграл от $|v_n(\zeta)|$ стремится к нулю в силу двух причин. При сколь угодно малом, но фиксированном δ интеграл по области $|\zeta|>\delta\sigma\sqrt{n}$ в силу (2.1) сходится к нулю. На интервале $|\zeta|<\delta\sigma\sqrt{n}$ подинтегральное выражение $|v_n|$ мало из-за поведения φ вблизи нуля. Последнее связано лишь с тем, что $\mu_1=0$ и $\mu_2=\sigma^2$. Если существуют моменты более высокого порядка, то мы можем использовать большее число членов разложения функции φ по формуле Тейлора и можем получить таким путем более точную информацию относительно скорости сходимости $f_n \to \mathfrak{n}$. К сожалению, все вычисления заметно усложняются, если взять более чем три члена асимптотического

[1]) Иногда многочленами Чебышева — Эрмита. Единой терминологии нет. Встречаются различные нормирующие множители, а вместо $e^{(-1/2)x^2}$ берут e^{-x^2}

[2]) В связи с этим условием см. примеры (5а, б) из гл. XV и задачу 17. гл. XI, 9.

разложения. Поэтому мы рассмотрим отдельно простейший и наиболее важный частный случай.

Теорема 1. *Допустим, что момент* μ_3 *существует и что для некоторого* $\nu \geqslant 1$ *функция* $|\varphi|^\nu$ *интегрируема. Тогда при* $n \geqslant \nu$ *существует плотность* f_n *и при* $n \to \infty$

$$f_n(x) - n(x) - \frac{\mu_3}{6\sigma^3\sqrt{n}}(x^3 - 3x)\,n(x) = o\left(\frac{1}{\sqrt{n}}\right) \tag{2.3}$$

равномерно относительно x.

Доказательство. По теореме обращения для интегралов Фурье [гл. XV, (3.3)] при $n \geqslant \nu$ левая часть равенства (2.3) определена и имеет норму Фурье

$$N_n = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left|\varphi^n\left(\frac{\zeta}{\sigma\sqrt{n}}\right) - e^{(-1/2)\zeta^2} - \frac{\mu_3}{6\sigma^3\sqrt{n}}(i\zeta)^3 e^{(-1/2)\zeta^2}\right| d\zeta. \tag{2.4}$$

Фиксируем произвольно $\delta > 0$. Так как φ^n — характеристическая функция распределения, имеющего плотность, то $|\varphi(\zeta)| < 1$ при $\zeta \neq 0$ и $\varphi(\zeta) \to 0$ при $\zeta \to \infty$ (лемма 4 из гл. XV, 1, и лемма 3 из гл. XV, 4). Поэтому существует $q_\delta < 1$ с тем свойством, что $|\varphi(\zeta)| < q_\delta$ для $|\zeta| \geqslant \delta$. Часть интеграла (2.4), взятая по области $|\zeta| > \delta\sigma\sqrt{n}$, меньше

$$q_\delta^{n-\nu}\int_{-\infty}^{+\infty}\left|\varphi\left(\frac{\zeta}{\sigma\sqrt{n}}\right)\right|^\nu d\zeta + \int_{|\zeta|>\delta\sigma\sqrt{n}} e^{(-1/2)\zeta^2}\left(1 + \left|\frac{\mu_3\zeta^3}{\sigma^3}\right|\right) d\zeta \tag{2.5}$$

и, таким образом, стремится к нулю быстрее любой степени числа $1/n$.

Полагая для краткости

$$\psi(\zeta) = \ln\varphi(\zeta) + \frac{1}{2}\sigma^2\zeta^2, \tag{2.6}$$

имеем

$$N_n = \frac{1}{2\pi}\int_{|\zeta|<\delta\sigma\sqrt{n}} e^{(-1/2)\zeta^2}\left|\exp\left(n\psi\left(\frac{\zeta}{\sigma\sqrt{n}}\right)\right) - 1 - \frac{\mu_3}{6\sigma^3\sqrt{n}}(i\zeta)^3\right| d\zeta + o\left(\frac{1}{n}\right). \tag{2.7}$$

Оценим подинтегральное выражение, используя неравенство

$$|e^\alpha - 1 - \beta| = |(e^\alpha - e^\beta) + (e^\beta - 1 - \beta)| \leqslant \left(|\alpha - \beta| + \frac{1}{2}\beta^2\right)e^\gamma, \tag{2.8}$$

где $\gamma \geqslant \max(|\alpha|, |\beta|)$ (справедливость этого неравенства для любых действительных или комплексных α и β очевидным об-

разом доказывается заменой e^{α} и e^{β} их разложениями в степенной ряд).

При данном $\varepsilon>0$ можно выбрать $\delta>0$ таким образом, что при $|\zeta|<\delta$

$$\left|\psi(\zeta)-\frac{1}{6}\mu_3(i\zeta)^3\right|<\varepsilon\sigma^3|\zeta|^3 \tag{2.9}$$

и

$$|\psi(\zeta)|<\frac{1}{4}\sigma^2\zeta^2, \quad \left|\frac{1}{6}\mu_3(i\zeta)^3\right|\leqslant\frac{1}{4}\sigma^2\zeta^2. \tag{2.10}$$

В самом деле, благодаря существованию μ_3 найдется окрестность нуля, в которой ψ трижды дифференцируема и (2.9) оказывается лишь формой записи разложения по формуле Тейлора.

Неравенство (2.8) показывает, что при нашем выборе δ подинтегральное выражение в (2.7) меньше, чем

$$e^{-\frac{1}{4}\zeta^2}\left(\frac{\varepsilon}{\sqrt{n}}\zeta^2+\frac{\mu_3^2}{72n}\zeta^6\right). \tag{2.11}$$

В силу произвольности ε имеем $N_n=o(1/\sqrt{n})$, так что (2.3), действительно, верно (см. задачу 3). ▶

Точно такие же рассуждения приводят к асимптотическим разложениям при наличии большего числа моментов. Однако их члены нельзя выразить простыми явными формулами. Поэтому мы отложим пока точное описание полиномов, входящих в эти разложения.

Теорема 2. *Предположим, что моменты* $\mu_3, \ldots, \mu_r$ *существуют и что для некоторого* $\nu\geqslant 1$ *функция* $|\varphi|^{\nu}$ *интегрируема. Тогда при* $n\geqslant\nu$ *существует* f_n *и при* $n\to\infty$

$$f_n(x)-n(x)-n(x)\sum_{k=3}^{r}n^{(-1/2)k+1}P_k(x)=o(n^{(-1/2)r+1}) \tag{2.12}$$

равномерно относительно x.

Здесь P_k *— многочлен с действительными коэффициентами, зависящими только от* $\mu_1, \ldots, \mu_k$ *(но не от* n *и* r *или других характеристик распределения* F).

Первые два члена имеют вид

$$P_3=\frac{\mu_3}{6\sigma^3}H_3, \quad P_4=\frac{\mu_3^2}{72\sigma^6}H_3+\frac{\mu_4-3\sigma^4}{24\sigma^4}H_4, \tag{2.13}$$

где H_k — полиномы Эрмита, определенные в (1.7). Разложение (2.12) называется (точнее, называлось) разложением Эджворта для f_n.

Доказательство. Ниже мы используем обозначение (2.6). Если p — многочлен с *действительными* коэффициентами $p_1, p_2, \ldots$, то

$$f_n - \mathfrak{n} - n\sum p_k H_k \tag{2.14}$$

имеет норму Фурье

$$N_n = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{(-1/2)\zeta^2}\left|\exp\left(n\psi\left(\frac{\zeta}{\sigma\sqrt{n}}\right)\right) - 1 - p(i\zeta)\right| d\zeta. \tag{2.15}$$

Теорема будет доказана построением подходящих полиномов p (чтобы не делать обозначения громоздкими, мы не будем писать индекс n).

Начнем с оценки подинтегрального выражения. Путь будет таким же, как и в предыдущем доказательстве. Разница лишь в том, что для аппроксимации ψ будут взяты члены формулы Тейлора со степенями аргумента до r-й включительно. Это аппроксимирующее выражение мы обозначим $\zeta^2\psi_r(\zeta)$. Таким образом, ψ_r — многочлен степени $r-2$ с $\psi_r(0)=0$. Он однозначно определяется требованием

$$\psi(\zeta) - \zeta^2\psi_r(\zeta) = o(|\zeta|^r), \qquad \zeta \to 0.$$

Положим теперь

$$p(\zeta) = \sum_{k=1}^{r-2}\frac{1}{k!}\left[\zeta^2\psi_r\left(\frac{\zeta}{\sqrt{n}}\right)\right]^k. \tag{2.16}$$

Тогда $p(i\zeta)$ будет многочленом с *действительными* коэффициентами, зависящими от n. С другой стороны, при фиксированном ζ p будет многочленом относительно $1/\sqrt{n}$, коэффициенты которого суть полиномы относительно ζ, $\mu_2, \ldots, \mu_r$ (их можно явно вычислить). Как и в предыдущем доказательстве, очевидно, что часть интеграла (2.15), взятая по области $|\zeta| > \delta\sigma\sqrt{n}$, стремится к нулю быстрее любой степени числа $1/n$. Рассмотрим поэтому подинтегральное выражение при $|\zeta| < \delta\sigma\sqrt{n}$. Чтобы оценить его, мы воспользуемся [вместо (2.8)] неравенством

$$\left|e^{\alpha} - 1 - \sum_1^{r-2}\beta^k\right| \leqslant |e^{\alpha} - e^{\beta}| + \left|e^{\beta} - 1 - \sum_1^{r-2}\beta^k\right| \leqslant$$

$$\leqslant e^{\gamma}\left(|\alpha-\beta| + \frac{1}{(r-1)!}|\beta|^{r-1}\right), \tag{2.17}$$

верным при $|\alpha| < \gamma$ и $|\beta| < \gamma$.

Подберем δ по аналогии с (2.9) таким образом, чтобы при $|\zeta|<\delta$ выполнялось неравенство

$$|\psi(\zeta)-\zeta^2\psi_r(\zeta)|\leqslant\varepsilon\sigma^r|\zeta|^r. \tag{2.18}$$

Мы можем предположить также (учитывая, что коэффициент при ζ в ψ_r равен $i^3\mu_3/6$), что при $|\zeta|<\delta$ и $a>1+|\mu_3|$

$$|\psi_r(\zeta)|<a|\zeta|<\frac{1}{4}\sigma^2. \tag{2.19}$$

Наконец, мы потребуем, чтобы при $|\zeta|<\delta$

$$|\psi(\zeta)|<\frac{1}{4}\sigma^2\zeta^2. \tag{2.20}$$

Тогда при $|\zeta|<\delta\sigma\sqrt{n}$ подинтегральное выражение в (2.15) меньше, чем

$$e^{-\frac{1}{4}\zeta^2}\left(\frac{\varepsilon|\zeta|^r}{n^{\frac{1}{2}r-1}}+\frac{a^{r-1}}{(r-1)!}\cdot\frac{|\zeta|^{3(r-1)}}{n^{\frac{1}{2}r-\frac{1}{2}}}\right). \tag{2.21}$$

В силу произвольности ε имеем $N_n=O\left(n^{-\frac{1}{2}r+1}\right)$.

Теперь мы должны определить действительные коэффициенты p_k, зависящие от n, такие, чтобы левая часть (2.12) была $o\left(n^{(-1/2)r+1}\right)$ равномерно по x. При фиксированном ζ правая часть (2.16) является многочленом относительно $1/\sqrt{n}$ Упорядочивая ее по возрастающим степеням $1/\sqrt{n}$, мы получаем разложение типа (2.12), с той лишь особенностью, что суммирование может идти и по индексам k, большим r. Но слагаемые, содержащие степени $1/n^k$ с $k>{}^1/_2r-1$, можно отбросить, и мы приходим к желаемому разложению (2.12).

Таким образом, явное вычисление полиномов P_k можно провести следующим образом. Полином ψ_r степени $r-2$ однозначно определяется из формулы Тейлора

$$\ln\varphi(\zeta)=\zeta^2\left[-\frac{1}{2}+\psi_2(\zeta)\right]+o(|\zeta|^r), \tag{2.22}$$

справедливой в окрестности нуля. Расположим слагаемые в (2.16) *соответственно степеням $1/\sqrt{n}$ Обозначим коэффициент при $n^{(-1/2)k+1}$ через $q_k(i\zeta)$. Тогда P_k — такой многочлен, что преобразование Фурье для $u(x)P_k(x)$ равно $e^{(-1/2)\zeta^2}q_k(i\zeta)$.*

§ 3. Сглаживание

Если выполняется условие (2.1), то, интегрируя асимптотическое разложение для плотностей f_n, получаем аналогичное асимптотическое разложение для распределений F_n. Чтобы справиться с более общим случаем, мы должны будем идти обход-

ным путем. Для оценки расхождения $F_n - \mathfrak{N}$ или какой-либо аналогичной функции Δ мы сначала применяем основанные на анализе преобразований Фурье методы предыдущего параграфа к некоторой аппроксимации функции Δ (обозначим ее ${}^T\Delta$). Затем мы прямыми методами оцениваем разность ${}^T\Delta - \Delta$. В этом параграфе мы дадим основные способы такой оценки.

Пусть V_T — распределение вероятностей с плотностью

$$v_T(x) = \frac{1}{\pi}\,\frac{1-\cos Tx}{Tx^2} \tag{3.1}$$

и характеристической функцией ω_T. Для $|\zeta| \leqslant T$ имеем

$$\omega_T(\zeta) = 1 - \frac{|\zeta|}{T}, \tag{3.2}$$

но эта явная форма не играет существенной роли. Важно лишь то, что $\omega_T(\zeta)$ *обращается в нуль* при $|\zeta| \geqslant T$. Последнее обстоятельство снимает все вопросы о сходимости.

Нас будут интересовать верхние границы для $|F_n - \mathfrak{N}|$ или, более общим образом, для функций вида $\Delta_n = F_n - G_n$. Эти функции мы будем аппроксимировать их свертками с V_T. Положим вообще ${}^T\Delta = V_T * \Delta$. Другими словами, для данной функции Δ мы определяем

$${}^T\Delta(t) = \int_{-\infty}^{+\infty} \Delta(t-x)\, v_T(x)\, dx. \tag{3.3}$$

Если Δ ограничена и непрерывна, то ${}^T\Delta \to \Delta$ при $T \to \infty$. Наша основная задача состоит в том, чтобы оценить максимум функции $|\Delta|$ через максимум функции $|{}^T\Delta|$.

Лемма 1. *Пусть F — распределение вероятностей и G — такая функция, что $G(-\infty) = 0$, $G(\infty) = 1$ и $|G'(x)| \leqslant m < \infty$. Положим*

$$\Delta(x) = F(x) - G(x) \tag{3.4}$$

и

$$\eta = \sup_x |\Delta(x)|, \quad \eta_T = \sup_x |{}^T\Delta(x)|. \tag{3.5}$$

Тогда

$$\eta_T \geqslant \frac{\eta}{2} - \frac{12m}{\pi T}. \tag{3.6}$$

Доказательство. Функция Δ обращается в нуль на бесконечности и пределы справа и слева ($\Delta(x_0+)$ и $\Delta(x_0-)$) существуют в каждой точке. Отсюда ясно видно, что найдется такая точка x_0, что или $|\Delta(x_0+)| = \eta$ или $|\Delta(x_0-)| = \eta$. Допустим, что $\Delta(x_0) = \eta$. Так как F не убывает, а модуль производной

от G не превосходит m, то

$$\Delta(x_0+s)\geqslant\eta-ms \qquad \text{для} \quad s>0. \tag{3.7}$$

Полагая

$$h=\frac{\eta}{2m}, \quad t=x_0+h, \quad x=h-s, \tag{3.8}$$

имеем

$$\Delta(t-x)\geqslant\frac{\eta}{2}+mx. \qquad \text{для} \quad |x|\leqslant h \tag{3.9}$$

Оценим интеграл (3.3), используя (3.9) и неравенство $\Delta(t-x)\geqslant-\eta$ при $|x|>h$. Интеграл от линейного члена обращается в нуль в силу симметрии. Далее плотность v_T приписывает области $|x|>h$ массу $\leqslant 4/(\pi Th)$. Поэтому

$${}^T\Delta(x_0)\geqslant\frac{\eta}{2}\left[1-\frac{4}{\pi Th}\right]-\eta\,\frac{4}{\pi Th}=\frac{\eta}{2}-\frac{6\eta}{\pi Th}. \tag{3.10}$$

Вспоминая определение h (3.8), видим, что из (3.10) вытекает (3.6). Доказательство закончено. ▶

В наших приложениях G будет иметь производную g, совпадающую или с нормальной плотностью u или с начальным отрезком асимптотического разложения, указанного в предыдущем параграфе. Во всяком случае преобразование Фурье γ, соответствующее g, имеет две непрерывные производные и $\gamma(0)=1$, $\gamma'(0)=0$. Очевидно, что преобразование Фурье свертки ${}^Tg=V_T*g$ равно $\gamma\omega_T$. Аналогично по теореме обращения преобразований Фурье (теорема 3 из гл. XV, 3) произведение $\varphi\omega_T$ является преобразованием Фурье плотности ${}^Tf$ распределения V_T*F. Иными словами,

$${}^Tf(x)-{}^Tg(x)=\frac{1}{2\pi}\int_{-T}^{T}e^{-i\zeta x}[\varphi(\zeta)-\gamma(\zeta)]\,\omega_T(\zeta)\,d\zeta. \tag{3.11}$$

Покажем, что это эквивалентно

$${}^T\Delta(x)=\frac{1}{2\pi}\int_{-T}^{T}e^{-i\zeta x}\frac{\varphi(\zeta)-\gamma(\zeta)}{-i\zeta x}\,\omega_T(\zeta)\,d\zeta. \tag{3.12}$$

Прежде всего, так как $\varphi(0)=\gamma(0)=1$ и $\varphi'(0)=\gamma'(0)=0$, дробь под знаком интеграла является непрерывной функцией, обращающейся в нуль, так что вопрос о сходимости не возникает. По лемме Римана — Лебега (лемма 4 из гл. XV, 4) правая часть стремится к нулю при $|x|\to\infty$. Формальное дифференцирование равенства (3.12) приводит к (3.11), так что правая и левая части равенства (3.12) могут отличаться только на постоянную. Так как оба выражения обращаются в нуль на бесконечности, они должны быть тождественными.

Комбинируя верхнюю границу для η_T, содержащуюся в (3.12), с (3.6), получаем верхнюю границу для η, а именно

$$|F(x)-G(x)|\leqslant\frac{1}{\pi}\int_{-T}^{T}\left|\frac{\varphi(\zeta)-\gamma(\zeta)}{\zeta}\right|d\zeta+\frac{24m}{\pi T}. \tag{3.13}$$

Так как это неравенство служит основой оценок в последующих двух параграфах, суммируем условия его применимости.

Лемма 2. *Пусть F — распределение вероятностей с нулевым математическим ожиданием и характеристической функцией φ. Предположим, что $F-G$ обращается в нуль на бесконечности и что G имеет производную g с $|g|\leqslant m$. Допустим, что γ — преобразование Фурье для g — дважды дифференцируемо и $\gamma(0)=1$, $\gamma'(0)=0$. Тогда для всех x и $T\geqslant 0$ выполняется* (3.13).

Мы дадим два *независимых друг от друга* применения этого неравенства. В следующем параграфе мы выведем интегральный вариант теорем § 2. В § 5 мы выведем знаменитое неравенство Берри — Эссеена для разности $F_n-\mathfrak{N}$.

§ 4. Асимптотические разложения для распределений

Формальный аналог теоремы 2.1, получаемый интегрированием равенства (2.3), утверждал бы, что

$$F_n(x)-\mathfrak{N}(x)-\frac{\mu_3}{6\sigma^3\sqrt{n}}(1-x^2)\,n(x)=o\left(\frac{1}{\sqrt{n}}\right). \tag{4.1}$$

Однако если F — решетчатое распределение, то скачки распределений F_n имеют порядок $1/\sqrt{n}$ [см. (5.12) в гл. XV], так что (4.1) не может быть верным для решетчатых распределений. Мы покажем, что (4.1) верно для *всех* других распределений и что ценой небольших видоизменений можно охватить и решетчатый случай. Для удобства мы эти два случая разделяем.

Теорема 1. *Если F — нерешетчатое распределение и если μ_3 существует, то* (4.1) *выполняется равномерно относительно x.*

Доказательство. Положим

$$G(x)=\mathfrak{N}(x)-\frac{\mu_3}{6\sigma^3\sqrt{n}}(x^2-1)\,n(x). \tag{4.2}$$

Тогда G удовлетворяет условиям предшествующей леммы с

$$\gamma(\zeta)=e^{(-1/2)\zeta^2}\left[1+\frac{\mu_3}{6\sigma^3\sqrt{n}}(i\zeta)^3\right]. \tag{4.3}$$

Мы применим теперь неравенство (3.13) с $T=a\sqrt{n}$, где постоянная a выбирается столь большой, что $24|G'(x)|<\varepsilon a$ для всех x.

Тогда

$$|F_n(x)-G(x)|\leqslant \int_{-a\sqrt{n}}^{a\sqrt{n}} \left|\frac{\varphi^n\left(\frac{\zeta}{\sigma\sqrt{n}}\right)-\gamma(\zeta)}{\zeta}\right| d\zeta+\frac{\varepsilon}{\sqrt{n}}. \quad (4.4)$$

Числитель совпадает с подинтегральным выражением в (2.4). Поэтому мы можем без изменений повторить доказательство теоремы 2.1. Так как распределение F предполагается нерешетчатым, то в соответствии с леммой 4 из гл. XV, 1, максимум $|\varphi(\zeta)|$ при $\delta\leqslant|\zeta|\leqslant a\sigma$ строго меньше единицы. Как и в § 2, отсюда выводим, что интеграл по области $|\zeta|>\delta\sigma\sqrt{n}$ стремится к нулю быстрее, чем любая степень числа $1/n$. При $|\zeta|<\delta\sigma\sqrt{n}$ мы можем использовать оценку (2.11) для числителя подинтегрального выражения в (4.4). Таким образом, правая часть в (4.4) меньше $1000\varepsilon/\sqrt{n}+o(1/n)$. Отсюда вытекает утверждение теоремы. ▶

Это рассуждение неприменимо к решетчатым распределениям, потому что их характеристические функции периодичны (и, стало быть, интеграл по области $|\zeta|>\delta\sigma\sqrt{n}$ не стремится к нулю). Но утверждения теоремы можно спасти естественным изменением формулировки, учитывающим решетчатый характер распределения. Функция распределения F является ступенчатой функцией, но мы будем аппроксимировать ее непрерывной функцией распределения $F^{\#}$ с полигональным графиком.

Пусть F сосредоточено на решетке точек b, $b\pm h$, $b\pm 2h$, ..., и пусть h является его *шагом*. Свертка распределения F с распределением, равномерным на интервале $-h/2<x<h/2$, определяется формулой

$$F^{\#}(x)=\frac{1}{h}\int_{-h/2}^{h/2} F(x-y)\,dy. \quad (4.5)$$

График функции распределения $F^{\#}$ представляет собой полигон с вершинами в точках $b+\left(n+\frac{1}{2}\right)h$. В этих точках $F^{\#}(x)=F(x)$, в то время как в точках решетки

$$F^{\#}(x)=\frac{1}{2}(F(x)+F(x-)). \quad (4.6)$$

Мы будем называть $F^{\#}$ *полигональным приближением к* F.

Теорема 2[1]). *Для решетчатых распределений асимптотическая формула* (4.1) *остается верной, если заменить* F_n *ее полигональным приближением.*

Это означает, в частности, что для точек решетки формула (4.1) выполняется, если F_n заменить средним арифметическим $(1/2)[F_n(x)+F_n(x-)]$.

Доказательство. Полигональное приближение $F_n^{\#}$ к F_n является сверткой распределения F_n с распределением, равномерным на $-h/2\sigma\sqrt{n} < x < h/2\sigma\sqrt{n}$. Обозначим через $G^{\#}$ свертку распределения G с тем же самым равномерным распределением. В (4.5) интеграл от линейного члена ряда Тейлора исчезает. Отсюда видно, что если $|F''(x)|<m$, то $|F^{\#}(x)-F(x)|<mh^2$. По этой причине разность $G^{\#}-G$ равномерно по x имеет порядок $O(1/n)$. Следовательно, достаточно доказать, что $F_n^{\#}-G^{\#}=o\left(1/\sqrt{n}\right)$. Но для этой разности неравенство (4.4) принимает вид

$$|F_n^{\#}(x)-G^{\#}(x)|\leqslant \int_{-a\sqrt{n}}^{a\sqrt{n}} \left|\frac{\varphi^n\left(\frac{\zeta}{\sigma\sqrt{n}}\right)-\gamma(\zeta)}{\zeta}\right| \cdot \left|\frac{\sin\frac{h\zeta}{2\sigma\sqrt{n}}}{\frac{h\zeta}{2\sigma\sqrt{n}}}\right| d\zeta+\frac{\varepsilon}{\sqrt{n}} \tag{4.7}$$

(новый множитель под интегралом — характеристическая функция равномерного распределения). Подинтегральное выражение здесь не больше, чем в (4.4). Остается лишь показать, что при фиксированном δ часть интеграла в (4.7), взятая по области $|\zeta|>\delta\sigma\sqrt{n}$, равна $o\left(1/\sqrt{n}\right)$. На самом деле она имеет порядок $O(1/n)$.

Для доказательства достаточно проверить, что

$$\int_{\delta}^{a/\sigma} \left|\varphi^n(y)\sin\frac{hy}{2}\right| dy = O\left(\frac{1}{n}\right). \tag{4.8}$$

По лемме 4 из гл. XV, 1, под интегралом стоит четная функция с периодом $2\pi/n$ и потому (4.8) будет верно, если верно,

---

[1]) Несмотря на различие в форме, эта теорема эквивалентна теореме Эссеена, в которой к (4.1) добавляется периодическая функция. Доказательство этой теоремы использует теорию рядов Фурье и сопровождается длинными вычислениями (см., например, книгу Гнеденко и Колмогорова (1954)). Внешне более изящно было бы скомбинировать формулировки теоремы 1.2 и их доказательства, допуская сглаживание произвольным равномерным распределением.

что

$$\int_0^{\pi/h} |\varphi^n(y)|\, y\, dy = O\left(\frac{1}{n}\right).$$

Последнее соотношение справедливо, так как в некоторой окрестности нуля $|\varphi(y)| < e^{(-1/4)\sigma^2 y^2}$, а за ее пределами подинтегральное выражение убывает быстрее любой степени числа $1/n$. ▶

Перейдем теперь к асимптотическим разложениям более высокого порядка. Доказательство соотношения (4.1) отличается от доказательства теоремы 2.1 только применением сглаживания, которое делает пределы интегрирования в (4.4) конечными. Точно такое же сглаживание можно применить и к разложениям, указанным в теореме 2.2, но при этом, очевидно, чтобы получился остаток $O(n^{(-1/2)r+1})$, мы должны взять $T \sim an^{1/2 r-1}$. Здесь возникает следующая трудность. Доказательство соотношения (4.1) использует тот факт, что максимум функции $\varphi(\zeta/\sigma\sqrt{n})$ | на интервале $\delta\sigma\sqrt{n} < |\zeta| < T$ меньше единицы. Для нерешетчатых распределений это всегда верно при $T = a\sqrt{n}$, но может быть неверно при T, растущем как более высокая степень числа n. Поэтому для вывода асимптотических разложений более высокого порядка мы вынуждены предположить, что

$$\limsup_{|\zeta|\to\infty} |\varphi(\zeta)| < 1; \tag{4.9}$$

для нерешетчатых распределений отсюда следует, что максимум q^δ функции $|\varphi(\zeta)|$ для $|\zeta| > \delta$ меньше 1. При дополнительном условии (4.9) метод, детально объясненный для (4.1), применим *без всяких изменений* к разложениям теоремы 2.2, что приводит к следующей теореме.

Теорема 3. *Если существуют моменты* $\mu_3, \ldots, \mu_r$ *и выполняется* (4.9), *то при* $n \to \infty$

$$F_n(x) - \mathfrak{N}(x) - \mathfrak{n}(x) \sum_{k=3}^{r} n^{(-1/2)k+1} R_k(x) = o(n^{(-1/2)r+1}) \tag{4.10}$$

равномерно по x. *Здесь* R_k — *многочлен, зависящий только от* $\mu_1, \ldots, \mu_r$, *но не от* n *или* r (*или других характеристик распределения* F).

Разложение (4.10) — это лишь интегральный вариант формулы (2.12), и полиномы R_k связаны с P_k из (2.12)

соотношением

$$\mathfrak{n}(x)\,P_k(x)=\frac{d}{dx}\,\mathfrak{n}(x)\,R_k(x). \tag{4.11}$$

Следовательно, нет нужды повторять их конструкцию. Заметим, что (4.9) выполняется для любого несингулярного распределения F.

Формула (4.10) называется *разложением Эджворта* для F. Если F имеет конечные моменты всех порядков, возникает искушение перейти к пределу при $r\to\infty$. Однако соответствующий бесконечный ряд не сходится ни при каком n (Г. Крамер показал, что он сходится при любом n тогда и только тогда, когда функция $e^{(1/4)x^2}$ интегрируема относительно F). Формальное разложение Эджворта не следует смешивать с разложением *по полиномам* Эрмита

$$F(x)-\mathfrak{N}(x)=\sum_{k=1}^{\infty}c_kH_k(x)\,e^{(-1/2)x^2}, \tag{4.12}$$

которое сходится при условии, что F имеет конечное математическое ожидание, но не имеет никакого вероятностного смысла. Например, даже если возможно разложить каждую функцию F_n в ряд типа (4.12), то тем не менее коэффициенты не дают нам никакой информации относительно скорости сходимости $F_n\to\mathfrak{N}$.

§ 5. Теорема Берри — Эссеена

Приводимая ниже важная теорема была доказана (различными путями) А. С. Берри (1941) и К. Г. Эссееном (1942).

Теорема 1. *Если F имеет нулевое математическое ожидание и*

$$\rho=\int_{-\infty}^{+\infty}|x|^3F\{dx\}<\infty, \tag{5.1}$$

то при всех x и n

$$|F_n(x)-\mathfrak{N}(x)|<\frac{33}{4}\cdot\frac{\rho}{\sigma^3\sqrt{n}}. \tag{5.2}$$

Поразительно здесь то, что верхняя граница зависит только от отношения ρ/σ^3 и не зависит от каких-либо других характеристик распределения. Теорема 4.1 дает лучший асимптотический результат, но скорость сходимости зависит от данного распределения F более сложным образом. Эти теоремы подчеркивают различные аспекты одной и той же ситуации, а их доказательства различаются лишь деталями. Ввиду важности последней теоремы мы повторим прежнее рассуждение, опираясь единственно на методы § 3. Множитель $^{33}/_4$ в (5.2) можно

уменьшить, но мы не будем пытаться получить лучшие оценки (или истинную границу C этого множителя) [1]).

Доказательство.[2]). При $G=\mathfrak{N}$ основное неравенство (3.13) принимает форму

$$|F_n(x)-\mathfrak{N}(x)|\leqslant\frac{1}{\pi}\int\limits_{-T}^{T}\left|\varphi^n\left(\frac{\zeta}{\sigma\sqrt{n}}\right)-e^{(-1/2)\zeta^2}\right|\frac{1}{|\zeta|}\,d\zeta+\frac{24}{T\pi\sqrt{2\pi}}. \tag{5.3}$$

В интервале, окружающем нуль, в котором $|1-\varphi(\zeta)|<1$, имеем

$$-\ln\varphi(\zeta)-\frac{1}{2}\sigma^2\zeta^2=\left(1-\varphi(\zeta)-\frac{1}{2}\sigma^2\zeta^2\right)+\sum_{k=2}^{\infty}\frac{1}{k}(1-\varphi(\zeta))^k. \tag{5.4}$$

Из неравенства между моментами следует, что $\sigma^6<\rho^2$, откуда

$$|1-\varphi(\zeta)|\leqslant\frac{1}{2}\sigma^2\zeta^2<\frac{1}{2}\quad\text{для}\quad|\zeta|<\frac{\sigma^2}{\rho}. \tag{5.5}$$

Первый член в правой части равенства (5.4) не превосходит $^1/_6\rho|\zeta|^3$ [см. формулу (4.1) из гл. XV]. Сравнивая ряд (5.4) с геометрической прогрессией со знаменателем $^1/_2$, мы получаем при $|\zeta|<\sigma^2/\rho$

$$\left|\ln\varphi(\zeta)+\frac{1}{2}\sigma^2\zeta^2\right|\leqslant\frac{1}{6}\rho|\zeta|^3+\frac{1}{4}\sigma^4\zeta^4\leqslant\frac{5}{12}\rho|\zeta|^3. \tag{5.6}$$

Пусть $T=(\sigma^3/\rho)/\sqrt{n}$. Используя (5.6) и неравенство $|e^t-1|\leqslant|t|e^{|t|}$, видим, что подинтегральное выражение в (5.3) не превосходит

$$e^{(-1/2)\zeta^2}\frac{\exp\left(\frac{5}{12}\frac{\rho}{\sigma^3}\frac{|\zeta|^3}{\sqrt{n}}\right)-1}{\zeta}\leqslant$$

$$\leqslant\frac{5}{12}\frac{\rho}{\sigma^3\sqrt{n}}\zeta^2\exp\left(-\frac{1}{2}\zeta^2+\frac{5}{12}\frac{\rho}{\sigma^3}\frac{|\zeta|^3}{\sqrt{n}}\right)\leqslant\frac{5}{12}\frac{\rho}{\sigma^3\sqrt{n}}\zeta^2e^{(-1/12)\zeta^2}. \tag{5.7}$$

Интеграл $\zeta^2e^{(-1/12)\zeta^2}$ равен $6\sqrt{12\pi}$. Поэтому из (5.3) вытекает

$$\frac{\sigma^3\sqrt{n}}{\rho}|F_n(x)-\mathfrak{N}(x)|\leqslant\frac{1}{\sqrt{2\pi}}\left(5\sqrt{6}+\frac{24}{\pi}\right)<\frac{12{,}5+8}{\sqrt{2\pi}}<8.2, \tag{5.8}$$

что доказывает (5.2). ▶

[1]) Берри утверждал, что $C\leqslant1{,}88$, но его вычисления содержали ошибку. Эссеен показал, что $C\leqslant7{,}59$. Неопубликованные вычисления дают неравенство $C\leqslant2{,}9$ (Эссеен, 1956) и $C\leqslant2{,}05$ (Д. Л. Воллес, 1958). См. задачу 6. [В. М. Золотарев показал, что $C\leqslant0{,}9051$ (1966). — *Прим. перев.*].

[2]) Лучшее и более простое доказательство намечено в задаче 5. Преимущество настоящего доказательства состоит в том, что оно применимо *без изменений* к случаю разнораспределенных слагаемых (см. теорему 7.1).

§ 6. Большие отклонения

При очень больших x как $F_n(x)$, так и $\mathfrak{N}(x)$ становятся близкими к единице и наши приближенные формулы становятся бесполезными. В действительности нам нужна *относительная* ошибка аппроксимации $1-F_n$ величиной $1-\mathfrak{N}$. Часто желательно бывает использовать соотношение

$$\frac{1-F_n(x)}{1-\mathfrak{N}(x)} \to 1, \tag{6.1}$$

когда и x и n стремятся к бесконечности. Это соотношение не может быть верно всегда. Так, для симметричного биномиального распределения числитель обращается в нуль для всех $x > \sqrt{n}$. Мы покажем, однако, что (6.1) верно, если x меняется вместе с n таким способом, что $xn^{-1/6} \to 0$. При этом предположим, что интеграл

$$f(\zeta) = \int_{-\infty}^{+\infty} e^{\zeta x} F\{dx\} \tag{6.2}$$

конечен при всех ζ из некоторого интервала $|\zeta| < \zeta_0$. [Это равносильно тому, что характеристическая функция $\varphi(\zeta) = f(i\zeta)$ аналитична в некоторой окрестности нуля; но мы предпочитаем иметь дело с вещественной функцией f.]

Теорема 1 [1]). *Если интеграл* (6.2) *сходится при* $|\zeta| < \zeta_0$ *и* x *меняется вместе с* n *так, что* $x = o(n^{1/6})$, *то верно* (6.1).

Заменяя x на $-x$, мы получаем двойственную теорему для левого конца. Теорема представляется достаточно общей, чтобы охватить все «практически интересные ситуации», но метод доказательства приводит к значительно более сильным результатам.

Для доказательства перейдем от функции f к ее логарифму. Равенство

$$\psi(\zeta) = \ln f(\zeta) = \sum \frac{\psi_k}{k!} \zeta^k \tag{6.3}$$

определяет функцию, аналитическую в некоторой окрестности нуля. Коэффициент ψ_k зависит только от $\mu_1, \ldots, \mu_k$ и носит название *семиинварианта порядка* k распределения F. В общем случае $\psi_1 = \mu_1$, $\psi_2 = \sigma^2$, $\ldots$. Мы предполагаем $\mu_1 = 0$, и потому $\psi_1 = 0$, $\psi_2 = \sigma^2$, $\psi_3 = \mu_3$, $\ldots$.

---

[1]) Обобщение на случай разнораспределенных слагаемых дано в теореме 7.4.

Доказательство основано на применении сопряженных распределений [1]). Распределение V, сопряженное к F, задается равенством

$$V\{dx\} = e^{-\psi(s)}e^{sx}F\{dx\}, \tag{6.4}$$

где параметр s изменяется в интервале сходимости функции ψ. Функция

$$v(\zeta) = \frac{f(\zeta+s)}{f(s)} \tag{6.5}$$

играет по отношению к V ту же роль, что и f по отношению к исходному распределению F. В частности, дифференцируя (6.5), видим, что V имеет математическое ожидание $\psi'(s)$ и дисперсию $\psi''(s)$.

Идею доказательства можно объяснить следующим образом. Как легко видеть или из (6.4), или из (6.5), распределения F^{n*} и V^{n*} также связаны соотношением (6.4). При этом нормирующая константа $e^{-\psi(s)}$ заменяется на $e^{-n\psi(s)}$. Распределение V^{n*} имеет математическое ожидание $n\psi'(s)$. Поэтому значение $F^{n*}(n\psi'(s)) = F_n(\sqrt{n}\,\psi'(s)/\sigma)$ может быть хорошо аппроксимировано с помощью центральной предельной теоремы. Выбирая s таким образом, чтобы

$$\frac{1}{\sigma}\psi'(s) = \frac{x}{\sqrt{n}}, \tag{6.6}$$

мы придем к (6.1) и другим аналогичным утверждениям. Идея доказательства проста, но вычисления затемняют ее.

Доказательство теоремы 1. По определению

$$1 - F^{n*}(n\psi'(s)) = e^{n\psi(s)} \int_{n\psi'(e)}^{\infty} e^{-sy}V^{n*}\{dy\}. \tag{6.7}$$

Разобьем доказательство на две части.

а) Сначала мы вычислим величину A_s, получаемую заменой V^{n*} нормальным распределением с тем же математическим ожиданием $n\psi'(s)$ и с той же самой дисперсией $n\psi''(s)$. Стандартная подстановка $y = n\psi'(s) + t\sqrt{n\psi''(s)}$ дает

$$A_s = e^{n[\psi(s)-s\psi'(s)]} \cdot \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-ts\sqrt{n\psi''(s)}-(1/2)t^2}\,dt. \tag{6.8}$$

[1]) Они уже были использованы в теории восстановления (гл. XI, 6) и при изучении случайных блужданий (гл. XII, 4).

Дополняя показатель под интегралом до полного квадрата, получаем

$$A_s = e^{n[\psi(s)-s\psi'(s)]} \cdot \rho\left(s\sqrt{n\psi''(s)}\right), \tag{6.9}$$

где

$$\rho(t) = e^{(1/2)t^2}[1-\mathfrak{N}(t)]. \tag{6.10}$$

Из **1**, гл. VII, (1.8) мы знаем, что при $t>0$

$$\frac{1}{t}-\frac{1}{t^3} < \sqrt{2\pi}\,\rho(t) < \frac{1}{t}. \tag{6.11}$$

Отсюда выводим, дифференцируя (6.10), что

$$0 \leqslant -\sqrt{2\pi}\,\rho'(t) < \frac{1}{t^2}. \tag{6.12}$$

Перейдем теперь к выбору параметра s. При достаточно малых значениях правой части уравнение (6.6) имеет единственный корень s, *который является аналитической функцией от* $x/\sqrt{n}$, *такой, что*

$$s \sim \frac{x}{\sigma\sqrt{n}}, \qquad \frac{x}{\sqrt{n}} \to 0. \tag{6.13}$$

Рассмотрим A_s как аналитическую функцию от x. Вблизи нуля мы имеем

$$n[\psi(x)-s\psi'(s)] = -\frac{1}{2}n\sigma^2 s^2 + \ldots,$$

и, следовательно,

$$A_s = e^{(-1/2)x^2}\rho(\bar{x})\left[1+O\left(\frac{x^3}{\sqrt{n}}\right)\right]. \tag{6.14}$$

Здесь

$$\bar{x} = s\sqrt{n\psi''(s)}. \tag{6.15}$$

Обычные вычисления показывают, что разложение разности $\bar{x}-x$ по степеням s начинается с кубического члена. Поэтому в силу (6.12) и (6.11)

$$|\rho(\bar{x})-\rho(x)| = O\left(\frac{x^3}{n}|\rho'(x)|\right) = O\left(\frac{x}{n}\right) = O\left(\frac{x^3}{n}\rho(x)\right). \tag{6.16}$$

Отсюда и из (6.14) получаем

$$A_s = [1-\mathfrak{N}(x)]\cdot\left[1+O\left(\frac{x^3}{\sqrt{n}}\right)\right]. \tag{6.17}$$

б) Если s — аналитическая функция от x, определенная равенством (6.6), то значение $1-F_n(x)$ дается формулой (6.7). Мы вычислили его приближенное значение заменой распределения V^{n*} соответствующим нормальным распределением, скажем G_3. Для s, близких к нулю, первые три момента распре-

деления V мало отличаются от соответствующих моментов распределения F, и теорема Берри — Эссеена гарантирует существование такой постоянной C, что $|V^{n*}(y) - G_s(y)| < C/\sqrt{n}$ для всех $|s| < s_0$. Интегрирование равенства (6.7) по частям показывает, что ошибка от подстановки G_s вместо V^{n*} не превосходит

$$2C\frac{1}{\sqrt{n}}\,e^{n[\psi(s)-s\psi'(s)]} = O\left(\frac{x}{\sqrt{n}}A_s\right). \tag{6.18}$$

Таким образом, мы не только доказали теорему, но и установили, что *если* $x \to \infty$ *и* $n \to \infty$, *то левая часть в* (6.1) *равна* $O(x^3/\sqrt{n})$. ▶

Доказательство можно провести без труда и в случае, когда условие $x = o(n^{1/6})$ заменяется условием $x = o(\sqrt{n})$. В самом деле, рассуждения (б) показывают, что *относительная* ошибка, возникающая при использовании нормальной аппроксимации для V^{n*}, всегда имеет $O(x/\sqrt{n})$. Нормальная аппроксимация приводит всегда к выражению (6.9). Оценка (6.16) остается верной. Мы получаем, следовательно, общую формулу

$$1 - F_n(x) = e^{n[\psi(s)-s\psi'(s)+(1/2)\psi'^2(s)]}[1 - \mathfrak{N}(x)]\cdot[1 + O(x/\sqrt{n})]. \tag{6.19}$$

Показатель можно представить степенным рядом относительно s, начинающимся с члена третьего порядка. Как и в (6.6), мы определяем теперь аналитическую функцию переменного z равенством $\psi'(s) = \sigma z$. Затем мы определяем *степенной ряд* λ соотношением

$$z^2\lambda(z) = \lambda_1 z^3 + \lambda_2 z^4 + \ldots = \psi(x) - s\psi'(s) + \frac{1}{2}\psi'^2(s). \tag{6.20}$$

Тогда верна

Теорема 2[1]). *Если в теореме* 1 *условие* $x = o(n^{1/6})$ *заменить условием* $1 < x = o(\sqrt{n})$, *то*

$$\frac{1 - F_n(x)}{1 - \mathfrak{N}(x)} = e^{x^2\lambda(x/\sqrt{n})}\left[1 + O\left(\frac{x}{\sqrt{n}}\right)\right]. \tag{6.21}$$

В частности, если x возрастает быстрее, чем $n^{1/6}$, но $x = o(n^{1/4})$, то достаточно учитывать только первый член степенного ряда λ. Мы получаем

$$\frac{1 - F_n(x)}{1 - \mathfrak{N}(x)} \sim e^{\lambda_1 x^3/\sqrt{n}}, \quad \lambda_1 = \frac{\mu_3}{6\sigma^3}. \tag{6.22}$$

---

[1]) Использование преобразования (6.4) в связи с центральной предельной теоремой принадлежит, по-видимому, Эшеру (1932). Теорема 2 доказана Г. Крамером (1938) и была затем распространена на разнораспределенные слагаемые Феллером (1943). Более новые результаты указаны в статьях В. В. Петрова [*УМН*, 9 (1954)] и В. Рихтера [Локальные теоремы для больших отклонений, *Теория вероятностей и ее применения*, 2 (1957)]. Последний рассматривает, правда, не функции распределения, а плотности. См. также теорему 4 из § 7.

Для $x=o(n^{3/10})$ имеем

$$\frac{1-F_n(x)}{1-\mathfrak{N}(x)}\sim\exp\left(\lambda_1\frac{x^3}{\sqrt{n}}+\lambda_2\frac{x^2}{n}\right),\quad \lambda_2=\frac{\sigma^2\psi_4-3\psi_3^2}{24\sigma^6} \tag{6.23}$$

и так далее. Теорема 2 интересна с общей точки зрения тем, что она вскрывает роль нормального распределения как первого приближения. [Формула (6.21) позволяет также объяснить кажущиеся загадочными явления, связанные со случаями, где неприменим закон повторного логарифма.]

§ 7. Различно распределенные слагаемые

Теория, развитая в предыдущих параграфах, легко переносится на случай различно распределенных слагаемых. Без сомнения, это верно, если не ставить целью отыскание наилучших условий. На самом деле все наши обозначения и рассуждения строились так, чтобы подготовить к решению этой задачи и потому не всегда были простейшими из возможных.

С этого момента *мы отказываемся от обозначений, принятых в* § 1. Вместо этого мы рассматриваем последовательность взаимно независимых случайных величин $\mathbf{X}_k$ с распределениями U_k. Мы предполагаем, что

$$\mathbf{E}(\mathbf{X}_k)=0,\quad \mathbf{E}(\mathbf{X}_k^2)=\sigma_k^2, \tag{7.1}$$

$$s_n^2=\sigma_1^2+\ldots+\sigma_n^2,\quad s_n\to\infty. \tag{7.2}$$

Чтобы облегчить сравнения, мы опять *обозначим через* F_n *распределение нормированной суммы* $(\mathbf{X}_1+\ldots+\mathbf{X}_n)/s_n$. Таким образом, F_n имеет тот же смысл, что и прежде, но аналитически определяется иначе:

$$F_n(x)=U_1*\ldots*U_n(s_nx). \tag{7.3}$$

Обозначим через ω_k и φ_n характеристические функции распределений U_k и F_n соответственно. Тогда

$$\varphi_n(\zeta)=\omega_1\left(\frac{\zeta}{s_n}\right)\ldots\omega_n\left(\frac{\zeta}{s_n}\right). \tag{7.4}$$

Эта функция играет роль $\varphi^n(\zeta/\sigma\sqrt{n})$. Части доказательства, опирающиеся на свойства интегралов Фурье, сохраняются без изменений. В вычислениях же с логарифмами мы должны быть уверены в том, что найдется единый интервал, где они определены. Чтобы гарантировать это, достаточно ввести некоторые условия «равномерности». Типичным примером служит

Теорема 1. (*Берри.*) *Если при* $k=1, 2, \ldots$

$$\frac{1}{\sigma_k^2}\mathbf{E}(|\mathbf{X}_k|^3)<\lambda, \tag{7.5}$$

то при всех n *и* x

$$|F_n(x)-\mathfrak{N}(x)|\leqslant\frac{33}{4}\cdot\frac{\lambda}{s_n} \tag{7.6}$$

(см. лучший результат в задаче 8).

Доказательство. Применим неравенство (5.3) с $T=s_n/\lambda$ [и, конечно, с $\varphi^n(\zeta|\sigma\sqrt{n})$, замененной на $\varphi_n(\zeta)$]. Из неравенства (7.5) вытекает, что внутри промежутка интегрирования каждая отдельная компонента $\omega_k(\zeta/s_n)$ удовлетворяет неравенствам типа (5.5), ранее имевшим место для $\varphi(\zeta/\sigma\sqrt{n})$. Поэтому никакие новые вычисления не требуются. ▶

Рассмотрим теперь разложение (4.1). Множитель $\mu_3/(\sigma^3\sqrt{n})$ соответствует коэффициенту при ζ^3 в разложении Тейлора функции $n\ln\varphi(\zeta/\sigma\sqrt{n})$. Последняя величина заменяется теперь суммой логарифмов, поэтому мы можем ожидать, что аналог разложения (4.1) имеет вид

$$F_n(x)-\mathfrak{N}(x)-\frac{\mu_3^{(n)}}{6s_n^3}(1-x^2)\,\mathfrak{n}(x)=o\left(\frac{n}{s_n^3}\right), \tag{7.7}$$

где

$$\mu_3^{(n)}=\sum_{k=1}^{n}\mathbf{E}(\mathbf{X}_k^3). \tag{7.8}$$

Мы докажем теперь справедливость равенства (7.7) при условиях, позволяющих избежать новой аргументации. Эти условия легко ослабить. В частности, (7.10) является «излишней роскошью» (см. задачи 10—13).

Теорема 2. *Пусть существуют такие константы* C *и* $q_\delta<1$, *что при всех* k

$$\mathbf{E}(|\mathbf{X}|^4)<C, \tag{7.9}$$

$$\max_{\zeta>\delta}|\omega_k(\zeta)|<q_\delta. \tag{7.10}$$

Тогда (7.7) *выполнено, причем равномерно относительно* x.

Доказательство. Напомним, что доказательство теоремы 4.1 начиналось с неравенства (4.4), которое в свою очередь представляет собой специальный случай общего неравенства (3.13). Здесь применима такая же процедура. Только чтобы получить остаточный член $o(n/s_n^3)$, мы должны взять $T=as_n^3/n$

(вместо $T=a\sqrt{n}$). Тогда отправной точкой становится неравенство

$$|F_n(x)-G(x)|\leqslant\frac{1}{\pi}\int\limits_{-as_n^3/n}^{as_n^3/n}\left|\frac{\varphi_n(\zeta)-\gamma(\zeta)}{\zeta}\right|d\zeta+\frac{\varepsilon n}{s_n^3}, \qquad (7.11)$$

где G обозначает левую часть равенства (7.7). (Здесь ε задано, а a — константа, зависящая от ε.)

Как следует из неравенств между моментами, условие (7.9) гарантирует равномерную ограниченность функций $\mathbf{E}(\mathbf{X}_k^2)$ и $\mathbf{E}(|\mathbf{X}_k|^3)$. Отсюда следует, что существует интервал $-\eta, \eta$, *не зависящий от* n, в котором $u_n(\zeta)=\log\varphi_n(\zeta)$ является функцией непрерывной вместе с четырьмя своими производными, причем она сама и ее производные ограничены равномерно по n. В этой обстановке мы можем без всяких изменений применить оценки из доказательства теоремы 4.1 [1]). ▶

Вследствие (7.9) величина n/s_n^3 по меньшей мере имеет порядок $1/s_n$, но может быть и значительно больше. Исчерпывающее изучение разложений более высоких порядков усложняется тем, что остаточный член складывается из нескольких компонент, относительная величина которых зависит от условий, наложенных на случайные величины $\mathbf{X}_k$ [2]). Мы предположим, что существуют две такие положительные постоянные c и C, что при всех j

$$c<\mathbf{E}(|\mathbf{X}|^k)<C \qquad k=1, \ldots, r+1. \qquad (7.12)$$

Возможно, что это условие излишне ограничительно. Но его большое преимущество в том, что при нем $c^2s_n^2<n<C^2s_n^2$. Таким образом, n и s_n^2 имеют один и тот же порядок, и мы можем произвести асимптотическое разложение по возрастающим степеням $1/\sqrt{n}$. На самом деле условие (7.12) дает возмож-

[1]) В прежнем доказательстве использовали только три производных функции $u=\log\varphi$, но оно было связано с выбором фиксированного интервала, на котором $|u'''(\zeta)-u'''(0)|<\varepsilon$. Условие (7.9) делает возможным подобный выбор для наших u_n и может быть ослаблено многими способами (см. задачу 9).

[2]) Например, при надлежащих ограничениях погрешность формулы (7.7) имеет вид $O(n^2s_n^{-6})+O(ns_n^{-4})$. Здесь главным может быть любое из слагаемых. В форме, предложенной Крамером, эти слагаемые соединены вместе, давая выражение $O\left(\sqrt{n}\,s_n^{-6}\sqrt{\left(\sum_1^n\mathbf{E}(\mathbf{X}_k^4)\right)^3}\right)$, которое может быть хуже, чем каждое из двух, указанных выше.

ность использовать (4.10) без изменений. Действительно, как и в предыдущем доказательстве, здесь можно определить в некоторой окрестности нуля функцию

$$\psi_n^{\#}(\zeta) = \frac{1}{n}\sum_{k=1}^{n} \ln \omega_k\left(\frac{\zeta\sqrt{n}}{s_n}\right). \tag{7.13}$$

Тогда

$$\ln \varphi_n(\zeta) = n\psi_n^{\#}\left(\frac{\zeta}{\sqrt{n}}\right), \tag{7.14}$$

т. е. мы приходим к точно такой же форме, как и в случае равных компонент. Отсюда следует, что можно повторить вычисления, которые привели к (4.10), заменяя лишь $\ln \varphi(\zeta)$ на $\psi_n^{\#}$ (и полагая $\sigma=1$). Конечно, коэффициенты полиномов R_k будут зависеть теперь не только от моментов, но и от n. Точно так же и в остаточный член войдет верхняя грань модуля r-й производной от $\psi_n^{\#}$ в окрестности нуля. Однако благодаря требованию равномерности (7.12) существует верхняя граница, не зависящая от n. Так мы приходим к следующей теореме.

Теорема 3[1]). *Допустим, что выполняются условия* (7.10) *и* (7.12). *Тогда имеет место разложение* (4.10) *с остаточным членом, равномерным относительно* x.

Полиномы R_k зависят от моментов, входящих в (7.12) и от n, при этом как функции от n они ограничены. Условие (7.10) легко ослабить (см. задачу 12). Нет нужды говорить, что подобным же образом можно получить *асимптотические разложения для плотностей.*

Предельная теорема 6.1 о вероятностях больших отклонений для последовательностей разнораспределенных случайных величин принимает следующий вид.

Теорема 4. *Предположим, что существует интервал* $\overline{-a, a}$, *на котором все характеристические функции* ω_k *аналитичны, и что*

$$\mathbf{E}(|\mathbf{X}_n|^3) \leqslant M\sigma_n^2, \tag{7.15}$$

где M *не зависит от* n.

Если x *и* n *меняются так, что* $s_n \to \infty$ *и* $x = o(\sqrt{s_n})$, *то*

$$\frac{1-F_n(x)}{1-\mathfrak{N}(x)} \to 1 \tag{7.16}$$

[*с ошибкой* $O(x^3/s_n)$ *равномерно по* x *при* $x>1$].

[1]) Это по существу теорема Крамера, но его условия были несколько более жесткими.

Доказательство. Мы можем повторить шаг за шагом доказательство § 7. Доставим себе удовольствие отметить один пункт, где требуются новые рассуждения. Определим на $-a, a$ аналитическую функцию (принимающую на этом интервале действительные значения)

$$\psi_n(s) = \frac{1}{n}\sum_{k=1}^{n} \ln \omega_k(-is). \tag{7.17}$$

В вычислениях ψ_n заменит ψ, а xs_n заменит $x\sigma\sqrt{n}$. Основное уравнение (6.6) примет форму

$$\psi_n'(s) = \frac{xs_n}{n}. \tag{7.18}$$

Мы рассмотрим x как независимую переменную, меняющуюся вместе с n, как указано в теореме. Из разложения Тейлора для ψ_n с тремя членами видно, что (7.18) имеет единственный непрерывный корень, такой, что $s = \frac{nx}{s_n} + O(s^2)$. Дальнейшие вычисления, как в § 6. ▶

§ 8. Задачи

1. Покажите, что полиномы Эрмита ортогональны с весом n, т. е.

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} H_m(x) H_n(x)\, dx = \begin{cases} 0 & \text{при } m \neq n, \\ n! & \text{при } m = n. \end{cases} \tag{8.1}$$

Обратно, если H_n — многочлен степени n с вещественными коэффициентами и выполняется (8.1), то H_n суть многочлены Эрмита, определенные формулой (1.7).

2. Если f — симметричная плотность с дисперсией σ^2 и интегрируемой характеристической функцией φ, то $\sqrt{n} f^{n*}(x) \to 1/\sqrt{2\pi\sigma}$ даже при отсутствии третьего момента. Докажите это двумя путями — выведите из теоремы 2.1 и получите независимо, используя тот же метод.

3. Если $\mu_4 < \infty$, то доказательство теоремы 2.1 приводит к остаточному члену $O(1/n)$. То же верно по отношению к теореме 4.1.

4. Лемму 3.1. можно улучшить, если F имеет плотность, ограниченную числом m. Обсудите детали.

5. Теорему Берри—Эссеена можно доказать с меньшими вычислениями, если использовать в (5.3) очевидное неравенство: при $|a| \leqslant c$ и $|b| \leqslant c$ имеем $|a^n - b^n| \leqslant n|a-b|c^{n-1}$. Примените этот метод с $T = 2\sqrt{n}\,\sigma^3/\rho$, что приводит к снижению константы с $^{33}/_4$ до $^{22}/_5$.

6. *Продолжение.* Покажите, что оценку теоремы Берри—Эссеена можно заменить на $\frac{5}{2}\frac{\rho}{\sigma^3\sqrt{n}} + \frac{C_\rho}{n}$, где C_ρ — константа, зависящая только от ρ.

7. Формула (6.7) влечет неравенство [1])

$$1 - F^{n*}(n\psi'(s)) \leqslant e^{n[\psi(s)-\psi'(s)]}.$$

8. Замените условие (7.5) теоремы Берри (§ 7) условием Линдеберга (используйте усечение).

9. Условие (7.9) в теореме 7.2 можно заменить более слабым: существуют такие постоянные a и M, что

$$\int_{|x|>a} |x|^3 U_k\{dx\} < M$$

для всех k. Аналогично можно усовершенствовать теорему 7.3 (см. сноску к теореме 7.2).

10. Условие (7.10) эквивалентно следующему: существуют такие положительные числа a и $\eta<1$, что $|\omega_k(\zeta)|<\eta$ для всех k и $\zeta>a$. [Используйте неравенство (9.1) из гл. XV. Не нужно никаких вычислений.]

11. Условие предыдущей задачи выполнено, если распределения U_k имеют плотности, которые в свою очередь имеют производные с $|u_k'| < M$.

12. Условие (7.10) теоремы 7.2 можно заменить условием: при любом δ

$$|\omega_1(\zeta) \ldots \omega_n(\zeta)| = o\left(\frac{1}{n}\right) \qquad (*)$$

равномерно для $\zeta \geqslant \delta>0$.

13. *Продолжение.* Условие (*) заведомо выполнено, если

$$\frac{1}{\log n}\sum_1^n |1-\omega_k(\zeta)| \to \infty$$

равномерно в некотором интервале $\zeta>a$.

---

[1]) Близкие неравенства получены Г. Черновым (Chernoff H., *Ann. Mach. Statist*, 23 (1952), 493—502).

ГЛАВА XVII

БЕЗГРАНИЧНО ДЕЛИМЫЕ РАСПРЕДЕЛЕНИЯ

В этой главе представлено ядро ставшей теперь классической области предельных теорем теории вероятностей. Эта область подобна водохранилищу, образовавшемуся из бесчисленных отдельных ручьев. Подготовительная работа проделана в § 1. Все дальнейшие результаты сравнительно просто вытекают из основных теорем этого параграфа. Отсюда логичный путь — прямо к схеме серий § 7. Однако мы снова выделим простейшие ситуации для того, чтобы облегчить восприятие важных вопросов.

Основные результаты этой главы были получены другими методами в гл. IX. Настоящее изложение не зависит от предшествующего и дает более детальную информацию.

Понятия безграничной делимости, устойчивости и т. д. читатель может найти в гл. VI.

§ 1. Теорема о сходимости

Пусть $\{F_n\}$ — последовательность распределений вероятностей с характеристическими функциями φ_n. Большая часть теории этой главы так или иначе связана с последовательностями функций вида

$$\psi_n(\zeta) = c_n[\varphi_n(\zeta) - 1 - i\beta_n\zeta], \tag{1.1}$$

где $c_n > 0$ и β_n — вещественная постоянная. Заметим, что $e^{c_n(\varphi_n - 1)}$ является характеристической функцией обобщенного распределения Пуассона [гл. XV, (2.4)] и e^{ψ_n} — характеристическая функция такого же распределения, но иначе центрированного. Функции типа (1.1) будем называть *квазихарактеристическими.*

В этом параграфе мы докажем важную теорему о сходимости, аналогичную теореме непрерывности для характеристических функций (теорема 2, гл. XV, 3). Особенность нового случая в том, что, как правило, $c_n \to \infty$. Это необходимо приводит к введению неограниченных мер. Однако дополнительных трудностей не возникает, так как мы будем иметь дело только с та-

кими мерами M, что $M\{I\}<\infty$ для любого конечного интервала I и что интегралы

$$M^+(x)=\int_{x^-}^{\infty}\frac{1}{y^2}M\{dy\},\quad M^-(-x)=\int_{-\infty}^{-x^+}\frac{1}{y^2}M\{dy\} \tag{1.2}$$

сходятся при всех $x>0$. Типичными примерами могут служить мера Лебега и более подходящая в данном случае мера $x^2F\{dx\}$, порождаемая распределением вероятностей F. Функция M^+ монотонна на $\overline{0,\ \infty}$ и обычно не ограничена в нуле. Она не определена на $\overline{-\infty,\ 0}$. Аналогично M^- существует только на $\overline{-\infty,\ 0}$.

Мы напомним определение 1 гл. VIII, 1: последовательность мер M_n сходится к мере M тогда и только тогда, когда $M_n\{I\}\to M\{I\}$ для любого *ограниченного* интервала непрерывности M. Суть нижеследующей теоремы в том, что сходимость последовательности (1.1) связывается с существованием такой меры M, что

$$c_nx^2F_n\{dx\}\to M\{dx\}. \tag{1.3}$$

Вне любой окрестности нуля (1.3) можно переписать в виде

$$c_nF_n\{dx\}\to\frac{1}{x^2}M\{dx\}. \tag{1.4}$$

Как и в случае собственной сходимости вероятностных распределений, мы будем требовать, чтобы «положительная масса не исчезала на бесконечности». Аналитически это сводится к требованию, чтобы (1.4) выполнялось и для неограниченных интервалов, т. е.

$$c_n[1-F_n(x)]\to M^+(x),\quad c_nF_n(-x)\to M^-(-x),\qquad x>0 \tag{1.5}$$

во всех точках непрерывности. Так как $M^+(\infty)=M^-(-\infty)=0$, то соотношение (1.5) верно тогда и только тогда, когда левые части равномерно малы при больших x. Это означает, что для любого $\varepsilon>0$ найдется $\tau>0$, такое, что

$$c_n[1-F_n(x)+F_n(-x)]<\varepsilon \tag{1.6}$$

при $x>\tau$ и при всех n.

Определение. *Назовем меру M канонической, если интеграл* (1.2) *сходится. Мы скажем, что меры $c_nx^2F_n\{dx\}$ собственно сходятся к M, если выполняются* (1.3) *и* (1.5). [*Условие* (1.5) *эквивалентно* (1.6).]

Примеры. а) Пусть F_n приписывает веса, равные $^1/_2$, каждой из двух точек $\pm 1/\sqrt{n}$. Тогда $nx^2F_n\{dx\}=F_n\{dx\}$, а $F_n\{dx\}$ сходится к вероятностной мере, сосредоточенной в нуле. Для $\beta_n=0$ мы имеем $\psi_n(\zeta)\to -^1/_2\zeta^2$. Заметьте, что предельная функция не ограничена.

б) Обозначим через F_n распределение Коши с плотностью $\frac{1}{\pi}\frac{n}{1+n^2x^2}$. Плотность меры $\pi nx^2F_n\{dx\}$ сходится к 1, и $\pi nx^2F_n\{dx\}$ собственно сходится к мере Лебега. Для $\beta_n=0$ имеем

$$\psi_n(\zeta)=\pi n\left[e^{-|\zeta|n^{-1}}-1\right]\to -\pi|\zeta|. \quad \blacktriangleright$$

Рассмотрим теперь произвольную последовательность распределений вероятностей F_n с характеристическими функциями φ_n. Выясним, при каких условиях

$$\psi_n(\zeta)\equiv c_n[\varphi_n(\zeta)-1-i\beta_n\zeta]\to\rho(\zeta) \tag{1.7}$$

при *всех* ζ. Нас будут интересовать только непрерывные предельные функции. Для краткости мы положим

$$b_n=\int_{-\infty}^{+\infty}\sin x\cdot F_n\{dx\}. \tag{1.8}$$

Теорема. *Непрерывная предельная функция* ρ *в* (1.7) *существует в том и только том случае, когда существуют каноническая мера* M *и число* b, *такие, что*

$$c_nx^2F_n\{dx\}\to M\{dx\} \tag{1.9}$$

в собственном смысле и

$$c_n(b_n-\beta_n)\to b. \tag{1.10}$$

В этом случае

$$\rho(\zeta)=\psi(\zeta)+ib\zeta, \tag{1.11}$$

где

$$\psi(\zeta)=\int_{-\infty}^{+\infty}\frac{e^{i\zeta x}-1-i\zeta\sin x}{x^2}M\{dx\}. \tag{1.12}$$

Мера M *однозначно определяется равенством* (1.12) [1]).

Пусть φ_n и c_n заданы. Из сказанного следует, что *если* $\{\psi_n\}$ *сходится при некотором выборе* β_n, *то они сходятся и при* $\beta_n=b_n$. Это не удивительно, поскольку такой выбор равносилен нормировке ψ_n условием $\operatorname{Im}\psi_n(1)=0$. Дополнительно о выборе постоянных β_n скажем в конце этого параграфа.

[1]) При фиксированном ζ подинтегральная функция считается непрерывной, принимающей значение $-^1/_2\zeta^2$ в нуле.

Пример. в) *Производные характеристических функций.* Пусть F — вероятностное распределение с характеристической функцией φ. Если F имеет математическое ожидание μ, то φ имеет производную φ' и $\varphi'(0)=i\mu$ (см. гл XV, 4). Обратное утверждение неверно, но из предшествующей теоремы вытекает, что[1]) $\varphi'(0)=i\mu$ *существует тогда и только тогда, когда при* $t\to\infty$

$$t[1-F(t)+F(-t)]\to 0 \tag{1.13}$$

и

$$\int_{-t}^{t} xF\{dx\}\to\mu. \tag{1.14}$$

Заметим, что $\varphi'(0)=i\mu$ тогда и только тогда, когда

$$\lim_{\varepsilon\to 0}\frac{\varphi(\varepsilon\zeta)-1-i\mu\varepsilon\zeta}{\varepsilon}=0. \tag{1.15}$$

Доказательство. Пусть ε пробегает последовательность $\varepsilon_1, \varepsilon_2 \ldots,\to$ $\to 0$. Тогда соотношение (1.15) можно рассматривать как частный случай (1.7) с $c_n=\varepsilon_n^{-1}$ и

$$F_n\{dx\}=F\left\{\varepsilon_n^{-1}\,dx\right\}.$$

Соответственно (1.15) выполняется тогда и только тогда, когда $\varepsilon_n^{-1}x^2F\left\{\varepsilon_n^{-1}\,dx\right\}\to 0$ в собственном смысле и

$$\varepsilon_n^{-1}\int_{-\infty}^{+\infty}\sin x\cdot F\left\{\varepsilon_n^{-1}\,dx\right\}=\varepsilon_n^{-1}\int_{-\infty}^{+\infty}\sin\varepsilon_n y\cdot F\{dy\}\to\mu \tag{1.16}$$

для любой последовательности $\varepsilon_n\to 0$, В силу (1.6) собственная сходимость имеет место тогда и только тогда, когда верно (1.13), и в этом случае (1.16), очевидно, эквивалентно (1.14). ▶

Доказательство теоремы мы разобьем на ряд лемм. Первая из них лишь формулирует определение собственной сходимости в терминах математических ожиданий. Мы будем постоянно использовать обозначение (1.7), так что

$$\psi_n(\zeta)=\int_{-\infty}^{+\infty}[e^{i\zeta x}-1-i\zeta\sin x]\,c_nF_n\{dx\}+ic_n(b_n-\beta_n)\zeta. \tag{1.17}$$

Лемма 1. *Если* $c_nx^2F_n\{dx\}\to M\{dx\}$ *в собственном смысле, то для любой непрерывной ограниченной функции* z, *для которой* $x^{-2}z(x)$ *непрерывна в нуле, имеет место*

$$c_n\int_{-\infty}^{+\infty}z(x)F_n\{dx\}\to\int_{-\infty}^{+\infty}x^{-2}z(x)M\{dx\}. \tag{1.18}$$

Доказательство. Пусть $|x|<t<\infty$ — интервал непрерывности для M. Тогда часть левого интеграла в (1.18),

[1]) Этот результат был доказан А. Зигмундом в 1947 г. (при некоторых ограничениях на гладкость φ) и в полной общности Питманом в 1956 г. Методы этих авторов различны.

взятая по этому интервалу, сходится к соответствующей части правого интеграла. В силу (1.6) можно выбрать t столь большим, что часть интегралов в (1.18), взятая по области $|x| \geqslant t$, меньше ε (при всех n). Этим доказано (1.18). ▶

Лемма 2. *Если* $\psi_n(\zeta) \to \rho(\zeta)$ *равномерно в области* $|\zeta| \leqslant \leqslant \zeta_0$, *то при* $0 < h \leqslant \zeta_0$

$$c_n \int_{-\infty}^{+\infty} \left(1 - \frac{\sin xh}{xh}\right) F_n\{dx\} \to -\frac{1}{2h} \int_{-h}^{h} \rho(\zeta)\, d\zeta. \tag{1.19}$$

Заметим, что никаких предположений относительно $\psi_n(\zeta)$ для $|\zeta| > \zeta_0$ не делается.

Доказательство. Умножая (1.17) на $1/2h$ и интегрируя по промежутку $-h \leqslant \zeta \leqslant h$, получаем (1.19). ▶

Лемма 3. *В условиях предыдущей леммы существуют каноническая мера* M *и последовательность* $n_1, n_2, \ldots \to \infty$, *такая, что* $c_{n_k} x^2 F_{n_k}\{dx\} \to M\{dx\}$ *в собственном смысле.*

Доказательство. Обозначим при любом n $M_n\{dx\} = c_n x^2 F_n\{dx\}$. Подинтегральное выражение слева в (1.19) в окрестности нуля эквивалентно $1/6 x^2 h^2$, а вне любой окрестности нуля — строго положительно. Отсюда следует, что для каждого конечного интервала I меры $M_n\{I\}$ ограничены. По теореме о выборе (теорема 2 из гл. VIII, 6) существует последовательность $\{n_k\}$, такая, что M_{n_k} сходится к предельной мере. Остается показать, что справедливо условие (1.6). Так как $\rho(0) = 0$, мы можем выбрать h столь малым, что правая часть (1.19) станет меньше $1/2\varepsilon$. В области $|xh| > 2$ подинтегральное выражение в левой части больше $1/2$, а при всех остальных x — положительно. Поэтому левая часть (1.19) не меньше, чем $\frac{1}{2} c_n [1 - F_n(2/h) + F_n(-2/h)]$, т. е. (1.6) выполняется при $x > 2/h$. ▶

Лемма 4. *Какова бы ни была каноническая мера* M, *равенство* (1.12) *определяет непрерывную функцию* ψ. *При этом различным каноническим мерам соответствуют различные функции.*

Доказательство. Интеграл в (1.12) сходится, так как подинтегральное выражение непрерывно, и его числитель ограничен, так как каждая каноническая мера удовлетворяет (1.2). Очевидно, что ψ непрерывна и при любом $h > 0$

$$\psi(\zeta) - \frac{\psi(\zeta + h) + \psi(\zeta - h)}{x^2} = \int_{-\infty}^{+\infty} e^{i\zeta x} \frac{1 - \cos xh}{x^2} M\{dx\}. \tag{1.20}$$

Справа стоит преобразование Фурье — Стилтьеса *ограниченной* меры $A_h\{dx\} = \frac{1-\cos xh}{x^2} M\{dx\}$. По теореме единственности для характеристических функций мера A_h однозначно определяется левой частью (1.20). В свою очередь мера A_h однозначно определяет меру M, за исключением, быть может, атомов в точке, где $1-\cos xh=0$. Но знание двух мер A_{h_1} и A_{h_2} с несоизмеримыми h_j определяет M однозначно. Таким образом, M однозначно определяется по ψ (см. задачи 1—2). ▶

Доказательство *теоремы.* Допустим, что $\psi_n(\zeta) \to \rho(\zeta)$ для всех ζ и что ρ — непрерывная функция. Так как функция e^{ψ_n} является характеристической, сходимость равномерна на каждом конечном интервале (теорема 2 из гл. XV, 3). Поэтому применимы леммы 2 и 3. Когда n пробегает последовательность $n_1, n_2, \ldots$, указанную в лемме 3, интегралы в (1.17) сходятся к $\psi(\zeta)$ в силу леммы 1. Мы видим, что предел имеет форму (1.11). Значение ψ (1) действительно, поэтому $b=\operatorname{Im}\rho(1)$. Таким образом, ψ и b определяются однозначно и не зависят от последовательности $\{n_k\}$. Это показывает, что сходимость ψ_n к непрерывному пределу ρ эквивалентна условиям (1.9) и (1.10) теоремы. ▶

Замечание *о других способах центрирования.* Выбор функции $\sin x$ в (1.8) и (1.12) не является единственно возможным. В самом деле, леммы 1—3 совершенно не зависят от этого выбора, а лемма 4 приложима к любой функции вида

$$\psi(\zeta) = \int_{-\infty}^{+\infty} \frac{e^{i\zeta x} - 1 - i\zeta\tau(x)}{x^2} M\{dx\}, \tag{1.21}$$

если только подинтегральное выражение непрерывно и числитель ограничен [1]). При любом приемлемом значении $\tau(x)$ доказательство теоремы осуществляется без изменения, если только $\sin x$ заменить на $\tau(x)$ в определении чисел b_n [см. (1.8)] и функций $\psi_n(\zeta)$ [см. (1.17)].

При специальных предположениях можно указать функции $\tau(x)$, которые предпочтительнее, чем $\sin x$. Так, если существуют математические ожидания m_n распределений F_n, то желательно заменить центрирующие константы b_n на m_n, т. е. иначе заменить $\sin x$ на x. При таком выборе числитель в (1.21) становится

[1]) См., например, (7.2). Центрирование с помощью $\sin x$ имеет ту отличительную черту, что каноническая форма ψ характеризуется «внутренним» свойством: ψ (1) вещественно.

неограниченным и для того, чтобы провести доказательство теоремы, мы должны усилить условие (1.6), потребовав, чтобы при всех достаточно больших t

$$c_n \int_{|x|>t} |x| F_n\{dx\} < \varepsilon. \tag{1.22}$$

При этом условии существует интеграл от $|x|^{-1}M\{dx\}$ по любому интервалу, не содержащему какой-либо окрестности нуля. В (1.18) достаточно потребовать, чтобы $|z(x)| = O(|x|)$, $x \to \pm\infty$. Из сказанного вытекает, что *если* (1.22) *выполняется, то теорема остается верной при замене* $\sin x$ *на* x *в* (1.12) *и в* (1.8). По той же причине, *если для всех достаточно малых* t

$$c_n \int_{|x|<t} |x| F_n\{dx\} < \varepsilon, \tag{1.23}$$

то $\sin x$ *можно заменить нулем.* Подинтегральное выражение в (1.21) уже не будет ограниченным в окрестности нуля, но мера M теперь такова, что интеграл от $x^{-1}M\{dx\}$ по любому конечному интервалу конечен.

§ 2. Безгранично делимые распределения

Мы по-прежнему будем использовать одни и те же описательные термины как для распределений, так и для их характеристических функций. Имея это в виду, переформулируем определение, данное в гл. VI, 3, следующим образом.

Определение. *Характеристическую функцию* ω *называют безгранично делимой тогда и только тогда, когда при каждом* n *существует характеристическая функция* ω_n, *такая, что*

$$\omega_n^n = \omega. \tag{2.1}$$

Ниже будет показано, что ничто не меняется, если заменить (2.1) формально более слабым условием

$$\varphi_n^n(\zeta) \to \omega(\zeta). \tag{2.2}$$

Соотношения подобного рода охватываются теорией предыдущего параграфа, так как оказывается, что (2.2) эквивалентно

$$n[\varphi_n(\zeta) - 1] \to \rho(\zeta). \tag{2.3}$$

Лемма. *Пусть* $\{\varphi_n\}$ *— последовательность характеристических функций. Если предел в правой части* (2.3) *является непрерывной функцией, то каждое из соотношений* (2.2) *и* (2.3) *влечет другое и* $\omega(\zeta) = e^{\rho(\zeta)}$.

Отсюда следует, в частности, что ω не имеет нулей.

Доказательство. Мы начнем с того, что напомним определение логарифма в комплексной области. Пусть $a=re^{i\theta}$ — тригонометрическое представление комплексного числа $a\neq 0$. Тогда $\log a=\log r+i\theta$, где $\log r$ — хорошо известное действительное значение логарифма положительного числа r. Аргумент θ определен только с точностью до кратных 2π, и в принципе эта неопределенность распространяется и на логарифмы. Тем не менее в любом интервале $|\zeta|<\zeta_0$, в котором $\varphi(\zeta)\neq 0$, характеристическая функция φ допускает *единственное* представление вида $\varphi(\zeta)=r(\zeta)e^{i\theta(\zeta)}$, такое, что θ *непрерывна* и $\theta(0)=0$. В таком интервале мы можем написать, не вызывая сомнений, $\log\varphi=\log r+i\theta$. Это *единственный* способ сделать $\log\varphi$ непрерывным и обращающимся в нуль в нуле. Мы будем использовать обозначение $\log\varphi$ только в указанном смысле. В любом интервале $|\zeta|<\zeta_0$, не содержащем нулей φ, непрерывная функция $\log\varphi$ вполне определена, однако определение перестает действовать, как только $\varphi(\zeta_0)=0$. Аналогичное замечание применимо и к дробным степеням φ [1]).

Перейдем к доказательству. Пусть верно (2.3). Тогда $\varphi_n(\zeta)\to 1$ при каждом фиксированном ζ и сходимость равномерна в каждом конечном интервале $|\zeta|<\zeta_0$ (теорема 2 из гл. XV, 3). Поэтому $\log\varphi_n(\zeta)$ определен при всех достаточно больших n. Так как при $|z-1|<1$

$$\log z=z-1+\frac{(z-1)^2}{2}+\ldots, \tag{2.4}$$

то (2.3) эквивалентно

$$\log\varphi_n^n(\zeta)=n\log\varphi_n(\zeta)\to\rho(\zeta). \tag{2.5}$$

Соответственно (2.3) влечет (2.2) и равенство $\omega=e^{\rho}$.

Пусть теперь выполняется (2.2). Тогда $\omega(0)=1$ и потому $\omega(\zeta)\neq 0$ в некотором интервале $|\zeta|\leqslant\zeta_0$. Так как сходимость характеристических функций равномерна в каждом конечном интервале (теорема 2 из гл. XV, 3), то при $|\zeta|\leqslant\zeta_0$ и достаточно больших n $\varphi_n(\zeta)\neq 0$. Поэтому мы можем в интервале $|\zeta|<\zeta_0$ перейти к логарифмам, откуда следует (2.3) с $\rho(\zeta)=\log\omega(\zeta)$ (причем сходимость равномерна). Теперь, используя леммы 2 и 3 § 1, мы видим, что существует последовательность

[1]) В гл. XV, 2 мы привели пример двух действительных характеристических функций, таких, что $\varphi_1^2=\varphi_2^2$. Он показывает, что при наличии у характеристической функции нулей ее квадратный корень не определен однозначно.

целых $n_1, n_2, \ldots \to \infty$, такая, что при *всех* ζ

$$n_k[\varphi_{n_k}(\zeta) - 1] \to \rho(\zeta), \tag{2.6}$$

где предельная функция непрерывна. Но отсюда вытекает, что $\varphi_n^n \to e^\rho$ по крайней мере для n, пробегающих последовательность $n_1, n_2, \ldots$. Предел $\rho = \log \omega$ в (2.6) — один и тот же для любой последовательности $\{n_k\}$, так что (2.2) верно при любом стремлении n к бесконечности. ▶

Отметим, что (2.3) — специальный случай соотношения (1.7) предшествующей теоремы. Поэтому ρ имеет вид

$$\rho(\zeta) = \psi(\zeta) + ib\zeta, \tag{2.7}$$

где

$$\psi(\zeta) = \int_{-\infty}^{+\infty} \frac{e^{i\zeta x} - 1 - i\zeta \sin x}{x^2} M\{dx\} \tag{2.8}$$

и M — каноническая мера. Таким образом, безгранично делимая характеристическая функция необходимо имеет вид $\omega = e^\rho$, где ρ определяется по (2.7) и (2.8). Докажем теперь, что, обратно, функция такого типа есть безгранично делимая характеристическая функция.

Допустим сперва, что каноническая мера M сосредоточена в области $|x| > \delta$. В этом случае мы можем представить M в виде

$$M\{dx\} = cx^2 G\{dx\},$$

где G — распределение вероятностей с характеристической функцией γ. Тогда $e^{c(\gamma-1)}$ будет характеристической функцией обобщенного пуассоновского распределения и потому безгранично делима. Она отличается от e^ρ только связанным с центрированием множителем $e^{ib\zeta}$. Следовательно, e^ρ безгранично делима, если мера M сосредоточена на $|x| > \delta > 0$. Обозначим теперь через M_δ меру, получаемую из M отбрасыванием массы, расположенной внутри интервала $|x| < \delta$, и через $\psi_\delta(\zeta)$ — соответствующий интеграл (2.8). Пусть $\sigma^2 \geqslant 0$ — масса, приписываемая распределением M точке 0. Тогда при $\delta \to 0$

$$-\frac{1}{2}\sigma^2\zeta^2 + \psi_\delta(\zeta) \to \psi(\zeta).$$

Левая часть есть логарифм характеристической функции. По теореме непрерывности то же самое верно и для предельной функции ψ. Далее $n^{-1}\psi$ снова имеет форму (2.8) с M, замененным на $n^{-1}M$, так что $n^{-1}\psi$ является логарифмом характеристической функции. Соответственно $\omega_n = e^{n^{-1}\psi}$ будет характери-

стической функцией и $\omega = e^{\psi}$ безгранично делима. Этим доказана

Теорема 1. *Для того чтобы* ω *была безгранично делимой характеристической функцией, необходимо и достаточно, чтобы существовали каноническая мера* M *и действительное число* b, *такие, что* $\omega = e^{\varrho}$, *где* ρ *определяется по* (2.7) *и* (2.8).

Эти формулы мы будем называть *каноническим представлением* ω, хотя функция $\sin x$ в центрирующем слагаемом будет заменяться на x или на 0, коль скоро это возможно. Мы будем говорить, что M определяет ω, что оправдано, если отвлечься от произвольной центрирующей константы b.

Пример. В качестве применения покажем, что *безгранично делимое распределение* U *имеет дисперсию тогда и только тогда, когда соответствующая каноническая мера* M *конечна.* В самом деле, по лемме 2 из гл. XV, 4 и ее следствию, U имеет дисперсию тогда и только тогда, когда существует вторая производная ψ''. Использованные ранее рассуждения показывают, что оправдано формальное дифференцирование, приводящее к равенству

$$-\psi''(\xi) = \int_{-\infty}^{+\infty} e^{i\xi x} M\{dx\}. \tag{2.9}$$

Обратно, если M — конечная мера, то существует, и притом только одна, функция ψ со второй производной (2.9) и с $\psi(0) = \psi'(0) = 0$. Эта функция имеет вид (2.8), и потому e^{ψ} является характеристической функцией безгранично делимого распределения с конечной дисперсией и нулевым математическим ожиданием.

Точно так же можно показать, что U *имеет момент порядка* $2k$ *в том и только том случае, когда интеграл от* $x^{2k-2}M\{dx\}$ *по всей прямой* $-\infty$, ∞ *конечен.* ▶

Следующая теорема лишь иначе формулирует отмеченный нами факт, что предел ω в (2.2) безгранично делим.

Теорема 2. *Пусть* $\{\varphi_n\}$ — *последовательность характеристических функций, такая, что в каждой точке* $\varphi_n^n \to \omega$. *Если* ω — *непрерывная функция, то она будет безгранично делимой характеристической функцией* [1]).

Легко описать ω в терминах распределений F_n, соответствующих φ_n. Пусть ρ определяется формулами (2.7) и (2.8).

[1]) Утверждение теоремы и ее доказательство сохраняются, если n пробегает какую-либо последовательность $n_1, n_2, \ldots \to \infty$.

Согласно теореме § 1, соотношение (2.3) выполняется тогда и только тогда, когда $nx^2F_n\{dx\} \to M\{dx\}$ в собственном смысле и $nb_n \to b$, где

$$b_n = \int_{-\infty}^{+\infty} \sin x \cdot F_n\{dx\}. \tag{2.10}$$

Теоретически этот критерий является исчерпывающим, но часто практически важен не вопрос сходимости $\varphi_n^n \to \omega$, а вопрос о существовании таких центрирующих констант β_n, что

$$\left(\varphi_n(\zeta)\, e^{-i\beta_n\zeta}\right)^n \to \omega(\zeta). \tag{2.11}$$

Обычно распределения F_n можно легко центрировать заранее таким образом, чтобы b_n были малы. Приводимый ниже критерий охватывает, как нам кажется, все практически интересные ситуации. В то же время он достаточно прост для применений. Ниже мы предполагаем, что $\omega = e^{\rho}$ есть безгранично делимая характеристическая функция, определяемая по (2.7) и (2.8).

Теорема 3. *Предположим, что*[1]) $\varphi_n(\zeta) \to 1$ *при всех* ζ *и* $nb_n^2 \to 0$. *Тогда три следующих утверждения эквивалентны друг другу*:

а) *имеет место* (2.11);

б) *выполняются соотношения*

$$n(b_n - \beta_n) \to b, \tag{2.12}$$

$$n[\varphi_n(\zeta) - 1 - ib_n\zeta] \to \rho(\zeta); \tag{2.13}$$

в) *выполняется* (2.12) *и* $nx^2F_n\{dx\} \to M\{dx\}$ *в собственном смысле.*

Доказательство. Достаточно рассмотреть случай $b = 0$. Тогда $\rho(\zeta) = \psi(\zeta)$. В силу последней леммы соотношение (2.11) эквивалентно

$$n[\varphi_n(\zeta)\, e^{-i\beta_n\zeta} - 1] \to \psi(\zeta). \tag{2.14}$$

Умножим обе стороны (2.14) на $e^{i\beta_n\zeta}$ и рассмотрим мнимые части при $\zeta = 1$. Так как $\psi(1)$ вещественно и $\beta_n \to 0$, мы получаем

$$n(b_n - \sin\beta_n) \to 0, \tag{2.15}$$

---

[1]) Условие $\varphi_n(\zeta) \to 1$ вводится только для того, чтобы избежать тривиальных усложнений следующего типа. Заменим $\varphi_n(\zeta)$ на $\varphi_n(\zeta)e^{i\pi\zeta}$. Тогда константы b_n останутся прежними, β_n заменится на $\beta_n + \pi$ и (2.12) не будет верно.

что, очевидно, равносильно (2.12). Из (2.14) следует также, что

$$n[\varphi_n(\zeta)e^{-ib_n\zeta}-1]\to\psi(\zeta). \tag{2.16}$$

Но левая часть равна

$$n[\varphi_n(\zeta)-1-ib_n\zeta]e^{-ib_n\zeta}+n[(1+ib_n\zeta)e^{-ib_n\zeta}-1]. \tag{2.17}$$

Второй член в (2.17) стремится к нулю, так как $nb_n^2\to 0$. Поэтому (2.13) выполняется. Таким образом, из (а) следует (б). Обратно, (б) влечет (2.16) и, следовательно, (2.14), что эквивалентно (2.11). Наконец, (б) и (в) равносильны по теореме § 1. ▶

Среди возможных следствий доказанных теорем стоит отметить такие: 1) *если* $\omega=e^{\psi}$ *безгранично делима, то при любом* $t>0$ и $\omega^t=b^{t\psi}$ *обладает этим же свойством*; 2) лемма показывает, что ω является пределом для $e^{n(\omega_n-1)}$, поэтому *каждое безгранично делимое распределение есть предел обобщенных пуассоновских распределений*; 3) *если последовательность* $\{\gamma_n\}$ *безгранично делимых характеристических функций сходится к непрерывной функции* γ, *то последняя также безгранично делима.* Последнее утверждение непосредственно вытекает из теоремы 2, так как каждая из функций γ_n может быть представлена в виде φ_n^n, где φ_n — характеристическая функция.

Замечание о других канонических представлениях. Представление с мерой M — не единственное, которое можно встретить в литературе. В первоначальных исследованиях П. Леви использовалась мера Λ, определяемая для $x\neq 0$ равенством $\Lambda\{dx\}=x^{-2}M\{dx\}$. Эта мера служит пределом для $nF_n\{dx\}$. Она конечна на интервалах $|x|>\delta>0$, но может быть неограниченной в окрестности нуля. Атом M в нуле, если он существует, не учитывается мерой Λ. В терминах этой меры (2.8) принимает вид

$$\psi(\zeta)=-\frac{1}{2}\sigma^2\zeta^2+ib\zeta+\lim_{\delta\to 0}\int_{|x|>\delta}[e^{i\zeta x}-1+i\zeta\sin x]\,\Lambda\{dx\}. \tag{2.18}$$

Это есть (с точностью до выбора центрирующей функции) *первоначальное каноническое представление, данное П. Леви.* Его основной недостаток в том, что полное описание свойств меры Λ оказывается слишком длинным.

Хинчин ввел *ограниченную* меру K, определяемую равенством $K\{dx\}=(1+x^2)^{-1}M\{dx\}$. Эта ограниченная мера может быть произвольной. Поэтому *каноническое представление по Хинчину* всего легче описать. Кроме того, оно имеет дело только с ограниченными мерами. Это — большое преимущество, однако искусственный выбор меры K приводит к трудностям в применениях. Устойчивые распределения и пример (3, е) показывают характер этих трудностей.

§ 3. Примеры. Специальные свойства

Мы укажем несколько специальных распределений, а затем рассмотрим такие свойства, как существование моментов или положительность. Эти свойства помещены в разделе «примеры» частично для того, чтобы сделать изложение более ясным, частично для того, чтобы подчеркнуть отсутствие связи между отдельными пунктами. Никакая часть материала этого параграфа не используется в дальнейшем. Дополнительные примеры можно найти в задачах 6, 7 и 18.

Примеры. а) *Нормальное распределение.* Если мера M сосредоточена в нуле и приписывает ему вес σ^2, то (2.8) превращается в равенство $\psi(\zeta) = -\frac{1}{2}\sigma^2\zeta^2$ и e^{ψ} будет нормальным распределением с нулевым математическим ожиданием и дисперсией σ^2.

б) *Распределение Пуассона.* Стандартное распределение Пуассона с математическим ожиданием α имеет характеристическую функцию $\omega = e^{\psi}$ с $\psi(\zeta) = \alpha(e^{i\zeta} - 1)$. Изменим параметры так, чтобы получить распределение, сосредоточенное в точках вида $-b + nh^{-1}$. При этом показатель степени превращается в $\rho(\zeta) = \alpha(e^{i\zeta h} - 1) - ib\zeta$, что представляет собой частный случай (2.7) с мерой M, сосредоточенной в одной точке h. Свойство, состоящее в том, что мера M сосредоточена в одной точке, характеризует класс нормальных распределений и класс пуассоновских распределений. Свертка конечного числа подобных распределений приводит к каноническим мерам с конечным числом атомов. В общем случае мера M может быть представлена, как предел подобных мер. Поэтому *любое безгранично делимое распределение может быть представлено как предел сверток конечного числа распределений Пуассона с нормальным распределением* [1]).

в) *Рандомизированное случайное блуждание.* В гл. II, (7.7) мы встретились с семейством арифметических распределений, приписывающих точкам $r = 0, \pm 1, \pm 2, \ldots$ вероятности

$$a_r(t) = \sqrt{\left(\frac{p}{q}\right)^r}\, e^{-t} I_r(2\sqrt{pq}\,t), \tag{3.1}$$

[1]) В терминологии функционального анализа класс безгранично делимых распределений можно описать как выпуклую оболочку совокупности нормальных и пуассоновских распределений. Они образуют *положительный конус*, а нормальные и пуассоновские распределения являются крайними точками в смысле Крейна и Мильмана.

где параметры p, q, t положительны, $p+q=1$ и I_r — функция Бесселя, определенная в гл. II, (7.1). Из того факта, что распределение $\{a_r\}$ удовлетворяет уравнению Чепмена — Колмогорова, вытекает его безграничная делимость. Соответствующую характеристическую функцию $\omega=e^{\psi}$ легко вычислить, так как она отличается от разложения Шлёмильха [гл. II, (7.8)] только заменой переменных $u=\sqrt{p/q}\,e^{-i\xi}$. Окончательный результат

$$\psi(\xi)=-t+t(pe^{i\xi}+qe^{-i\xi}) \tag{3.2}$$

показывает, что $\{a_r(t)\}$ можно рассматривать как *распределение разности двух независимых случайных величин, распределенных по закону Пуассона* с математическими ожиданиями pt и qt соответственно. Безграничная делимость очевидна.

г) *Показательное распределение и гамма-распределения.* Показательное распределение с математическим ожиданием, равным единице, имеет характеристическую функцию $\omega(\zeta)=(1-i\zeta)^{-1}$. Здесь $\psi=\log\omega$ существует. Очевидно, что $-i\psi'=\omega$, т. е.

$$\psi'(\zeta)=i\int_0^{\infty}e^{i\zeta x}e^{-x}\,dx. \tag{3.3}$$

Интегрируя, получаем

$$\psi(\xi)=\int_0^{\infty}\frac{e^{i\zeta x}-1}{x}e^{-x}\,dx, \tag{3.4}$$

откуда видно, что ω безгранично делима и соответствует канонической мере $M\{dx\}=xe^{-x}\,dx$. Гамма-плотность $g_t(x)=e^{-x}x^{t-1}/\Gamma(t)$ имеет характеристическую функцию $e^{t\psi}$, соответствующую мере tM.

д) *Законы Коши, Бесселя и гиперболического косинуса.* Из вида характеристических функций в п. 6, 8, 9 табл. XV, 2 легко вывести их безграничную делимость. Для плотности $f(x)=\frac{1}{\pi\operatorname{ch}x}$ с характеристической функцией $\omega(\zeta)=\frac{1}{\operatorname{ch}(\pi\zeta/2)}$ имеем $-\psi''=(-\log\omega)''=\frac{1}{8}\pi\omega^2$, следовательно, $-\psi''$ является преобразованием Фурье — Стильтьеса конечной меры $M\{dx\}=\frac{1}{4}\pi^2f^{2*}\,dx$. Сравнение с (2.8) показывает, что M есть каноническая мера, определяющая f. Далее f^{2*} была вычислена в задаче 6 из гл. II, 9. Мы получаем

$$M\{dx\}=\frac{x}{e^x-e^{-x}}\,dx.$$

Мы покажем в § 4, что для закона Коши

$$M\{dx\} = 2\pi^{-1}\,dx.$$

Из примера гл. XIII, (7, г) вытекает, что для распределения Бесселя $M\{dx\} = txe^{-x}I_0(x)\,dx$.

е) *Пример П. Леви.* Функция

$$\psi(\zeta) = 2\sum_{k=-\infty}^{+\infty} 2^{-k}[\cos 2^k\zeta - 1] \tag{3.5}$$

имеет вид (2.8) с симметричной мерой M, приписывающей вес 2^k точкам $\pm 2^k$, $k=0, \pm 1, \pm 2, \ldots$. Характеристическая функция $\omega = e^{\psi}$ обладает любопытным свойством: $\omega^2(\zeta) = \omega(2\zeta)$. Эта функция была бы устойчивой, если бы при каждом n было $\omega^n(\zeta) = \omega(a_n\zeta)$, но последнее соотношение выполняется только для $n=2, 4, 8, \ldots$. В терминологии § 9 ω принадлежит своей собственной области *частичного притяжения*, но она не имеет области притяжения (см. задачу 9).

ж) *Рандомизация и подчиненность* [1]). Если e^{ψ} безгранично делима, то при любом $s>0$ $e^{s\psi}$ также безгранично делима. Каково бы ни было распределение G, сосредоточенное на $0, \infty$, мы можем построить, рандомизируя параметр s, новую характеристическую функцию вида

$$\int_0^{\infty} e^{s\psi(\zeta)}G\{ds\}. \tag{3.6}$$

Это новая характеристическая функция, но она не обязана быть безгранично делимой. Обозначим преобразование Лапласа функции G через γ. Так как $|e^{\psi}| \leqslant 1$, то вещественная часть ψ не положительна и (3.6) равно $\gamma(-\psi(\zeta))$. Выражения этого типа всегда представляют собой, таким образом, характеристические функции. Допустим теперь, что G безгранично делима. Тогда $\gamma(\lambda) = e^{-\rho(\lambda)}$, и потому $\omega_t = e^{-t\rho(-\psi(\zeta))}$ является характеристической функцией при любом $t>0$. Отсюда следует, что $e^{-t\rho(-\psi(\zeta))}$ *есть безгранично делимая характеристическая функция.*

В терминологии подчиненных марковских процессов ω_t представляет собой характеристическую функцию процесса $\mathbf{X}(\mathbf{Y}(t))$, где $\mathbf{X}(t)$ имеет характеристическую функцию $e^{t\psi(\zeta)}$ и $\mathbf{Y}(t)$ имеет преобразование Лапласа $e^{-t\rho(\lambda)}$. Мы дали простое аналитическое доказательство того, что метод, описанный в гл. X, 7, действительно приводит к безгранично делимым распределениям.

[1]) Этот пример связан с преобразованиями Лапласа.

з) *Разложения.* Каждое представление $M=M_1+M_2$ меры M как суммы двух мер индуцирует представление $\omega=e^{\psi_1}e^{\psi_2}$, где сомножители безгранично делимы. В качестве типичного примера рассмотрим разложение M на меру, сосредоточенную на $|x|\leqslant\eta$, и меру, сосредоточенную на $|x|>\eta>0$. В первом интервале центрирующую функцию $\sin x$ естественно заменить на x, во втором заменим $x^{-2}M\{dx\}$ на $N\{dx\}$. Иными словами, мы полагаем

$$\psi_1(\zeta)=\int_{|x|\leqslant\eta}\frac{e^{i\zeta x}-1-i\zeta x}{x^2}M\{dx\}, \tag{3.7}$$

$$\psi_2(\zeta)=\int_{|x|>\eta}(e^{i\zeta x}-1)N\{dx\} \tag{3.8}$$

и

$$\psi(\zeta)=\psi_1(\zeta)+\psi_2(\zeta)+i\beta\zeta. \tag{3.9}$$

Как мы неоднократно замечали, e^{ψ_2} является характеристической функцией *обобщенного распределения Пуассона* U_2, индуцированного таким распределением вероятностей, что $F\{dx\}=cN\{dx\}$. Характеристическая функция e^{ψ_1} бесконечно дифференцируема и соответствует распределению, имеющему конечные моменты всех порядков. *Ограниченные* меры M и N и действительное число β могут быть выбраны произвольно, и предшествующие формулы дадут нам безгранично делимую характеристическую функцию e^{ψ}. Мы проиллюстрируем сейчас пользу подобного представления.

и) *Положительные случайные величины.* Мы покажем, что распределение U с характеристической функцией e^{ψ} *сосредоточено на* $\overline{0,\infty}$ *тогда и только тогда, когда* [1])

$$\psi(\zeta)=\int_0^\infty\frac{e^{i\zeta x}-1}{x}P\{dx\}+ib\zeta, \tag{3.10}$$

где $b\geqslant 0$ и P — мера на открытом интервале $\overline{0,\infty}$, такая, что функция $(1+x)^{-1}$ интегрируема по отношению к P. Допустим, что U сосредоточено на $\overline{0,\infty}$. Используя обозначения предыдущего примера, возьмем точку роста a распределения F, порождающего обобщенное распределение Пуассона U_2. Тогда a,

[1]) Это было замечено П. Леви и доказано непосредственными вычислениями Г. Бакстером и Дж. Шапиро. Заметим, что (3.10) отличается только формой записи от преобразования Лапласа гл. XIII, (7.3).

$2a$, $3a$, ... будут точками роста U_2 и если b — точка роста U_1, то $b+na+\beta$ будет точкой роста для U (см. лемму 1 в гл. V, 4). Таким образом, $a \geqslant 0$, так что N сосредоточено на $\overline{0, \infty}$. Аналогичное рассуждение показывает, что U не может иметь нормальной компоненты, следовательно, M не имеет атома в нуле. Мы видим, что $e^{\psi(\zeta)}$ является пределом при $\eta \to 0$ функций $e^{\psi_2(\zeta)+ib\zeta}$. Для распределений, соответствующих этим характеристическим функциям, β служит точкой роста. Поэтому величина β должна стремиться к неотрицательному пределу. Вводя обозначение $N\{dx\}=x^{-1}P\{dx\}$, получаем (3.10), так что высказанное условие необходимо. Но мы уже видели, что распределение U с характеристической функцией e^{ψ} есть предел распределений, сосредоточенных на $\overline{0, \infty}$. Это доказывает достаточность нашего условия.

к) *Асимптотическое поведение.* Разложение, указанное в примере (з), полезно при анализе поведения хвостов распределения U. Распределение U_1 с характеристической функцией e^{ψ_1} имеет моменты всех порядков, так что $1-U_1(x)$ стремится к нулю быстрее любой степени x^{-n}. При весьма слабых условиях регулярности отсюда следует, что $1-U(x) \sim 1-U_2(x)$. Поэтому мы можем ограничиться только обобщенным распределением Пуассона. Последнее может быть разложено на два таких же распределения, сосредоточенных на $\overline{-\infty, 0}$ и $\overline{0, \infty}$ соответственно. Мы не будем развивать здесь эту линию дальше, так как асимптотическое поведение односторонних распределений лучше исследовать с помощью преобразований Лапласа и тауберовых теорем. Подобные рассуждения показывают, например, непосредственно, что результат простой задачи 16 из гл. XIII, 11 верен и в общем случае: *если* $P(\overline{0, x}) \sim x^c L(x)$, *где* $0<c<1$ *и* L *медленно меняется на бесконечности, то* $1-U(x) \sim \frac{1}{1-c}\, x^{c-1}L(x)$. ▶

§ 4. Устойчивые характеристические функции

В соответствии с определением гл. VI, 1 характеристическая функция ω называется устойчивой в том и только том случае, когда $\omega^n(\zeta)=\omega(a_n\zeta)^{ib_n\zeta}$, т. е. когда ω^n и ω отличаются только параметрами сдвига и масштаба. Отсюда вытекает безграничная делимость, и поэтому мы можем начать с канонического представления $\omega=e^{\psi}$, где ψ определяется по (2.8). *Функция* ω *будет устойчивой в том и только том случае, когда для каждого*

n найдутся вещественные постоянные $a_n>0$ *и* b_n, *такие, что*

$$n\psi(\zeta)=\psi(a_n\zeta)+ib_n\zeta. \tag{4.1}$$

Мы найдем явное выражение[1]) для всех таких функций ψ.

Подстановка $x=a_ny$ в (2.8) показывает, что для (4.1) необходимо соотношение

$$nM\{dy\}=a_n^2M\{a_n^{-1}\,dy\}. \tag{4.2}$$

Положим $\mu(x)=x^{-2}M\{\overline{-x,\ x}\}$. Тогда (4.2) принимает вид

$$n\mu(a_nx)=\mu(x). \tag{4.3}$$

Теперь легко проверить (это, впрочем, известно из леммы 3 гл. VIII, 2), что $a_n\to\infty$ и $a_{n+1}/a_n\to 1$. Поэтому из простой леммы 2 гл. VIII, 8 следует, что μ должно быть пропорционально некоторой степени x, т. е. $M\{\overline{-x,\ x}\}=Cx^{2-\alpha}$. Здесь $0<\alpha\leqslant 2$, так как функция $M\{\overline{-x,\ x}\}$ должна быть неубывающей и так как $x^{-2}M\{dx\}$ интегрируемо на $\overline{1,\ \infty}$. Значению $\alpha=2$ соответствует мера, сосредоточенная в нуле. При $\alpha<2$ нуль есть точка непрерывности для M и применение тех же рассуждений к $M\{\overline{0,\ x}\}$ и $M\{\overline{-x,\ 0}\}$ показывает, что справедлива

Лемма 1. *Каноническая мера M, определяющая устойчивое распределение, или сосредоточена в нуле, или*

$$M\{\overline{-x,\ 0}\}=Cqx^{2-\alpha},\quad M\{\overline{0,\ x}\}=Cpx^{2-\alpha},\quad x>0, \tag{4.4}$$

где $0<\alpha<2$, $p\geqslant 0$, $q\geqslant 0$ *и* $p+q=1$.

Мере M, сосредоточенной в нуле, соответствует нормальное распределение [пример (3, а)]. Для вычисления функции e_ψ, определяемой по (4.4), мы могли бы использовать стандартную форму (2.8). Однако, принимая во внимание специальный вид меры M, предпочтительнее заменить центрирующую функцию $\sin x$ центрирующей функцией

$$\tau_\alpha(x)=\begin{cases} x, & \text{если } \alpha>1,\\ \sin x, & \text{если } \alpha=1,\\ 0, & \text{если } \alpha<1.\end{cases} \tag{4.5}$$

Тогда

$$\psi(\zeta)=\int_{-\infty}^{+\infty}\frac{e^{i\zeta x}-1-i\zeta\tau_\alpha(x)}{x^2}\,M\{dx\}. \tag{4.6}$$

[1]) Другой метод нахождения смотри в задачах 10—11.

При $\alpha \neq 1$ и $\zeta > 0$ подстановка $y = \zeta x$ показывает, что $\psi(\zeta) = a\zeta^\alpha$ и $\psi(-\zeta) = \bar{a}\zeta^\alpha$, где a — комплексная постоянная и $\bar{a}$ сопряжено с a.

Численное значение a представляет второстепенный интерес; мы будем пользоваться им, но вычисление отложим. Чтобы суммировать получающиеся результаты, введем соглашение: *в символах «$\pm$» и «$\mp$» выбирается верхний знак при $\zeta > 0$ и нижний при $\zeta < 0$.*

Лемма 2. *Пусть мера M определена как в* (4.4). *Тогда при* $\alpha < 1$

$$\psi(\zeta) = -|\zeta|^\alpha C \cdot \frac{\Gamma(1-\alpha)}{\alpha}\left[\cos\frac{\pi\alpha}{2} \mp i(p-q)\sin\frac{\pi\alpha}{2}\right]. \tag{4.7}$$

При $1 < \alpha < 2$

$$\psi(\zeta) = |\zeta|^\alpha C \frac{\Gamma(2-\alpha)}{1-\alpha}\left[\cos\frac{\pi\alpha}{2} \mp i(p-q)\sin\frac{\pi\alpha}{2}\right]. \tag{4.8}$$

Наконец, при $\alpha = 1$

$$\psi(\zeta) = -|\zeta| C\left[\frac{\pi}{2} \pm i\log|\zeta|\right]. \tag{4.9}$$

Если мера M сосредоточена в нуле и приписывает ему вес C, то $\psi(\zeta) = -\frac{1}{2}C\zeta^2$.

Чтобы получить простую стандартную форму [1]), мы выберем нормирующий множитель C так, чтобы вещественная часть $\psi(\zeta)$ стала равной $-|\zeta|^\alpha$. Полагая $p - q = \delta$, мы получим тогда

$$\psi_{\alpha,\delta}(\zeta) = \begin{cases} -|\zeta|^\alpha \cdot \left[1 \mp i\delta\operatorname{tg}\frac{\pi\alpha}{2}\right], & \text{если } \alpha \neq 1, \\ -|\zeta|\left[1 \pm i\delta\log|\zeta|\right], & \text{если } \alpha = 1. \end{cases} \tag{4.10}$$

Требование устойчивости (4.1) выполняется при $a_n = n^{1/\alpha}$ и $b_n = 0$, за исключением случая $\alpha = 1$, когда $a_n = n$ и $b_n = -\frac{2}{\pi}\log n$.

Мы нашли, таким образом, все устойчивые характеристические функции. Основная часть доказанного результата может быть выделена следующим образом.

---

[1]) Иная стандартная форма указана в § 6.

Теорема. *Логарифм любой устойчивой характеристической функции имеет вид* [1]) $a\psi_{\alpha,\delta}(\zeta)+ib\zeta$, *где* $0<\alpha\leqslant 2$, $-1\leqslant\delta\leqslant 1$, *и* $\psi_{\alpha,\delta}$ *определено по* (4.10).

Здесь подразумевается, что $a>0$ и b вещественно. Примечательная черта этой теоремы — ограничение на значения δ: $-1\leqslant\delta\leqslant 1$.

Сравнивая с примером (3, и), получаем следующий результат.

Лемма 3. *Устойчивое распределение сосредоточено на* $\overline{0,\infty}$ *тогда и только тогда, когда* $\alpha<1$, $\delta=1$ *и* $b\geqslant 0$.

(Поведение хвостов описано в теореме 3 § 5; разложение для плотностей см. в § 6. О распределениях, сосредоточенных на $\overline{0,\infty}$, см. теорему 1, гл. XIII, 6).

Добавление. Вычисление интеграла (4.6)

Достаточно рассмотреть случай $\zeta>0$. Мы докажем, что при $\alpha<1$

$$\int_0^\infty \frac{e^{i\zeta x}-1}{x^{\alpha+1}}\,dx=-\zeta^\alpha\frac{\Gamma(1-\alpha)}{\alpha}e^{-i\pi\alpha/2}. \tag{4.11}$$

Аналогичный интеграл, взятый по интервалу $\overline{-\infty,0}$, вычисляется заменой i на $-i$, и комбинация двух интегралов дает (4.7). Рассмотрим интеграл (4.11) как предел при $\lambda\to 0+$ интегралов

$$\int_0^\infty \frac{e^{i\zeta x-\lambda x}-1}{x^{\alpha+1}}\,dx=-\frac{\lambda-i\zeta}{\alpha}\int_0^\infty e^{i\zeta x-\lambda x}x^{-\alpha}\,dx. \tag{4.12}$$

С точностью до постоянного множителя последний интеграл представляет собой характеристическую функцию гамма-плотности. По таблице гл. XV, 2 (п. 6) находим, что величина (4.12) равна

$$-\frac{\Gamma(1-\alpha)}{\alpha}\cdot(\lambda-i\zeta)^\alpha=-\frac{\Gamma(1-\alpha)}{\alpha}(\lambda^2+\zeta^2)^{\alpha/2}\cdot e^{i\theta\alpha}, \tag{4.13}$$

где θ — аргумент $\lambda-i\zeta$, т. е. $\operatorname{tg}\theta=-\zeta/\alpha$. Очевидно, что при $\lambda\to 0$ $\theta\to-\pi/2$, и мы видим, что (4.11) верно.

При $1<\alpha<2$ дифференцирование по ζ приводит интеграл (4.6) к виду, только что рассмотренному. Здесь (4.8) проверяется элементарными выкладками. При $\alpha=1$ вещественная часть подинтегрального выражения равна $(1-\cos\zeta x)/x^2$ и соответствующий интеграл можно найти в таблице гл. XV, 2 (п. 5). Остается оценить только мнимую часть, а именно

$$\int_0^\infty \frac{\sin\zeta x-\zeta\sin x}{x^2}\,dx=\lim_{\varepsilon\to 0}\left[\int_\varepsilon^\infty\frac{\sin\zeta x}{x^2}\,dx-\zeta\int_\varepsilon^\infty\frac{\sin x}{x^2}\,dx\right]. \tag{4.14}$$

[1]) Нормальная характеристическая функция получается из (4.10) при $\alpha=2$, так как коэффициент при δ обращается в нуль

Подстановка $\zeta x = y$ приводит подинтегральное выражение в первом интеграле к тому же виду, что и во втором интеграле. Поэтому (4.14) принимает форму

$$\zeta \lim_{\varepsilon \to 0} \int_{\varepsilon\zeta}^{\varepsilon} \frac{\sin x}{x^2}\, dx = \zeta \int_{\varepsilon\zeta}^{\varepsilon} \frac{dx}{x} = -\zeta \log \zeta. \tag{4.15}$$

(Здесь использован тот факт, что $\frac{\sin x}{x} \to 1$ при $x \to 0$.) Соотношение (4.9) доказано.

§ 5. Области притяжения

Назовем характеристическую функцию *унитарной*, если $|\varphi(\zeta)| = 1$ для всех ζ, т. е. если соответствующее распределение сосредоточено в одной точке [см. лемму гл. XV, (5.4)]. На языке характеристических функций определение 2 гл. VI, 1 звучит следующим образом.

Характеристическая функция φ распределения F принадлежит области притяжения неунитарной характеристической функции γ, если существуют параметры $a_n > 0$ и β_n, такие, что

$$\left(\varphi\left(\frac{\zeta}{a_n}\right) e^{-i\beta_n \zeta}\right)^n \to \gamma(\zeta). \tag{5.1}$$

В гл. VI, 1 было показано, что предельная функция γ необходимо устойчива. Однако этот результат легко доказать заново, тем самым сохранив независимость излагаемой теории от материала гл. VI, 1. Для сравнения с теоремами § 1 и 2 положим

$$\varphi_n(\zeta) = \varphi\left(\frac{\zeta}{a_n}\right), \quad F_n(x) = F(a_n x). \tag{5.2}$$

Кроме того, положим при $x > 0$

$$\mu(x) = \int_{-x}^{x} y^2 F\{dy\}. \tag{5.3}$$

Теорема 1. *Для того чтобы F принадлежало некоторой области притяжения, необходимо, чтобы для μ выполнялось соотношение*

$$\mu(x) \sim x^{2-\alpha} L(x), \tag{5.4}$$

где $0 < \alpha \leqslant 2$ и L медленно меняется на бесконечности, т. е. при любом фиксированном $x > 0$

$$\frac{L(tx)}{L(t)} \to 1, \qquad t \to \infty. \tag{5.5}$$

(i) *Если $\alpha = 2$, то F принадлежит области притяжения нормального распределения.*

(ii) *Если* $\alpha<2$ *и при* $x\to\infty$

$$\frac{1-F(x)}{1-F(x)+F(-x)}\to p,\qquad \frac{F(-x)}{1-F(x)+F(-x)}\to q, \tag{5.6}$$

то F *принадлежит области притяжения безгранично делимого распределения, определяемого канонической мерой* M, *такой, что для* $x>0$

$$M\{\overline{-x,\ x}\}=Cx^{2-\alpha}, \tag{5.7}$$

$$M^{+}(x)=pCx^{-\alpha},\quad M^{-}(-x)=qCx^{-\alpha}. \tag{5.8}$$

Если (5.6) *не выполняется, то* F *не принадлежит никакой области притяжения.*

(iii) *Если* $\alpha<2$, *то условие* (5.4) *равносильно*

$$1-F(x)+F(-x)\sim\frac{2-\alpha}{\alpha}x^{-\alpha}L(x),\quad x\to\infty. \tag{5.9}$$

В предыдущем параграфе было показано, что каноническим мерам вида (5.7) и (5.8) соответствуют устойчивые распределения. Поэтому мы можем изложить содержание теоремы 1 по-иному.

Теорема 1а. 1) *F принадлежит области притяжения нормального закона тогда и только тогда, когда* μ *медленно меняется (в частности, когда существует второй момент).*

2) *F принадлежит области притяжения не нормального устойчивого закона тогда и только тогда, когда выполняются* (5.4) *и* (5.6), *и* $0<\alpha<2$. *Условие* (5.4) *эквивалентно* (5.9).

3) *Только устойчивые распределения имеют область притяжения.*

Функция, стремящаяся к конечному пределу, меняется медленно. Поэтому из первой части теоремы вытекает, что каждое распределение, имеющее дисперсию, принадлежит области притяжения нормального распределения. Однако эта область включает и распределения, не имеющие дисперсии [см. пример (4, а) гл. VIII].

Полезно, может быть, напомнить теорему 2 гл. VIII, 9, в соответствии с которой при $0<\alpha\leqslant 2$ условие (5.4) эквивалентно условию

$$\lim_{t\to\infty}\frac{t^2[1-F(t)+F(-t)]}{\mu(t)}\to\frac{2-\alpha}{\alpha}. \tag{5.10}$$

Мы выведем теорему 1 из следующей леммы.

Лемма 1. *Условия теоремы* 1 *необходимы и достаточны для существования констант* a_n *и* β_n, *таких, что*

$$n[\varphi_n(\zeta) - 1 - i\beta_n\zeta] \to \rho(\zeta), \tag{5.11}$$

где функция ρ *непрерывна.*

Доказательство. Допустим, что (5.11) выполняется. По теореме § 1 существует каноническая мера M, такая, что $nx^2F_n\{dx\} \to M\{dx\}$ в собственном смысле. Условия (1.3) и (1.4) для собственной сходимости сводятся в настоящем случае к тому, что во всех точках непрерывности $x>0$

$$\frac{n}{a_n^2}\,\mu(a_n x) \to M\{\overline{-x,\ x}\}, \tag{5.12}$$

$$n[1 - F(a_n x)] \to M^+(x), \quad nF(-a_n x) \to M^-(-x). \tag{5.13}$$

По лемме 3, гл. VIII, 2 $a_{n+1}/a_n \to 1$, а лемма 2, гл. VIII, 8 показывает, что (5.12) и (5.13) влекут (5.7) и (5.8). Таким образом, условия теоремы необходимы.

Допустим теперь, что выполняется (5.4). Так как L — медленно меняющаяся функция, то можно найти такие константы a_n, что

$$\frac{n}{a_n^\alpha}\,L(a_n) \to C, \tag{5.14}$$

где $C>0$ — заданная константа. Из условий теоремы тогда вытекают соотношения (5.12) и (5.13) с функцией M, определяемой (5.7) и (5.8). Это означает, что $nx^2F\{a_n dx\} \to M\{dx\}$ в собственном смысле, и по теореме § 1 наши условия достаточны для (5.11). Доказательство закончено. ▷

Лемма 2. *Распределение* F, *удовлетворяющее условиям теоремы* 1, *обладает абсолютными моментами любого порядка* $\beta<\alpha$.

Доказательство. Ясно [1]), что при $t \to \infty$ и любом $\varepsilon>0$

$$1 - F(t) + F(-t) = O(t^{-\alpha+\varepsilon}). \tag{5.15}$$

Наше утверждение вытекает теперь из формулы гл. V, (6.3) для моментов. ▷

[1]) Если t пробегает последовательность 2, 2^2, 2^3, ..., то (5.15) легко проверяется с помощью (5.10). Так как левая часть монотонна, то (5.15) верно при любом стремлении t к ∞ (другое доказательство см. в гл. VIII, 9).

Доказательство *теоремы* 1. Для доказательства теоремы нам достаточно установить, что исходное соотношение (5.1) эквивалентно условию (5.11) леммы 1. В случае симметричных распределений мы можем принять $\beta_n=0$, и эквивалентность выражений (5.1) и (5.11) вытекает из леммы § 2. Поэтому для симметричных распределений теорема верна.

Более общим образом условия (5.1) и (5.11) эквивалентны, когда применима теорема 3 § 2. В этой теореме предполагается, что

$$nb_n^2 \to 0, \quad \text{где} \quad b_n = \int_{-\infty}^{+\infty} \sin \frac{x}{a_n} F\{dx\}, \tag{5.16}$$

и что $\varphi(a_n^{-1}\zeta) \to 1$. Последнее условие выполнено, так как $a_n \to \infty$. Остается доказать, что мы можем принять условие $nb_n^2 \to 0$.

Пусть F принадлежит некоторой области притяжения. Тогда то же самое верно и для симметризованного распределения ${}^0F$, и нормирующие коэффициенты a_n будут теми же самыми. Но для ${}^0F$ теорема уже доказана, так что a_n удовлетворяют соотношению типа (5.14). Более того, ${}^0F$ обладает абсолютными моментами любого порядка $\beta<\alpha$. Поэтому то же самое верно и для F [см. гл. V, (6.3)]. В частности, при $\alpha>1$ существует математическое ожидание, и в этом случае мы допускаем, что F *имеет нулевое математическое ожидание.*

Теперь $|\sin x|<|x|^\beta$ при $0<\beta<1$ и $|\sin x - x|<2|x|^\beta$ при $1\leqslant\beta\leqslant 2$. Выберем $\beta<\alpha$ при $\alpha\leqslant 1$ и $1<\beta<\alpha$ при $\alpha>1$. Обозначим через c_β абсолютный момент порядка β распределения F. Тогда $|b_n|<2c_\beta a_n^{-\beta}$ и в силу (5.14)

$$nb_n^2 = O\left(\frac{a_n^{\alpha-2\beta}}{L(a_n)}\right). \tag{5.17}$$

При $2\beta>\alpha$ правая часть стремится к нулю, т. е. выполняется (5.16). ▶

Теорема 1 описывает предельные распределения в терминах канонической меры M. Явный вид соответствующих характеристических функций был указан в лемме 2 § 4. Мы переформулируем теперь наши результаты, особенно выделив способ выбора нормирующих постоянных.

Теорема 2. *Пусть F удовлетворяет условиям теоремы* 1. *При $\alpha>1$ мы предполагаем, что F имеет нулевое математическое ожидание.*

Пусть $\gamma=e^{\psi}$, *где* ψ *определено формулами* (4.7)—(4.9) *при* $a<2$, *и* $\psi(\zeta)=-\frac{1}{2}\zeta^2$ *при* $\alpha=2$. *Пусть* a_n *определены по* (5.14).

Тогда $\varphi^n(\zeta/a_n)-\gamma(\zeta)$ *при* $\alpha\neq 1$ *и* $\left(\varphi(\zeta/a_n)e^{-ib_n\zeta}\right)^n\to\gamma(\zeta)$ *при* $\alpha=1$.

Доказательство. Формулы (4.7)—(4.9) вытекают из описанного в теореме 1 строения канонической меры, поэтому в комментариях нуждается лишь способ выбора центрирующих постоянных. Будем придерживаться тех же обозначений, что и в стандартной форме безгранично делимых распределений [см. (2.7) и (2.8)]. Тогда центрирующие постоянные в (4.7)—(4.9) имеют вид $b=0$ при $\alpha=1$ и

$$b=\int_{-\infty}^{+\infty}\frac{\sin x}{x^2}M\{dx\}\quad\text{или}\quad b=\int_{-\infty}^{+\infty}\frac{\sin x-x}{x}M\{dx\}\tag{5.18}$$

в зависимости от того, что $\alpha<1$ *или* $\alpha>1$. Наши центрирующие постоянные в (5.1) имеют вид $\beta_n=0$ при $\alpha\neq 1$ и $\beta_n=b_n$ при $\alpha=1$. Мы должны доказать (2.12), а именно, что $n(b_n-\beta_n)\to b$. При $\alpha=1$ обе стороны равны нулю. При $\alpha>1$ наше утверждение состоит в том, что

$$n\int_{-\infty}^{+\infty}\frac{\sin x-x}{x^2}F\{a_n\,dx\}\to\int_{-\infty}^{+\infty}\frac{\sin x-x}{x^2}M\{dx\}.\tag{5.19}$$

Эта сходимость очевидна для интегралов, взятых по любому конечному промежутку, а по теореме 2, гл. VIII, 9 интегралы, взятые по области $|x|>t$, равномерно малы при больших t. Аналогичные соображения применимы и при $\alpha<1$. ▷

Соотношение (5.14) можно использовать двумя способами. При данном распределении F оно позволяет определить числа a_n, а в случае, когда a_n известны, с его помощью можно сделать определенные заключения относительно L. Например, каждое устойчивое распределение принадлежит своей собственной области притяжения с $a_n=n^{1/\alpha}$. Из (5.14) следует, что в этом случае $L(x)\to 1$ при $x\to\infty$. Сравнение (5.6) и (5.9) показывает, что верна

Теорема 3. *Если* G *— устойчивое распределение с характеристической функцией* e^{ψ} *из леммы* 4.2, *то при* $x\to\infty$

$$x^{\alpha}[1-G(x)]\to Cp\frac{2-\alpha}{\alpha},\quad x^{\alpha}G(-x)\to Cq\frac{2-\alpha}{\alpha}.\tag{5.20}$$

(Это утверждение верно и при $\alpha=2$.)

Заключительные замечания. Область притяжения нормального закона не следует смешивать с введенным Б В Гнеденко понятием «области *нормального притяжения устойчивого распределения G с показателем* α» Говорят, что распределение F принадлежит указанной области, если оно принадлежит области притяжения G при нормирующих коэффициентах $a_n = n^{1/\alpha}$. Выделение этой области представляло собой первоначально серьезную задачу, однако в нашей обстановке решение дается условием (5 14), определяющим нормирующие постоянные Если G удовлетворяет условиям теоремы 1, то *распределение F принадлежит области «нормального притяжения» G в том и только том случае, когда* $x^{\alpha}[1-F(x)] \to Cp$ *и* $x^{\alpha}F(-x) \to Cq$ *при* $x \to \infty$. Здесь $c>0$ — некоторая постоянная (Отметим попутно, используя введенную терминологию, что нормальное распределение обладает областью ненормального притяжения)

§ 6 *. Устойчивые плотности

Невозможно, по-видимому, выразить устойчивые плотности в «замкнутой» форме. Разложения в ряды для них были даны независимо друг от друга Феллером (1952) и Бергстремом (1953). В этих работах неявно содержатся результаты, установленные позже более сложными методами. Вывод этих разложений представляет собой хороший пример использования формулы обращения Фурье (хотя при этом применяется интегрирование в комплексной области). Мы не будем рассматривать случай $\alpha = 1$.

При $\zeta > 0$ мы можем представить устойчивые характеристические функции в форме $e^{-a\zeta^{\alpha}}$, где a — некоторое комплексное число. В стандартной форме (4.10) вещественная часть a равна 1. Для наших целей удобнее считать, что a по модулю равно единице, именно $a = e^{i\pi\gamma/2}$, где γ — вещественное число. Таким образом, мы полагаем

$$\psi(\zeta) = -|\zeta|^{\alpha} \cdot e^{\pm i\pi\gamma/2}, \tag{6.1}$$

где (как и в § 4) в символе «$\pm$» верхний знак выбирается при $\zeta > 0$, а нижний — при $\zeta < 0$. *Для того чтобы функция* e^{ψ} *была устойчивой характеристической функцией, необходимы и достаточны условия*

$$|\gamma| \leqslant \begin{cases} \alpha, & \text{если } 0 < \alpha < 1, \\ 2-\alpha, & \text{если } 1 < \alpha < 2. \end{cases} \tag{6.2}$$

Так как e^{ψ} абсолютно интегрируема, то соответствующее распределение имеет плотность, которую мы обозначим $p(x; \alpha, \gamma)$. Мы вычислим ее, используя формулу обращения для интегралов Фурье гл. XV, (3.5). Так как p — вещественная функция

*) Этот параграф посвящен специальным вопросам. Его следует пропустить при первом чтении.

и величина $\psi(-\zeta)$ сопряжена с $\psi(\zeta)$, мы получаем

$$p(x;\ \alpha,\ \gamma)=\frac{1}{\pi}\operatorname{Re}\int_0^\infty e^{-ix\zeta-\zeta^\alpha e^{i\pi\gamma/2}}\,d\zeta. \tag{6.3}$$

Так как

$$p(-x;\ \alpha,\ \gamma)=p(x;\ \alpha,\ -\gamma), \tag{6.4}$$

то эту функцию достаточно вычислить лишь для $x>0$.

а) *Случай* $\alpha<1$. Рассмотрим подинтегральное выражение как функцию комплексного переменного ζ. При $x>0$ и $\operatorname{Im}\zeta\to-\infty$ подинтегральное выражение стремится к нулю благодаря тому, что главную роль в показателе играет линейный член. Мы можем заменить путь интегрирования отрицательной частью мнимой оси, что сводится к подстановке $\zeta=\frac{t}{x}e^{-i\pi/2}$ и проведению дальнейших выкладок так, как если бы все коэффициенты были вещественны. Новое подинтегральное выражение имеет вид e^{-t-ct^α}. Используя разложение e^{-ct^α} в ряд и формулу для гамма-функции, мы получаем сразу

$$p(-x;\ \alpha,\ \gamma)=\operatorname{Re}\frac{-i}{\pi x}\sum_{k=0}^{\infty}\frac{\Gamma(k\alpha+1)}{k!}\left(-x^{-\alpha}\exp\left[i\frac{\pi}{2}(\gamma-a)\right]\right)^k. \tag{6.5}$$

б) *Случай* $1<\alpha<2$. Формальную подстановку

$$\zeta=t^{\alpha-1}\exp\left(-i\frac{\pi\gamma}{2\alpha}\right)$$

легко обосновать так же, как это было сделано при $\alpha<1$. Новое подинтегральное выражение имеет вид $e^{-t-ct^{\alpha-1}}t^{\alpha-1-1}$. Разлагая $e^{-ct^{\alpha-1}}$ в степенной ряд, получаем

$$p(-x;\ \alpha,\ \gamma)=$$

$$=\frac{1}{\alpha\pi}\operatorname{Re}\exp\left(-i\frac{\pi\gamma}{2\alpha}\right)\sum_{n=0}^{\infty}\frac{\Gamma\left(\frac{n+1}{\alpha}\right)}{n!}\left(-ix\exp\left[-i\frac{\pi\gamma}{2\alpha}\right]\right)^n. \tag{6.6}$$

Заменяя индекс суммирования n на $k-1$ и применяя известную рекуррентную формулу $\Gamma(s+1)=s\Gamma(s)$, имеем

$$p(-x;\ \alpha,\ \gamma)=$$

$$=\frac{1}{\pi x}\operatorname{Re}i\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha^{-1}+1)}{k!}\left(-x\exp\left[-i\frac{\pi}{2\alpha}(\gamma-\alpha)\right]\right)^k. \tag{6.7}$$

Таким образом, нами доказана

Лемма 1. *Для* $x>0$ *и* $0<\alpha<1$

$$p(x;\ \alpha,\ \gamma)=\frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha+1)}{k!}(-x^{-\alpha})^k\sin\frac{k\pi}{2}(\gamma-\alpha). \quad (6.8)$$

Для $x>0$ *и* $1<\alpha<2$

$$p(x;\ \alpha,\ \gamma)=\frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha^{-1}+1)}{k!}(-x)^k\sin\frac{k\pi}{2\alpha}(\gamma-\alpha). \quad (6.9)$$

Для $x<0$ *значения* p *определяются в соответствии с* (6.4).

Отметим, что (6.8) дает асимптотическое разложение для $x\to\infty$. Попутно из доказанных формул вытекает следующий любопытный результат.

Лемма 2. *Если* $\alpha<1$ *и* $x>0$, *то*

$$\frac{1}{x^{\alpha+1}}\,p\left(\frac{1}{x^{\alpha}};\ \frac{1}{\alpha},\ \gamma\right)=p(x;\ \alpha,\ \gamma^*), \quad (6.10)$$

где $\gamma^*=\alpha(\gamma+1)-1$.

Простая проверка показывает, что значения γ^* попадают в интервал, указанный в (6.2). Тождество (6.10) было впервые отмечено (с другим доказательством) В. М. Золотаревым.

§ 7*. Схема серий

Понятие схемы серий было описано в гл. VI, 3 следующим образом. Каждому n соответствует конечное число, скажем r_n, независимых случайных величин $\mathbf{X}_{k,n}$ $(k=1,2,\ldots,r_n)$ с распределениями $F_{k,n}$ и характеристическими функциями $\varphi_{k,n}$. Мы образуем «суммы по строкам» $\mathbf{S}_n=\mathbf{X}_{1,n}+\ldots\mathbf{X}_{r_n,n}$. Обозначим их распределения и характеристические функции через U_n и ω_n соответственно. По причинам, объясненным в гл. VI, 3, мы интересуемся в первую очередь схемами, где влияние отдельных компонент асимптотически пренебрежимо. Чтобы обеспечить это свойство, мы ввели условие гл. VI, (3.3): величины $\mathbf{X}_{k,n}$ должны стремиться к нулю по вероятности равномерно относительно $k=1,\ldots,r_n$. В терминах характеристических функций это означает, что при любых числах $\varepsilon>0$ и $\zeta_0>0$ для всех достаточно больших n

$$|1-\varphi_{k,n}(\zeta)|<\varepsilon \quad \textit{при} \quad |\zeta|<\zeta_0,\quad k=1,\ \ldots,\ r_n. \quad (7.1)$$

Такие схемы мы назовем *нулевыми схемами.*

*) В подлиннике «треугольные схемы» (triangular arrays). — *Прим. перев.*

В § 1 и 2 мы имели дело со схемами серий, в которых распределения $F_{k,n}$ не зависели от k. Такие схемы автоматически оказываются нулевыми. Условие (7.1) дает возможность применить прежние рассуждения почти без всяких изменений.

Теорема 1. *Пусть* $\{\mathbf{X}_{k,n}\}$ *— нулевая схема и* $\{\beta_n\}$ *— последовательность вещественных постоянных. Если распределения случайных величин* $\mathbf{S}_n+\beta_n$ *сходятся к распределению вероятностей* U*, то последнее оказывается безгранично делимым.*

Мы объединим доказательство этой теоремы с выводом необходимых и достаточных условий сходимости. Для этого применим новый способ центрирования, отчасти для разнообразия, отчасти потому, что он упрощает общий критерий сходимости.

Центрирование математическими ожиданиями будет, конечно, наилучшим, коль скоро оно применимо. В общем случае мы введем *урезанные математические ожидания*. Положим

$$\tau(x)=\begin{cases}-1 & \text{при} \quad x\leqslant -1,\\ x & \text{при} \quad -1\leqslant x\leqslant 1,\\ 1 & \text{при} \quad x\geqslant 1.\end{cases} \tag{7.2}$$

Мы будем использовать $\mathbf{E}(\tau(\mathbf{X}))$ вместо математического ожидания $\mathbf{X}$. Так как функция τ непрерывна и монотонна, то $\mathbf{E}(\tau(\mathbf{X}+t))$ будет монотонной и непрерывной функцией от t. Поэтому найдется такое значение t, что $\mathbf{E}(\tau(\mathbf{X}+t))=0$. Для нашей схемы серий это означает, что найдутся такие постоянные $t_{k,n}$, для которых $\mathbf{E}(\tau(\mathbf{X}_{k,n}+t_{k,n}))=0$ при всех k и n. Очевидно, что $\{\mathbf{X}_{k,n}+t_{k,n}\}$ снова будет нулевой схемой и что теорему достаточно доказать для этой схемы. Таким образом, если мы положим

$$b_{k,n}=\mathbf{E}(\tau(\mathbf{X}_{k,n})), \tag{7.3}$$

то можно, не ограничивая общности, считать $b_{k,n}=0$ для всех k и n. В этом случае величина

$$A_n=\sum_{k=1}^{r_n}\mathbf{E}(\tau^2(\mathbf{X}_{k,n})) \tag{7.4}$$

будет служить заменой для дисперсии суммы $\mathbf{S}_n$.

Доказательство теоремы 1. В предположении, что $b_{k,n}=0$, мы имеем

$$\varphi_{k,n}(\zeta)-1=\int_{-\infty}^{+\infty}[e^{i\zeta x}-1-i\zeta\tau(x)]F_{k,n}\{dx\}. \tag{7.5}$$

При $|x|\leqslant 1$ подинтегральное выражение по модулю не превосходит $(1/2)\zeta^2x^2$, а при $|x|\geqslant 1$ не превосходит $2+|\zeta|$. Таким образом, при всех ζ

$$\sum_{k=1}^{r_n}|\varphi_{k,n}(\zeta)-1|\leqslant(2+\zeta+\zeta^2)A_n. \tag{7.6}$$

Ввиду (7.1) мы можем (по крайней мере для достаточно больших n) перейти к логарифмам [1]). Тогда мы увидим, что при фиксированном ζ и $n\to\infty$

$$\sum_{k=1}^{r_n}\log\varphi_{k,n}(\zeta)=\sum_{k=1}^{r_n}[\varphi_{k,n}(\zeta)-1]+o(A_n). \tag{7.7}$$

Обозначим характеристическую функцию предельного распределения U через ω. Левая часть (7.7) представляет собой логарифм характеристической функции $\mathbf{S}_n$. Поэтому

$$\sum_{k=1}^{r_n}\log\varphi_{k,n}(\zeta)+i\beta_n\zeta\to\log\omega(\zeta) \tag{7.8}$$

равномерно в каждом интервале $|\zeta|\leqslant\zeta_0$, где ω не имеет нулей. Умножим обе части (7.8) на $1/(2h)$ и проинтегрируем результат по отрезку $|\zeta|<h<\zeta_0$. Интеграл от $i\beta_n\zeta$ при этом исчезнет, и из (7.7) мы получаем

$$\sum_{k=1}^{r_n}\int_{-\infty}^{+\infty}\left(1-\frac{\sin xh}{xh}\right)F_{k,n}\{dx\}+o(A_n)\to-\frac{1}{2h}\int_{-h}^{h}\log\omega(\zeta)\,d\zeta. \tag{7.9}$$

Подинтегральное выражение в левой части больше, чем $1/10h^2x^2$ при $|x|\leqslant 1$ и больше $1/2$ при $|x|\geqslant 1$. В любом случае оно больше $1/10h^2\tau^2(x)$. Следовательно, сумма в левой части (7.9) больше, чем $1/10h^2A_n$, и из (7.9) вытекает, что величины A_n ограничены. Но тогда (7.7) принимает вид

$$\sum_{k=1}^{r_n}[\varphi_{k,n}(\zeta)-1]+i\beta_n\zeta\to\log\omega(\zeta). \tag{7.10}$$

Левая часть является квазихарактеристической функцией вида (1.1) с

$$F_n=\frac{1}{r_n}\sum_{k=1}^{r_n}F_{k,n},\quad c_n=r_n. \tag{7.11}$$

Таким образом, функция ω безгранично делима по теореме § 1. ▶

[1]) Относительно логарифмов характеристических функций см. замечание в начале первого доказательства § 2.

Мы доказали теорему. Более того, как мы знаем, из (7.10) вытекает, что

$$\sum_{k=1}^{r_n} x^2 F_{k,n}\{dx\} \to M\{dx\}, \tag{7.12}$$

где M — каноническая мера и сходимость понимается в собственном смысле. Обратно, если это верно, то из (7.5) с очевидностью вытекает, что выполняется (7.10) с $\beta_n = 0$ и

$$\log \omega(\zeta) = \int_{-\infty}^{+\infty} \frac{e^{i\zeta x} - 1 - i\zeta\tau(x)}{x^2} M\{dx\} \tag{7.13}$$

(это специальный случай леммы 1.1). Таким образом мы установили

Критерий. *Пусть* $\{\mathbf{X}_{k,n}\}$ *— нулевая схема серий, такая, что* $b_{k,n} = 0$ *при всех* k *и* n.

Если имеет место собственная сходимость (7.12), *то распределения сумм* $\mathbf{S}_n$ *сходятся к безгранично делимому распределению с характеристической функцией* (7.13). *В противном случае распределения* $\mathbf{S}_n + \beta_n$ *не сходятся в собственном смысле ни при каком выборе постоянных* β_n.

В применениях некоторые усложнения вносит принятая нами центрировка. Поэтому мы сформулируем наш критерий заново. В таком виде он будет применим к произвольным нулевым схемам. Функции M^+ и M^- были определены в (1.2).

Теорема 2. *Пусть* $\{\mathbf{X}_{k,n}\}$ *— произвольная нулевая схема. Если существует каноническая мера* M, *такая, что во всех точках непрерывности* $x > 0$,

$$\sum_{k=1}^{r_n} [1 - F_{k,n}(x)] \to M^+(x), \quad \sum_{k=1}^{r_n} F_{k,n}(-x) \to M^-(-x) \tag{7.14}$$

и

$$\sum_{k=1}^{r_n} \operatorname{Var}(\tau(\mathbf{X}_{k,n})) \to M\{\overline{-1, 1}\} + M^+(1) + M^-(-1), \tag{7.15}$$

то распределения сумм для схемы серий $\{\mathbf{X}_{k,n} - b_{k,n}\}$ *сходятся к распределению с характеристической функцией* (7.13).

В противном случае ни при каких константах β_n *распределения* $\{\mathbf{S}_n + \beta_n\}$ *не сходятся к распределению вероятностей.*

Доказательство. При $b_{k,n} = 0$ условия (7.14) и (7.15) эквивалентны собственной сходимости в (7.12). Следовательно, для схем, центрированных так, что $\mathbf{E}(\tau(\mathbf{X}_{k,n})) = 0$, утверждение теоремы является лишь иной формой нашего критерия.

Рассмотрим теперь *произвольную* нулевую схему $\{\mathbf{X}_{k,n}\}$. Положим $\mathbf{Y}_{k,n}=\mathbf{X}_{k,n}-\theta_{k,n}$, где константы $\theta_{k,n}$ выбраны так, что $\mathbf{E}(\tau(\mathbf{Y}_{k,n}))=0$. Тогда $\theta_{k,n}\to 0$ равномерно по $k(k=1,\ldots,r_n)$. Поэтому $\{\mathbf{Y}_{k,n}\}$ — нулевая схема, к которой применима наша теорема. Мы должны показать, что условия теоремы полностью эквивалентны соответствующим условиям для схемы $\{\mathbf{Y}_{k,n}\}$. Функция распределения $\mathbf{Y}_{k,n}$ равна $F_{k,n}(x+\theta_{k,n})$, и соотношение (7.14) очевидным образом выполняется, если в нем заменить x на $x+\theta_{k,n}$. Таким образом, соотношения (7.14) и соответствующие соотношения для $\{\mathbf{Y}_{k,n}\}$ равносильны, и потому условия (7.14) необходимы. Для завершения доказательства достаточно проверить, что из них следует

$$\sum_{k=1}^{r_n}\operatorname{Var}(\tau(\mathbf{Y}_{k,n}))\sim\sum_{k=1}^{r_n}\operatorname{Var}(\tau(\mathbf{X}_{k,n})) \tag{7.16}$$

и

$$\sum_{k=1}^{r_n}(b_{k,n}-\theta_{k,n})\to 0. \tag{7.17}$$

Положим

$$\tau(\mathbf{Y}_{k,n})-\tau(\mathbf{X}_{k,n})+\theta_{k,n}=\mathbf{Z}_{k,n}. \tag{7.18}$$

Функция $\tau(x-\theta)-\tau(x)+\theta$ равна нулю всегда, за исключением случая, когда $|x|$ лежит между $1-\theta$ и $1+\theta$. Для этих значений она не превосходит $|\theta|$. Отсюда выводим, что при $n\to\infty$

$$\Sigma\mathbf{E}(|\mathbf{Z}_{k,n}|)\to 0.$$

Последнее соотношение сильнее, очевидно, чем (7.17). Легко видеть, что из него вытекает (7.16).

Пример. *Роль центрирования.* Пусть величины $\mathbf{X}_{k,n}$ $(k=1,2,\ldots,n)$ распределены нормально с математическим ожиданием $n^{-1/4}$ и дисперсией n^{-1}. Тогда величины $\mathbf{S}_n-n^{3/4}$ распределены нормально и существует предельное распределение. Однако $\Sigma\mathbf{E}(\mathbf{X}_{k,n}^2)=1+\sqrt{n}\to\infty$. Это показывает, что в теореме 1 существенно используется введенный там способ центрирования. В иных обстоятельствах члены высшего порядка в разложении логарифмов не будут пренебрежимо малыми (см. также задачи 16 и 17).

§ 8*. Класс *L*

Мы проиллюстрируем силу теоремы (7.1), приведя простое доказательство одной теоремы, установленной П. Леви. Мы еще раз вернемся к частным суммам $\mathbf{S}_n=\mathbf{X}_1+\cdots+\mathbf{X}_n$

*) Здесь рассматриваются специальные вопросы.

последовательности взаимно независимых случайных величин. Однако здесь в отличие от § 5 распределение F_n случайной величины $\mathbf{X}_n$ может зависеть от n. Положим $\mathbf{S}_n^* = (\mathbf{S}_n - b_n)/a_n$. Мы желаем описать все возможные предельные распределения для $\{\mathbf{S}_n^*\}$ при условиях

$$a_n \to \infty, \quad \frac{a_{n+1}}{a_n} \to 1. \tag{8.1}$$

Первое условие исключает сходящиеся ряды $\Sigma \mathbf{X}_k$, которые рассматриваются в § 10. Ситуации, которые исключает второе условие, легче всего поможет уяснить

Пример. Пусть $\mathbf{X}_n$ имеет показательное распределение с математическим ожиданием $n!$. Положим, $a_n = n!$ и $b_n = 0$. Тогда очевидно, что распределение $\mathbf{S}_n^*$ сходится к показательному распределению с математическим ожиданием 1, но эта сходимость имеет место исключительно благодаря подавляющей роли слагаемого $\mathbf{X}_n$. ▶

Следуя Хинчину, обычно говорят, что *распределение принадлежит классу L, если оно является предельным распределением для последовательности* $\{\mathbf{S}_n^*\}$, *удовлетворяющей условиям* (8.1).

При таком определении не ясно, что все распределения класса L безгранично делимы. Мы выведем этот факт из следующей леммы.

Лемма. *Характеристическая функция* ω *принадлежит классу L тогда и только тогда, когда для любого* s $(0<s<1)$ *отношение* $\omega(\zeta)/\omega(s\zeta)$ *есть характеристическая функция.*

[Отсюда следует, что $\omega(\zeta) \neq 0$ при всех ζ.]

Доказательство. а) *Необходимость.* Обозначим характеристическую функцию $\mathbf{S}_n^*$ через ω_n. Пусть $n > m$. Случайная величина $\mathbf{S}_n^*$ может быть представлена в виде суммы $(a_m/a_n)\mathbf{S}_m^*$ и случайной величины, зависящей только от $\mathbf{X}_{m+1}, \ldots, \mathbf{X}_n$ и нормирующих постоянных. Следовательно,

$$\omega_n(\zeta) = \omega_m\left(\frac{a_m}{a_n}\zeta\right) \cdot \varphi_{m,n}(\zeta), \tag{8.2}$$

где $\varphi_{m,n}$ — характеристическая функция. Пусть теперь $n \to \infty$ и $m \to \infty$ так, что $a_m/a_n \to s < 1$ [этого возможно добиться в силу (8.1)]. Левая часть (8.2) сходится к $\omega(\zeta)$, а первый множитель справа — к $\omega(s\zeta)$ (так как сходимость характеристических функций равномерна в каждом конечном интервале [теорема 2, гл. XV, 3]). Из этого мы выводим прежде всего, что ω не имеет

нулей. В самом деле, функции $\varphi_{m,n}$ ограничены, поэтому равенство $\omega(\zeta_0)=0$ влекло бы $\omega(s\zeta_0)=0$ и вообще $\omega(s^k\zeta_0)=0$ при всех $k>0$, в то время как $\omega(s^k\zeta_0)\to 1$. Соответственно отношение $\omega(\zeta)/\omega(s\zeta)$ будет непрерывной функцией, предельной для характеристических функций $\varphi_{m,n}$, а потому оно будет характеристической функцией.

б) *Достаточность.* Мы начнем с тождества

$$\omega(n\zeta)=\omega(\zeta)\cdot\frac{\omega(2\zeta)}{\omega(\zeta)}\cdot\ \cdots\ \cdot\frac{\omega(n\zeta)}{\omega((n-1)\zeta)}. \tag{8.3}$$

В условиях леммы множитель $\dfrac{\omega(k\zeta)}{\omega((k-1)\zeta)}$ будет характеристической функцией некоторой случайной величины $\mathbf{X}_k$. Поэтому $\omega(\zeta)$ — характеристическая функция для $(\mathbf{X}_1+\cdots+\mathbf{X}_n)/n$. ▶

Мы не только доказали теорему, но и установили, что ω является характеристической функцией n-й суммы в схеме серий. Условие (7.1), выделяющее нулевые схемы, тривиальным образом выполняется. Следовательно, функция ω будет безгранично делимой. Чтобы отыскать каноническую меру M, определяющую ω, отметим, что отношение $\omega(\zeta)/\omega(s\zeta)$ также безгранично делимо. Это видно из формулы (8.3). Каноническая мера N, соответствующая $\omega(\zeta)/\omega(s\zeta)$, связана с M тождеством

$$N\{dx\}=M\{dx\}-s^2M\{s^{-1}dx\}. \tag{8.4}$$

В терминах функций M^+ и M^-, введенных в (1.2), это соотношение имеет вид

$$N^+(x)=M^+(x)-M^+(x/s),\quad N^-(-x)=M^-(-x)-M^-(-x/s). \tag{8.5}$$

Мы показали, что если каноническая мера M определяет характеристическую функцию ω класса L, то функции N^+ и N^-, определенные в (8.5), должны быть монотонны при каждом $0<s<1$. Обратно, если это верно, то (8.4) определяет каноническую меру, соответствующую $\omega(\zeta)/\omega(s\zeta)$. Нами доказана, таким образом,

Теорема. *Характеристическая функция ω принадлежит классу L тогда и только тогда, когда она безгранично делима и ее каноническая мера M такова, что две функции* (8.5) *монотонны при любом фиксированном* $0<s<1$.

Примечание. Легко проверить, что указанные функции монотонны тогда и только тогда, когда $M^+(e^x)$ и $M^-(-e^x)$ суть *выпуклые* функции.

§ 9*. Частичное притяжение. «Универсальные законы»

Как мы уже видели, распределение F не обязано принадлежать какой-либо области притяжения. Поэтому возникает вопрос о том, существуют ли какие-нибудь общие законы асимптотического поведения последовательности $\{F^{n*}\}$ сверток F с собой. К сожалению, можно встретить практически любое мыслимое поведение. Мы опишем здесь несколько возможностей главным образом по причине их неожиданного характера.

Говорят, что характеристическая функция φ *принадлежит области частичного притяжения* γ, *если существуют такие нормирующие постоянные* a_r *и* b_r *и такая последовательность натуральных чисел* $n_r \to \infty$, *что*

$$\left[\varphi\left(\frac{\zeta}{a_r}\right) e^{-ib_r\zeta}\right]^{n_r} \to \gamma(\zeta). \tag{9.1}$$

Здесь предполагается, что $|\gamma|$ не равен тождественно единице, т. е. что соответствующее распределение не сосредоточено в одной точке. Таким образом, соотношение (9.1) обобщает понятие области притяжения, позволяя рассматривать пределы подпоследовательностей.

По теореме 2.1 предел γ *будет обязательно безгранично делимым*. Приводимые ниже примеры показывают, что возможны оба крайних случая: *существуют как распределения, не принадлежащие никакой области частичного притяжения, так и распределения, принадлежащие области частичного притяжения любого безгранично делимого закона.*

Примеры. а) В примере (3, е) указана характеристическая функция φ, которая не является устойчивой, но принадлежит своей собственной области частичного притяжения.

б) *Симметричное распределение с медленно меняющимися хвостами не принадлежит ни одной области частичного притяжения.* Допустим, что функция $L(x) = 1 - F(x) + F(-x)$ медленно меняется на бесконечности [т. е. выполняется (5.5)]. По теореме 2, гл. VIII, 9 в этом случае

$$U(x) = \int_{-x}^{x} y^2 F\{dy\} = o(x^2 L(x)), \qquad x \to \infty. \tag{9.2}$$

Согласно критерию § 7, для того чтобы распределение F принадлежало какой-либо области частичного притяжения, необхо-

*) Здесь рассматриваются специальные вопросы.

димо и достаточно условие: для n, пробегающих некоторую подпоследовательность, функции $n[1-F(a_n x)+F(-a_n x)]$ и $na_n^{-2}U(a_n x)$ сходятся во всех точках непрерывности. Из первого условия следует $nL(a_n)\sim 1$, а из второго $nL(a_n)\to\infty$.

в) *Безгранично делимое распределение не обязано принадлежать своей собственной области частичного притяжения.* В самом деле, из доказательства теоремы 2.1 следует, что если φ принадлежит области притяжения γ, то тем же свойством обладает безгранично делимая функция распределения $e^{\varphi-1}$. Последний пример показывает, что функция $e^{\varphi-1}$ может не принадлежать ни одной области частичного притяжения.

г) Мы докажем сейчас одно утверждение, которое подготовит нас к странностям последующих примеров. Рассмотрим произвольную последовательность безгранично делимых характеристических функций $\omega_r=e^{\psi_r}$ с ограниченными показателями. Положим

$$\lambda(\zeta)=\sum_{k=1}^{\infty}\frac{1}{n_k}\psi_k(a_k\zeta). \tag{9.3}$$

Тогда возможно подобрать постоянные $a_k>0$ и целые n_k так, что при всех ζ при $r\to\infty$

$$n_r\lambda\left(\frac{\zeta}{a_r}\right)-\psi_r(\zeta)\to 0. \tag{9.4}$$

Доказательство. Выберем в качестве $\{n_k\}$ монотонную последовательность целых чисел, возрастающую столь быстро, что $n_k/n_{k-1}-2^k\max|\psi_k|$. Левая часть (9.4) не превосходит в этом случае

$$n_r\sum_{k=1}^{r-1}\left|\psi_k\left(\frac{a_k}{a_r}\zeta\right)\right|+\sum_{k=r+1}^{\infty}\frac{1}{2^k}. \tag{9.5}$$

Коэффициенты a_r мы будем выбирать один за другим по следующему правилу. Положим $a_1=1$. При данных $a_1,\ldots,a_{r-1}$ выберем a_r столь большим, чтобы величина (9.5) была меньше $1/r$ для всех $|\zeta|<r$. Это можно сделать, так как первая сумма непрерывно зависит от ζ и равна нулю при $\zeta=0$.

д) *Каждая безгранично делимая характеристическая функция $\omega=e^{\psi}$ имеет область частичного притяжения.* В самом деле, мы знаем, что функция ω может быть представлена как предел последовательности характеристических функций $\omega_k=e^{\psi_k}$ обобщенного пуассоновского типа. Определим λ по (9.3) и положим $\varphi=e^{\lambda}$. Тогда φ будет характеристической функцией и, согласно (9.4),

$$\lim\varphi^{n_r}\left(\frac{\zeta}{a_r}\right)=\lim e^{\psi_r(\zeta)}=\omega(\zeta). \tag{9.6}$$

е) *Другие возможности.* Возьмем две безгранично делимые характеристические функции e^{α} и e^{β}. Выберем ψ_k с четными индексами так, что $\psi_{2k} \to \alpha$, а ψ_k с нечетными индексами так, что $\psi_{2k+1} \to \beta$. Тогда $\varphi = e^{\lambda}$ *принадлежит области частичного притяжения каждого распределения с характеристической функцией вида* $e^{p\alpha+q\beta-1}$, *где* $p+q=1$, *и не принадлежит никаким другим областям частичного притяжения.* Этот пример легко обобщить. В терминах выпуклых множеств можно утверждать, что существует распределение F, принадлежащее области частичного притяжения любого распределения, расположенного в *выпуклой оболочке n заданных безгранично делимых распределений.*

ж) *Пусть дана последовательность безгранично делимых характеристических функций* e^{α_1}, e^{α_2}, ... *Тогда существует функция* $\varphi = e^{\lambda}$, *принадлежащая области частичного притяжения каждой из этих функций.* Разобьем последовательность целых чисел на бесконечное число подпоследовательностей (например, включая в n-ю подпоследовательность все те целые числа, которые делятся на 2^{n-1}, но не делятся на 2^n). Выберем ψ_r в примере (г) так, что $\psi_r \to \alpha_n$ при r, пробегающем n-ю подпоследовательность. В этом случае, как показывает (9.4), функция $\varphi = e^{\lambda}$ обладает требуемым свойством.

з) *«Универсальные законы» Деблина. Существуют функции* φ, *принадлежащие области частичного притяжения любой безгранично делимой функции* ω. Действительно, из определения (9.1) видно, что если φ принадлежит области частичного притяжения каждой из функций $\omega_1, \omega_2, \ldots$ и $\omega_n \to \omega$, то φ принадлежит и области частичного притяжения ω. Далее рассмотрим безгранично делимые характеристические функции, у которых соответствующие канонические меры сосредоточены в конечном числе рациональных точек и приписывают каждой из них рациональный вес. Множество этих функций счетно. Расположим их в последовательность e^{α_1}, e^{α_2}, ... Тогда *каждая* безгранично делимая функция ω будет пределом некоторой подпоследовательности последовательности $\{e^{\alpha_k}\}$. Характеристическая функция φ, построенная по методу предыдущего примера, принадлежит области частичного притяжения каждой из функций α_k, а потому и области частичного притяжения ω.

[**Примечание.** Последний результат был получен в замечательной статье Деблина (1940), продолжившей работу Хинчина (1937). Технические трудности, связанные с этой задачей, в то время были весьма значительными. Явление, описанное в примере (б), было в частных случаях отмечено Гнеденко, Хинчиным и Леви. Интересно отметить те трудности, с которыми сталкиваются в отдельных примерах, если не принимают во внимание основной роли медленной изменчивости.]

§ 10 *. Бесконечные свертки

Пусть $\mathbf{X}_1, \mathbf{X}_2, \ldots$ — независимые случайные величины с характеристическими функциями $\varphi_1, \varphi_2, \ldots$ Как и в (7.2), мы определим «непрерывную урезающую функцию» τ равенствами: $\tau(x)=x$ при $|x|\leqslant 1$ и $\tau(x)=\pm 1$ при $|x|\geqslant 1$. Основная теорема о бесконечных свертках утверждает: *распределения частных сумм* $\mathbf{X}_1+\cdots+\mathbf{X}_n$ *сходятся к некоторому распределению вероятностей* U *тогда и только тогда, когда*

$$\sum_{k=1}^{\infty} \operatorname{Var}(\tau(\mathbf{X}_k))<\infty, \quad \sum_{k=1}^{\infty} P\{|\mathbf{X}_k|>1\}<\infty \tag{10.1}$$

и

$$\sum_{k=1}^{n} \mathbf{E}(\tau(\mathbf{X}_k)) \to b, \tag{10.2}$$

где b — некоторое число.

Частный случай конечных дисперсий был разобран в гл. VIII, 5 с указанием примеров и применений. В общей форме теорема появилась в гл. IX, 9, где сформулированный результат был обобщен (была доказана сходимость ряда $\sum \mathbf{X}_n$ — «теорема о трех рядах»). Было показано, что упомянутая теорема является простым следствием основных теорем, касающихся схемы серий, и поэтому нет необходимости повторять сделанные рассуждения [1]). Поэтому мы удовлетворимся примерами, иллюстрирующими использование характеристических функций.

Примеры. а) *Факторизация равномерного распределения.* Пусть $\mathbf{X}_k=\pm 2^{-k}$ с вероятностью 1/2. В примере (11, в) гл. I было объяснено, что $\Sigma \mathbf{X}_k$ можно интерпретировать как «число, выбранное наудачу на отрезке от -1 до 1». Это равносильно утверждению, что характеристическая функция $(\sin\zeta)/\zeta$ равномерного распределения представима в виде бесконечного произведения характеристических функций $\cos(\zeta/2^k)$. Аналитическое доказательство основано на тождестве

$$\frac{\sin\zeta}{\zeta}=\cos\frac{\zeta}{2}\cdot\cos\frac{\zeta}{4}\cdot\ldots\cdot\cos\frac{\zeta}{2^n}\cdot\frac{\sin(\zeta/2^n)}{\zeta/2^n}, \tag{10.3}$$

*) Здесь рассматриваются специальные вопросы.

[1]) *Прямая* проверка того, что (10.1) и (10.2) обеспечивают равномерную в каждом конечном интервале сходимость произведений $\varphi_1, \varphi_2, \ldots$, представляет собой хорошее упражнение. (Необходимые оценки приведены в начале § 7.) *Необходимость* этих условий менее очевидна. Однако она легко вытекает из того факта, что схема серий с n-й строкой $\mathbf{X}_n, \mathbf{X}_{n+1}, \ldots, \mathbf{X}_{n+r_n}$ должна удовлетворять условиям теоремы 7.2 с $M=0$.

которое доказывается по индукции, исходя из формулы $\sin 2\alpha = 2\sin\alpha\cos\alpha$. При $n\to\infty$ последний множитель сходится к 1 равномерно в каждом конечном интервале.

Отметим, что произведение четных членов снова соответствует сумме независимых случайных величин. Мы знаем из примера (11, г) гл. I, что эта сумма имеет *сингулярное распределение* канторовского типа [1]) [см. задачи 5,7 и 18].

б) Пусть $\mathbf{Y}_k$ имеет плотность $\frac{1}{2}e^{-|x|}$ с характеристической функцией $1/(1+\zeta^2)$. Ряд $\sum \mathbf{Y}_k/k$ сходится. Для соответствующей характеристической функции $\frac{\pi\zeta}{\operatorname{sh}\pi\zeta}$ мы получаем каноническое разложение на множители (sh — гиперболический синус). Используя задачу 7, гл. XV, 9, находим, что *плотность* $\sum \mathbf{Y}_k/k$ *равна* $\frac{1}{2+e^x+e^{-x}} = \frac{1}{4\,(\operatorname{ch}(x/2))^2}$.

§ 11. Многомерный случай

Развитая в этой главе теория без существенных изменений переносится на многомерный случай. Мы не будем останавливаться на деталях. В канонической форме для безгранично делимых распределений лучше выделять нормальную компоненту и рассматривать только канонические меры, не имеющие атома в начале координат. Прежние формулы не потребуют никаких изменений, если интерпретировать ζx как скалярное произведение (см. гл. XV, 7). Для определенности мы выпишем формулу для двумерного случая.

Мера M является канонической, если по отношению к ней интегрируема функция $1/(1+x_1^2+x_2^2)$ и если она не имеет атома в начале координат. Выберем подходящую *одномерную* центрирующую функцию, скажем $\tau(x)=\sin x$, или функцию, указанную в (7.2). Положим

$$\psi(\zeta_1,\zeta_2) = \int \frac{e^{i(\zeta_1 x_1+\zeta_2 x_2)} - 1 - i\zeta_1\tau(x_1) - i\zeta_2\tau(x_2)}{x_1^2+x_2^2} M\{dx\}, \quad (11.1)$$

где интеграл распространяется на всю плоскость. Тогда $\omega = e^{\psi}$ будет двумерной безгранично делимой характеристической функцией. Наиболее общая безгранично делимая характеристическая

[1]) Шоке́ дал привлекательное геометрическое доказательство, применимое и к более общим бесконечным сверткам. Оно приведено в статье Tortrat A., *J. Math. Pures. Appl.*, **39** (1960), 231—273.

функция получается умножением на нормальную характеристическую функцию.

Переход к полярным координатам может сделать ситуацию легче воспринимаемой. Положим

$$\zeta_1 = \varphi\cos\varphi, \quad \zeta_2 = \rho\sin\varphi, \quad x = r\cos\theta, \quad y = r\sin\theta. \tag{11.2}$$

Определим каноническую меру в полярных координатах следующим образом. Выберем при каждом θ, $-\pi<\theta\leqslant\pi$, одномерную каноническую меру Λ_θ, сосредоточенную на $\overline{0, \infty}$; выберем далее конечную меру W на $-\pi<\theta\leqslant\pi$ (на окружности). Тогда M можно определить с помощью рандомизации по параметру θ, и (с тривиальными изменениями центрирующей функции) мы можем переписать (11.1) в форме

$$\psi(\zeta_1, \zeta_2) = \int_{-\pi}^{\pi} W\{d\theta\} \int_{0+}^{\infty} \frac{e^{ir\rho\cos(\varphi-\theta)} - 1 - i\rho\tau(r)\cos(\varphi-\theta)}{r^2} \Lambda_\theta\{dr\}. \tag{11.3}$$

(В такой форме, добавляя к Λ_θ атом в нуле, можно охватить и случай нормальной компоненты.)

Пример. *Устойчивые распределения.* Положим, по аналогии с одномерным случаем, $\Lambda_\theta\{dr\} = r^{-\alpha+1}\,dr$. Можно было бы добавить произвольный множитель C_θ, но это свелось бы к замене меры W. Как мы видели в § 4, при такой мере Λ_θ (11.3) принимает вид

$$\psi(\zeta_1, \zeta_2) = -C\rho^\alpha \int_{-1}^{\pi} |\cos(\varphi-\theta)|^\alpha \left(1 \mp \operatorname{tg}\frac{\pi}{2^\alpha}\right) W\{d\theta\}, \tag{11.4}$$

где верхний (нижний) знак выбирается для $\varphi-\theta>0$ ($\varphi-\theta<0$). Из (11.4) ясно, что e^ψ — строго устойчивая характеристическая функция. Так же, как и в § 4, можно установить, что других строго устойчивых функций нет. В $\mathscr{R}^2$ существуют, однако, устойчивые распределения $e^{\psi_{1,\delta}}$ с логарифмическим членом в показателе.

При $\alpha=1$ и равномерном распределении W мы получаем характеристическую функцию $e^{-\alpha\varphi}$ *симметричного распределения Коши в* $\mathscr{R}^2$ [см. пример (7, г) гл. XV и задачи 20—22]. ▶

§ 12. Задачи

1. При доказательстве леммы 1.4 было установлено, что если ψ представляется в канонической форме (1.12), то функция

$$\psi(\xi) - \frac{\psi(\xi+h)+\psi(\xi-h)}{2} = \chi(\xi) \tag{12.1}$$

пропорциональна некоторой характеристической функции Докажите обратное если ψ непрерывна и $\chi(\zeta)/\chi(0)$ — характеристическая функция (при любом выборе h), то ψ отличается от (1 12) только на линейную функцию

Указание Докажите, что все решения однородного уравнения линейны.

2. Функция ψ представляет собой логарифм безгранично делимой характеристической функции тогда и только тогда, когда при каждом $\delta>0$ функция

$$\psi(\zeta)-\frac{1}{2\delta}\int_{-\delta}^{\delta}\psi(\zeta-s)\,ds=\chi(\zeta) \tag{12.2}$$

пропорциональна некоторой характеристической функции и $\psi(0)=0$, $\psi(-\zeta)=\overline{\psi(\zeta)}$.

Указание Отличие этой задачи от предыдущей задачи и леммы 1 4 в том, что вместо распределения, сосредоточенного в $\pm h$, здесь используется равномерное распределение См следующую задачу

3. Пусть R — произвольное симметричное распределение вероятностей с конечной дисперсией Если ω — безгранично делимая характеристическая функция и $\psi=\log\omega$, то функция

$$\chi=\psi-R*\psi$$

пропорциональна некоторой характеристическои функции Верно и обратное утверждение (см задачи 1 и 2)

4 Пусть ω — безгранично делимая характеристическая функция. Тогда существуют такие постоянные a и b, что $|\log\omega(\zeta)|<a+b\zeta^2$ при всех ζ.

5. *Дробовой эффект в вакуумных лампах* В примере (3 з) гл VI мы имели дело со схемои серии, в которой $\mathbf{X}_{k,n}$ обладает характеристической функцией

$$\varphi_{k,n}(\zeta)=1+\alpha h\,[e^{i\zeta I(kh)}-1],$$

где $h=1/\sqrt{n}$. Покажите, что характеристические функции сумм $\mathbf{S}_n=\mathbf{X}_{1,n}+\ldots+\mathbf{X}_{n,n}$ сходятся к пределу e^{ψ}, где

$$\psi(\zeta)=\alpha\int_0^{\infty}[e^{i\zeta I(x)}-1]\,dx,$$

e^{ψ} является характеристической функциеи случайной величины $\mathbf{X}(t)$. Дифференцированием получаем *теорему Кемпбелла* гл. VI, (3,5).

6. Пусть $U=\Sigma\mathbf{X}_n/n$, где случайные величины $\mathbf{X}_k$ независимы и имеют одну и гу же плотность $^1/_2e^{-|x|}$. Покажите [1]), что случайная величина U безгранично делима и что соответствующая каноническая мера равна

$$M\{dx\}=|x|\,\frac{e^{-|x|}}{1-e^{-|x|}}\,dx.$$

---

[1]) Характеристическая функция ω равна бесконечному произведению, которое совпадает с хорошо известным каноническим произведением для функции $\dfrac{\pi(\zeta)}{e^{\pi|\zeta|}-e^{-\pi|\zeta|}}$.

[Не требуется никаких вычислений, кроме суммирования геометрической прогрессии.]

7. Пусть $\mathbf{P}(s)=\Sigma p_k s^k$, где $p_k \geqslant 0$ и $\Sigma p_k=1$. Допустим, что $\mathbf{P}(0)>0$ и что $\log \frac{\mathbf{P}(s)}{\mathbf{P}(0)}$ разлагается в степенной ряд с положительными коэффициентами. Если φ — характеристическая функция какого-либо распределения F, то $\mathbf{P}(\varphi)$ — безгранично делимая характеристическая функция. Опишите соответствующую каноническую меру M в терминах F^{n*}.

Интересный частный случай: если $0 \leqslant a<b<1$, то функция

$$\frac{1-b}{1-a}\cdot\frac{1-a\varphi}{1-b\varphi}$$

будет безгранично делимой характеристической функцией.

7а. Интерпретируйте $\mathbf{P}(\varphi)$ в терминах рандомизации и подчиненных процессов, принимая во внимание тот факт, что $\mathbf{P}$ — производящая функция безгранично делимой целочисленной случайной величины.

8. Пусть $\mathbf{Y}$ — *положительная* устойчивая случайная величина с показателем $\alpha<1$, и пусть $\mathbf{X}$ — устойчивая величина с показателем β. Докажите, что $\mathbf{XY}^{1/\alpha}$ — *устойчивая величина с показателем* $\alpha\beta$ [см. пример (2, ж) гл. VI. О характеристических функциях произведений см. гл. XV, (9.4)].

9. Пусть ω — такая характеристическая функция, что $\omega^2(\zeta)=\omega(a\zeta)$ и $\omega^3(\zeta)=\omega(b\zeta)$. Тогда ω устойчива.

[Пример (3, е) показывает, что одного первого условия недостаточно. Показатели степени 2 и 3 могут быть заменены любыми двумя взаимно простыми числами.]

10. Покажите, что простая лемма 2 гл. VIII, 8 применима не только к монотонным функциям, а и к логарифмам характеристических функций. Выведите отсюда, что если при всех n $\omega^n(\zeta)=\omega(a_n\zeta)$, то $\log\omega(\zeta)=A\zeta^\alpha$ при $\zeta>0$, где A — комплексная постоянная.

11. *Продолжение.* Используя результат задачи 23 из гл. VIII, 10, покажите (не применяя теорию устойчивых распределений), что если ω — устойчивая характеристическая функция, то при $\zeta>0$ или $\log\omega(\zeta)=A\zeta^\alpha+ib\zeta$, или $\log\omega(\zeta)=A\zeta+ib\zeta\log\zeta$, где b — вещественное число.

12. Пусть F сосредоточено на $\overline{0,\infty}$ и $1-F(x)\sim x^{-\alpha}L(x)$ с $0\leqslant\alpha<1$, где L — медленно меняющаяся на бесконечности функция. Докажите, что $1-\varphi(\zeta)\sim A\zeta^\alpha L(1/\zeta)$ при $\zeta\to 0+$.

13. *Продолжение.* Опираясь на результаты § 5, докажите *обратное* утверждение, а также докажите, что $A=-\Gamma(1-\alpha)e^{i\pi\alpha/2}$.

14. *Продолжение.* Докажите индукцией по k следующее утверждение: для того чтобы при $x\to\infty$ было $1-F(x)\sim ax^{-\alpha}L(x)$, где L — медленно меняющаяся функция и $k<\alpha<k+1$, необходимо и достаточно, чтобы при $\zeta\to 0+$ было

$$\varphi(\zeta)-1-\frac{\mu_1(i\zeta)}{1!}-\ldots-\frac{\mu_k(i\zeta)^k}{k!}\sim A\zeta^\alpha L\left(\frac{1}{\zeta^{-1}}\right). \qquad (*)$$

Тогда автоматически $A=-a\Gamma(k-\alpha)e^{-i\frac{1}{2}\pi\alpha}$, где $\Gamma(k-\alpha)$ определяется по индукции из формулы $\Gamma(2-\alpha)=(1-\alpha)\Gamma(1-\alpha)$.

15. Сформулируйте (слабый) закон больших чисел для схемы серий как частный случай общей теоремы § 7.

16. Пусть $\{\mathbf{X}_{k,\,n}\}$ — нулевая схема, в которой суммы по строкам имеют предельное распределение с канонической мерой M. Тогда при $x>0$

$$\mathbf{P}\left\{\max\left[\mathbf{X}_{1,\,n},\ \ldots,\ \mathbf{X}_{r_n,\,n}\right]\leqslant x\right\}\to e^{-M^{+}(x)}.$$

Сформулируйте обратное утверждение.

17. Пусть $\{\mathbf{X}_{k,\,n}\}$ — нулевая схема из симметричных случайных величин, в которой суммы по строкам имеют предельное распределение с канонической мерой M, имеющей атом веса σ^2 в нуле. Покажите, что распределение случайной величины $\mathbf{S}_n^{*}=\sum\mathbf{X}_{k,\,n}^2-\sigma^2$ сходится к распределению, определяемому мерой M_{*} без атома в нуле и такой, что $M_{*}^{+}(x)=2M^{+}(\sqrt{x})$ при $x>0$.

18. Пусть $0<r_j<1$ и $\Sigma r_j<\infty$. При любых вещественных a_j бесконечное произведение

$$\frac{1-r_1}{1-r_1e^{ia_1\zeta}}\cdot\frac{1-r_2}{1-r_2e^{ia_2\zeta}}\cdots$$

сходится и представляет собой безгранично делимую характеристическую функцию. *Указание.* В соответствии с задачей 7 каждый множитель безгранично делим.

19. Используйте метод примера (9, г) для построения распределения F, такого, что $\limsup F^{n*}(x)=1$ и $\liminf F^{n*}(x)=0$ во всех точках.

20. Возьмем в (11.4) в качестве W равномерное распределение. Тогда

$$\psi(\zeta_1,\,\zeta_2)=-c\left[\zeta_1^2+\zeta_2^2\right]^{1/2\alpha},$$

так что e^{ψ} — симметричное устойчивое распределение.

21. Возьмем в (11.4) в качестве W распределение с весами $^1/_4$ в четырех точках 0, π, $^1/_2\pi$, $-{}^1/_2\pi$. Тогда (11.4) определяет двумерную характеристическую функцию пары *независимых* одномерных устойчивых случайных величин.

22. Пусть W в (11.4) сосредоточено в двух точках σ и $\sigma+\pi$. Тогда (11.4) задает *вырожденную* характеристическую функцию такой пары, что

$$\mathbf{X}_1\sin\sigma-\mathbf{X}_2\cos\sigma=0.$$

Вообще любой дискретной мере W соответствует свертка вырожденных распределений. Объясните (11.4) с помощью предельного перехода.

ГЛАВА XVIII

ПРИМЕНЕНИЕ МЕТОДОВ ФУРЬЕ К СЛУЧАЙНЫМ БЛУЖДАНИЯМ

В этой главе в значительной степени повторяются вопросы, уже разобранные в гл. XII (по этой причине применения сведены до минимума). Была предпринята серьезная попытка построить эту главу независимо от других и сделать ее доступной при минимуме предварительных сведений (исключая анализ Фурье из гл. XV). Излагаемая теория совсем не связана с содержанием двух предшествующих глав.

§ 1. Основное тождество

Всюду в этой главе $\mathbf{X}_1, \mathbf{X}_2, \ldots$ обозначают взаимно независимые случайные величины с одним и тем же распределением F и характеристической функцией φ. Как обычно, мы полагаем $\mathbf{S}_0 = 0$ и $\mathbf{S}_n = \mathbf{X}_1 + \cdots + \mathbf{X}_n$; последовательность $\{\mathbf{S}_n\}$ образует случайное блуждание, порожденное F.

Пусть A — произвольное множество на вещественной прямой и A' — его дополнение (в большинстве применений A' будет конечным или бесконечным интервалом). Если I — подмножество (интервал) из A' и если

$$\mathbf{S}_1 \in A, \ \ldots, \ \mathbf{S}_{n-1} \in A, \quad \mathbf{S}_n \in I \quad (I \subset A'), \tag{1.1}$$

то мы скажем, что *вход* [1]) в A' *происходит в момент* n *в точке* I.

Так как вход в A' не обязан быть достоверным событием, то *момент вхождения* $\mathbf{N}$ может оказаться несобственной случайной величиной, равно как и *точка* $\mathbf{S}_N$ *первого вхождения*. Для совместного распределения пары $(\mathbf{N}, \mathbf{S}_N)$ мы примем обозначение

$$\mathbf{P}\{\mathbf{N} = n, \ \mathbf{S}_n \in I\} = H_n\{I\}, \quad n = 1, 2, \ldots \tag{1.2}$$

Таким образом, $H_n\{I\}$ есть вероятность события (1.1). Дополнительное соглашение, что $H_n\{I\} = 0$ при $I \subset A$ позволяет определить распределение (1.2) для всех I на числовой прямой. Вероятности (1.2) будут называться *вероятностями достижения*. Их изучение неотделимо от изучения случайного блуждания до

[1]) Иногда мы будем говорить «достижение A'». — *Прим. перев.*

момента первого вхождения в A', т. е. иными словами, от *случайного блуждания, ограниченного множеством* A. Положим для $I \subset A$ и $n=1, 2, \ldots$

$$G_n\{I\} = \mathbf{P}\{\mathbf{S}_1 \in A, \ldots, \mathbf{S}_{n-1} \in A, \mathbf{S}_n \in I\}, \tag{1.3}$$

т. е. $G_n(I)$ есть вероятность того, что в момент n блуждание приводит в $I \subset A$ и что до момента n не произошло вхождения в A'. Мы распространим это определение на все множества I, считая $G_n\{I\}=0$ при $I \subset A'$.

Отметим, что, согласно нашему определению,

$$G_n\{A\} = 1 - \mathbf{P}\{\mathbf{N} \leqslant n\}. \tag{1.4}$$

Случайная величина $\mathbf{N}$ будет собственной тогда и только тогда, когда эта величина стремится к нулю при $n \to \infty$. Рассматривая положение $\mathbf{S}_n$ блуждающей точки в моменты $n=1, 2, \ldots$, видим, что при $I \subset A'$

$$H_{n+1}\{I\} = \int_A G_n\{dy\}\, F\{I-y\}, \tag{1.5а}$$

а при $I \subset A$

$$G_{n+1}\{I\} = \int_A G_n\{dy\}\, F\{I-y\}. \tag{1.5б}$$

Условимся теперь обозначать буквой G_0 *распределение вероятностей, сосредоточенное в нуле*. Тогда соотношения (1.5) верны при $n=0, 1, 2, \ldots$ и с их помощью можно последовательно вычислить все вероятности H_n и G_n. Эти два соотношения можно объединить в одно. Разобьем произвольное множество I на числовой прямой на компоненты IA' и IA и применим (1.5) к этим компонентам. Вспоминая, что H_n и G_n сосредоточены на A' и A соответственно, получаем для $n=0, 1, \ldots$ и произвольного I

$$H_{n+1}\{I\} + G_{n+1}\{I\} = \int_A G_n\{dy\}\, F\{I-y\}. \tag{1.6}$$

Интеграл в правой части представляет собой свертку $G_n * F$.

Частный случай $A=\overline{0, \infty}$ разбирался в гл. XII, 3. Соотношение гл. XII, (3.5) играло там ту же роль, что (1.5) здесь. Мы могли бы повторить здесь прежний путь и вывести интегральное уравнение типа Винера — Хопфа, аналогичное гл. XII, (3.9). Оно снова имело бы только одно вероятностное решение (хотя эта единственность решения не абсолютна). Однако на этот раз предпочтительнее опираться на мощный метод анализа Фурье.

Нас интересует распределение пары $(\mathbf{N}, \mathbf{S}_{\mathbf{N}})$. Так как $\mathbf{N}$ — целочисленная случайная величина, то мы будем использовать производящую функцию для $\mathbf{N}$ и характеристическую для

$\mathbf{S}_N$. Соответственно этому положим

$$\chi(s,\zeta)=\sum_{n=1}^{\infty}s^n\int_{A'}e^{i\zeta x}H_n\{dx\},\quad \gamma(s,\zeta)=\sum_{n=0}^{\infty}s^n\int_{A}e^{i\zeta x}G_n\{dx\}. \tag{1.7}$$

(Для ясности здесь указана существенная часть области интегрирования; можно было бы взять интеграл в пределах от $-\infty$ до ∞. Во втором ряде член с $n=0$ равен 1.)

Так как полные массы распределений G_n и H_n не превосходят единицы, то оба ряда в (1.7) сходятся во всяком случае при $|s|<1$. Для таких s мы выводим из (1.6) *основное тождество*

$$\chi+\gamma=1+s\gamma\varphi \tag{1.8}$$

или

$$1-\chi=\gamma[1-s\varphi]. \tag{1.9}$$

(Другое доказательство см. в задаче 10.)

В принципе χ и γ могут быть вычислены, исходя из (1.5), и поэтому тождество (1.9) может на первый взгляд показаться избыточным. В действительности прямые вычисления редко можно довести до конца. В то же время непосредственно из (1.9) можно извлечь много ценной информации.

Пример. Пусть F обозначает двустороннее показательное распределение с характеристической функцией $\varphi(\zeta)=1/(1+\zeta^2)$, и пусть $A=-a, a$. В силу свойств показательного распределения, обсуждавшихся в первой главе, кажется правдоподобным, что распределения H_n должны быть показательными и одинаковыми. Если это было бы верно, то по причинам симметрии мы имели бы соотношение

$$\chi(s,\zeta)=P(s)\frac{e^{i\zeta a}}{1-i\zeta}+P(s)\frac{e^{-i\zeta a}}{1+i\zeta}, \tag{1.10}$$

где $2P(s)$ — производящая функция для момента $\mathbf{N}_1$ первого вхождения во множество $|x|\geqslant a$. Так как правая часть (1.9) обращается в нуль при $\zeta=\pm i\sqrt{1-s}$, то для этих ζ величина (1.10) должна быть равна единице. Таким образом,

$$P(s)=\left[\frac{e^{-a\sqrt{1-s}}}{1+\sqrt{1-s}}+\frac{e^{a\sqrt{1-s}}}{1-\sqrt{1-s}}\right]^{-1}. \tag{1.11}$$

Исходное допущение легко проверить, опираясь на (1.5а), и без всяких вычислений: если F имеет показательную плотность, то то же самое будет верно для H_{n+1} при любом виде G_n. Этим строго обосновывается окончательный результат, содержащийся в формулах (1.10) и (1.11). Прямое вычисление было бы сложнее. (Другие примеры см. в задачах 3—6.) ▶

§ 2 *. Конечные интервалы. Вальдовская аппроксимация

Теорема. *Пусть* $A=\overline{-a, b}$ *— конечный интервал, содержащий точку нуль, и пусть пара* $(\mathbf{N}, \mathbf{S}_{\mathbf{N}})$ *соответствует достижению дополнения* A' *к* A.

Тогда случайные величины $\mathbf{N}$ *и* $\mathbf{S}_{\mathbf{N}}$ *будут собственными. Ряд для производящей функции*

$$\sum_{n=0}^{\infty} s^n \mathbf{P}\{\mathbf{N}>n\} = \sum_{n=0}^{\infty} s^n G_n\{A\} \tag{2.1}$$

сходится при некотором [1]) $s>1$, *и потому* $\mathbf{N}$ *имеет моменты всех порядков. Величина* $\mathbf{S}_{\mathbf{N}}$ *имеет математическое ожидание тогда и только тогда, когда распределение* F, *определяющее случайное блуждание, имеет математическое ожидание* μ. *В последнем случае*

$$\mathbf{E}(\mathbf{S}_{\mathbf{N}}) = \mu \cdot \mathbf{E}(\mathbf{N}). \tag{2.2}$$

Доказательство. Пусть r — фиксированное число. При $\nu r<n<(\nu+1)r$ тривиально верно неравенство

$$G_n\{A\} \leqslant \mathbf{P}\{-a < \mathbf{S}_{kr} < b \quad \text{при} \quad k=1, \ldots, \nu\}. \tag{2.3}$$

Событие, стоящее в правой части, может произойти только если $|\mathbf{S}_{kr} - \mathbf{S}_{(k-1)r}|<a+b$. Эти ν событий взаимно независимы и имеют одну и ту же вероятность $\eta=\mathbf{P}\{\mathbf{S}_r|<a+b\}$. Таким образом, $G_n(A)\leqslant\eta^{\nu}\leqslant\eta^{n/r-1}$. Следовательно, ряд в правой части (2.1) во всяком случае сходится при $0<s<\eta^{-1/r}$. Величину $\eta^{-1/r}$ можно сделать больше единицы, выбирая r достаточно большим. Из сказанного следует, в частности, что $G_n\{A\}\to 0$. В силу (1.4) отсюда вытекает, что $\mathbf{N}$—собственная случайная величина.

Далее, при $t>0$

$$\mathbf{P}\{|\mathbf{S}_{\mathbf{N}}|>t+a+b\} \leqslant$$

$$\leqslant \sum_{n=1}^{\infty} \mathbf{P}\{\mathbf{N}=n\}\,\mathbf{P}\{|\mathbf{X}_n|>t\} = \mathbf{P}\{|\mathbf{X}_1|>t\}. \tag{2.4}$$

Математическое ожидание $\mu=\mathbf{E}(\mathbf{X}_1)$ существует тогда и только тогда, когда правая часть интегрируема на $\overline{0, \infty}$. Если она интегрируема, то интегрируема и левая часть и $\mathbf{E}(\mathbf{S}_{\mathbf{N}})$ существует. С другой стороны, неравенство $|\mathbf{X}_1|>t+a+b$ влечет неравенство $|\mathbf{S}_{\mathbf{N}}|>t$. Поэтому из существования $\mathbf{E}(\mathbf{S}_{\mathbf{N}})$ следует суще-

*) Параграф не используется в дальнейшем.

[1]) Это утверждение известно статистикам как лемма Стейна. Другое доказательство см. в задаче 7.

ствование $\mu = E(X_1)$. Мы знаем, что (1.9) выполняется для всех s, для которых ряд (2.1) сходится. Если математические ожидания существуют, то

$$E(\mathbf{S}_N) = \frac{\partial \chi(1, 0)}{\partial \zeta} = \varphi'(0) \cdot \gamma(1, 0) = \mu \cdot \mathbf{E}(\mathbf{N}). \qquad (2.5) \blacktriangleright$$

Тождество (2.2) было получено (при слабых ограничениях) А. Вальдом. Для случая полуконечного интервала оно согласуется с гл. XII, (2.7). Данное в указаном разделе доказательство (с привлечением мартингалов) без изменений приложимо и к случаю конечного A. Это доказательство можно видоизменить так, чтобы оно приводило к тождеству (2.7).

Чтобы избежать использования мнимого аргумента, входящего в характеристическую функцию, введем обозначение

$$f(\lambda) = \int_{-\infty}^{+\infty} e^{-\lambda x} F\{dx\}. \qquad (2.6)$$

Тождество Вальда[1]). *Если* $f(\lambda) < \infty$ *при* $\lambda_0 < \lambda < \lambda_1$, *то в этом интервале*

$$\mathbf{E}(f^{-\mathbf{N}}(\lambda) \cdot e^{-\lambda \mathbf{S}_N}) = \chi(f^{-1}(\lambda), i\lambda) = 1. \qquad (2.7)$$

Доказательство. Из самого вывода тождества (1.9) очевидно, что оно выполняется при тех мнимых значениях аргумента ζ, при которых γ и χ аналитичны. Формальная подстановка в (1.9) $s = f^{-1}(\lambda)$ и $\zeta = i\lambda$ приводит к (2.7). Чтобы оправдать ее, достаточно показать, что ряд (2.1) сходится для $s = f^{-1}(\lambda)$. Теперь если $f(\eta) < \infty$, то

$$G_n\{A\} \leqslant \mathbf{P}\{-a < \mathbf{S}_n < b\} \leqslant$$

$$\leqslant e^{(a+b)|\eta|} \int_{-a}^{b} e^{-nx} F^{n*}\{dx\} \leqslant e^{(a+b)|\eta|} \cdot f^n(\eta). \qquad (2.8)$$

Поэтому ряд (2.1) сходится в точке $s = f^{-1}(\lambda)$, если $f(\lambda) > f(\eta)$. Так как выбор η находится в нашем распоряжении, то тем самым (2.7) доказано для всех точек, кроме тех, где f достигает своего минимума. Но, будучи выпуклой функцией, f может иметь не

[1]) Вальд применял (2.7) в связи с последовательным анализом. Это было еще до 1945 г. и до того времени, когда началось систематическое изучение общих случайных блужданий. Естественно поэтому, что его условия были жесткими, а методы — трудными. Однако их влияние на статистическую литературу все еще продолжается. Рассуждения в тексте используют одну идею Миллера (H. D. Miller (1961)).

более одной точки минимума. Для нее (2.7) верно по непрерывности. ▶

Покажем, как применяется (2.7) к построению *оценок для распределения N*. Положим

$$p_k = \mathbf{P}\{\mathbf{N}=k,\ \mathbf{S}_\mathbf{N} \geqslant b\}, \quad q_k = \mathbf{P}\{\mathbf{N}=k,\ \mathbf{S}_\mathbf{N} \leqslant -a\} \tag{2.9}$$

и обозначим соответствующие производящие функции через $P(s)$ и $Q(s)$. (Тогда $P+Q$ будет производящей функцией для $\mathbf{N}$.) Допустим теперь, что a и b велики по сравнению с математическим ожиданием и дисперсией F. Тогда $\mathbf{S}_\mathbf{N}$, по-видимому, будет близко к b или $-a$. Если бы только эти числа были возможными значениями $\mathbf{S}_\mathbf{N}$, то тождество (2.7) имело бы вид

$$P\left(\frac{1}{\hat{f}(\lambda)}\right)e^{-\lambda b} + Q\left(\frac{1}{\hat{f}(\lambda)}\right)e^{\lambda a} = 1. \tag{2.10}$$

Естественно ожидать, что при сделанных предположениях (2.10) будет приближенно выполняться. Функция $\hat{f}$ выпукла, и обычно можно найти такой интервал $s_0<s<s_1$, что на нем уравнение

$$s\hat{f}(\lambda) = 1 \tag{2.11}$$

имеет два корня $\lambda_1(s)$ и $\lambda_2(s)$, непрерывно зависящих от s. Подставляя их в (2.10), мы получаем *два линейных уравнения для производящих функций* P и Q и таким образом можем найти (по крайней мере приближенно) распределение $\mathbf{N}$ и вероятности выхода вправо или влево.

Примеры. а) *Биномиальное случайное блуждание.* Пусть F имеет атомы веса p и q в точках $+1$ и -1 соответственно $(p+q=1)$. Тогда $-a$ и b действительно будут единственно возможными значениями $\mathbf{S}_\mathbf{N}$, и потому описанный выше метод будет точным. Уравнение (2.11) имеет два корня, такие, что $\tau_j = e^{-\lambda_j}$ определяются равенствами

$$\tau_1(s) = \frac{1+\sqrt{1-4pqs^2}}{2ps}, \quad \tau_2(s) = \frac{1-\sqrt{1-4pqs^2}}{2ps}. \tag{2.12}$$

Упомянутый метод приводит к формулам

$$P(s) = \frac{\tau_2^a - \tau_1^a}{\tau_2^{a+b} - \tau_1^{a+b}}, \quad Q(s) = \left(\frac{q}{p}\right)^a \frac{\tau_2^b - \tau_1^b}{\tau_2^{a+b} - \tau_1^{a+b}}. \tag{2.13}$$

Легко видеть, что этот результат лишь обозначениями отличается от формул **1**, гл. XIV, (4.11) для вероятностей разорения [1]) и что линейные уравнения, полученные из (2.10), эквивалентны граничным условиям, использованным в **1**, гл. XIV, 4.

[1]) Достаточно заменить U_z на Q, z на a и a на $a+b$ соответственно.

б) *Арифметические распределения* с конечным числом атомов приводят к алгебраическому уравнению (2.11) для $e^{-\lambda}$. В **1**, гл. XIV, 8 было описано, как можно найти *точные* выражения для P и Q, используя все корни этого уравнения. Более того, было указано, как простое видоизменение нашей несложной аппроксимации приводит к *точным неравенствам* для вероятностей p_k и q_k. Этот метод годится и для произвольных распределений, сосредоточенных на конечном интервале.

§ 3. Факторизация Винера — Хопфа

Вернемся к обозначениям § 1 и выберем $A=\overline{-\infty, 0}$. Тогда $(\mathbf{N}, \mathbf{S_N})$ будет точкой первого входа в открытый интервал $A'=\overline{0, \infty}$ и

$$\chi(s, \zeta) = \mathbf{E}(s^{\mathbf{N}} e^{i\zeta \mathbf{S_N}}). \tag{3.1}$$

Следующая лемма может быть не так интересна сама по себе, но она дает ключ ко всем дальнейшим результатам. [Это лишь иная формулировка гл. XII, (9.3).]

Лемма 1[1]). *При* $|s|<1$

$$\log \frac{1}{1-\chi(s, \zeta)} = \sum_{n=1}^{\infty} \frac{s^n}{n} \int_{0+}^{\infty} e^{i\zeta x} F^{n*}\{dx\}. \tag{3.2}$$

Доказательство. Переходя в (1.9) к логарифмам[2]), получаем

$$\sum_{1}^{\infty} \frac{s^n}{n} \varphi^n(\zeta) = \log \frac{1}{1-\chi(s, \zeta)} + \log \gamma(s, \zeta). \tag{3.3}$$

Предполагая s достаточно малым, разложим логарифмы в ряд по степеням χ и $(\gamma - 1)$ соответственно. Первое разложение имеет положительные коэффициенты, так что первый логарифм представляет собой преобразование Фурье — Стильтьеса некоторой мерой, сосредоточенной на $\overline{0, \infty}$. С другой стороны, $\gamma - 1$ есть разность преобразований, соответствующих двум мерам,

[1]) Принадлежит Г. Бакстеру (G. Baxter). Упрощенное (но все еще технически сложное) доказательство было дано Спицером (Spitzer F., *Trans. Amer. Math. Soc.*, **94** (1960), 150—169).

[2]) Логарифм характеристической функции определяется однозначно по непрерывности в окрестности начала, где она не имеет нулей (см. первое доказательство гл. XVII, 2). При $|s|<1$ ни один из членов в (3.3) не может обратиться в нуль.

сосредоточенным на $\overline{-\infty, 0}$. То же самое ноэтому верно и для $\log \gamma$. Левая часть (3.3) есть преобразование меры $\sum n^{-1} s^n \overline{F^{n*}}$. Приравнивая друг другу компоненты, сосредоточенные на $0, \infty$, приходим к (3.2). Из аналитичности обеих частей этого соотношения вытекает, что оно верно при всех $|s|<1$. ▷

В качестве простого следствия получается

Лемма 2. *Пусть* $(\tilde{\mathbf{N}}, \mathbf{S}_{\tilde{\mathbf{N}}})$ — *точка первого входа в* $\overline{-\infty, 0}$. *Определим* $\tilde{\chi}$ *по аналогии с* (3.1). *Тогда при* $|s|<1$

$$\log \gamma(s, \zeta) = \log \frac{1}{1-\tilde{\chi}(s, \zeta)} = \sum_{n=1}^{\infty} \frac{s^n}{n} \int_{-\infty}^{0+} e^{i\zeta x} F^{n*}\{dx\}. \quad (3.4)$$

Доказательство. Равенство крайних членов устанавливается вычитанием (3.2) из (3.3). Другое тождество получается заменой $\overline{0, \infty}$ на $\overline{-\infty, 0}$ в лемме 1. ▶

Подставляя (3.4) в (1.9), получаем при $|s|<1$

$$1 - s\varphi(\zeta) = [1 - \chi(s, \zeta)] \cdot [1 - \tilde{\chi}(s, \zeta)]. \quad (3.5)$$

Это тождество по непрерывности распространяется и на значение $s=1$. Положим для краткости $\alpha(\zeta) = \chi(1, \zeta)$ и $\tilde{\alpha}(\zeta) = \tilde{\chi}(1, \zeta)$. Тогда α и $\tilde{\alpha}$ будут преобразованиями Фурье — Стильтьеса случайных величин (возможно, несобственных) $\mathbf{S}_{\mathbf{N}}$ и $\mathbf{S}_{\tilde{\mathbf{N}}}$. Мы имеем

$$1 - \varphi = (1 - \alpha)(1 - \tilde{\alpha}). \quad (3.6)$$

Положительная и отрицательная оси входят в эту формулу несколько несимметрично. Но это легко поправить. Чтобы не отклоняться в сторону, мы отложим соответствующее рассуждение до § 6. Здесь мы сформулируем теорему о факторизации в простейшей форме.

Теорема о факторизации. *Для произвольной характеристической функции* φ *при* $|s| \leqslant 1$ *имеют место тождества* (3.5) *и* (3.6).

Примечательная черта этой факторизации в том, что в нее входят два (возможно, несобственных) распределения, сосредоточенных на двух непересекающихся интервалах.

Аргументы, приводящие к этой теореме, показывают также, что факторизация единственна и определяет распределения лестничных величин. До того как мы углубимся в обсуждение этой

теоремы и ее связей с результатами гл. XII, мы рассмотрим несколько примеров.

Примеры. а) *Биномиальное случайное блуждание.* Для классического случайного блуждания с $p \geqslant q$ факторизация (3.5) имеет вид

$$1 - pe^{i\zeta} - qe^{-i\zeta} = (1 - e^{i\zeta})(1 - q - qe^{-i\zeta}). \tag{3.7}$$

Отсюда видно, что вход в $\overline{-\infty, 0}$ имеет вероятность $2q$ и

$$\gamma(1, \zeta) = \frac{1}{p - qe^{-i\zeta}} = \frac{1}{p} \sum_0^\infty \left(\frac{q}{p}\right)^k e^{-ik\zeta}. \tag{3.8}$$

В соответствии с определением γ (3.8) можно интерпретировать следующим образом: *математическое ожидание числа заходов в состояние $k \leqslant 0$ до первого входа в $\overline{0, \infty}$ равно* $\frac{1}{p}\left(\frac{q}{p}\right)^k$. В частности, при $p = q = 1/2$ получаем занятный результат: до получения первой положительной суммы $\mathbf{S}_n$ каждая точка $k \leqslant 0$ получается в среднем дважды. [См. пример (2,б) гл. XII и задачи 1 и 2 в гл. XII, 10.]

б) *Двустороннее показательное распределение.* Если F имеет плотность $\frac{1}{2}e^{-|x|}$, то (3.6) принимает форму

$$1 - \frac{1}{1+\zeta^2} = \left(1 - \frac{1}{1-i\zeta}\right)\left(1 - \frac{1}{1+i\zeta}\right). \tag{3.9}$$

Справа встречаются характеристические функции односторонних показательных распределений.

в) *Формула Хинчина — Полячека.* Пусть F представляется в виде свертки сосредоточенного на $\overline{0, \infty}$ показательного распределения с математическим ожиданием $1/a$ и некоторого распределения B, сосредоточенного на $\overline{-\infty, 0}$. Обозначим характеристическую функцию B через β, а соответствующее математическое ожидание через $-b < 0$. Допустим, что математическое ожидание F, равное $a^{-1} - b$, положительно. Тогда (3.6) принимает вид

$$1 - \frac{a}{a - i\zeta}\beta(\zeta) = \left(1 - \frac{a}{a - i\zeta}\right)\left(1 - a\,\frac{1 - \beta(\zeta)}{i\zeta}\right). \tag{3.10}$$

Последний множитель соответствует несобственному распределению с *плотностью* $aB(x)$ и полной массой ab. Этот же результат был получен прямыми методами в гл. XII, 4 (см. задачу 8).

г) *Ограниченные арифметические распределения.* В используемой здесь терминологии мы можем сказать, что вычисления, проделанные в примере (4, в) гл. XII, приводят к факторизации (3.5) для случая F, сосредоточенного в конечном числе целых точек.

д) *Распределение лестничной случайной величины в показательном случайном блуждании.* В случайном блуждании, связанном с одним важным примером теории очередей [гл. VI, (9,д)], распределение F представлялось в виде свертки двух показательных распределений и

$$\varphi(\zeta)=\frac{a}{a+i\zeta}\cdot\frac{b}{b-i\zeta}. \tag{3.11}$$

К счастью, в этом примере можно произвести факторизацию явно для всех $s<1$. В самом деле, как легко проверить,

$$1-s\varphi(\zeta)=\Big(1-\frac{p(s)}{a+i\zeta}\Big)\Big(1-\frac{q(s)}{b-i\xi}\Big), \tag{3.12}$$

где

$$2p(s)=a+b-\sqrt{(a+b)^2-4abs}. \tag{3.13}$$

При $\zeta=0$ отсюда выводим, что *возрастающие и убывающие лестничные моменты* имеют производящие функции $b^{-1}p(s)$ и $a^{-1}p(s)$ соответственно.

К (3.12) можно прийти эвристически. Отсутствие памяти, внутренне присущее показательному распределению, имеет следствием то, что при каждом n функция H_n имеет плотность вида $b_n\cdot e^{-bx}$. В этом случае χ должно иметь форму $B(s)/(b-i\zeta)$. По той же причине естественно ожидать, что $\chi(s,\zeta)=A(s)/(a+i\zeta)$. Подставляя эти выражения в (3.5), видим, что $A(s)=B(s)=p(s)$. ▶

§ 4. Обсуждение результатов. Применения

Отметим, что результаты предыдущего параграфа лишь обозначениями отличаются от результатов, полученных комбинаторными методами в гл. XII. [Факторизация (3.6) эквивалентна гл. XII, (3.11), в то время как (3.3) совпадает с гл. XII, (9.3).] Последующие замечания должны пояснить связь между двумя подходами.

Тождество (3.4), записанное для положительной полуоси, дает

$$\frac{1}{1-\chi(s,\zeta)}=\tilde{\gamma}(s,\zeta), \tag{4.1}$$

где $\tilde{\gamma}$ — комбинация производящей функции и преобразований Фурье — Стильтьеса для

$$\tilde{G}_n\{I\} = \mathbf{P}\{\mathbf{S}_1 \geqslant 0, \ldots, \mathbf{S}_n \geqslant 0, \ \mathbf{S}_n \in I\}. \tag{4.2}$$

Мы покажем сейчас, что левая часть (4.1) построена аналогично по функциям

$$L_n\{I\} = \mathbf{P}\{0 < \mathbf{S}_n, \ \mathbf{S}_1 < \mathbf{S}_n, \ldots, \mathbf{S}_{n-1} < \mathbf{S}_n \in I\}. \tag{4.3}$$

В самом деле, (4.3) означает лишь, что $(n, \mathbf{S}_n)$ является лестничной точкой (как они были определены в гл. VI, 8; там же было показано, что эти точки образуют процесс восстановления). Теперь χ представляет собой преобразование для первой лестничной точки, а χ^k — для k-й. Поэтому $\sum \chi^k$ соответствует вероятностям $L_n\{I\}$ попадания какой-либо лестничной точки в I.

Имея совпадающие преобразования, распределения (4.2) и (4.3) сами совпадают. Это уже было доказано в гл. XII, 2 комбинаторными методами, так что здесь мы имеем новое комбинаторное доказательство леммы 2 [1]). Наглядная и элементарная природа отмеченного принципа двойственности вполне объясняет его привлекательность. Кроме того, комбинаторные методы можно иногда применять в ситуациях, не поддающихся аналитическим методам (тому примером могут быть симметрично зависимые случайные величины). С другой стороны, использование преобразований Фурье сильно упрощает формулы и более тонкие рассуждения, включающие использование тауберовых теорем, были бы невозможны без преобразований Фурье.

Речь о применениях мы начнем с новой интерпретации $\chi(s, \zeta)$, которая играет важную роль в теории очередей [см. гл. VI, (9.5)].

Теорема [2]). *Характеристическая функция для*

$$\mathbf{U}_n = \max[0, \mathbf{S}_1, \ldots, \mathbf{S}_n] \tag{4.4}$$

равна коэффициенту при s^n *в разложении функции*

$$\frac{\gamma(s, 0)}{1 - \chi(s, \zeta)} = \frac{1}{1-s} \exp \sum_{n=1}^{\infty} \frac{s^n}{n} \int_0^{\infty} [e^{i\zeta x} - 1] F^{n*}\{dx\}. \tag{4.5}$$

Доказательство. Существует однозначно определяемый *первый* индекс $0 \leqslant \nu \leqslant n$, такой, что $\mathbf{S}_\nu = \mathbf{U}_n$. Он выделяется

[1]) По историческим причинам аналитическое доказательство обычно использует более глубокие методы комплексного переменного. Этим оба подхода разобщаются больше, чем это необходимо и естественно.

[2]) Соотношение (4.5) по существу совпадает с гл. XII, (2,6). Уточнение см. в теореме 5.2 и задаче 9.

следующими требованиями. Во-первых, $\mathbf{S}_0 > \mathbf{S}_j$ для $j < \nu$, и, во-вторых, при $k > \nu$ разности $\mathbf{S}_k - \mathbf{S}_\nu$ неположительны. Первое условие касается только величин $\mathbf{X}_1, \ldots, \mathbf{X}_\nu$, второе — величин $\mathbf{X}_{\nu+1}, \ldots, \mathbf{X}_n$. Поэтому

$$\mathbf{P}\{\mathbf{U}_n \in I\} = \sum_{\nu=0}^{n} L_\nu\{I\}\, G_{n-\nu}\{A\}, \tag{4.6}$$

где L_0 обозначает распределение вероятностей, сосредоточенное в нуле. Из мультипликативного свойства производящих функций следует, что $\sum s^n \mathbf{P}\{\mathbf{U}_n \in I\}$ имеет преобразование Фурье — Стильтьеса, равное левой части (4.5). Справедливость (4.5) вытекает теперь из (3.2) и (3.4). ▷

В гл. XII, 8 было показано, каким образом соотношение (4.6) может быть использовано для отыскания распределения вероятностей случайной величины $\mathbf{U}_n$ — *положения* максимума и для вывода предельной формы этого распределения — закона арксинуса.

Перейдем к формулам для математических ожиданий нескольких случайных величин. Напомним в этой связи, что существует два существенно различных типа случайных блужданий (см. гл. XII, 3).

а) *Уход к* $+\infty$ имеет место, когда $\mathbf{S}_{\widetilde{\mathbf{N}}}$ — точка достижения полупрямой $\overline{-\infty, 0}$ имеет несобственное распределение, т. е. когда $\tilde{\chi}(1, 0) < 1$. Для момента $\mathbf{N}$ достижения полупрямой $\overline{0, \infty}$ мы получаем, исходя из самого определения

$$\mathbf{E}(\mathbf{N}) = \sum_{n=0}^{\infty} \mathbf{P}\{\mathbf{N} > n\} = \sum_{0}^{\infty} G_n\{\overline{-\infty, 0}\} = \gamma(1, 0). \tag{4.7}$$

Полагая для краткости

$$a_n = \mathbf{P}\{\mathbf{S}_n > 0\}, \tag{4.8}$$

выводим из (3.4) при $s = 1$ и $\zeta = 0$

$$\mathbf{E}(\mathbf{N}) = \exp \sum_{n=1}^{\infty} \frac{1 - a_n}{n} < \infty. \tag{4.9}$$

Если все математические ожидания существуют, то, дифференцируя (1.9), находим

$$-\chi'(1, 0) = -\gamma(1, 0) \cdot \varphi'(0),$$

то есть

$$\mathbf{E}(\mathbf{S}_{\mathbf{N}}) = \mathbf{E}(\mathbf{N}) \cdot \mathbf{E}(\mathbf{X}_1). \tag{4.10}$$

Это соотношение было установлено в гл. XII, 2, где также было показано, что $\mathbf{E}(S_N)$ существует в том и только том случае когда $\mathbf{E}(X_1)$ существует и положительно.

б) *Осциллирующее случайное блуждание.* Если $\alpha(0) = \tilde{\alpha}(0) = 1$, то с вероятностью единица происходит хотя бы одно (а следовательно, и бесконечно много) попадание на каждую полуось. Из (4.7) заключаем, что $\mathbf{E}(N) = \infty$. Устойчивость колебаний приводит к тому, что моменты достижения полуосей имеют бесконечное математическое ожидание (так как это происходит при бросании монеты).

При каких условиях точки S_N и $S_{\tilde{N}}$ имеют конечные математические ожидания? Если $\mathbf{E}(S_N) < \infty$, то характеристическая функция α дифференцируема. Из (3.6) ясно, что существование производных α' и $\tilde{\alpha}'$ влечет как дифференцируемость φ, так и равенство $\varphi'(0) = 0$. Поделим теперь (3.6) на ζ^2 и устремим ζ к нулю. Тогда правая часть стремится к $\alpha'(0)Z'(0) = -\mathbf{E}(S_N) \times \mathbf{E}(\tilde{S}_N)$. Как было показано в гл. XV, (4.8), из ограниченности правой части вытекает, что F имеет второй момент. Таким образом, $\mathbf{E}(X_j) = \varphi'(0) = 0$ и X_j *имеет конечную дисперсию* σ^2, *такую, что*

$$\frac{1}{2}\sigma^2 = \mathbf{E}(S_N)\,\mathbf{E}(S_{\tilde{N}}). \tag{4.11}$$

§ 5 *. Уточнения

Основная цель этого параграфа — проиллюстрировать методы, пригодные для явных вычислений и для использования тауберовых теорем. Мы начнем с уточнения предыдущего результата.

Теорема 1 [1]). *Если* $\mathbf{E}(X_1) = 0$ *и* $\mathbf{E}(X_1^2) = \sigma^2 < \infty$, *то*

$$\mathbf{E}(S_N) = \frac{\sigma}{\sqrt{2}}\, e^{-c}, \quad \mathbf{E}(S_{\tilde{N}}) = \frac{\sigma}{\sqrt{2}}\, e^{c}, \tag{5.1}$$

где

$$c = \sum_{n=1}^{\infty} \frac{1}{n}\left(a_n - \frac{1}{2}\right), \tag{5.2}$$

причем последний ряд сходится (по крайней мере условно).

---

*) Материал не используется в дальнейшем.

[1]) Впервые доказана (более глубокими методами) Спицером (см. сноску в § 3). Сходимость (5.2) важна в связи с уточненным законом арксинуса [гл. XII, (8.7)].

Доказательство. Дифференцируя (3.2) по ζ и полагая $\zeta=0$, получим

$$-i\chi'(s,0)\cdot\frac{\sqrt{1-s}}{1-\chi(s,0)}=\sqrt{1-s}\sum_{n=1}^{\infty}\frac{s^n}{n}\int_0^{\infty}xF^{n*}\{dx\}. \tag{5.3}$$

По центральной предельной теореме распределение $F^{n*}(\sigma\sqrt{n}\,x)$ сходится к нормальному, причем таким образом, что сходятся и все моменты порядка $\leqslant 2$. Отсюда легко следует, что коэффициент при s^n в последнем ряде эквивалентен $\sigma/\sqrt{2\pi n}$. По тауберовой теореме 5 гл. XIII, 5 заключаем, что

$$-i\chi'(s,0)\frac{\sqrt{1-s}}{1-\chi(s,0)}\to\frac{\sigma}{\sqrt{2}}. \tag{5.4}$$

Далее $-i\chi'(s,0)\to\mathbf{E}(S_N)$ при $s\to 1$. Предел может быть конечным или бесконечным, но во всяком случае он отличен от нуля. Отсюда следует, что дробь, стоящая в левой части (5.4), стремится к конечному пределу. Величина, обратная этой дроби, встречается в соотношении двойственности для $\tilde{\chi}$. Поэтому указанный предел отличен от нуля. Но в силу (3.2)

$$\log\frac{\sqrt{1-s}}{1-\chi(s,0)}=\sum_{n=1}^{\infty}\frac{s^n}{n}\left(a_n-\frac{1}{2}\right). \tag{5.5}$$

Мы доказали, что правая часть имеет конечный предел c. Отсюда и из элементарных тауберовых теорем для степенных рядов [1]) вытекает (5.2). Доказательство закончено.

Теорема 2. *Если* $\mathbf{E}(\mathbf{X}_1)=\mu<0$, *то точки максимумов* $\mathbf{U}_n$ *имеют предельное распределение, преобразование Лапласа которого задается формулой*

$$\omega(\lambda)=\exp\sum_{1}^{\infty}\frac{1}{n}\int_0^{\infty}\left[e^{-\lambda x}-1\right]F^{n*}\{dx\}. \tag{5.6}$$

Доказательство. Формула (4.5) остается верной и при $\zeta=i\lambda$, $\lambda\geqslant 0$. Преобразование Лапласа для $\mathbf{U}_n$ равно коэффициенту при s^n в

$$\frac{1}{1-s}\exp\sum_{n=1}^{\infty}\frac{s^n}{n}\int_0^{\infty}\left[e^{-\lambda x}-1\right]F^{n*}\{dx\}. \tag{5.7}$$

[1]) См., например, гл. I книги Титчмарша «Теория функций», ГИТТЛ, М., 1951.

Как легко видеть, условие $\mu<0$ влечет сходимость [1]) последнего ряда при $s=1$, так что при $s\to 1$ величина (5.7) $\sim \omega(\lambda)(1-s)^{-1}$, где ω определена по (5.6). Величины U_n не убывают. Интегрирование по частям тривиально показывает, что соответствующие преобразования Лапласа образуют монотонную последовательность. Тауберова теорема 5 гл. XIII, 5 гарантирует сходимость коэффициентов при s^n в (5.7) к $\omega(\lambda)$. ▶

§ 6. Возвращения в нуль

В предыдущих вычислениях можно отметить небольшую асимметрию, связанную с тем обстоятельством, что отрицательная полуось бралась замкнутой. Это не играет никакой роли, если исходное распределение F непрерывно, а в общем случае положение можно исправить. Положим

$$r_n=\mathbf{P}\{\mathbf{S}_0\leqslant 0,\ \ldots,\ \mathbf{S}_{n-1}\leqslant 0,\ \mathbf{S}_n=0\}.\qquad n\geqslant 0, \tag{6.1}$$

$$f_n=\mathbf{P}\{\mathbf{S}_1<0,\ \ldots,\ \mathbf{S}_{n-1}<0,\ \mathbf{S}_n=0\},\qquad n\geqslant 1. \tag{6.2}$$

Событие (6.1) означает возвращение в нуль до попадания на $\overline{0,\infty}$. После возвращения в нуль восстанавливается исходная ситуация. Другими словами, (6.1) определяет рекуррентное событие и (6.2) задает распределение для его времени возвращения. Так как $\sum f_n\leqslant \mathbf{P}\{\mathbf{X}_1\leqslant 0\}$, то рекуррентное событие *невозвратно* (исключая тот тривиальный случай, когда распределение F сосредоточено в нуле). Пусть $r(s)$ и $f(s)$ обозначают производящие функции для $\{r_n\}$ и $\{f_n\}$. Согласно основному соотношению для рекуррентных событий,

$$r(s)=\frac{1}{1-f(s)}. \tag{6.3}$$

Теперь r_n — масса, приписываемая нулю мерой G_n, и потому $\gamma(s,\zeta)=r(s)+\gamma_1(s,\zeta)$, где γ_1 — преобразование Фурье — Стильтьеса меры, сосредоточенной на *открытом* интервале $\overline{-\infty,0}$. Выделение множителя $r(s)$ приводит (3.5) к симметричной форме

$$1-s\varphi(\zeta)=[1-f(s)]\cdot[1-\chi(s,\zeta)]\cdot[1-\chi^-(s,\zeta)], \tag{6.4}$$

где χ^- — преобразование для точки достижения открытого интервала $\overline{-\infty,0}$. Для $s=1$ получаем

$$1-\varphi(\zeta)=[1-f(1)]\cdot[1-\alpha(\zeta)]\cdot[1-\alpha^-(\zeta)], \tag{6.5}$$

[1]) В самом деле, подставим в (3.2) $\zeta=0$ и $s=1$. Получим, что при уходе случайного блуждания в $-\infty$ ряд $\sum n^{-1}P\{S_n>0\}$ сходится. Другие доказательства см. в теореме гл. XII, 7.2 и задаче 16 гл. XII, 9.

где α и α^- — суть преобразования для точек достижения открытых полупрямых. Такова *симметричная запись факторизации Винера — Хопфа.*

Пусть теперь желательно считать замкнутой правую полупрямую. Тогда достаточно соединить два первых множителя в (6.5). Получится выражение

$$f(1)+[1-f(1)]\alpha(\xi) \tag{6.6}$$

для преобразования точки достижения полупрямой $\overline{0,\infty}$. Эта формула имеет очевидное вероятностное истолкование: с вероятностью $f(1)$ первое достижение происходит в нуле и с вероятностью $1-f(1)$ первые достижения полупрямых $\overline{0,\infty}$ и $\overline{0,\infty}$ одинаковы. [Этот факт был использован в гл. XII, (1.10).] По причинам симметрии из наших формул следует, что вероятности f_n и r_n не изменяются, если все неравенства (6.1) и (6.2) заменить обратными. (Этот факт был установлен в гл. XII, 2 обращением порядка случайных величин.) В заключение уместно заметить, что рассуждения, приводящие к (3.2), применимы и к подсчету массы, сосредоточенной в нуле. Можно показать, что

$$\log\frac{1}{1-f(s)}=\sum_{n=1}^{\infty}\frac{s^n}{n}\mathbf{P}\{\mathbf{S}_n=0\}. \tag{6.7}$$

[Это то же самое, что и XII, (9.6).]

§ 7. Критерии возвратности

Материал этого параграфа не связан с предшествующей теорией; он посвящен методу, развитому Чжуном и Фуксом (1950) и позволяющему определить, является ли случайное блуждание возвратным или невозвратным. Несмотря на наличие указанных в гл. XII критериев и методов, метод, основанный на анализе Фурье, сохраняет и методологический, и исторический [1]) интерес. Кроме того, в настоящее время это единственный метод, применимый в многомерном случае. В дальнейшем F обозначает одномерное распределение с характеристической функцией $\varphi=u+iv$. За соответствующими определениями читатель отсылается к гл. VI, 10.

[1]) Тот факт, что условие $\mathbf{E}(\mathbf{X}_j)=0$ влечет возвратность, был установлен Чжуном и Фуксом. Интересно отметить, что в 1950 г. эта задача доставляла серьезные затруднения и многие попытки решить её закончились неудачей. Внимание к этой задаче было привлечено неожиданным открытием обманчиво «безобидного» случайного блуждания, в котором $\mathbf{P}\left\{\mathbf{S}_n>\frac{n}{\log n}\right\}\to 1$. (См. **1**, гл. X, 3 и задачу 15 в **1** гл. X. 8.)

Критерий. *Распределение F будет невозвратным в том и только том случае, когда при некотором $\alpha>0$ интеграл*

$$\int_0^a \frac{1-su}{(1-su)^2+s^2v^2}\,d\zeta \tag{7.1}$$

остается ограниченным при $s\to 1$.

(Мы увидим, что в противном случае интеграл стремится к ∞.)

Доказательство. Определим для $0<s<1$ конечную меру

$$U_s=\sum_{n=0}^{\infty} s^n F^{n*}. \tag{7.2}$$

Распределение F будет невозвратным тогда и только тогда, когда $U_s\{I\}$ остается ограниченным для некоторого открытого интервала, содержащего нуль (в этом случае $U_s\{I\}$ остается ограниченным для любого конечного интервала I). Равенство Парсеваля гл. XV, (3.2), примененное к F^{n*} и треугольной плотности (номер 4 в гл. XV, 2), дает

$$2\int_{-\infty}^{+\infty}\frac{1-\cos ax}{a^2x^2}F^{n*}\{dx\}=\frac{1}{a}\int_{-a}^{a}\left(1-\frac{|\zeta|}{a}\right)\varphi^n(\zeta)\,d\zeta. \tag{7.3}$$

Следовательно,

$$2\int_{-\infty}^{+\infty}\frac{1-\cos ax}{a^2x^2}U_s\{dx\}=\frac{1}{a}\int_{-a}^{a}\left(1-\frac{|\zeta|}{a}\right)\frac{d\zeta}{1-s\varphi(\zeta)}=$$

$$=\frac{2}{a}\int_0^a\left(1-\frac{\zeta}{a}\right)\cdot\frac{1-su}{(1-su)^2+s^2v^2}\,d\zeta \tag{7.4}$$

(так как действительная часть — четная функция). Пусть I означает интервал $|x|<2/a$ На I подинтегральное выражение в крайнем левом интеграле (7.4) будет $>1/3$, так что величина $U_s\{I\}$ ограничена. Поэтому условия теоремы *достаточны.* Ее необходимость устанавливается сходным образом на основе равенства Парсеваля с характеристической функцией $1-\frac{|\zeta|}{a}$ (номер 5 в гл. XV, 2). При этом (7.4) заменяется соотношением

$$\int_{-a}^{a}\left(1-\frac{|x|}{a}\right)U_s\{dx\}=\frac{2}{\pi}\int_0^{\infty}\frac{1-\cos a\zeta}{a\zeta^2}\cdot\frac{1-su}{(1-su)^2+s^2v^2}\,d\zeta. \tag{7.5}$$

Для невозвратного распределения F левая часть остается ограниченной, так что интеграл (7.1) с $\alpha=1/2a$ ограничен. ▶

Применения. а) *Пусть существует производная*[1]) $\varphi'(0)=i\mu$. *Тогда F невозвратно при $\mu\neq 0$ и возвратно при $\mu=0$.* В частности, если F имеет математическое ожидание μ, то условие $\mu=0$ необходимо и достаточно для возвратности.

Для доказательства заметим, что при фиксированных $\alpha>0$, $\eta>0$ и $s\to 1$

$$\int_0^a \frac{1-s}{(1-s)^2+\eta^2\zeta^2}\,d\zeta=\frac{1}{\eta}\operatorname{arctg}\frac{\eta a}{1-s}\to\frac{\pi}{2}\cdot\frac{1}{\eta}. \tag{7.6}$$

Так как u — четная функция, то $u'(0)=0$ и $v'(0)=\mu$. При $\mu\neq 0$ берем $\eta=2|\mu|$ и берем α столь малым, чтобы при $|s|>1/2$ подинтегральное выражение в (7.1) не превосходило умноженного на фиксированную константу подинтегрального выражения в (7.6). Так как интеграл (7.6) ограничен, то F невозвратно. Пусть теперь $\mu=0$. Тогда мы можем выбрать η сколь угодно малым, а $\alpha>0$ таким, чтобы отношение подинтегральных выражений в (7.1) и (7.6) превосходило некоторую положительную константу. Так как интеграл (7.1) может быть сделан сколь угодно большим, то F возвратно. [Более сильное утверждение см. в задаче 11.]

б) Из указанного критерия ясно видно, что F возвратно, если для любого $\alpha>0$

$$\int_0^a \frac{1-u}{(1-u)^2+v^2}\,d\zeta=\infty, \tag{7.7}$$

и невозвратно, если при некотором $\alpha>0$

$$\int_0^a \frac{d\zeta}{1-u}<\infty. \tag{7.8}$$

Применения см. в задачах 11—14.

в) *Невырожденное двумерное распределение с нулевыми математическими ожиданиями и конечными дисперсиями компонент является возвратным.* Для доказательства заметим, что наш критерий без изменений применим к двумерным характеристическим функциям. Положим $r^2=\zeta_1^2+\zeta_2^2$. Из формулы Тейлора видно, что в некоторой окрестности нуля разность $1-u$

[1]) По поводу существования $\varphi'(0)$ см. пример (1, в), гл. XVII.

лежит между двумя положительными кратными r^2 и потому подинтегральное выражение в (7.7) больше некоторого кратного $1/r^2$.

г) *Каждое трехмерное распределение невозвратно.* В самом деле, из формулы Тейлора для косинуса ясно видно, что для любой характеристической функции величина $(1-u)/r^2$ больше некоторой положительной постоянной. Трехмерный аналог (7.8) следует сравнить с интегралом от $1/r^2$, который в трехмерном случае расходится.

§ 8. Задачи

1. Распространите пример § 1 на случай несимметричного интервала $-a, b$. (Выведите два линейных уравнения для двух производящих функций, соответствующие двум границам. Явное решение затруднительно.)

2. Покажите, что в примере § 1 $\mathbf{E}(\mathbf{N}) = 2P'(1) = 1 - \frac{a}{2} + \frac{a^2}{4}$. *Указание.* Используйте дифференцирование (1.10).

Задачи 3—6 *касаются симметричного биномиального блуждания,* т. е. $\varphi(\zeta) = \cos\zeta$. Используются обозначения § 1.

3. Пусть A состоит из двух точек: 0 и 1. Покажите элементарным путем, что

$$\chi(s, \zeta) = \frac{1}{1 - \frac{1}{4}s^2}\left(\frac{s}{2}e^{-i\zeta} + \frac{s^2}{4}e^{2i\zeta}\right)$$

и

$$\gamma(s, \zeta) = \frac{1}{1 - \frac{1}{4}s^2}\left(1 + \frac{s}{2}e^{i\zeta}\right).$$

Проверьте (1.9).

4. Если в предыдущем примере поменять ролями A и A', то

$$\chi(s, \zeta) = \frac{s}{2}e^{i\zeta} + \frac{1}{2}(1 - \sqrt{1 - s^2})$$

и

$$\gamma(s, \zeta) = \left[1 - \frac{1 - \sqrt{1 - s^2}}{s}e^{-i\zeta}\right]^{-1}.$$

Дайте вероятностное объяснение.

5. Пусть A состоит из одной точки 0. Тогда χ зависит только от s, и γ будет суммой двух степенных рядов, расположенных по степеням $e^{i\zeta}$ и $e^{-i\zeta}$ соответственно. Используя эту информацию, найдите χ и γ непосредственно из (1.9).

6. Если A' состоит из одной точки 0, то $\chi = s\varphi$ и $\gamma = 1$.

7. Докажите сходимость (2.1), показав, что урезание F в точках $\pm(a+b)$ не влияет на задачу, так что применимы неравенства типа (2.8).

8. Допустим, что в примере (3, в) (формула Хинчина — Полячека) $ab > 1$. Покажите, что существует единственное положительное число $\varkappa$, лежащее

между 0 и a, такое, что

$$\alpha\beta(-i\varkappa) = a - \varkappa.$$

Докажите, что

$$\tilde{\chi}(\zeta) = \frac{a - \varkappa - \alpha\beta(\zeta - i\varkappa)}{i\zeta}.$$

[Сравните с примерами 5,а—б, гл. XIII.]

9. Пусть $\mathbf{U}_n = \max[0, \mathbf{S}_1, \ldots, \mathbf{S}_n]$ и $\mathbf{V}_n = \mathbf{S}_n - \mathbf{U}_n$. Слегка изменяя рассуждения, использованные при выводе (4.5), покажите, что двумерная характеристическая функция для пары $(\mathbf{U}_n, \mathbf{V}_n)$ равна коэффициенту при s^n в [1])

$$\exp\sum_{n=1}^{\infty}\frac{s^n}{n}\left[\int_0^{\infty}(e^{i\zeta_1 x}-1)\,F^{n*}\{dx\} + \int_{-\infty}^{0}(e^{i\zeta_2 x}-1)\,F^{n*}\{dx\}\right].$$

10. *Другое доказательство тождества* (1.9). Покажите, что (в обозначениях § 1)

$$F^{n*}\{I\} = \sum_{k=1}^{n}\int_{A'} H_k\{dy\}\,F^{(n-k)*}\{I-y\} + G_n\{I\} \qquad (*)$$

а) прямыми вероятностными рассуждениями,

б) по индукции.

Возьмите $U_s = \sum_0^{\infty} s^n F^{n*}$. Покажите, что $(*)$ равносильно (1.9).

11. Допустим, что в некоторой окрестности нуля $|1-\varphi(\zeta)| < A\cdot|\zeta|$. Тогда F возвратно, за исключением случая, когда оно имеет математическое ожидание $\mu \neq 0$

Указание. Интеграл (7.7) превосходит $\int_0^a d\zeta \int_{-1/\zeta}^{1/\zeta} x^2 F\{dx\}$. Подстановка $\zeta = 1/t$ и перемена порядка интегрирования показывают, что этот интеграл расходится (за исключением случая, когда существует μ).

12. Используя критерий (7.8), покажите, что если при некотором $\rho > 0$ и $t \to \infty$

$$t^{-1-\rho}\int_{-t}^{t} x^2 F\{dx\} \to \infty,$$

то распределение F невозвратно [1]).

13. Распределение с характеристической функцией $\varphi(\zeta) = e^{-1}\sum\frac{1}{n!}\cos(n!\,\zeta)$ невозвратно.

Указание. Используйте (7.8) и замену переменных $\zeta = \frac{1}{n!}\,t$.

14. Асимметричное устойчивое распределение с показателем $\alpha = 1$ невозвратно. Однако распределение Коши возвратно.

---

[1]) Результат получен впервые аналитическим путем Спицером, Spitzer F., *Trans. Amer. Math. Soc.*, **82** (1956), 323—339.

[2]) Отсюда видно, что при слабых условиях регулярности F невозвратно, если абсолютный момент какого-либо порядка $\rho < 1$ расходится. Затруднения, возникающие при отказе от условий регулярности, показаны в статье Shepp L. A., *Bull. Amer. Math. Soc.*, **70** (1964), 540—542.

ГЛАВА XIX

ГАРМОНИЧЕСКИЙ АНАЛИЗ

В этой главе даны дополнения к теории характеристических функций, изложенной в гл. XV, также даны применения гармонического анализа к случайным процессам и стохастическим интегралам. Обсуждение формулы суммирования Пуассона в § 5 практически не зависит от остального материала. Глава как целое не зависит от гл. XVI—XVIII.

§ 1. Равенство Парсеваля

Пусть U — распределение вероятностей с характеристической функцией

$$\omega(\zeta) = \int_{-\infty}^{+\infty} e^{i\zeta x} U\{dx\}. \tag{1.1}$$

Интегрируя это выражение по какому-либо другому распределению F, получаем

$$\int_{-\infty}^{+\infty} \omega(\zeta) F\{d\zeta\} = \int_{-\infty}^{+\infty} \varphi(x) U\{dx\}, \tag{1.2}$$

где φ — характеристическая функция F. Это одна из форм *равенства Парсеваля*, из которого были выведены основные результаты гл. XV, 3. Удивительно количество новой информации, которую можно получить, переписывая равенство Парсеваля в эквивалентных формах и рассматривая специальные случаи. Этот метод, который мы постоянно будем использовать, можно проиллюстрировать простым примером (имеющим и самостоятельный интерес).

Пример. Формула

$$\int_{-\infty}^{+\infty} e^{-ia\zeta} \omega(\zeta) F\{d\zeta\} = \int_{-\infty}^{+\infty} \varphi(x) U\{a + dx\} \tag{1.3}$$

отличается от (1.2) только обозначениями. Рассмотрим частный случай, когда F — равномерное распределение на $-t, t$ и $\varphi(x) = \frac{\sin tx}{tx}$. Эта функция не превосходит по абсолютной величине единицы и стремится к нулю при $t \to \infty$ при всех $x \neq 0$. По теореме об ограниченной сходимости получаем

$$U(a) - U(a-) = \lim_{t\to\infty} \frac{1}{2t} \int_{-t}^{t} e^{-ia\zeta} \omega(\zeta)\, d\zeta. \tag{1.4}$$

Эта формула позволяет узнать, будет ли a точкой непрерывности и, если a — атом, позволяет определить его вес. Наиболее интересный результат получается применением (1.4) к симметризованному распределению ${}^0U$ с характеристической функцией $|\omega|^2$. Если $p_1, p_2, \ldots$ суть веса атомов U, то ${}^0U$ имеет атом веса $\sum p_k^2$ в нуле (см. задачу 8 в гл. V, 11). Поэтому

$$\frac{1}{2t} \int_{-t}^{t} |\omega(\zeta)|^2 d\zeta \to \sum p_k^2. \tag{1.5}$$

Эта формула показывает, в частности, что характеристические функции непрерывных распределений «в среднем» малы. ▶

Приведем теперь полезный и «гибкий» вариант формулы Парсеваля (1.2). *Пусть A и B — произвольные распределения вероятностей с характеристическими функциями α и β соответственно. Тогда*

$$\int_{-\infty}^{+\infty} \int \omega(s-t)\, A\{ds\}\, B\{dt\} = \int_{-\infty}^{+\infty} \alpha(x)\overline{\beta(x)}\, U\{dx\}, \tag{1.6}$$

где $\overline{\beta}$ — величина, комплексно сопряженная с β. Справедливость (1.6) проверяется непосредственно интегрированием

$$\omega(s-t) = \int_{-\infty}^{+\infty} e^{i(s-t)x} U\{dx\} \tag{1.7}$$

по A и B. Может создаться обманчивое впечатление, что (1.6) общее (1.2). *На самом деле* (1.6) *есть специальный случай равенства Парсеваля* (1.2), *соответствующий* $F = A * {}^-B$, где ${}^-B$ — распределение с характеристической функцией $\overline{\beta}$ (т. е. ${}^-B(x) = 1 - B(-x)$ во всех точках непрерывности). Действительно, F имеет характеристическую функцию $\varphi = \alpha\overline{\beta}$, так что правые части (1.2) и (1.6) идентичны. Левые части отличаются лишь формой записи. Лучше всего это можно показать, вводя две не-

зависимые случайные величины $\mathbf{X}$ и $\mathbf{Y}$ с распределениями A и B соответственно. Тогда левая часть (1.6) представляет собою прямое определение математического ожидания $\mathbf{E}(\omega(\mathbf{X}-\mathbf{Y}))$, а левая часть (1.2) выражает это математическое ожидание в терминах распределения F разности $\mathbf{X}-\mathbf{Y}$. (Мы еще вернемся к формуле Парсеваля в § 7.)

§ 2. Положительно определенные функции

В 1932 г. С. Бохнер доказал важную теорему, которая делает возможным описание класса характеристических функций по их «внутренним» свойствам. Путь рассуждений указывается следующим простым критерием.

Лемма 1. *Пусть ω — ограниченная непрерывная комплекснозначная функция, интегрируемая*[1]) *на $-\infty$, ∞. Определим и равенством*

$$u(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\xi x}\omega(\xi)\,d\xi. \tag{2.1}$$

Для того чтобы функция ω была характеристической, необходимо и достаточно, чтобы было $\omega(0)=1$ и $u(x)\geqslant 0$ при всех x. При этих условиях и является плотностью вероятности, соответствующей ω.

Доказательство. Теорема об обращении интеграла Фурье [гл. XV, (3.3)] показывает, что высказанные условия необходимы. Выберем теперь произвольную четную плотность f с интегрируемой характеристической функцией $\varphi\geqslant 0$. Умножим (2.1) на $\varphi(tx)e^{iax}$ и проинтегрируем по x. Применяя к паре f, φ формулу обращения гл. XV, (3.5), получаем

$$\int_{-\infty}^{+\infty} u(x)\varphi(tx)e^{iax}\,dx=\int_{-\infty}^{+\infty}\omega(\xi)f\left(\frac{\xi-a}{t}\right)\frac{d\xi}{t}. \tag{2.2}$$

Правая часть представляет собой математическое ожидание ω по отношению к некоторому распределению вероятностей и потому не превосходит максимума $|\omega|$. При $a=0$ подинтегральное выражение слева не отрицательно и стремится к $u(x)$ при $t\to 0$. Из ограниченности интеграла вытекает *интегрируемость* u. Переходя в (2.2) к пределу при $t\to 0$, получаем

$$\int_{-\infty}^{+\infty} u(x)e^{iax}\,dx=\omega(a) \tag{2.3}$$

[1]) Как и всегда, это означает *абсолютную* интегрируемость.

(в левой части применяется теорема об ограниченной сходимости; в правой части распределение, по которому интегрируется ω, сходится к распределению, сосредоточенному в точке a). Полагая $a=0$, видим, что u — плотность вероятности, которой соответствует характеристическая функция ω. ▷

Условие интегрируемости, появляющееся в лемме, выглядит более ограничительным, чем оно в действительности является. В самом деле, по теореме непрерывности непрерывная функция ω будет характеристической в том и только том случае, когда при каждом $\varepsilon>0$ $\omega(\zeta)e^{-\varepsilon\zeta^2}$ есть характеристическая функция. Из леммы следует, что ограниченная и непрерывная функция с $\omega(0)=1$ является характеристической тогда и только тогда, когда при всех x и $\varepsilon>0$

$$\int_{-\infty}^{+\infty} e^{-i\zeta x}\omega(\zeta)\,e^{-\varepsilon\zeta^2}\,d\zeta \geqslant 0. \tag{2.4}$$

Этот критерий имеет совершенно общий характер, однако его трудно применять в частных случаях. К его недостаткам относится и произвольный выбор «множителя сходимости» $e^{-\varepsilon\zeta^2}$. Поэтому мы сформулируем критерий в новой форме.

Лемма 2. *Ограниченная непрерывная функция ω будет характеристической в том и только том случае, если $\omega(0)=1$ и для любого распределения вероятностей A и всех x*

$$\int_{-\infty}^{+\infty} e^{-i\zeta x}\omega(\zeta)\,{}^0A\,\{d\zeta\} \geqslant 0, \tag{2.5}$$

где ${}^0A=A*{}^-A$ — распределение, полученное симметризацией.

Доказательство. а) *Необходимость.* Пусть α — характеристическая функция A. Тогда ${}^0A$ имеет характеристическую функцию $|\alpha|^2$, и необходимость (2.5) вытекает из равенства Парсеваля (1.3).

б) *Достаточность.* Как показывает (2.4), условие (2.5) является достаточным уже в том случае, когда в качестве A выбираются нормальные распределения с произвольными дисперсиями. ▷

Мы видели, что (2.5) можно переписать в виде (1.6) с $B=A$. В частности, если A сосредоточено в конечном числе точек $t_1, t_2, \ldots, t_n$ и приписывает им веса $p_1, p_2, \ldots, p_n$, то (2.5) при-

нимает форму

$$\sum_{j,k} \omega(t_j - t_k)\, e^{-ix(t_j - t_k)}\, p_j p_k \geqslant 0. \tag{2.6}$$

Если это неравенство верно при любом выборе t_j и p_j, то (2.5) выполняется для всех дискретных распределений с конечным числом атомов. Так как каждое распределение может быть представлено как предел подобных дискретных распределений, условие (2.6) оказывается необходимым и достаточным. Обозначим $z_j = p_j e^{-ixt_j}$, тогда оно принимает вид

$$\sum_{j,k} \omega(t_j - t_k)\, z_j \bar{z}_k \geqslant 0. \tag{2.7}$$

Для окончательной формулировки нашего критерия введем один часто используемый термин.

Определение. *Комплекснозначная функция*[1]) *вещественного переменного t называется положительно определенной, если* (2.7) *выполняется при любом выборе вещественных чисел $t_1, \ldots, t_n$ и комплексных чисел $z_1, \ldots, z_n$.*

Теорема. (*Бохнер.*) *Непрерывная функция от ω является характеристической функцией некоторого распределения вероятностей в том и только в том случае, когда она положительно определена и $\omega(0) = 1$.*

Доказательство. Мы показали раньше, что эти условия необходимы и что они достаточны, если ω ограничена. Доказательство заканчивается ссылкой на следующую далее лемму, которая утверждает, что все положительно определенные функции ограничены. ▶

Лемма 3. *Если ω положительно определена, то*

$$\omega(0) \geqslant 0, \quad |\omega(t)| \leqslant \omega(0). \tag{2.8}$$

Доказательство. Пусть $t_2 = 0$ и $z_2 = 1$. При $n = 2$ получаем из (2.7)

$$\omega(0)[1 + |z_1|^2] + \omega(t_1) z_1 + \omega(-t_1)\bar{z}_1 \geqslant 0. \tag{2.9}$$

Полагая $z_1 = 0$, видим, что $\omega(0) \geqslant 0$. Кроме того, (2.9) показывает, что $\omega(-t_1)$ комплексно сопряжено с $\omega(t_1)$. При $z_1 = \lambda \cdot \overline{\omega(t_1)}$ левая часть (2.9) становится квадратным много-

[1]) Об «обобщенных» положительно определенных функциях (эквивалентных распределениям в смысле Л. Шварца) в произвольных пространствах см. Гельфанд и Виленкин (1964).

членом относительно λ, не принимающим отрицательных значений. Поэтому его дискриминант не положителен, откуда и вытекает второе из неравенств (2.8). ▶

§ 3. Стационарные процессы

К важным следствиям приводит применение последней теоремы к случайным процессам со стационарными ковариациями. Под этим понимается семейство случайных величин $\{\mathbf{X}_t\}$, определенных на некотором пространстве при всех $-\infty<t<\infty$ и таких, что

$$\operatorname{Cov}(\mathbf{X}_{s+t},\ \mathbf{X}_s)=\rho(t) \tag{3.1}$$

при любых s. До сих пор мы рассматривали только действительные случайные величины, но теперь мы введем *комплексные* случайные величины, что сделает обозначения более простыми и более симметричными. Комплексная случайная величина — это просто пара действительных случайных величин, записанная в форме $\mathbf{X}=\mathbf{U}+i\mathbf{V}$. Мы не делаем никаких предположений относительно совместного распределения величин $\mathbf{U}$ и $\mathbf{V}$. Случайная величина $\bar{\mathbf{X}}=\mathbf{U}-i\mathbf{V}$ называется комплексно сопряженной с $\mathbf{X}$, и произведение $\mathbf{X}\bar{\mathbf{X}}$ играет здесь ту же роль, что и $\mathbf{X}^2$ для вещественных случайных величин. Отсюда некоторая асимметрия в определении дисперсии и ковариации: *если*

$$\mathbf{E}(\mathbf{X})=\mathbf{E}(\mathbf{Y})=0,$$

то мы полагаем

$$\operatorname{Cov}(\mathbf{X},\ \mathbf{Y})=\mathbf{E}(\mathbf{X}\bar{\mathbf{Y}}). \tag{3.2}$$

Тогда $\operatorname{Var}(\mathbf{X})=\mathbf{E}(|\mathbf{X}|^2)\geqslant 0$.

Теорема. *Пусть* $\{\mathbf{X}_t\}$ *— такое семейство случайных величин, что функция*

$$\rho(t)=\mathbf{E}(\mathbf{X}_{t+s}\bar{\mathbf{X}}_s) \tag{3.3}$$

не зависит от s *и непрерывна* [1]. *Тогда* ρ *положительно определена и потому*

$$\rho(t)=\int_{-\infty}^{+\infty} e^{i\lambda t}R\{d\lambda\}, \tag{3.4}$$

где R *— мера на вещественной прямой, приписывающая всей прямой массу* $\rho(0)$.

[1]) Непрерывность существенна: для взаимно независимых случайных величин $\mathbf{X}_t$ мы имеем $\rho(t)=0$ для всех $t\neq 0$; эта функция не допускает представления (3.4) (см. задачу 2).

Если случайные величины $\mathbf{X}_t$ *вещественны, то мера* R *симметрична и*

$$\rho(t) = \int_{-\infty}^{+\infty} \cos \lambda t R\{d\lambda\}. \tag{3.5}$$

Доказательство. Выберем произвольно точки $t_1, \ldots, t_n$ и комплексные числа $z_1, \ldots, z_n$. Тогда

$$\sum \rho(t_j - t_n) z_j \bar{z}_k = \sum \mathbf{E}(\mathbf{X}_{t_j} \bar{\mathbf{X}}_{t_k}) z_j \bar{z}_k =$$

$$= \mathbf{E}\left(\sum \mathbf{X}_{t_j} z_j \bar{\mathbf{X}}_{t_k} \bar{z}_k\right) = \mathbf{E}\left(\left|\sum \mathbf{X}_{t_j} z_j\right|^2\right) \geqslant 0, \tag{3.6}$$

и (3.4) вытекает из критерия предшествующего параграфа. Если ρ вещественно, то соотношение (3.4) выполняется также и для «отраженной» меры, получаемой заменой x на $-x$. По теореме единственности R симметрично. ▶

Меру R называют *спектральной мерой*[1]) случайного процесса. Множество ее точек роста называют *спектром* $\{\mathbf{X}_t\}$. В большинстве применений случайные величины центрированы так, что $\mathbf{E}\{\mathbf{X}_t\}=0$. В этом случае $\rho(t) = \mathrm{Cov}\,(\mathbf{X}_{t+s}, \mathbf{X}_s)$. По этой причине ρ называют обычно *ковариационной функцией* процесса. Центрирование $\mathbf{X}_t$ никак не влияет на те свойства процесса, которые мы будем изучать.

Примеры. а) Пусть $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ — попарно некоррелированные случайные величины с нулевыми математическими ожиданиями и дисперсиями $\sigma_1^2, \ldots, \sigma_n^2$. Положим

$$\mathbf{X}_t = \mathbf{Z}_1 e^{i\lambda_1 t} + \ldots + \mathbf{Z}_n e^{i\lambda_n t}, \tag{3.7}$$

где числа $\lambda_1, \ldots, \lambda_n$ вещественны. Тогда

$$\rho(t) = \sigma_1^2 e^{i\lambda_1 t} + \ldots + \sigma_n^2 e^{i\lambda_n t}, \tag{3.8}$$

так что мера R сосредоточена в n точках $\lambda_1, \ldots, \lambda_n$. Мы увидим позже, что общие процессы можно трактовать как предельный случай процессов, указанных в этом примере.

Если процесс (3.7) — вещественный, его можно представить в виде

$$\mathbf{X}_t = \mathbf{U}_1 \cos \nu_1 t + \ldots + \mathbf{U}_r \cos \nu_r t + \mathbf{V}_1 \sin \nu_1 t + \ldots + \mathbf{V}_r \sin \nu_r t, \tag{3.9}$$

где $\mathbf{U}_j$ и $\mathbf{V}_j$ — вещественные попарно некоррелированные случайные величины и

$$\mathbf{E}(\mathbf{U}_j^2) = \mathbf{E}(\mathbf{V}_j^2) = \sigma_j^2.$$

[1]) В теории связи говорят также о «спектре мощностей».

Типичный пример см. в гл. III, (7.22). Соответствующая ковариация $\rho_t = \sigma_1^2 \cos \nu_1 t + \ldots + \sigma_r^2 \cos \nu_r t$. Аналогичное замечание применимо и к другим примерам.

б) *Марковские процессы.* Если случайные величины $\mathbf{X}_t$ нормальны и процесс марковский, то $\rho(t) = e^{-a|t|}$ [см. гл. III, (8.14)]. Спектральная мера пропорциональна распределению Коши.

в) Пусть $\mathbf{X}_t = \mathbf{Z}e^{it\mathbf{Y}}$, где $\mathbf{Y}$ и $\mathbf{Z}$ — независимые *вещественные* случайные величины и $\mathbf{E}(\mathbf{Z}) = 0$. Для такого процесса мера R пропорциональна распределению вероятностей $\mathbf{Y}$. ▶

Теоретически не играет роли вопрос о том, описывается ли процесс в терминах ковариационной функции $\rho(t)$ или эквивалентным образом в терминах спектральной меры R. Однако с практической точки зрения описание в терминах спектральной меры обычно проще и предпочтительнее. В применениях к теории связи спектральное описание имеет технические преимущества при проектировании приборов и измерениях, но мы не будем задерживаться на этом обстоятельстве. Более важным преимуществом с нашей точки зрения является то, что линейные операции над $\mathbf{X}_t$ (часто называемые фильтрами) лучше описываются в терминах R, чем в терминах ρ.

г) *Линейные операции.* Начнем с простейшего примера случайных величин $\mathbf{Y}_t$, определяемых равенством

$$\mathbf{Y}_t = \sum c_k \mathbf{X}_{t-\tau_k}, \tag{3.10}$$

где c_k и τ_k — постоянные (τ_k вещественны) и сумма содержит конечное число членов. Ковариационная функция для $\mathbf{Y}_t$ определяется двойной суммой

$$\rho_{\mathbf{Y}}(t) = \sum c_j \bar{c}_k \rho(t - \tau_j + \tau_k), \tag{3.11}$$

а соответствующая спектральная мера $R_{\mathbf{Y}}$ равна

$$R_{\mathbf{Y}}\{d\lambda\} = \left|\sum c_j e^{-i\tau_j \lambda}\right|^2 R\{d\lambda\}. \tag{3.12}$$

В отличие от (3.11) последнее соотношение допускает интуитивную интерпретацию: составляющая R, соответствующая частоте λ, умножается на «частотную характеристику фильтра», определяемую данным преобразованием (3.10).

Этот пример имеет значительно более широкую область применимости, чем это может показаться на первый взгляд. Причина этого в том, что интегралы и производные суть пределы сумм вида (3.10) и потому аналогичное замечание применимо и к ним. Пусть, например, $\mathbf{X}_t$ — сигнал, подаваемый на вход стандартной электрической цепи. Тогда выходной сигнал $\mathbf{Y}_t$ может быть представлен в виде некоторого интеграла, содержа-

щего $\mathbf{X}_t$, и спектральная мера R_Y опять может быть выражена через R и частотную характеристику. Последняя зависит от строения цепи. Поэтому наш результат может быть использован в двух направлениях: именно, чтобы описать процесс на выходе и чтобы строить цепи, выходной сигнал которых обладает заданными свойствами. ▶

Перейдем к теореме, обратной к доказанной выше. Мы покажем, что *для любой меры R на числовой прямой найдется стационарный процесс $\mathbf{X}_t$ со спектральной мерой R.* Поскольку преобразование $\mathbf{X}_t \to a\mathbf{X}_t$ переводит R в a^2R, мы можем, не ограничивая общности, считать R вероятностной мерой. Возьмем в качестве выборочного пространства λ-ось вместе с мерой R, и обозначим $\mathbf{X}_t$ случайную величину $\mathbf{X}_t(\lambda) = e^{it\lambda}$. Тогда

$$\mathbf{E}(\mathbf{X}_{t+s}\overline{\mathbf{X}}_s) = \rho(t). \tag{3.13}$$

Мы построили, таким образом, *модель стационарного процесса с заданной спектральной мерой.* Приятной неожиданностью является тот факт, что в подобной модели возможно выбрать в качестве *выборочного пространства вещественную прямую.* Мы вернемся к этой модели в § 8.

Легко видоизменить модель так, чтобы получить случайные величины с нулевыми математическими ожиданиями. Пусть $\mathbf{Y}$ — случайная величина, не зависящая от всех $\mathbf{X}_t$ и принимающая значение ± 1 с вероятностью по $1/2$ каждое. Положим $\mathbf{X}'_t = \mathbf{Y}\mathbf{X}_t$. Тогда $\mathbf{E}(\mathbf{X}'_t) = 0$ и $\mathbf{E}(\mathbf{X}'_{t+s}\overline{\mathbf{X}}'_s) = \mathbf{E}(\mathbf{Y}^2)\mathbf{E}(\mathbf{X}_{t+s}\overline{\mathbf{X}}_s)$. Таким образом, $\{\mathbf{X}'_t\}$ *является стационарным процессом с ковариационной функцией* ρ.

§ 4. Ряды Фурье

Арифметическое распределение, приписывающее точке n вероятность φ_n, имеет периодическую, с периодом 2π, характеристическую функцию

$$\varphi(\zeta) = \sum_{-\infty}^{+\infty} \varphi_n e^{in\zeta}. \tag{4.1}$$

Вероятности φ_n выражаются с помощью формулы обращения

$$\varphi_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\zeta}\varphi(\zeta)\,d\zeta, \tag{4.2}$$

которая легко выводится из (4.1) [см. гл. XV, (3.14)].

Возьмем теперь произвольную функцию φ с периодом 2π и определим числа φ_k с помощью (4.2). Задача состоит в том,

чтобы узнать, является ли φ характеристической функцией, т. е. задают ли числа φ_n распределение вероятностей. Метод опирается на исследование семейства функций f_r, определенных при $0<r<1$ равенством

$$f_r(\zeta)=\sum_{-\infty}^{+\infty}\varphi_n r^{|n|}e^{in\zeta}. \tag{4.3}$$

Несмотря на их простоту, наши рассуждения приведут к важным результатам относительно рядов Фурье и характеристических функций распределений, сосредоточенных на конечных интервалах.

В последующем удобнее рассматривать основной интервал $-\pi$, π как круг (т. е. отождествить точки π и $-\pi$). Для интегрируемой φ число φ_k *будет называться* k-м *коэффициентом Фурье функции* φ. Ряд (4.1) называют «формальным рядом Фурье», соответствующим φ. Он не обязан сходиться, но так как последовательность $\{\varphi_n\}$ ограничена, то ряд (4.3) сходится к непрерывной (и, более того, дифференцируемой) функции f_r. При $r\to 1$ функции f_r могут стремиться к пределу ψ, даже если ряд (4.1) расходится. В этом случае говорят, что ряд «суммируем по Абелю» к функции ψ.

Примеры. а) Пусть $\varphi_n=1$ для $n=0, 1, 2, \ldots$, но $\varphi_n=0$ для $n<0$. Члены ряда (4.1) не стремятся к нулю, и потому ряд расходится при любом ζ. С другой стороны, правая часть (4.3) есть сумма членов геометрической прогрессии, следовательно,

$$f_r(\zeta)=\frac{1}{1-re^{i\zeta}}. \tag{4.4}$$

Предел при $r\to 1$ существует во всех точках, кроме $\zeta=0$.

б) Важный частный случай (4.3) получается, когда $\varphi_n={}^1/_2\pi$ при всех n;

$$p_r(t)=\frac{1}{2\pi}\sum_{-\infty}^{+\infty}r^{|n|}e^{int}. \tag{4.5}$$

Сумма членов с $n\geqslant 0$ указана в (4.4). По соображениям симметрии получаем

$$2\pi p_r(t)=\frac{1}{1-re^{it}}+\frac{1}{1-re^{-it}}-1 \tag{4.6}$$

или

$$p_r(t)=\frac{1}{2\pi}\cdot\frac{1-r^2}{1+r^2-2r\cos t}. \tag{4.7}$$

Эта функция постоянно используется в теории гармонических функций [где $p_r(t-\zeta)$ называют *ядром Пуассона*]. Для удоб-

ства ссылок мы сформулируем ее основное свойство в следующей лемме.

Лемма. *При фиксированном* $0<r<1$ *функция* p_r *есть плотность некоторого распределения* P_r *на единичной окружности. При* $r \to 1$ *последнее сходится к распределению вероятностей, сосредоточенному в точке с* $t=0$.

Доказательство. Ясно, что $p_r \geqslant 0$. Из (4.5) видно, что интеграл от p_r по $-\pi$, π равен единице, так как при $n \neq 0$ интеграл от e^{int} равен нулю. При $r \to 1$ в каждом открытом интервале, не содержащем $t=0$, $p_r(t)$ ограниченно сходится к нулю. Следовательно, P_r имеет предельное распределение, сосредоточенное в $t=0$. ▶

Теорема 1. *Непрерывная функция* φ, *имеющая период* 2π, *будет характеристической тогда и только тогда, когда ее коэффициенты Фурье* (4.2) *неотрицательны и* $\varphi(0)=1$. *Если эти условия выполнены, ряд Фурье* (4.1) *сходится к* φ.

[Иными словами, формальный ряд Фурье с неотрицательными коэффициентами (4.2) сходится к непрерывной функции в том и только том случае, когда $\sum \varphi_k < \infty$. Последнее условие влечет (4.1).]

Доказательство. В силу (4.2) и (4.5) функция f_r из (4.3) может быть представлена в виде

$$f_r(\zeta) = \int_{-\pi}^{\pi} \varphi(t) \cdot p_r(\zeta - t)\, dt. \tag{4.8}$$

Справа стоит свертка φ и распределения вероятностей P_r, поэтому при $r \to 1$

$$f_r(\zeta) \to \varphi(\zeta). \tag{4.9}$$

Далее если m — верхняя грань $|\varphi|$, то

$$f_r(0) = \sum_{-\infty}^{+\infty} \varphi_n r^{|n|} \leqslant m. \tag{4.10}$$

Так как члены последнего ряда неотрицательны, то переходом к пределу при $r \to 1$ получаем неравенство $\sum \varphi_n \leqslant m$. Поэтому ряд $\sum \varphi_n e^{in\zeta}$ сходится и из (4.3) с очевидностью вытекает, что $f_r(\zeta)$ сходится к его сумме. Таким образом, (4.1) верно. Доказательство закончено. ▶

Заметим, что (4.9) — прямое следствие свойств свертки и не зависит от предположения о неотрицательности коэффициентов φ_n. То есть нами попутно доказана

Теорема 2[1]). *Если φ непрерывна и имеет период 2π, то* (4.9) *выполняется равномерно относительно ζ.*

(Обобщение на случай разрывных функций см. в задачах 5 и 6.)

Следствие 1. (*Фейер.*) *Любая непрерывная периодическая функция φ может быть представлена как предел равномерно сходящейся последовательности тригонометрических многочленов.*

Другими словами, для любого $\varepsilon>0$ найдутся числа $a_{-N}, \ldots, \ldots, a_N$, такие, что

$$\left|\varphi(\zeta)-\sum_{n=-N}^{N} a_n e^{in\zeta}\right|<\varepsilon \tag{4.11}$$

при всех ζ.

Доказательство. Выберем r так, что $|f_r-\varphi|<\frac{1}{2}\varepsilon$. Так как ряд в (4.3) мажорируется геометрической прогрессией, то можно выбрать N столь большим, что сумма всех членов с $n>N$ при всех ζ по абсолютной величине меньше $\frac{1}{2}\varepsilon$. ▶

Полноты ради укажем следующий результат.

Следствие 2. *Две интегрируемые периодические функции с одинаковыми коэффициентами Фурье могут отличаться друг от друга только на множестве меры нуль (т. е. их интегралы совпадают).*

Доказательство. Для интегрируемой периодической функции φ с коэффициентами Фурье φ_n положим

$$\Phi(x)=\int_{-\pi}^{x}[\varphi(t)-\varphi_0]\,dt. \tag{4.12}$$

---

[1]) Теорема может быть сформулирована по-другому: *ряд Фурье непрерывной периодической функции φ суммируем по Абелю к φ.* Теорема (и метод доказательства) легко распространяется на другие методы суммирования.

Указанное явление было впервые замечено Л. Фейером (для суммирования по Чезаро, см. задачу 8) в то время, когда расходящиеся ряды все еще представлялись чем-то таинственным. Открытие Фейера было сенсационным и по историческим причинам в учебниках до сих пор употребляют суммирование по Чезаро, хотя метод Абеля более удобен и делает доказательства единообразными.

Тогда Φ непрерывна и периодична, и интегрирование по частям показывает, что ее коэффициенты Фурье равны $\Phi_n = -i\varphi_n/n$. В силу последней теоремы Φ однозначно определяется по φ_n. ▷

§ 5*. Формула суммирования Пуассона

Пусть f — плотность вероятности с интегрируемой характеристической функцией φ[1]). При слабых условиях регулярности (например, если существует второй момент) для любого $\lambda > 0$ выполняется соотношение

$$\sum_{-\infty}^{+\infty} f(n\pi/\lambda) < \infty. \tag{5.1}$$

Мы можем поэтому ввести арифметическую меру, приписывающую вес $f(n\pi/\lambda)$ точке $n\pi/\lambda$. Ее преобразование Фурье—Стильтьеса будет рядом Фурье с периодом 2λ. Следующая ниже теорема показывает, что этот ряд может быть выражен и в терминах характеристической функции φ. На первый взгляд результат может показаться только забавой, но его частные случаи оказываются важными и полезными.

Теорема 1. *Пусть плотность вероятности f имеет интегрируемую характеристическую функцию φ. Тогда при каждом фиксированном λ*

$$\sum_{-\infty}^{+\infty} \varphi(\zeta + 2k\lambda) = \frac{\pi}{\lambda} \sum_{-\infty}^{+\infty} f(n\pi/\lambda)\, e^{in(\pi/\lambda)\zeta}, \tag{5.2}$$

если только справа или слева стоит непрерывная функция. В этом случае выполняется и (5.1).

Доказательство[2]). Левая часть представляет собой функцию с периодом 2λ. Вычислим формально ее коэффициенты Фурье, а заодно покажем, что эта функция интегрируема. Переделывая формулу (4.2) для нашего случая (интервал $-\lambda$, λ),

---

*) Здесь рассматриваются важные, но специальные вопросы. Содержание параграфа не используется в дальнейшем и практически не зависит от предыдущих параграфов.

[1]) Формула обращения интегралов Фурье [гл. XI, (3.5)] показывает, что интегрируемость φ влечет существование непрерывной плотности, и поэтому всюду в этом разделе подразумевается, что f непрерывна.

[2]) Иное доказательство см. в задачах 12—13. Задача 13 показывает, что условие непрерывности не является излишним.

получаем для n-го коэффициента Фурье левой части выражение

$$\frac{1}{2\lambda}\sum_{k=-\infty}^{+\infty}\int_{-\lambda}^{\lambda}\varphi(\zeta+2k\lambda)\,e^{-in(\pi/\lambda)\zeta}\,d\zeta=$$

$$=\frac{1}{2\lambda}\sum_{k=-\infty}^{+\infty}\int_{(2k-1)\lambda}^{(2k+1)\lambda}\varphi(s)\,e^{-in(\pi/\lambda)s}\,ds. \qquad (5.3)$$

Интервалы $(2k-1)\lambda \leqslant s < (2k+1)\lambda$ попарно не пересекаются и в сумме покрывают всю вещественную ось. Поэтому правая часть равна (по теореме обращения интегралов Фурье гл. XI, 3)

$$\frac{1}{2\lambda}\int_{-\infty}^{+\infty}\varphi(s)\,e^{-in(\pi/\lambda)s}\,ds=\frac{\pi}{\lambda}f(n\pi/\lambda). \qquad (5.4)$$

Далее в теореме 1 предыдущего раздела было показано, что функция с неотрицательными коэффициентами Фурье непрерывна тогда и только тогда, когда ряд из этих коэффициентов сходится. Далее непрерывная функция однозначно определяется своими коэффициентами Фурье, так что (5.1) и (5.2) верны. ▶

Рассмотрим интересный частный случай, когда характеристическая функция φ тождественно равна нулю вне интервала $\overline{-a,\ a}$, где $a\leqslant\lambda$. Для $-\lambda<\zeta<\lambda$ левая часть (5.2) сводится к $\varphi(\zeta)$, так как все члены с $k\neq 0$ исчезают. Поэтому левая часть служит периодическим продолжением φ и, следовательно, непрерывна. При $\zeta=0$ получаем

$$\sum_{-\infty}^{+\infty}\frac{\pi}{\lambda}f\left(n\frac{\pi}{\lambda}\right)=1, \qquad (5.5)$$

откуда видно, что (5.2) представляет собой характеристическую функцию. Иными словами, *если характеристическая функция* φ *равна нулю для* $|\zeta|>a$, *то все ее периодические продолжения с периодом* $\lambda\geqslant a$ *снова будут характеристическими функциями.* Мы получаем, таким образом, метод построения бесконечного числа *характеристических функций, совпадающих на некотором конечном интервале.*

Этот результат используется в теории передачи сообщений (под названием «sampling theorem») и приписывается то Найквисту, то Шеннону. Мы придадим ему форму теоремы.

Теорема 2. *Допустим, что плотность вероятности* f *имеет характеристическую функцию, равную нулю вне* $\overline{-a,\ a}$. *Тогда* f

однозначно определяется по значениям $\frac{\pi}{\lambda} f\left(n\frac{\pi}{\lambda}\right)$ *при любом фиксированном* $\lambda \geqslant a$. *Эти значения индуцируют арифметическое распределение* [1]).

В следующем параграфе будет доказана теорема, в некотором смысле двойственная только что приведенной.

Примеры. а) Рассмотрим плотность $f(x) = (1 - \cos x)/\pi x^2$ с характеристической функцией $\varphi(\zeta) = 1 - |\zeta|$ для $|\zeta| \leqslant 1$ и $\varphi(\zeta) = 0$ для $|\zeta| > 1$. При $\lambda = 1$ из (5.5) получаем хорошо известное соотношение

$$\frac{1}{2} + \frac{4}{\pi^2} \sum_{\nu=0}^{\infty} \frac{1}{(2\nu+1)^2} = 1. \tag{5.6}$$

Периодическое продолжение φ с периодом $2\lambda = 2$ изображено на рис. 2 в гл. XI, 2. Выбирая бо́льший период, можно построить сколько угодно других характеристических функций. Например, значению $\lambda = 2$ соответствует характеристическая функция, представленная на том же рис. 2 ломаной линией с вершинами -3, -1, A, 1, 3, C, 5, 7, E, 9, ... Для каждого λ мы получаем тождество, аналогичное (5.6).

[То обстоятельство, что эти периодические продолжения суть характеристические функции, было использовано в гл. XV, 2 (iii) для построения примера двух действительных характеристических функций с $\varphi_1^2 = \varphi_2^2$.]

б) Простой пример к (5.2) указывается в задаче 11. ▶

Как часто бывает в подобных ситуациях, (5.2) можно переписать в форме, *кажущейся* более общей. В самом деле, применяя (5.2) к плотности $f(x+s)$, мы получаем *другую запись*

---

[1]) *Явная формула для* $f(x)$ может быть получена следующим путем. Внутри интервала $-\lambda \leqslant \zeta \leqslant \lambda$ характеристическая функция φ совпадает с характеристической функцией арифметического распределения $\{(\pi/\lambda) f(n\pi/\lambda)\}$, но вне этого интервала φ всюду равна 0. По формуле обращения

$$f(x) = \frac{1}{2\pi}\frac{\pi}{\lambda} \sum_{-\infty}^{+\infty} \int_{-\lambda}^{\lambda} f\left(n\frac{\pi}{\lambda}\right) e^{in(\pi/\lambda)\zeta - i\zeta x}\, d\zeta = \frac{\sin \lambda x}{\lambda} \sum_{-\infty}^{+\infty} f\left(n\frac{\pi}{\lambda}\right) \frac{(-1)^n}{x - n\pi/\lambda}.$$

Эта формула применима в более широком классе случаев, чем можно думать на основании теоремы 2. [См. теорему 16 в книге Whittaker J. M., Interpolatory function theory, Cambridge Tracts 33, 1935. Многомерный аналог см. Petersen D. P., Middleton D., *Information and Control*, **5** (1962), 279—323.]

формулы суммирования Пуассона:

$$\sum_{-\infty}^{+\infty} \varphi(\zeta+2k\lambda)\, e^{-is(\zeta+2k\lambda)} = \frac{\pi}{\lambda} \sum_{-\infty}^{+\infty} f\left(n\frac{\pi}{\lambda}+s\right) e^{in(\pi/\lambda)\zeta}. \quad (5.7)$$

в) Для нормальной плотности и $\zeta=\lambda$ из (5.7) выводим

$$\sum_{-\infty}^{+\infty} e^{-\frac{1}{2}(2k+1)^2\lambda^2} \cos(2k+1)\lambda s = \frac{\pi}{\lambda} \sum_{-\infty}^{+\infty} (-1)^k n\left(\frac{(2k+1)\pi}{\lambda}+s\right). \quad (5.8)$$

Это — знаменитая формула теории тэта-функций. Мы ссылались на нее в гл. X, 5. [Беря производные по x в гл. X, (5.8) и гл. X, (5.9), мы видим, что эквивалентность этих выражений следует из (5.8) с $\lambda=(\pi/a)\sqrt{t}$ и $s=x\sqrt{t}$. Доказательство, данное в гл. X, 5, элементарнее.]

г) Для плотности $f(x)=\pi^{-1}(1+x^2)^{-1}$ с характеристической функцией $\varphi(\zeta)=e^{-|\zeta|}$ мы получаем из (5.2), полагая $\zeta=0$,

$$\frac{e^{\lambda}+e^{-\lambda}}{e^{\lambda}-e^{-\lambda}} = \sum_{n=-\infty}^{+\infty} \frac{\lambda}{\lambda^2+n^2\pi^2}. \quad (5.9)$$

Это не что иное, как *разложение гиперболического котангенса на элементарные дроби.*

д) *Плотности на окружности* длины 2π можно получить, «наматывая» действительную ось на окружность, как это объяснено в гл. II, 8. При этом плотности f на прямой ставится в соответствие плотность на окружности, задаваемая рядом в правой части (5.7) с $\lambda=1/2$ и $\zeta=0$. Соотношение (5.7) дает представление этой плотности в терминах первоначальной характеристической функции.

Для частного случая нормальной плотности с нулевым математическим ожиданием и дисперсией t получаем

$$\frac{1}{\sqrt{2\pi t}} \sum_{-\infty}^{+\infty} \exp\left(-\frac{1}{2t}(s+2n\pi)^2\right) = \frac{1}{2\pi} \sum_{-\infty}^{+\infty} e^{-\frac{1}{2}n^2t} \cos ns. \quad (5.10)$$

Слева стоит плотность на окружности, по правая часть дает для нее более полезное выражение. Оно показывает, что эти плотности обладают полугрупповым свойством. ▶

§ 6. Положительно определенные последовательности

Пусть F — распределение вероятностей на открытом интервале $-\pi$, π и φ — его характеристическая функция. Как и в § 4, мы отождествим точки $-\pi$ и π и будем интерпретировать F

как распределение на окружности. По аналогии с (4, г) мы определяем *коэффициенты Фурье* φ_k распределения F равенством

$$\varphi_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}F\{dt\}, \qquad k=0, \pm 1, \ldots, \tag{6.1}$$

так что $2\pi\varphi_k=\varphi(-k)$. Покажем, что распределение F однозначно восстанавливается по коэффициентам φ_k. Отвлекаясь от тривиального изменения масштаба, мы видим, что наше утверждение эквивалентно следующему: *распределение, сосредоточенное на* $\overline{-\lambda, \lambda}$, *однозначно определяется по значениям* $\varphi(n\pi/\lambda)$, принимаемым характеристической функцией в точках с абсциссами, кратными π/λ. Этим объясняется формальное различие между спектральной теорией стационарных последовательностей $\{\mathbf{X}_n\}$ и спектральной теорией стационарных процессов $\{\mathbf{X}_t\}$ (с непрерывным временем). (Высказанное выше утверждение двойственно теореме 2 предыдущего параграфа, в соответствии с которой характеристическая функция, равная нулю вне $\overline{-\lambda, \lambda}$, однозначно определяется по значениям $f(n\pi/\lambda)$ соответствующей плотности.

Теорема 1. *Распределение F на окружности однозначно определяется своими коэффициентами Фурье φ_k.*

Доказательство. Как и в (4.3), положим при $0<r<1$

$$f_r(\zeta) = \sum_{-\infty}^{+\infty} \varphi_n \cdot r^{|n|} \cdot e^{in\zeta}. \tag{6.2}$$

Вычисления, подобные тем, которые привели к (4.8), показывают, что теперь

$$f_r(\zeta) = \int_{-\pi}^{\pi} p_r(\zeta - t)F\{dt\}, \tag{6.3}$$

где p_r — плотность распределения P_r (см. лемму в § 4). Таким образом, f_r есть плотность свертки $P_r * F$, которая сходится к F при $r \to 1$. Поэтому F восстанавливается по f_r (см. также задачу 10). ▶

Лемма 1. *Пусть $\{\varphi_n\}$ — произвольная ограниченная последовательность комплексных чисел. Для того чтобы существовала мера F с коэффициентами Фурье φ_n, необходимо и достаточно, чтобы при каждом $r<1$ функции f_r [см. (6.2)] удовлетворяли условию*

$$f_r(\zeta) > 0. \tag{6.4}$$

Доказательство. Необходимость, очевидно, следует из (6.3) и положительности p_r. Умножая теперь (6.2) на $e^{-ik\zeta}$ и интегрируя, получаем

$$\varphi_k \cdot r^{|k|} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_r(\zeta)\, e^{-ik\zeta}\, d\zeta. \tag{6.5}$$

Полагая $k=0$, видим, что $\varphi_0>0$. Не ограничивая общности, допустим, что $\varphi_0=1/2\pi$. Тогда (6.5) показывает, что f_r есть плотность некоторого распределения F_r на окружности и что $\varphi_k r^{|k|}$ есть k-й коэффициент Фурье F_r. По теореме о выборе найдется такая последовательность значений $r \to 1$, что F_r сходятся к некоторому распределению F. Из (6.5) вытекает, очевидно, что числа φ_k удовлетворяют (6.1). Доказательство закончено. ▶

Поступая так же, как и в § 2, мы установим аналог теоремы Бохнера (принадлежащий Герглотцу).

Определение. *Последовательность* $\{\varphi_k\}$ *называется положительно определенной, если при любом выборе чисел* $z_1, \ldots, z_n$

$$\sum_{j,\,k} \varphi_{j-k} z_j \bar{z}_k \geqslant 0. \tag{6.6}$$

Лемма 2. *Если последовательность* $\{\varphi_n\}$ *положительно определена, то* $\varphi_0 \geqslant 0$ *и* $|\varphi_n| \leqslant \varphi_0$.

Доказательство повторяет доказательство леммы 2.3 (см. также задачу 14).

Теорема 2. *Последовательность* $\{\varphi_n\}$ *представляет собой последовательность коэффициентов Фурье некоторой меры* F *на окружности в том и только том случае, когда она положительно определена.*

Доказательство. а) Простые вычисления показывают, что для φ_k, задаваемых (6.1), левая часть (6.6) равна интегралу по мере F от $\frac{1}{2\pi}\left|\sum e^{-ijt} z_j\right|^2$. Поэтому условие (6.6) необходимо.

б) Для доказательства достаточности положим $z_k = r^k e^{ikt}$ при $k \geqslant 0$ и $z_k = 0$ при $k<0$. Хотя последовательность z_k бесконечна, соотношение (6.6) для нее выполняется (что можно установить простым предельным переходом). Левая часть (6.6) при этом равна

$$\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \varphi_{j-k} r^{j+k} e^{i(j-k)t} = \sum_{n=-\infty}^{+\infty} \varphi_n e^{int} \sum_{k=0}^{\infty} r^{|n|+2k}. \tag{6.7}$$

Внутренняя сумма справа равна $r^{|n|}(1-r^2)$, так что функция (6.7) равна $f_r(t)/(1-r^2)$. По лемме 1 числа φ_k суть коэффициенты Фурье некоторой меры. ▶

Из теоремы 2 сразу получаем следующий аналог теоремы § 3.

Теорема 3. *Пусть* $\{\mathbf{X}_n\}$ *— последовательность случайных величин, заданных на некотором пространстве вероятностей, и пусть*

$$\rho_n = \mathbf{E}(\mathbf{X}_{n+\nu}\mathbf{X}_\nu) \tag{6.8}$$

не зависит от ν. *Тогда существует, и притом только одна мера* R *на окружности, такая, что числа* ρ_n *суть ее коэффициенты Фурье.*

Доказательство. Очевидно,

$$\sum \rho_{j-k} z_j \bar{z}_k = \sum \mathbf{E}(\mathbf{X}_j z_j \bar{\mathbf{X}}_k \bar{z}_k) = \mathbf{E}\left(\left|\sum \mathbf{X}_j z_j\right|^2\right), \tag{6.9}$$

т. е. последовательность $\{\rho_n\}$ положительно определена. ▶

Как и в § 3, построением случайных величин с заданными ковариациями можно доказать утверждение, обратное теореме 3. Мы вернемся к этому в § 8.

Примеры. а) Последовательности независимых случайных величин соответствует равномерная спектральная мера.

б) *Марковские процессы.* Из построений § 4 видно, что плотности, определенной при фиксированном r и действительных θ как $p_r(t-\theta)$, соответствуют коэффициенты Фурье

$$\rho_n = r^{|n|} e^{in\theta}. \tag{6.10}$$

В гл. III, 8 было показано, что нормальная стационарная марковская последовательность имеет ковариации вида

$$\rho_n = r^{|n|} \quad \text{с} \quad 0 \leqslant r \leqslant 1.$$

Сходные рассуждения показывают, что ковариации произвольной стационарной марковской последовательности имеют вид (6.10). При $r<1$ спектральная мера имеет плотность $p_r(t-\theta)$, а при $r=1$ она сосредоточена в точке θ. ▶

§ 7. L^2-теория

Для целей теории вероятностей мы ввели характеристические функции, определив их как некоторые преобразования мер. Однако в равной степени естественны и другие подходы к гармоническому анализу. В частности, возможно ввести преобразования Фурье для функций (а не для мер!). Формула обращения интегралов Фурье делает правдоподобным предположение о том,

что на этом пути можно достичь бóльшей симметрии. Оказывается, что простота и красота теории становятся наибольшими, когда рассматриваются лишь функции с интегрируемым квадратом. Мы разберем сейчас этот случай. Он интересен и сам по себе и по тем обширным применениям, которые он находит в теории случайных процессов.

Определим *норму* $\|u\| \geqslant 0$ комплекснозначной функции и действительного аргумента x равенством

$$\|u\|^2 = \int_{-\infty}^{+\infty} |u(x)|^2 dx. \tag{7.1}$$

Две функции, отличающиеся на множестве меры нуль, мы будем рассматривать как идентичные (другими словами, мы будем иметь дело с классами эквивалентности, но присоединимся к обычному и безвредному неправильному словоупотреблению). При этом соглашении $\|u\| = 0$ тогда и только тогда, когда $u = 0$. Класс всех функций с конечной нормой обозначим L^2. *Расстояние* между u и v из L^2 определяется как $\|u - v\|$. При таком определении L^2 становится метрическим пространством и последовательность функций u_n из L^2 сходится в этой метрике [1]) тогда и только тогда, когда $\|u_n - u\| \to 0$. Эта сходимость [2]) будет обозначаться символами $u = \text{l. i. m. } u_n$ или $u_n \xrightarrow{\text{l. i. m.}} u$. Отметим без доказательства, что метрическое пространство L^2 полно в том смысле, что каждая последовательность Коши [3]) $\{u_n\}$ имеет (единственный) предел $u \in L^2$.

Скалярное произведение (u, v) двух функций определяется равенством

$$(u, v) = \int_{-\infty}^{+\infty} u(x)\,\overline{v(x)}\,dx. \tag{7.2}$$

---

[1]) Этот тип сходимости называют «сходимостью в среднем квадратичном».

[2]) Поточечная сходимость u_n к пределу v не влечет $u_n \to v$ [см. пример (2, г) гл. IV]. Однако если известно, что $u = \text{l. i. m. } u_n$ существует, то $u = v$. В самом деле, по лемме Фату

$$\int_{-\infty}^{+\infty} |u(x) - v(x)|^2 dx \leqslant \lim \int_{-\infty}^{+\infty} |u(x) - u_n(x)|^2 dx = 0.$$

[3]) $\{u_n\}$ называют последовательностью Коши, если $\|u_n - u_m\| \to 0$ при $n, m \to \infty$.

Оно существует для любых двух функций из L^2, так как по неравенству Шварца

$$|(u, v)|^2 \leqslant \|u\| \cdot \|v\|. \tag{7.3}$$

В частности, $(u, u) = \|u\|^2$. При указанном определении скалярного произведения L^2 становится *гильбертовым пространством*. Аналогия между скалярным произведением (7.2) и ковариацией двух случайных величин с нулевыми математическими ожиданиями очевидна и позже будет нами использована.

После этих предварительных замечаний перейдем к нашей основной задаче — определению преобразований вида

$$\hat{u}(\zeta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u(x) e^{i\zeta x} dx. \tag{7.4}$$

В случае когда u — плотность вероятности, $\hat{u}$ лишь множителем $\sqrt{2\pi}$ отличается от характеристической функции. Чтобы избежать недоразумений, назовем $\hat{u}$ *преобразованием Планшереля функции* u. Определение (7.4) применимо только к интегрируемым функциям. Однако мы расширим область определения u до всего пространства L^2. Для краткости назовем функцию u, принадлежащую L^2, «хорошей», если и u, и $\hat{u}$ интегрируемы.

Сначала мы покажем, что для «хороших» функций справедлива *формула обращения*

$$u(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{u}(\zeta) e^{-i\zeta t} d\zeta. \tag{7.5}$$

(Это легко вывести из предыдущих результатов, но доказательство столь просто, что его стоит привести.) Умножая (7.4) на $\frac{1}{\sqrt{2\pi}} e^{-i\zeta t - \frac{1}{2}\varepsilon^2\zeta^2}$ и интегрируя, получим

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{u}(\zeta) e^{-i\zeta t - \frac{1}{2}\varepsilon^2\zeta^2} d\zeta = \int_{-\infty}^{+\infty} u(x) \mathfrak{n}\left(\frac{t-x}{\varepsilon}\right) \frac{dx}{\varepsilon}, \tag{7.6}$$

где $\mathfrak{n}$ обозначает стандартную нормальную плотность. При $\varepsilon \to 0$ интеграл слева стремится к интегралу в (7.5), в то время как свертка справа стремится к $u(t)$. Таким образом, (7.5) для «хороших» функций выполняется.

Пусть теперь v — другая «хорошая» функция. Умножим (7.5) на функцию $\bar{v}$, комплексно сопряженную с v, и произведем интегрирование. Мы получим соотношение

$$(\hat{u}, \hat{v}) = (u, v), \tag{7.7}$$

которое будем называть *равенством Парсеваля*. Для $v=u$ оно сводится к

$$\|\hat{u}\| = \|u\|, \tag{7.8}$$

откуда видно, что преобразование Планшереля на множестве «хороших» функций *изометрично*.

Теперь все готово для последнего шага, а именно для распространения преобразования Планшереля на все L^2. Предположим, что для некоторой функции $u \in L^2$ существует сходящаяся к ней в принятой нами метрике последовательность «хороших» функций $u^{(n)}$. В силу (7.8) преобразования $\hat{u}^{(n)}$ образуют последовательность Коши и $\hat{u}^{(n)} \to \hat{u}$, если функция u — «хорошая». Если u не является «хорошей», то мы желаем *определить* $\hat{u}$ как единственный предел последовательности $\hat{u}^{(n)}$. Иными словами, мы желаем определить преобразование Планшереля по принципу:

$$\textit{если } u^{(n)} \to u, \textit{ то мы полагаем } \hat{u} = \text{l. i. m.}\, \hat{u}^{(n)}. \tag{7.9}$$

Так как две последовательности Коши можно соединить в одну, предел $\hat{u}$ не зависит от выбора последовательности $\{u^{(n)}\}$. Чтобы оправдать принцип (7.9), остается показать, что он согласуется с первоначальным определением (7.4), когда функция u интегрируема. Для таких функций положим $u_\varepsilon(x) = u(x)\, e^{-\varepsilon^2 x^2/2}$ и обозначим $\hat{u}_\varepsilon$ преобразование Планшереля u_ε. Беря в (7.4) свертку с нормальной плотностью, мы видим, что равенство (7.6) остается верным, если u и $\hat{u}$ поменять местами. Левая часть тогда становится равной $\hat{u}_\varepsilon(t)$, так что $\hat{u}_\varepsilon$ представляется как свертка ограниченной непрерывной функции $\hat{u}$ с нормальной плотностью. Отсюда следует, что $\hat{u}_\varepsilon$ интегрируема и $\hat{u}_\varepsilon(t) \to \hat{u}(t)$ при $\varepsilon \to 0$. Функции u_ε — «хорошие» и $u_\varepsilon \xrightarrow{\text{l. m.}} u$. Поэтому $\hat{u}_\varepsilon \xrightarrow{\text{l. m.}} \hat{u}$ (см. примечание 2 на стр. 714). Тем самым доказано, что принцип (7.9) совместим с (7.4).

Так как множество интегрируемых функций плотно в L^2, мы доказали заодно, что каждая функция из L^2 может быть представлена как предел «хороших» функций. Соответственно *каждая функция из L^2 обладает преобразованием Планшереля $\hat{u}$, определенным по* (7.9). При этом (7.8) и (7.9) выполняются всегда. Далее $u(-t)$ служит преобразованием Планшереля для $\hat{u}$, т. е. каждая функция из L^2 представляется как преобразование некоторой другой функции. Другими словами, *отображение $u \to \hat{u}$ является изометричным отображением L^2 на себя.*

Примеры. а) Функция $u(x)=(\sin x)/x$ не интегрируема, но принадлежит L^2. Так как u — характеристическая функция равномерной на $-1, 1$ плотности, то следует ожидать, что

$$\hat{u}(\zeta)=\begin{cases}\sqrt{2\pi} & \text{при } |\zeta|<1,\\ 0 & \text{при } |\zeta|>1.\end{cases} \tag{7.10}$$

Для формальной проверки рассмотрим последовательность характеристических функций $u^{(n)}(x)=u(x)\,e^{-x^2/n}$, сходящуюся к u. Преобразования $\hat{u}^{(n)}$ представляются в виде свертки функции (7.10) и нормальных плотностей и потому сходятся к (7.10).

б) Интегрируемая в квадрате функция u оказывается автоматически интегрируемой на любом *конечном* интервале. Если определить функции $u^{(n)}$, полагая $u^{(n)}(x)=u(x)$ при $|x|<n$ и $u^{(n)}(x)=0$ при $|x|\geqslant n$, то $u^{(n)}$ будут интегрируемыми и $u^{(n)}\to u$ в нашей метрике. Далее,

$$\hat{u}^{(n)}(\zeta)=\frac{1}{\sqrt{2\pi}}\int_{-n}^{n}u(x)\,e^{i\zeta x}\,dx. \tag{7.11}$$

Следовательно, $\hat{u}$ является пределом преобразований (7.11) в том смысле, что

$$\|\hat{u}-\hat{u}^{(n)}\|\to 0.$$

Наряду с аппроксимацией, получаемой «урезанием», можно было бы с таким же успехом использовать и аппроксимацию $u(x)\,e^{-x2/n}$ из предыдущего примера. ▶

Подведем итог: *преобразование Планшереля $\hat{u}$ определено для каждой функции $u\in L^2$, при этом всегда выполняется равенство Парсеваля* (7.7). Преобразование Планшереля «кососимметрично» в том смысле, что *преобразование $\hat{u}$ равно $\bar{u}$.*

Как уже упоминалось в другой связи, равенство Парсеваля (7.7) можно переписать в других формах, которые *выглядят* более общими. Одна из возможностей состоит в замене $\hat{v}(x)$ на $\hat{v}(x)\,e^{-i\lambda x}$ (преобразование функции $v(x+\lambda)$). В специальном случае $u=v$ получаем

$$\int_{-\infty}^{+\infty}|\hat{u}(x)|^2\,e^{i\lambda x}\,dx=\int_{-\infty}^{+\infty}u(x)\,\overline{u(x+\lambda)}\,dx. \tag{7.12}$$

Это замечательное соотношение широко используется в теории прогнозирования случайных процессов. Пусть дана произвольная интегрируемая функция $f\geqslant 0$. Тогда можно выбрать $\hat{u}$ так, что $|\hat{u}|^2=f$ и из (7.12) вытекает следующее утверждение:

функция φ *является характеристической функцией для некоторой плотности вероятностей тогда и только тогда, когда* [1])

$$\varphi(\lambda)=\int_{-\infty}^{+\infty} u(x)\,\overline{u(x+\lambda)}\,dx, \tag{7.13}$$

где $\|u\|=1$. Выбор u возможен не единственным способом. (Решение одной из задач теории прогнозирования случайных процессов связано с возможностью выбрать функцию u, равной нулю на полупрямой.)

Изложенная выше L^2-теория интегралов Фурье без существенных изменений переносится на *ряды Фурье*. Норма $\|u\|$ и скалярное произведение определяются так же, как в (7.1) и (7.2) с той лишь разницей, что интегралы берутся по «основному» промежутку $-\pi$, π. Роль «хороших» функций играют конечные тригонометрические многочлены вида

$$u(x)=\sum u_n e^{inx}. \tag{7.14}$$

Их коэффициенты Фурье равны, очевидно,

$$u_k=\frac{1}{2\pi}\int_{-\pi}^{\pi} u(x)\,e^{-inx}\,dx. \tag{7.15}$$

Каждой «хорошей» функции соответствует конечная последовательность $\{u_n\}$ ее коэффициентов и обратно, каждой конечной последовательности комплексных чисел соответствует «хорошая» функция. Соотношения (7.14) и (7.15) определяют преобразование Фурье $\widehat{u}=\{u_n\}$ и обратное преобразование. Умножение и интегрирование показывают, что для двух «хороших» функций

$$\int_{-\pi}^{\pi} u(x)\,\overline{v(x)}\,dx=2\pi\sum u_k\bar{v}_k. \tag{7.16}$$

Рассмотрим теперь гильбертово пространство $\mathfrak{H}$ бесконечных последовательностей $\widehat{u}=\{u_n\}$, $\widehat{v}=\{v_n\}$ и т. д. Определим норму и скалярное произведение равенствами

$$\|\widehat{u}\|=\sum|u_n|^2,\quad (\widehat{u},\widehat{v})=\sum u_n\bar{v}_n. \tag{7.17}$$

Это пространство обладает свойствами, сходными со свойствами L^2. В частности, конечные последовательности образуют всюду

[1]) Этот специальный случай равенства Парсеваля встречается еще в классической работе Винера, но часто называется *критерием Хинчина* (и рассматривается так, как будто он требует отдельного доказательства).

плотное множество. Можно показать, что существует взаимно однозначное соответствие между последовательностями $\hat{u}=\{u_n\}$ из $\mathfrak{H}$ и функциями u из $L^2(-\pi, \pi)$. *Каждой последовательности* $\{u_n\}$, *для которой* $\sum|u_n|^2<\infty$, *соответствует функция с интегрируемым квадратом и коэффициентами Фурье* u_n. *Верно и обратное утверждение.* Отображение $u \leftrightarrow \{u_n\}$ снова *изометрично.* Снова имеет место *равенство Парсеваля* $(u, v)=(\hat{u}, \hat{v})$. Ряд Фурье (7.14) не обязан сходиться, но его частные суммы $\sum_{-n}^{n} u_k e^{ikx}$ образуют последовательность непрерывных функций, сходящуюся к u в метрике L^2. Это же верно и по отношению к другим непрерывным аппроксимациям: так, $\sum u_k r^{|k|} e^{ikx}$ сходится к u при $r \to 1$.

Рассмотрим, как и выше, специальный случай равенства Парсеваля

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} u(x)\,\bar{v}(x)\,e^{-inx}\,dx=\sum_k u_{k+n}\bar{v}_k. \tag{7.18}$$

При $u=v$ получаем, что *последовательность* $\{\varphi_n\}$ *является последовательностью коэффициентов Фурье некоторой плотности вероятностей на* $-\pi, \pi$ *в том и только том случае, когда она допускает представление*

$$\varphi_n=\sum u_{k+n}\bar{u}_k, \quad \textit{где} \quad \sum|u_k|^2=1. \tag{7.19}$$

Ковариация такого вида встречалась в гл. III, (7.4) (см. также задачу 17).

§ 8. Случайные процессы и стохастические интегралы

Ради упрощения обозначений мы рассмотрим в этом параграфе лишь *последовательности* $\{\mathbf{X}_n\}$ случайных величин. Однако, и это станет очевидным из дальнейшего, все изложение пригодно и для семейств, зависящих от непрерывного параметра. Отличие лишь в том, что спектральные меры не будут сосредоточены на конечном промежутке, и суммы должны быть заменены интегралами.

Пусть $\{\mathbf{X}_n\}$ обозначает бесконечную в обе стороны последовательность случайных величин, заданных на некотором вероятностном пространстве $\mathfrak{S}$ и имеющих конечные вторые моменты. Эта последовательность предполагается стационарной «в широком смысле» [1]), т. е.

$$\mathbf{E}(\mathbf{X}_{n+\nu}\bar{\mathbf{X}}_\nu)=\rho_n$$

[1]) У автора — «in restricted sense». — *Прим. перев.*

не зависит от ν. По теореме 6.3 существует единственная мера R на окружности $-\pi$, π, такая, что

$$\rho_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-inx} R\{dx\}, \quad n = 0, \pm 1, \ldots. \tag{8.1}$$

Мы рассмотрим теперь более детально отмеченную еще в § 3 идею о том, что окружность вместе со спектральной мерой R могут быть использованы для построения *конкретного представления случайного процесса* $\{\mathbf{X}_n\}$. (Представления, пригодного во всяком случае для изучения свойств, зависящих только от вторых моментов.) Мы будем применять терминологию гильбертовых пространств и рассмотрим два гильбертовых пространства.

а) *Пространство* L_R^2. Построим пространство функций, определенных на окружности, повторяя (с заменой прямой на окружность $-\pi$, π и меры Лебега на меру R) рассуждения, касающиеся L^2 (см. § 7). Норма и скалярное произведение (комплексно-значных) функций на окружности определяются при этом равенствами

$$\|u\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |u(x)|^2 R\{dx\}, \quad (u, v) = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(x)\overline{v(x)} R\{dx\}. \tag{8.2}$$

Примем основное соглашение: *считать две функции равными, если они отличаются друг от друга на множестве R-меры нуль.* Последствия этого соглашения серьезны. Так, если R сосредоточено в двух точках 0 и 1, то «функция» в нашем понимании вполне определяется своими значениями в этих точках. Например, функция $\sin n\pi x$ равна нулю. Однако даже в этих крайних случаях использование обычных формул для непрерывных функций и ссылки на их графики не принесут никакого вреда. Точно так же осмысленно (и упрощает язык) применение термина «ступенчатая функция».

Пространство Гильберта L_R^2 *состоит из всех функций, заданных на окружности и имеющих конечную норму.* Если $\|u_n - u\| \to 0$, то мы назовем последовательность $\{u_n\}$ сходящейся к u в нашей метрике (или в среднем квадратичном по отношению к мере R). Гильбертово пространство L_R^2 есть полное метрическое пространство. Непрерывные функции образуют в нем плотное множество (определения см. в § 7).

б) *Гильбертово пространство* $\mathfrak{H}$, *натянутое на* $\{\mathbf{X}_n\}$. Обозначим символом $\mathfrak{H}_0$ совокупность случайных величин с конечными

вторыми моментами, заданных на произвольном, но фиксированном пространстве выборок $\mathfrak{S}$. По неравенству Шварца $\mathbf{E}(\mathbf{U}\bar{\mathbf{V}})$ существует для любой пары таких случайных величин. Естественно обобщить (8.2), рассматривая вместо круга пространство $\mathfrak{S}$ и беря вместо меры R основное распределение вероятностей в $\mathfrak{S}$. Мы опять условимся отождествлять две случайные величины, если они отличаются друг от друга на множестве меры пуль. Определим скалярное произведение как $\mathbf{E}(\mathbf{U}\bar{\mathbf{V}})$ и норму $\mathbf{U}$ как положительный корень из $\mathbf{E}(\mathbf{U}\bar{\mathbf{U}})$. Тогда $\mathfrak{H}_0$ превращается в гильбертово пространство. Это — полное метрическое пространство. Последовательность принадлежащих ему случайных величин $\mathbf{U}_n$ называют сходящейся к $\mathbf{U}$, если $\mathbf{E}(|\mathbf{U}_n - \mathbf{U}|^2) \to 0$.

Изучая последовательность $\{\mathbf{X}_n\}$, обычно интересуются только случайными величинами, которые являются функциями $\mathbf{X}_k$, а во многих случаях ограничиваются лишь линейными функциями, т. е. рассматривают конечные линейные комбинации $\sum a_k \mathbf{X}_{n_k}$ и их пределы. Получаемые таким образом случайные величины образуют подпространство $\mathfrak{H}$ пространства $\mathfrak{H}_0$, называемое *гильбертовым пространством, натянутым на величины* $\mathbf{X}_k$. В нем скалярные произведения, нормы и сходимость определены, как только что было описано, и $\mathfrak{H}$ является полным метрическим пространством.

В настоящем контексте математические ожидания $\mathbf{E}(\mathbf{X}_n)$ не играют никакой роли, но, так как термин «ковариация» звучит лучше, чем «скалярное произведение», мы примем обычное предположение, что $\mathbf{E}(\mathbf{X}_n) = 0$. Единственная цель этого предположения — добиться совпадения ρ_n и ковариации. Никакое центрирование не нужно, если согласиться назвать $\mathbf{E}(\mathbf{X}\bar{\mathbf{Y}})$ ковариацией $\mathbf{X}$ и $\mathbf{Y}$. Мы подходим теперь к решающему моменту. Мы покажем, что для наших целей возможно взять пространство L^2_R (простое для интуитивного восприятия) *в качестве конкретной модели* $\mathfrak{H}$. В самом деле, по определению ковариация $\rho_{j-k} = \mathrm{Cov}\,(\mathbf{X}_j, \mathbf{X}_k)$ любой пары $\mathbf{X}$ равна скалярному произведению функций e^{ijx} и e^{ikx} в L^2_R. Отсюда следует, что ковариация двух линейных комбинаций $\mathbf{U} = \sum a_j \cdot \mathbf{X}_{n_j}$ и $\mathbf{V} = \sum b_k \mathbf{X}_{n_k}$ равна скалярному произведению соответствующих линейных комбинаций $u = \sum a_j e^{in_j x}$ и $v = \sum b_k e^{in_k x}$. Определения сходимости в рассматриваемых пространствах позволяют распространить отображение на все случайные величины. Мы получаем, таким образом, важный результат: отображение $\mathbf{X}_k \leftrightarrow e^{ikx}$ *порождает взаимно однозначное соответствие между случайными величи-*

нами из $\mathfrak{H}$ *и функциями из* L^2_R; *это соответствие сохраняет скалярные произведения и нормы* (и, следовательно, сходимость). Иными словами, эти пространства изометричны [1]). Теперь мы в состоянии изучать $\mathfrak{H}$ и $\{\mathbf{X}_n\}$, рассматривая лишь конкретное пространство L^2_R. Эта процедура, помогая интуиции, в то же время имеет и теоретические преимущества. Так как функции на окружности — объект хорошо знакомый, то сравнительно просто построить последовательность $\{u^{(n)}\}$ с заданными структурными свойствами. Функциям $u^{(n)}$ соответствуют некоторые случайные величины $\mathbf{Z}_n$, определенные на пространстве выборок $\mathfrak{S}$. Если известны коэффициенты Фурье $u^{(n)}$, то $\mathbf{Z}_n$ возможно выразить явно в виде предела конечных линейных комбинаций величин $\mathbf{X}_k$. Если совместные распределения $\mathbf{X}$ нормальны, то и распределение $\mathbf{Z}$ будет нормальным.

На практике указанная процедура часто производится в обратном порядке. Если данный процесс $\{\mathbf{X}_k\}$ — сложный, то наша цель — выразить его в терминах величин $\mathbf{Z}_n$, образующих более простой случайный процесс. Несколько простых примеров лучше пояснят задачу, чем теоретические рассуждения.

Примеры. а) *Представление* $\{\mathbf{X}_n\}$ *в терминах независимых случайных величин.* Рассмотрим случай, когда спектральная мера R имеет плотность r. Для простоты допустим, что r строго положительна и непрерывна [2]). Выберем функцию γ так, что

$$|\gamma(x)|^2 = r(x). \tag{8.3}$$

Ряд Фурье γ сходится в L^2-норме, как это было объяснено в § 7. Обозначая коэффициенты Фурье γ через γ_k, имеем $\Sigma|\gamma_k|^2<\infty$ и,

[1]) Читателям, знакомым с теорией гильбертовых пространств, полезно иметь в виду связь изложенного со стандартной *спектральной теоремой для унитарных операторов*. Линейный оператор, отображающий $\mathfrak{H}$ на себя таким образом, что $\mathbf{X}_n \to \mathbf{X}_{n+1}$, называют *оператором сдвига*, и L^2_R служит моделью области действия этого оператора. Обратно, пусть дан произвольный унитарный оператор T на гильбертовом пространстве $\mathfrak{H}_0$ и произвольный элемент $\mathbf{X}_0 \in \mathfrak{H}_0$. Тогда последовательность элементов $\mathbf{X}_n = T^n \mathbf{X}_0$ можно рассматривать как стационарную последовательность и T — как оператор сдвига в подпространстве $\mathfrak{H}$, натянутом на эту последовательность. Если $\mathbf{X}_0$ можно выбрать так, что $\mathfrak{H} = \mathfrak{H}_0$, мы получаем стандартную спектральную теорему для T (с тем отличием, что для «разложения единицы» взято конкретное представление, основанное на выборе $\mathbf{X}_0$). Если $\mathfrak{H} \subset \mathfrak{H}_0$, то $\mathfrak{H}_0$ можно представить как прямую сумму двух инвариантных подпространств и можно повторить рассуждения. Таким путем возможно прийти к общей теории спектральных разложений, включая теорию кратности спектра.

[2]) Эти ограничения используются лишь затем, чтобы избежать скучных объяснений того, как понимать $r(x)/\gamma(x)$ при $r(x)=\gamma(x)=0$ и как понимать сходимость рядов.

по равенству Парсеваля (7.19),

$$\rho_n = \sum_{k=-\infty}^{+\infty} \gamma_{k+n}\bar{\gamma}_k. \tag{8.4}$$

Рассмотрим теперь бесконечную в обе стороны последовательность $u^{(n)}$, определенную соотношением

$$u^{(n)}(x) = \frac{e^{inx}}{\gamma(x)}. \tag{8.5}$$

Подставляя $u^{(n)}$ в (8.2), видим, что

$$\|u^{(n)}\| = 1 \text{ и } (u^{(n)}, u^{(m)}) = 0 \tag{8.6}$$

при $m \neq n$. Отсюда вытекает, что соответствующие случайные величины $\mathbf{Z}_n$ некоррелированы и имеют единичные дисперсии. В частности, *если* $\mathbf{X}_k$ *нормальны, то* $\mathbf{Z}_k$ *взаимно независимы.*

Интересно отметить, что пространство, натянутое на случайные величины $\mathbf{X}_k$, содержит стационарную последовательность $\{\mathbf{Z}_n\}$ из некоррелированных величин. Явное выражение $\mathbf{Z}_n$ в терминах $\mathbf{X}_k$ может быть получено из разложения Фурье функций $u^{(n)}$. Однако предпочтительнее идти в обратном направлении — выразить $\mathbf{X}_k$ в терминах $\mathbf{Z}_n$, так как структура последовательности $\{\mathbf{Z}_n\}$ проще, чем структура $\{\mathbf{X}_k\}$. Мы имеем

$$\sum_{n=-N}^{N} \gamma_n u^{(n+k)}(x) = \frac{e^{ikx}}{\gamma(x)} \sum_{n=-N}^{N} \gamma_n e^{inx}. \tag{8.7}$$

Сумма справа представляет собой отрезок ряда Фурье для γ и сходится (в метрике гильбертова пространства) к γ. Отсюда следует, что величина (8.7) стремится к e^{ihx}. При нашем отображении $u^{(n+k)}$ соответствует $\mathbf{Z}_{n+k}$, так что ряд $\Sigma\gamma_n\mathbf{Z}_{n+k}$ сходится и мы можем написать

$$\mathbf{X}_k = \sum_{n=-\infty}^{+\infty} \gamma_n \mathbf{Z}_{n+k}. \tag{8.8}$$

Мы получили, таким образом, *явное выражение* $\mathbf{X}_k$ *в форме «скользящих средних» стационарной последовательности некоррелированных величин* $\mathbf{Z}_n$.

Очевидно, что представление (8.8) не единственно. Возникает естественный вопрос, возможно ли выразить $\mathbf{X}_k$ только через $\mathbf{Z}_k$, $\mathbf{Z}_{k-1}$, $\mathbf{Z}_{k-2}$, ... (через «прошлое» процесса $\{\mathbf{Z}_n\}$) или, что то же самое, возможно ли функцию γ в (8.3) выбрать так, что $\gamma_n = 0$ при $n \geqslant 1$. Эта проблема имеет фундаментальный характер для теории прогнозирования случайных процессов, но она

лежит вне рамок этой книги. Типичные примеры см. в гл. III (7.5) и задаче 18.

б) *Соответствующий процесс с независимыми приращениями.* Для каждого t, $-\pi<t\leqslant\pi$ положим

$$y_t(x)=\begin{cases}1 & \text{при } x\leqslant t\\ 0 & \text{при } x>t\end{cases} \tag{8.9}$$

и обозначим $\mathbf{Y}_t$ соответствующую случайную величину в $\mathfrak{H}$. Очевидно, что ковариация приращений $\mathbf{Y}_t-Y_s$ для непересекающихся интервалов равна 0; более того, $D(\mathbf{Y}_t)=R\{-\pi, t\}$, так что $\mathbf{Y}_t$ — процесс с *некоррелированными приращениями и дисперсией, определяемой по* R. Если $\mathbf{X}_t$ нормально распределены, то приращения процесса $\mathbf{Y}_t$ будут независимы.

Итак, для каждой стационарной последовательности $\{\mathbf{X}_k\}$ существует связанный с ней указанным образом процесс с некоррелированными приращениями. Явное выражение для $\mathbf{Y}_t$ в терминах $\mathbf{X}_k$ можно получить стандартным путем — разлагая функцию y_t в (8.9) в ряд Фурье. Однако снова предпочтительнее идти в обратном направлении.

в) *Стохастические интегралы.* Представление какой-либо случайной величины $\mathbf{U}$ в терминах $\mathbf{X}_k$ определяется (как мы видели) разложением Фурье функции, соответствующей $\mathbf{U}$. По сравнению с этим представление в терминах $\mathbf{Y}_t$ может показаться «слишком простым». Оно связано с графиком функции, соответствующей рассматриваемой случайной величине. Для простоты мы предположим эту функцию непрерывной.

Рассмотрим сначала «ступенчатую функцию» w, т. е. функцию вида

$$w=a_1y_{t_1}+a_2\left(y_{t_2}-y_{t_1}\right)+\ \dots\ +a_n\left(y_\pi-y_{t_{n-1}}\right), \tag{8.10}$$

где a_j — постоянные и $-\pi<t_1<t_2<\ldots<t_{n-1}<\pi$. Соответствующая случайная величина получается заменой в (8.10) J_{t_j} на $\mathbf{Y}_{t_j}$. Теперь каждая непрерывная функция $\mathbf{W}$ может быть равномерно аппроксимирована ступенчатыми функциями $w^{(n)}$ вида (8.10). Равномерная сходимость $w^{(n)}$ к w влечет сходимость в L^2_R-норме, а следовательно, и сходимость соответствующих случайных величин $\mathbf{W}^{(n)}$ к $\mathbf{W}$. Это дает нам рецепт отыскания образа $\mathbf{W}$ произвольной непрерывной функции w с помощью простого предельного перехода: надо аппроксимировать w ступенчатыми функциями типа (8.10) и заменить j_{t_j} на $\mathbf{Y}_{t_j}$. Напомним, что (8.10) определяет функцию, а не число (так же как и w яв-

ляется функцией, а не числом). Но (8.10) *выглядит* как риманова сумма, и формально наша процедура напоминает определение интеграла Римана. Поэтому общепринятым стало обозначение

$$\mathbf{W} = \int_{-\pi}^{\pi} w(t)\, d\mathbf{Y}_t, \tag{8.11}$$

указывающее на описанный выше предельный переход. Случайная величина (8.11) называется *стохастическим интегралом* от непрерывной функции w. Название это до некоторой степени произвольно, и обозначение — не более как краткое выражение предельного перехода, который был строго определен выше. По определению функция e^{int} соответствует случайной величине $\mathbf{X}_n$, и потому мы можем написать

$$\mathbf{X}_n = \int_{-\pi}^{\pi} e^{int}\, d\mathbf{Y}_t. \tag{8.12}$$

Это есть основное *спектральное представление произвольной стационарной последовательности* $\{\mathbf{X}_n\}$ *в терминах соответствующего процесса с некоррелированными приращениями.*

Обозначение стохастического интеграла, пожалуй, более наглядно, чем логично, но мы не будем касаться этого вопроса. Наша цель состояла в том, чтобы показать, как это полезное понятие и важное представление (8.12) могут быть получены средствами анализа Фурье. Этим иллюстрируются возможности канонического отображения, использованного в этом разделе (и введенного впервые Крамером). ▶

Изложенная теория использует только вторые моменты $\{\mathbf{X}_n\}$ и практически применима лишь в случаях, где эти моменты в самом деле значимы. Таков случай, когда процесс является нормальным, так как нормальные распределения полностью определяются своими ковариациями. В других применениях можно верить, что процесс «не слишком отклоняется от нормального» (подобно тому как регрессионный анализ основывался на вере в универсальную приложимость методов, развитых для нормального закона). К сожалению, один лишь факт существования прекрасной теории никоим образом не оправдывает этой веры. Так, в примере (3, в) выборочные функции процесса строго периодичны. Будущее отдельной реализации вполне определяется ее заданием на полном периоде. Однако методы прогноза, опирающиеся на L^2-теорию, не принимают во внимание этот факт и отождествляют все процессы с одной и той же

спектральной мерой. Наблюдатель, получающий последовательность ..., 1, —1, 1, —1, ..., может однозначно предсказать следующий член ряда, но L^2-методы заставляют его предсказывать появление нуля. Эти методы, таким образом, не являются универсально приложимыми. Однако они дают идеальное средство изучения нормальных процессов.

§ 9. Задачи

1. *Любопытные примеры характеристических функций.* Пусть $\tau_h(x) = 1 - \frac{|x|}{h}$ для $|x| \leqslant h$ и $\tau_h(x) = 0$ для $|x| > h$. Положим

$$\alpha(x) = \sum_{n=-\infty}^{+\infty} a_n \tau_h(x-n). \tag{9.1}$$

Когда a_n вещественны и $h=1$, *график* α *представляет собой ломаную линию с вершинами* (n, a_n). При $h<1/2$ график α состоит из отрезков оси x и боковых сторон равнобедренных треугольников с вершинами (n, a_n).

Покажите без вычислений, что из $\Sigma|\alpha_n| < \infty$ вытекает

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} \alpha(x)\, e^{-i\zeta x}\, dx = \frac{h}{\pi}\,\frac{1-\cos\zeta}{\zeta^2} \sum_{-\infty}^{+\infty} a_n e^{-in\zeta}. \tag{9.2}$$

Выведите отсюда, что α *является характеристической функцией в том и только том случае, когда* $a_0=1$ *и сумма ряда в* (9.2) *неотрицательна* при всех ζ. Приведите примеры. Обобщите результат на другие τ_h. (См. задачи 15—16.)

2. Ковариационная функция ρ [см. (3.1)] непрерывна всюду тогда и только тогда, когда она непрерывна в нуле. Последнее верно в том и только том случае, когда $\mathbf{E}((\mathbf{X}_t - \mathbf{X}_0)^2) \to 0$ при $t \to 0$.

3. Пусть $\{\mathbf{X}_t\}$ — стационарный процесс со спектральной мерой R. Положим для $h>0$

$$\mathbf{X}_t^{(h)} = \frac{1}{h}[\mathbf{X}_{t+h} - \mathbf{X}_t].$$

а) Найдите спектральную меру этого процесса и покажите, что предел при $h \to 0$ существует тогда и только тогда, когда R имеет второй момент (т. е. $x^2 R\{dx\}$ определяет конечную меру).

б) В последнем случае при фиксированном t и $h \to 0$, $\varepsilon \to 0$ $\operatorname{Var}\left(\mathbf{X}_t^{(h)} - \mathbf{X}_t^{(\varepsilon)}\right) \to 0$ (в терминологии гильбертовых пространств, применявшейся в § 8, это значит, что при $h \to 0$ $\left\{\mathbf{X}_t^{(h)}\right\}$ — последовательность Коши, и потому существует предел $\mathbf{X}_t'$).

4. Если φ — характеристическая функция распределения F, то

$$f_\lambda(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\lambda|\zeta|}\varphi(\zeta)\, e^{-i\zeta x}\, d\zeta$$

есть плотность некоторого распределения F_λ, причем $F_\lambda \to F$ [это аналог (4.3)].

5. *К теореме* 4.2. Если φ непрерывна всюду, кроме нуля, где она имеет разрыв первого рода, то (4.9) выполняется при всех

$$\zeta \neq 0 \text{ и } f_r(0) \to \frac{1}{2}[\varphi(0+) - \varphi(0-)].$$

6. *Продолжение.* Если φ может быть представлена как разность двух монотонных функций, то во всех точках $f_r(\zeta) \to \frac{1}{2}[\varphi(\zeta+) - \varphi(\zeta-)]$.

7. Ограниченная периодическая функция φ с неотрицательными коэффициентами Фурье φ_n необходимо непрерывна и $\Sigma\varphi_n < \infty$. Пример $\varphi_n = 1/n$ показывает, что утверждение неверно, если предполагается только интегрируемость φ.

8. *Суммируемость по Чезаро.* Заменим (4.3) на

$$f_r(\zeta) = \Sigma\varphi_n a_n e^{in\zeta},$$

где $a_n = 1 - \frac{|n|}{2N+1}$ при $|n| \leqslant 2N$ и $a_n = 0$ при $|n| > 2N$. Покажите, что с заменой $p_r(t)$ на

$$q_N(t) = \frac{1}{2N+1} \frac{\sin^2\left(N + \frac{1}{2}\right)t}{\sin^2 \frac{1}{2}t}$$

при этом сохраняется теория § 4 ($q_N(t)$ снова плотность вероятности).

9. *Продолжение.* Докажите более общее утверждение, что эта теория сохраняется, если в качестве a_n взять коэффициенты Фурье симметризованной плотности вероятности, заданной на окружности.

10. Покажите, что следствие 4.2 вытекает из теоремы 6.1 и что, обратно, эта теорема может быть доказана методами, использованными в (4.12).

11. Покажите, используя формулу суммирования Пуассона (5.2), что

$$\sum_{-\infty}^{+\infty}\left\{\mathfrak{n}\left(\frac{y-x+2k\lambda}{\sqrt{t}}\right) + \mathfrak{n}\left(\frac{y+x+2k\lambda}{\sqrt{t}}\right)\right\} = \\ = \frac{1}{a}\sum_{-\infty}^{+\infty}\exp\left(-\frac{1}{2}tn^2\frac{\pi^2}{\lambda^2}\right)\cos\frac{n\pi}{\lambda}x \cdot \cos\frac{n\pi}{\lambda}y,$$

где $\mathfrak{n}$ обозначает стандартную нормальную плотность [это — решение задачи об отражающем экране, см. гл. X, (5, e)].

12. *Другой вывод формулы суммирования Пуассона.* Пусть φ — характеристическая функция распределения F. Используя пример § 4 (и без всяких вычислений), покажите, что при $0 < r < 1$

$$\frac{1}{2\pi}\sum_{n=-\infty}^{+\infty}\varphi(\zeta+n)\,r^{|n|}e^{tn\lambda} = \int_{-\infty}^{+\infty}e^{it\zeta x}p_r(x+\lambda)\,F\{dx\}. \tag{9.3}$$

Отсюда следует, что *левая часть есть характеристическая функция.* Полагая $r\to 1$, выведите отсюда, что если F имеет плотность f, то

$$\frac{1}{2\pi}\sum_{n=-\infty}^{+\infty}\varphi(\zeta+n)\,e^{in\lambda}=\sum_{k=-\infty}^{+\infty}e^{i\zeta(-\lambda+2k\pi)}f(-\lambda+2k\pi) \tag{9.4}$$

в предположении $\Sigma f(-\lambda+2k\pi)<\infty$. Покажите, что (9.4) равносильно (5.7).

12а. *Продолжение.* Полагая $\zeta=\lambda=0$, выведите из (9.4), что ряд $\frac{1}{2\pi}\Sigma\varphi(n)$ суммируем по Абелю к $\Sigma f(2k\pi)$ в предположении, что последний ряд сходится.

13. *К теореме 5.1.* Пусть $g(x)=\frac{1}{2}e^{-|x|}$ и

$$f(x)=\frac{6}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n}\,g(n(x-n)).$$

Покажите, что f — плотность вероятности с характеристической функцией

$$\varphi(\xi)=\frac{6}{\pi^2}\sum\frac{1}{n^2+\xi^2}\,e^{in\xi}.$$

Выведите (без вычислений) из (5.2), что $\varphi(\xi)\geqslant 0$ при всех ξ и что φ интегрируема (см. следствие теоремы 3 из гл. XV, 3). Мы получили, следовательно, пример *непрерывной плотности f с интегрируемой характеристической функцией φ и такой, что $\Sigma f(n)=\infty$* (пример принадлежит Б. Вейссу).

14. Последовательность $\{\varphi_n\}$ положительно определена тогда и только тогда, когда последовательность $\{\varphi_n r^{|n|}\}$ положительно определена при любом r, $0<r<1$. Для этого необходимо и достаточно условие $\Sigma\varphi_n r^{|n|}e^{in\lambda}\geqslant 0$ для всех λ.

15. Выведите (без вычислений) из задач 1 и 14 следующую теорему (отмеченную Л. Шеппом). *Пусть $\{\varphi_n\}$ — положительно определенная последовательность. Обозначим через α кусочно-линейную функцию с вершинами графика (n,φ_n). Тогда α — положительно определена.*

16. *Пусть φ — характеристическая функция. Обозначим через α кусочно-линейную функцию с вершинами графика $(n,\varphi(n))$. Тогда α — характеристическая функция* (это лишь иная формулировка задачи 15 и частный случай задачи 1 с $h=1$). Используйте задачу 1 и опишите другие любопытные характеристические функции, которые можно получить из φ.

17. При $r<1$ ковариации (6.9) марковских последовательностей удовлетворяют (7.19) с $u_k=\sqrt{1-r^2}\,r^k e^{ik\theta}$ при $k\geqslant 0$ и $u_k=0$ при $k<0$. Найдите другие представления.

18. *Продолжение.* Пусть $\{\mathbf{X}_n\}$ — марковская последовательность с ковариациями $\rho_n=r^{|n|}e^{in\theta}$. Если $r<1$, то $\mathbf{X}_n=\sqrt{1-r^2}\sum_{k=0}^{\infty}r^k e^{ik\theta}\mathbf{Z}_{n-k}$, где величины $\mathbf{Z}_k$ не коррелированы. При $r=1$ имеем $\mathbf{X}_n=e^{in\theta}\mathbf{Z}_0$.

ЛИТЕРАТУРА

Некоторые книги в приводимом списке появились (или должны появиться) после завершения работы над этим томом

А. Общие курсы

Hennequin P L, Tortrat A, Théorie des Probabilités of Quelques Applications, Masson, Paris, 1965, 457

Krickeberg K, Wahrscheinlichkeitstheorie, Taubner, Stuttgart, 1963, 200

Лоэв М, Теория вероятностей, ИЛ, М, 1962

Неве Ж, Математические основы теории вероятностей (готовится к печати в 1968 г).

Б. Специальные темы

Bochner S, Harmonic Analysis and the Theory of Probability, Univ of California Press, 1955, 176

Гельфанд И М, Виленкин Н Я, Обобщенные функции Выпуск четвертый Некоторые применения гармонического анализа Оснащенные гильбертовы пространства, Физматгиз, М, 1961

Гренандер У, Теория вероятностей и алгебраические структуры, «Мир», М.

Lukacs E, Characteristic Functions Griffin, London, 1960, 216

Lukacs E, Laha R G, Applications of Characteristic Functions Griffin, London, 1954, 202

В. Случайные процессы (теория)

Чжун Кай-лай, Однородные цепи Маркова, «Мир», М, 1964

Дынкин Е Б, Основания теории марковских процессов, Физматгиз, М, 1959

Дынкин Е Б, Марковские процессы, Физматгиз, М, 1963

Ито К, Макин Х, Диффузионные процессы и их траектории, «Мир», М, 1967

Kemperman J H B, The Passage Problem for a Stationary Markov Chain. University of Chicago Press, 1961, 127

Levy Paul, Processus Stochastiques et Mouvement Brownien 2nd ed Gauthier—Villars, Paris, 1965, 438 pp

Спицер Ф, Принципы случайного блуждания, «Мир», М, 1968.

Скороход А. В., Случайные процессы с независимыми приращениями, «Наука», М., 1964.

Яглом А. М., Введение в теорию стационарных случайных функций. Выйдет в изд. «Наука» в 1969 г.

Г. Случайные процессы (приложения и примеры)

Barucha-Reid A. T., Elements of the Theory of Stochastic Processes and Their Applications. McGraw-Hill, New York, 1960, 468.

Beneš V. E., General Stohastic Processes in the Theory of Queues. Addison—Wesley, Reading, 1963, 88.

Grenander U., Rosenblatt M., Statistical Analysis of Stationary Time Series, John Wiley, New York, 1957, 300.

Хинчин А. Я., Математические методы в теории массового обслуживания, Физматгиз, М., 1963.

Prabhu N. U., Stochastic Processes. Macmillan, New York, 1965, 233.

Prabhu N. U., Queues and Inventories, John Wiley, New York, 1965, 275.

Riordan J., Stochastic Service Systems, John Wiley, New York, 1962, 139.

Wax N. (editor), Selected Papers on Noise and Stochastic Processes, Dover, New York, 1954, 337.

Bartlett M. S., Stochastic Population Models in Ecology and Epidemology, 1960, 90.

Кокс Д., Смит У., Теория очередей, «Мир», М., 1966.

Hannan E. J., Time Series Analysis, 1960, 152.

Moran P. A. P., The Theory of Storage, 1959, 111.

Takacs L., Stochastic Processes, 1960, 137.

Д. Книги, имеющие исторический интерес

Крамер Г., Математические методы статистики, ИЛ, М., 1948.

Дуб Дж., Вероятностные процессы, ИЛ, М., 1956.

Гнеденко Б. В., Колмогоров А. И., Предельные теоремы для серий независимых случайных величин, Гостехиздат, М., 1949.

Колмогоров А. Н., Основы понятия теории вероятностей, Гостехиздат, М., 1936.

Levy P., Calcul des Probabilités. Gauthier — Villars, Paris, 1925, 350.

Levy P., Théorie de l'Addition des Variables Aléatoires. Gauthier — Villars, Paris, 384.

Е. Полугруппы и общий анализ

Хилле Э., Филлипс Р., Функциональный анализ и полугруппы, ИЛ, М., 1962.

Karlin S., Studden W., Tchebycheffian Systems and Applications to Analysis. Interscience, New York, 1966.

Иосида К., Функциональный анализ, «Мир», М., 1967.

ОТВЕТЫ НА ЗАДАЧИ

Глава I

1. (i) $\frac{\alpha}{3}\frac{1}{\sqrt[3]{x^2}}e^{-\alpha x^{1/3}}$ (ii) $\frac{\alpha}{2}e^{-\alpha(x-3)/2}$ при $x>3$

(iii) $\frac{\alpha}{2}e^{-\alpha|x|}$ при всех x

(iv) $\alpha e^{-\alpha x}$ при $x>0$

(v) $\alpha\left(1+\frac{1}{3}\frac{1}{\sqrt[3]{x^2}}\right)e^{-\alpha x-\alpha x^{1/3}}$

(vi) $\alpha e^{-\alpha x}+\frac{\alpha}{3}\frac{1}{\sqrt[3]{x^2}}e^{-\alpha x^{1/3}}-\alpha\left(1+\frac{1}{3}\cdot\frac{1}{\sqrt[3]{x^2}}\right)e^{-\alpha x-\alpha x^{1/3}}$.

2. (i) $\frac{1}{6}\frac{1}{\sqrt[3]{x^2}}$ при $|x|<1$

(ii) $\frac{1}{4}$ при $1<t<5$

(iii) $\frac{1}{2}\left(1-\frac{|x|}{2}\right)$ при $|x|<2$

(iv) $1-\frac{x}{2}$ при $0<x<2$

(v) $\frac{1}{4}-\frac{1}{3}x^{1/3}+\frac{1}{12}x^{-2/3}$ при $|x|<1$

(vi) $\frac{1}{4}+\frac{1}{3}x^{1/3}+\frac{1}{12}x^{-2/3}$ при $|x|<1$.

3. (i) $h^{-1}(1-e^{-\alpha x})$ при $0<x<h$
и $h^{-1}(e^{\alpha h}-1)e^{-\alpha x}$ при $x>h$
(ii) $h^{-1}(1-e^{-\alpha(x+h)})$ при $-h<x<0$
и $h^{-1}(1-e^{-\alpha h})e^{-\alpha x}$ при $x>0$.

4. (i) $1-\frac{1}{3}h$, если $h<1$, но $\frac{2}{3}\frac{1}{\sqrt{h}}$, если $h>1$

(ii) $\sqrt{\alpha\pi}\,e^{\frac{1}{4}\alpha}(1-\mathfrak{N}(\sqrt{\alpha/2}))$, где $\mathfrak{N}$ есть стандартное нормальное распределение.

5. (i) $1-x^{-1}$ при $x>1$;
(ii) $x^2(x+1)^{-2}$ при $x>0$.

8. $\mathbf{P}\{Z\leqslant x\}=1-e^{-\alpha x}$ при $x<t$ и $=1$ при $x>t$.

15. $p=\sum_{0}^{m-1}\binom{n+k-1}{k}2^{-n-k+1}$. При $m=1$, $n=2$ получаем $p=\frac{1}{4}$.

17. $nt^{n-1}-(n-1)t^n$.

18. (i) $2\int_0^1 dx\int_x^1(1-z)\,dz=\frac{1}{3}$
(ii) *Плотность* равна $2t-t^2$ при $0<t\leqslant 1$ и $(2-t)^2$ при $1\leqslant t<2$
(iii) Плотность равна $2t^2$ при $0<t<1$ и снова $(2-t)^2$ при $1\leqslant t<2$.

19. $2\int_0^1 x(1-x)\,dx=\frac{1}{3}$. Две из шести перестановок приводят к пересечению.

20. $\mathbf{X}_{11}$: $4\log\frac{1}{4x}$ при $x<\frac{1}{4}$; $\mathbf{X}_{12}$ и $\mathbf{X}_{21}$: $4\log 2$ при $x<\frac{1}{4}$; $4\log\frac{1}{2x}$ при $\frac{1}{4}<x<\frac{1}{2}$; $\mathbf{X}_{22}$: $4\log 4x$ при $\frac{1}{4}<x<\frac{1}{2}$; $4\log\frac{1}{x}$ при $\frac{1}{2}<x<1$. Математические ожидания равны $\frac{1}{16}$, $\frac{3}{16}$, $\frac{9}{16}$.

22. $q_n(t)=nt^{-n}\int_0^{t-h}(t-x-h)^{n-1}q_{n-1}(t-x-h)\,dx=t^{-n}(t-(n-1)h)_+^n$.

25. Функции распределения $\frac{2}{\pi}\arcsin\frac{1}{2}x$ и $\frac{1}{4}x^2$; плотности $\frac{2}{\pi}\frac{1}{\sqrt{4-x^2}}$ и $\frac{1}{2}x$ при $0<x<2$.

26. $2\pi^{-1}\arcsin\frac{1}{2}x$.

28. а) $\log\frac{1}{x}$,
б) $\frac{2}{\pi}\log\frac{1+\sqrt{1-x^2}}{x}$, где $0<x<1$.

29. $\frac{4}{\pi}\int_{0<\cos\theta<x}\sin^2\theta\,d\theta=\frac{4}{\pi}\int_0^x\sqrt{1-y^2}\,dy=2\pi\left[\arcsin x+x\sqrt{1-x^2}\right]$, где $0<x<1$.

30. $F(t) = \frac{2}{\pi} \int\limits_0^{\pi/2} V\left(\frac{t}{\cos\theta}\right)(1 - \cos 2\theta)\, d\theta.$

Глава II

4. $g^* g(x) = \frac{1}{4} e^{-|x|}(1 + |x|)$

$g^{3*}(x) = \frac{1}{16} e^{-|x|}(3 + 3|x| + x^2)$

$g^{4*}(x) = \frac{1}{32} e^{-|x|}\left(5 + 5|x| + 2x^2 + \frac{1}{3}|x|^3\right).$

10. $\lambda e^{-\lambda t} - \mu e^{-\mu t} + (\mu - \lambda)\, e^{-(\mu+\lambda)t}.$

12. Плотность равна $1 - \frac{1}{2}t^2$ при $0 < t < 1$ и $\frac{1}{2}(2-t)^2$ при всех $1 < t < 2$. Математическое ожидание равно $\frac{7}{12}$.

Глава III

6. а) e^{-x} и $1 - e^{-x} - e^{-y} + e^{-x-y-axy}$ при $x > 0$, $y > 0$.

б) $\mathbf{E}(Y \mid X) = \frac{1 + a + ax}{(1 + ax)^2}$,

$\mathrm{Var}(Y \mid X) = \frac{1}{(1+ax)^2} + \frac{2a}{(1+ax)^3} - \frac{a^2}{(1+ax)^4}$.

7. Если f имеет математическое ожидание μ и дисперсию σ^2, то $\mathbf{E}(X) = \mathbf{E}(Y) = \frac{1}{2}\mu$, $\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{1}{3}\sigma^2 + \frac{1}{12}\mu^2$,

$\mathrm{Cov}(X, Y) = \frac{1}{6}\sigma^2 - \frac{1}{12}\mu^2$.

8. Плотность равна $2x_2$ в единичном квадрате. Для выборки объема n она равна $(n-1)!\, x_2 x_3^2 \ldots x_{n-1}^{n-2}$ в $(n-1)$-мерном единичном кубе.

9. $\frac{1}{3} e^{-(x+y)}$ при $y > x > 0$ и $\frac{1}{3} e^{-y+2x}$ при $y > x$, $x < 0$. Если $y < x$, поменяйте x и y местами.

11. Двумерная нормальная плотность с дисперсиями m, n и ковариацией $\sqrt{m/n}$. Условная плотность, очевидно, имеет математическое ожидание $\frac{m}{n}t$ и дисперсию $m \cdot \frac{n-m}{n}$.

12. Сумма $X_1^2 + \ldots + X_n^2$ имеет гамма-плотность $f_{\frac{1}{2}, \frac{n}{2}}$ [см. гл. II, 2.2].

Поэтому из (3.1)

$$u_t = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-m}{2}\right)} \left(\frac{x}{t}\right)^{\frac{1}{2}(m-1)} \left(1-\frac{x}{t}\right)^{\frac{1}{2}(n-m)-1} \frac{1}{t}.$$

При $m=2$, $n=4$ мы получим пример (3.а).

14. а) $4xy$, когда $x+y<1$, $x>0$, $y>0$
$4xy-4(x+y-1)^2$, когда $x+y>1$, $0<x$, $y<1$
$4x(2-x-y)$, когда $y>1$, $x+y<2$, $x>0$
$4y(2-x-y)$, когда $x>1$, $x+y<2$, $y>0$.
б) $2(1-x-y)^2$ при $0<x$, $y<1$, $x+y<1$
$2(1-x)^2$ при $x>0$, $y<0$, $x+y>0$
$2(1+y)^2$ при $x>0$, $y<0$, $x+y<0$.
При $x<0$ по симметрии.

15. $2\frac{r}{\pi^2}\left(\arccos\frac{r}{2}-\frac{r}{2}\sqrt{1-\frac{r^2}{4}}\right)$.

16. $\int\limits_0^\infty f(\rho)\rho\,d\rho \int\limits_0^{2\pi} g\left(\sqrt{r^2+\rho^2-2r\rho\cos\theta}\right)d\theta$.

17. а) $\mathbf{X}_n = \left(\mathbf{U}\cos\frac{1}{2}\pi n + \mathbf{V}\sin\frac{1}{2}\pi n\right)\lambda$
б) $\lambda\mathbf{U}+\mu\mathbf{V}(-1)^n$
в) $\lambda\left(\mathbf{U}\cos\frac{1}{2}\pi n+\mathbf{V}\sin\frac{1}{2}\pi n\right)+\mu\mathbf{W}$.

18. а) $\mathrm{Var}(\mathbf{Y}_{n+1})-\mathrm{Var}(\mathbf{Y}_n)=\mathrm{Var}(\mathbf{C}_n)-2\,\mathrm{Cov}(\mathbf{Y}_n,\mathbf{C}_n)+1$, откуда
б) $\alpha^2-2\alpha\sigma\rho+1=0$
в) $\sigma=\frac{1}{2}\left(\alpha+\frac{1}{\alpha}\right)$, $\mathbf{Y}_{n+1}=a\frac{1}{ap}+\delta+\sum\limits_{k=0}^{n-1}(1-p)^k\mathbf{X}_{n-k}+p^n\mathbf{Y}_0$.

19. $\sigma^2\geqslant\frac{1}{4}\left(\alpha+\frac{1}{\alpha}\right)^2+N$.

Глава VI

7. Не обязательно. Необходимое условие $n[1-F(\varepsilon)+F(-\varepsilon)]\to 0$.

8. При $x>0$ плотности равны 1 и $\frac{1}{2}(1-e^{-2x})$.

9. $qV(x)=1-pe^{-qct}$. Число моментов восстановления всегда распределено геометрически.

14. V представляет собой свертку Z и показательного распределения.

15. $V = A + B * V$, где $A\{dx\} = [1 - G(x)]\, F\{dx\}$ и $B\{dx\} = G(x)\, F\{dx\}$

18. Плотность арксинуса $g(y) = \dfrac{1}{\pi} \dfrac{1}{\sqrt{y(1-y)}}$.

Глава VII

6. а) $\binom{n}{k} p^k (1-p)^{n-k}$. Распределение F сосредоточено в p.

б) $\dfrac{1}{n+1}$, плотность $f(x) = 1$.

в) $\dfrac{2(k+1)}{(n+1)(n+2)}$, плотность $2x$.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Абелевы теоремы 511
Абсолютная непрерывность 174, 176
Абсолютно беспристрастная последовательность 265, 295
— монотонные функции 279, 280, 505
Аварии 78, 228
Автобусы (очереди на автобусы, ожидание) 37, 76, 88, 236
Аддитивные функции множества 136
Алгебра множеств 143, 144, 150
— —, порождения 204
Арифметическое распределение 173; см. также Решетчатое распределение
Арксинуса распределения 70
— — в процессах восстановления 539
— — в случайном блуждании 483
— — обобщенное 604
Арцела — Асколи теорема 329
Асимптотическая пренебрежимость 223, 594
Асимптотически несмещенная оценка 277
— плотное (множество) 184
Атомические меры 172
Атомы (мер) 172
— при свертке 183—184

Банаха пространство 313, 414, 557
Башелье процесс; см. Броуновское движение
Безгранично делимые полугруппы 524
— — распределения 220, 351, 516, 632
— — — в $\mathscr{R}^r$ 670, 674
Бернулли испытания и случайный выбор 53
Берштейна полиномы (в $\mathscr{R}^2$ 300) 278
Бесконечная дифференцируемость 312, 353
Бесконечные свертки 324, 379, 669, 674
Бесселевы функции 79—80, 598
— —, безграничная делимость 221, 518, 646
— — в стохастических процессах 81, 385
— —, Лапласа преобразование 503, 504, 550
— —, распределения, связанные с ними 80 и сл., 187, 207, 208
— —, характеристическая функция 573
Бета-интеграл 67
— плотность 70
— — в задаче восстановления 538
Борелевская алгебра 143, 150
Борелевское множество 143
— —, аппроксимация 145, 156
— —, измеримое по 146
— —, соглашения об обозначениях 160
Бохнера интеграл 521—522
Броуновское движение 127, 394, 544
— — в $\mathscr{R}^r$ 218, 407
— —, непрерывность траекторий 226, 365, 395
— —, первое прохождение 217, 403, 545
— —, подчиненность 411
— — с упругой силой 397
Бэровские функции 134, 139, 148, 161
Бюффона задача об игле 83

Вальда тождество 469, 492, 679
Вейбула распределение 73
Вейерштрасса теорема об аппроксимации 278
Вероятность вхождения; см. Вероятность достижения
Вероятностные меры и пространства 142, 168
Ветвящиеся процессы 298, 507, 542
Взаимные пары функций 575
Видимость в x-направлении 23
Винера процесс; см. Броуновское движение
Винера — Хопфа интегральное уравнение 470
— — факторизация 456, 465, 494, 681, 690
Вложенные процессы восстановления 241, 446
Водохранилище; см. Хранение запасов
Возвращение в нуль 490, 689
Возраст; см. Длительность
Вполне монотонные функции 282, 504, 517
— — — абстрактные 520
— — —, интерполяция 531
Вращения 103, 113
Время возвращения 231
— — максимальное 238, 453
— жизни; см. Длительность, Время возвращения
Выборка, медиана, экстремумы 32
— среднее значение, дисперсия 111

Выпуклые функции 193
— — и мартингалы 271
Вырожденное распределение 108, 113
Вырожденные процессы 118

Гамеля уравнение 358, 366
Гамма-плотность 24, 66
— распределение 23, 66, 500
— —, безграничная делимость 220, 355, 518, 645
— — как предел порядковых статистик 39
— —, подчинение 412
— —, приближение 276
— — рандомизированное 80
— функция 66, 86
Гармонические функции 298
Гейгера счетчики 238, 240, 439, 453, 536
Гёльдера неравенство 195
Гильбертовы пространства 715, 720
Гиперболические функции и плотности 87, 574, 603, 645, 670, 710
Гравитационные поля 216, 272
Граничные условия 400, 546
— значения 299
Грина функция 396, 545, 567
Группировка данных 17
Гюйгенса принцип 71

Двойные орбиты 51
Двойственность 461
Двустороннее преобразование Лапласа 499
Двусторонняя показательная плотность 69
— — —, факторизация 682
— — —, характеристическая функция 573
Дисперсия 18, 65, 171
— инфинитезимальная 397
— условная 96
Диффузия в генетике 398
— в $\mathscr{R}^1$ 394, 532, 544, 567
— в $\mathscr{R}^r$ 407
— с упругой силой 397
Длина случайных цепей 262
Длительность, диффузия 404—405
—, мертвый период 238
—, период занятости 452, 553
—, процесс восстановления 234, 273, 441
—, процесс размножения 561
Дробная доля 186, 327
Дробовый эффект 223, 672

Естественный масштаб в диффузии 395

Жордана разложения 173

Задача о разорении 228, 250, 388, 538
— — —, оценки 444, 479
Задержки в движении 237, 445, 454, 544, 567
Законы больших чисел 274, 290, 346, 502, 586
— — — для мартингалов 299
— — — для серий 378, 674
— — — для стационарных последовательностей 301
— — —, обратная теорема 292
— нуля или единицы 157
Заражение 79
Значащие цифры 85

Иенсена неравенство 193, 271
Измеримое пространство 144
Измеримость 146
Изометрия 722
Индикатор множества 132
Интегральное уравнение Абеля 51
Интегрирование по частям 189
Интервал непрерывности 303
Инфинитезимальная скорость и дисперсия 397
Ионизация 385
Исчезающие на бесконечности функции 303
«Исчезновение», случайные блуждания 490

Каноническая мера 633
Кантора тип распределения 54, 177, 670
— — — (свертки) 186
Квазиустойчивость 215
Квазихарактеристические функции 632
Ковариация 91
—, матрица 107
— процесса 116, 700, 725
Колмогорова дифференциальные уравнения 293, 553
— неравенство 296, 301
— Смирнова теорема 57, 405
— теорема о трех рядах 379
Компактность в схеме серий 370

Конкордантные (согласованные) функции 266, 298
Координатные случайные величины 17, 91
Корреляция 91
Коши — Леви теорема 595, 600
— распределение 71, 87, 292, 599
— — в $\mathscr{R}^r$ 94, 97, 128, 599, 671
— — и броуновское движение 218, 411
— —, случайное блуждание 259, 693
— —, устойчивость 216, 634, 645
Крамера оценка для вероятности разорения 229, 444, 471, 480
Критерий значимости 201
Кронекера лемма 299
— символ 554
— ядро 260, 554
Круг, распределение на нем 46, 83, 88, 104
— — равномерное на нем 327, 333
—, теорема покрытия 44
Кэмпбелла теорема 224, 347, 672

Лапласа второй закон распределения 70
— преобразование 288, 495, 534
Лапласа преобразование в $\mathscr{R}^r$ 519
— — двустороннее 499
— — для полугрупп 520
— Стильтьеса преобразование 489
Лебега интеграл 135
— мера 51, 142, 159
— Никодима теорема 176
— пополнение 159
— Стильтьеса интеграл 136, 139, 166
— теорема о разложении 179
Леви каноническое представление 643
— Крамера теорема 600
— метрика 345
— Парето распределение 219
—, примеры 272, 382, 646
Лежандра формула удвоения 86
Лестничные величины 241, 457, 481
— — в процессе ожидания 248
Линдеберга условие для диффузии 395
— условия 320, 346, 592
Линейные операции над стохастическими процессами 702
— приращения в скачкообразных процессах 387
Линейный функционал 151
Логарифм комплесных чисел 639
Логарифмическое распределение 86
Логистическое распределение и развитие 73
Локально компактные пространства 152, 156, 303
Ляпунова условие 346

Мажорированная сходимость 141
Максвелла распределение 49, 68, 103
Максимальная оценка 479
— характеристическая функция 685, 694
— частная сумма 242, 250, 383, 470, 477, 487
Максимальное время возвращения 238, 453
Максимальный член 207, 215, 337, 347
— —, влияние на сумму 532; см. также Порядковые статистики, Рекордные значения
Марковские процессы с дискретным временем 122, 129, 258, 274
— — — — — и мартингалы 298
— — — — —, спектральная функция 713, 728
Марковские процессы с дискретным временем, эргодические теоремы 330
— — с непрерывным временем 125, 383, 700
— — — — — в процессах восстановления 436
— — — — — в счетных пространствах 553
— — — — — и полугруппы 413, 521
— — — — —, эргодические теоремы 446, 564
Марковское свойство 22; см. Строго марковское свойство
Мартингалы 265, 295
— в процессах восстановления 431
— в случайных блужданиях 466
—, неравенства 295, 296, 301
Масштабные параметры 64, 171
Медиана 32, 172
Медленно меняющиеся функции; см. Правильно меняющиеся функции
Мера скорости (speed measure) 395
«Мертвый» период 240
Метод ограничений 403
Метрики 345; см. также Банахово пространство, Гильбертово пространство
Минимальное решение, Винера — Хопфа уравнение 469
— —, дифференциальное уравнение

Колмогорова 555
— —, диффузия 402
— —, полумарковские процессы 568
— —, скачкообразные процессы 393
— —, уравнение восстановления 255
Миттаг-Леффлёра функция 519
Момент регенерации 231
Моменты 18, 170, 189
— восстановления 231, 432
— в процессах восстановления 442
—, неравенство 195
—, проблема однозначности 283, 290, 532, 527
— — — в $\mathscr{R}^r$ 606
—, производящие функции 499
—, сходимость 306, 328
—, Хаусдорфа проблема 281
Монотонной сходимости принцип 140
— — свойства 414

Найквиста формула 708
Направляющий процесс 411
Невозвратные случайные блуждания 253
Невозвратные случайные блуждания, свойства 257, 465, 483, 690, 694
— — —, уравнение восстановления и теорема для них 252 449, 450, 494
Невозможность систем игры 270, 271
Независимость случайных величин 19, 92, 95, 153, 169
— — —, комплексные величины 570
— — —, критерий для нее 170
Независимые приращения 123, 224, 352, 365, 380, 724
— —, разрывы 366, 380
Неймана тождество 82
Нелинейное восстановление 454
Непосредственная интегрируемость по Риману 426
Непрерывность полугрупп 416; см. также Фиксированные разрывы
Несмещенная оценка 277
Несобственные (дефектные) распределения 162, 164, 260
— — — в процессах восстановления 234
Неудачи 29
Норма 313, 413, 714, 720
—, топология, порождаемая 345
Нормальное распределение 65, 87, 574
— — в $\mathscr{R}^r$ 108, 597
— — вырожденное 112
— — двумерное 93, 96, 129
Нормальное распределение марковское 122
— —, область притяжения 377, 652
— —, характеристические свойства 87, 102, 110
— — частное 128, 206
Нормальные полугруппы 360, 368, 381
Нормальный стохастический процесс 114, 719
Носитель 64
Нулевая схема серий 222, 659
Нулевое множество 158, 175, 714, 720

Область притяжения 214
— —, критерий для нее 373, 382, 514, 652
— — нормальная 657
— — частичная 382, 646, 656
Обобщенный пуассоновский процесс 225, 365, 388
— — — и полугруппы 357, 360
— — —, подчинение в нем 412
— — —, разорение в нем 228, 250, 538
Образование колонн 59, 567
Обратное уравнение для диффузии 396, 398 и сл., 407, 545 и сл.
— — — полугрупп 417, 420 и т. д.
— — — полумарковских процессов 568
— —, минимальное решение 392, 555 и т. д.
— —, скачкообразный процесс 390, 554 и т. д.
Обрывающийся (terminating) или невозвратный (transient) процесс 234, 440 и сл.
Обслуживание больных 229
Обслуживающие устройства; см. Очереди
Одновершинные [унимодальные] распределения 197, 603
— —, свертки 208
Оператор перехода 414
— сдвига 722
Операторы, связанные с распределениями 311
Операционное время 227, 409
Орбиты двойные 51
Орнстейна — Уленбека процесс 127, 397
Ортогональные матрицы 114
Остаточные события 157

Остающееся время ожидания 236
— — —, предельная теорема 434, 436, 453, 540
Осциллирующие случайные блуждания 258, 464, 687
Отношение правдоподобия 267
Отражающий экран 402, 406, 532
Оценка 60
Очереди 75, 76, 87, 88, 248
— параллельные 31, 32, 35, 60; см. также Периоды занятости
Ошибки округления 37, 84

Парадокс контроля 235, 436
Парадоксы 24, 38, 235, 436
Параметры расположения 64, 171
Парето распределение 70
Парсеваля равенство 530, 695, 716, 719
— — как критерий Хинчина 718
Первое возвращение 459, 491, 566, 689
Первые прохождения, диффузия 217, 403, 544
— —, марковские цепи 562
— —, процесс размножения и гибели 82, 88, 551, 565
Периодограмма 101
Периоды занятости 246, 251, 553
— — и ветвящийся процесс 542, 517
Петербургская игра 293
Пирсона система распределений 67, 70
Планшереля преобразование 715
— теорема 582
Плотность 16, 64, 174
Повторные математические ожидания 203
Поглощающие экраны 402 и т. д., 532, 546 и т. д
Поглощение (физическое) 41, 48, 385
Подчиненные процессы 408, 518, 525, 646, 673
— — и показательная формула 419
Пойа кригерий 577, 581
— урновая схема 266, 285, 297
Показательная формула для полугрупп 287, 409, 417
Показательное распределение 21, 58, 98
— — двумерное 128
— — двустороннее 69, 187
— — и равномерное распределение 62, 98
Показательное распределение как предел 39, 62, 434; см. также Гамма-распределения
Положительно определенные матрицы 107
— — *последовательности* 710
— — функции 697
Полугруппы 286, 413, 520
— со сверткой 353
— устойчивые 366
Полумарковский процесс 553, 568
Полумартингалы 271
Полярные координаты 92
Пополнение мер 159
Популяция, рост 399, 401
—, случайное рассеивание 319
Порядковые статистики 32, 36, 129
— —, предельные теоремы 39, 62, 87
— —, применение к статическому оцениванию 60
Потенциал 558
Потеря энергии 385, 387; см. Столкновение частиц
Потерянные вызовы 239, 566
Правильно меняющиеся функции 334, 348
Приложение к астрономии 51, 115 170—171, 261, 272, 388
Принцип инвариантности 400
— отражения 218, 547, 548
Притяжение; см. Область притяжения
Проекция 90
— случайных векторов 47, 49, 62
Произведение мер и *пространств* 153
Производящие операторы 354, 420, 522, 545
— функции, интерпретация с помощью «исчезновения» 490
— —, характеризация 279
Произвольная остановка 270
Пропуски (пробелы) большие 236, 445, 536
— малые 273
Прохождение света через вещество 41, 47, 62
— — в звездных системах 261, 388
— — и принцип Гюйгенса 71
— — через сферу 47
Процесс авторегрессии 116, 123
— восстановления с запаздыванием 234, 431
Прочность на разрыв 22
Прошедшее время ожидания 236, 453
— — —, предельные теоремы 436, 540
Прямое уравнение, диффузия 400, 544

Прямое уравнение, диффузия, минимальное решение 393, 556
— —, полугруппы 417
— —, скачкообразный процесс 387, 391
— — — — (счетность) 554
Прямоугольная плотность 36, 70
Псевдопуассоновский процесс 409
— —, показательная формула 417
— —, полугруппа 526
— — с линейными приращениями 387
Пуанкаре задача о рулетке 84
Пуассона формула суммирования 85, 405, 707, 727
— ядро 704
Пуассоновские ансамбли 28; см. также Гравитационные поля
— распределения, аппроксимация с их помощью 346
— — как крайние точки 644
— —, разность 207, 645
— —, формула Тейлора 286
Пуассоновский процесс 24
— — как предел в процессах восстановления 434
— —, supremum 299

Равномерное распределение 36, 327
Равностепенная непрерывность 307, 329
Радиация звездная 261
Радона — Никодима производная 174
— — теорема 176
Развитие логистическое 73
Разложение единицы 722
Размножения и гибели процессы 549, 567
— — — — и диффузия 567
— — — —, периоды занятости в них 552; см. также Рандомизованные случайные блуждания
— процесс 60, 325, 559
— — двусторонний 561
Разности (обозначения) 277
Разрывы фиксированные 227, 380
Рандомизированное случайное блуждание 81, 549, 644
Рандомизация и подчиненные процессы 409
—, полугруппы 288, 419
—, симметрично зависимые случайные величины 283
Распределение роста 97
Расслоение 78
Расстояние между распределениями 345
— по вариации 345
Расхождение; см. Эмпирическое распределение
Регистрация 240, 439
Регрессия 94, 112
Регулярное стохастическое ядро 331
Резольвенты 495, 527, 557
— и полная монотонность 529
Рекордные значения 29, 58, 59; см. также Порядковые статистики
Решающая функция 269
Решетчатые распределения 173, 184
— —, характеристические функции 572, 583
— —, центральная предельная теорема для 591, 617
Римана — Лебега теорема 586
Рисса теорема о представлении 152
Рулетка 37, 84

Самовосстанавливающиеся совокупности 234
Свертка ядер 261
Свертки, бесконечные 324, 379, 669, 674
— —, поведение F^{n*} 347
— на окружности 86, 333
— плотностей 19, 65, 95
— распределений 179
— — (сингулярных распределений) 186
Свободный пробег 22, 29
Сглаживание 115, 181, 613
Сдвиг, оператор 287, 313
—, полугруппы 355, 421
— принцип 499
Сепарабельность 328
Сервостохастический процесс 130
Серии 62; см. также Рекордные значения
Сжатие 414, 521
Сильная непрерывность в нуле 416, 521
— сходимость 314, 416
Симметризация 186
—, неравенства 188
Симметрично зависимые величины 283, 489
— — —, центральная предельная теорема 346
Симметричные события 157
Сингулярные (вырожденные) распределения 54, 137, 177
— —, свертки 186, 670

Скалярное произведение 714
Скачкообразный процесс 389
— — с бесконечным числом скачков 393, 554
Скользящее среднее 115, 723
Скрытая периодичность 101
Слабая сходимость 304
Случайная величина 16, 91, 145, 165
— — комплексная 569
Случайное рассеивание 319
Случайные блуждания в $\mathscr{R}^1$ 240, 252, 456, 575
— — — — и эмпирическое распределение 56, 62
— — — — простые (Бернулли) 270, 381, 460, 462, 491, 503, 680
— — — — сопряженные 474
— — — в $\mathscr{R}^r$ 50, 598
— — — —, центральная предельная теорема 319
— импульсы 240
— направления 46, 61, 62
— —, сложение 48, 71, 186, 262, 598
— разбиения 37, 98, 101; см. Теоремы о покрытии
— — (расщепления частиц) 40, 61, 129
— суммы 75, 199, 208
— —, характеристическая функция 576
— —, центральная предельная теорема для них 323, 605
— цепи 261
Случайный вектор 48, 50; см. также Случайные направления
— выбор 14, 36, 32
— — и бросание монеты 53; см. также Теоремы о покрытии, Случайные разбиения
Смеси 74, 98, 196, 208
—, преобразование 503
Снедекора плотность 69, 70
Снос в диффузии 397
— в случайных блужданиях 465, 686
Собственное распределение 164
— —, сходимость 304, 344, 633
Совместное распределение 91
Совпадения 273
Сопряженные случайные блуждания 474
Спектр мощностей 701
Спектральная мера 701, 726 (заданная 703)
— —, представление для унитарных операторов 722
Статистика 32
Стационарность, мера и вероятность 262, 274, 332
—, приращение 225, 352
—, процесс 114, 129, 700, 710 (закон больших чисел 301) (Ср. Эргодическая теорема)
Стационарный режим 27, 263, 388
— — в процессах восстановления 434
Столкновение частиц 261, 385, 387
Стохастическая ограниченность 310, 369
Стохастические интегралы 724
Стохастическое разностное уравнение 128
— ядро 199, 259, 330
Строго марковское свойство 38
— устойчивые распределения 211
Структура (алгебраическая) 414
Стьюдента плотность 69, 72
Субмартингал 271, 296
Субстохастическое ядро 260
Сумма случайных векторов 48, 186, 262, 598
Суммируемость по Абелю 704, 706, 728
— — Чезаро 706, 727
Суперпозиция процессов восстановления 434
Схемы серий 222, 272, 369, 659, 674
Сходимость в среднем квадратичном 714
— мер 304, 326, 344
— операторов 314, 353, 414
— по вероятности, 290, 309
— сильная 314, 416
— слабая 305
Счетно-аддитивные функции (σ-аддитивные) 137, 150—151, 163
Счетчики; см. Гейгера счетчики, Очереди

Тауберовы теоремы 508
— —, применение 486, 514, 535, 540, 687
Тейлора формула обобщенная 286
Телефонные вызовы 78, 228
— — потерянные 239, 266; см. также Периоды занятости
Теорема непрерывности, квазихарактеристические функции 634
— —, Лапласа преобразования 496
— —, полугруппы 327
— —, характеристические функции 580, 582, 604

Теорема об аппроксимации в среднем 141
— о продолжении 149, 156
— — среднем значении 138
— — трех рядах 379
Теоремы о выборе 325
— — покрытии 42, 61, 101, 273, 538
Теория надежности 73
— страхования 229; см. также Задача о разорении
Типы распределений 64, 171
— —, сходимость 307
Точка роста 172, 184
Точки первого вхождения в процессах восстановления 236, 434, 435, 492; см. также Лестничные величины
— — — в случайных блужданиях 490, 492, 675
Транспортные задачи 59, 237, 445, 454, 544, 567
Треугольная плотность 43, 70, 573
Треугольника неравенство 313
Тэта-функции 405, 406, 710

Универсальные законы Дёблина 568
Унимодальность по Хинчину 603
Унитарные операторы в гильбертовом пространстве 722
Упругая сила 397
Уравнение восстановления 232, 425, 436, 453
— — для всей прямой 253, 448, 494
— —, процесс 25, 230, 272, 431, 453
Уравнения восстановления, теоремы 424, 427, 441
Условные распределения и математические ожидания 95, 200
Устойчивые полугруппы 366
— распределения в процессах восстановления 440
— — — $\mathcal{R}^2$ 647, 671
— —, плотности 657
— —, подчиненность 412, 518
— —, положительные 514, 673
— —, произведения 219, 519
— — с показателем $1/2$ 72, 87, 216, 381, 503

Фату теорема 140
Фиксированные разрывы 380
Фильтр 116, 702
Фишера Z-статистика 69
Фоккера — Планка уравнение 356, 386, 391
Формула обращения и проблема моментов 281
— —, преобразование Лапласа 288, 506, 507, 530
— —, характеристические функции 581, 584, 599
— удвоения 86
Фубини теорема 141, 155, 181
Функции интервалов 136, 162
Функционал линейный 151
Фурье коэффициенты 705, 708
— обращение 581, 583
— ряд 703, 719, 727
— преобразование 570, 608; см. также Планшереля преобразование
— Стильтьеса преобразование 570

Характеристические функции 569
— — в $\mathcal{R}^r$ 596
— —, логарифм 639
— — периодические 703, 706
— —, производная в нуле 635
— —, факторизация 578, 669, 708; см. также Пуассона формула суммирования
Хи-квадрат 68
Хилле — Иосида теорема 526, 545
Хинчина критерий 718
— Полячека формула 478, 539, 683, 693
Хольцмарка распределение 215, 217, 272
Хранение и управление запасами 228, 229, 247

Центральная предельная теорема 315, 346, 350, 588, 605
— — —, необходимые и достаточные условия 379, 654, 662
— — —, приложения 257, 264, 319
— — —, процессы восстановления 438
Центрирование 64, 171, 660
— безгранично делимых распределений 637
Цепи, прочность 22
— случайные 261
Цифры, распределение 52, 85

Частицы, притяжение 382, 646, 666
—, расщепление 40, 61, 129
— — быстрыми частицами 385, 387; ср. Столкновения частиц
—, упорядочение 107, 167
Частные (маргинальные) распределения 90, 169, 196
— —, заданные 206

Частотная характеристика фильтра 702
Чебышева неравенство 191, 276
— — обобщение 299
— — — на мартингалы 301
— Эрмита многочлен 609
Чепмена — Колмогорова уравнение, дискретное время 127
— — —, непрерывное время 384, 556
— — —, полугруппы 416, 528
Шаг 174
Шварца неравенство 192
Шлемильха формула 82
Шум 346, см также Дробовой эффект

Эджворта разложение 611, 620
Эквивалентные функции 158, 714
Экраны, см Поглощающие экраны, Граничные условия
Эмпирические распределения 55
Эндоморфизм 414
Эргодические стохастические ядра 332
— теоремы, дискретное время 263, 274, 330
— —, непрерывное время 446, 562, см Стационарный режим
Эрлангова плотность 66
Эрмита полиномы 609, 630
— разложение 620

Ядра 199, 259, 330
Якоби тэта-функция 405, 406, 710
— σ-аддитивность 137
— алгебра 143
— конечная мера 144

ИМЕННОЙ УКАЗАТЕЛЬ

Амбарцумян В А 261, 388
Андре (Andrè D) 218
Аннекен (Hennequin P L) 131, 729

Байес (Bayes T) 77
Бакстер (Baxter G) 472, 481, 490, 647, 681
Бартлетт (Bartlett M S) 730
Баруча-Рид (Barucha-Reid A T) 399, 730
Бенеш (Beneš E V) 446, 730
Бенфорд (Benford F) 85
Бергстрем (Bergstrom H) 657
Бернштейн С Н 103, 505
Берри (Berry A C) 607, 620, 621, 625, 627, 630
Бикель (Bickel P J) 455
Биллингслей (Billingsley P) 57, 323, 406
Блекуэлл (Blackwell D) 424, 449
Блюм (Blum J R) 346
Боль (Bohl) 327
Ботте (Botts T A) 131
Бохнер (Bochner S) 383, 411, 520, 697, 699, 712, 729
Брело (Brelot M) 299
Будро (Boudreau P E) 248
Бурбаки (Bourbaki N) 131
Бюльман (Buhlman H) 285

Вальд (Wald A) 283, 465, 474
Вебер (Weber H) 504
Вейль (Weyl H) 327
Вейсс (Weiss B) 728
Виленкин Н Я 699, 729
Винер (Wiener N) 426, 718
Винтнер (Wintner A) 209
Волд (Wold H) 597
Вольфовиц (Wolfowitz J) 252, 333, 447, 449

Гальтон (Galton F) 97
Гейтлер (Heitler W) 385
Гельфанд И М 699, 729
Герглотц (Herglotz G) 712
Гири (Geary R C) 111
Гнеденко Б В 57, 62, 337, 607, 618, 657, 668, 730
Гренандер (Grenander U) 102, 729, 730
Гринвуд (Greenwood M) 79
Гриффин (Griffin J S) 248
Гуд (Good D J) 542
Гумбел (Gumbel E J) 206

Дарлинг (Darling D A) 533
Дворецкий А 333
Дерман (Derman C) 564
Деблин (Doblin W) 215, 668
Джильберт (Gilbert E) 273
Донскер (Donsker M) 57, 406
Дуб (Doob J L) 130, 206, 266 297, 299 431, 730
Дынкин Е Б 383, 395, 540, 729

Зигмунд (Zygmund A) 635
Золотарев В М 659

Иахав (Yahav J. A) 455
Ибрагимов И А 209
Ито (Ito K) 395, 400, 729

Кантор (Cantor G) 326
Карамата (Karamata J) 215, 302, 334, 339, 512
Карлеман (Carleman T) 282, 587
Карлин (Carlin S) 283, 449, 730
Кац (Kac M) 103, 248, 406
Кельвин (Kelvin) 403
Кемперман (Kemperman J H B) 472, 729
Кендалл (Kendall D C) 244, 288, 542
Кенуй (Quenouille M H) 50
Кимура (Kimura M) 399
Кингман (Kingman J F C) 231
Кифер (Kiefer J) 252, 447
Кокс (Cox D R) 730
Колмогоров А Н 155, 156, 206, 224, 292, 299, 395, 607, 618, 730
Королюк В С 56, 57, 62
Коши (Cauchy A) 215, 581
Крамер (Cramer H) 597, 607, 620, 628, 625, 730
Крикеберг (Krickeberg K) 729

Ламперти (Lamperti A) 238
Ландау Л Д 385
Ландау (Landau F) 405, 513
Лаха (Laha R G) 729
Лебег (Lebeque H) 152
Леви Поль (Levy Paul) 215, 224, 226, 266, 320, 365, 380, 400, 568, 647, 663, 668, 729, 730
Ле Кам (Le Cam L) 346
Линдеберг (Lindeberg J W) 585
Линдли (Lindley D V) 244, 456
Литтл (Little J D C) 544
Литтльвуд (Littlewood J F) 195, 512
Лоэв (Loeve M) 131, 132, 266 285, 297, 323, 383, 729
Лукач (Lukacs F) 111 729
Лундберг (Lundberg F) 229

Маккин (McKean H P) 395 400, 729
Макшейн (McShane E J) 131
Мандельбройт (Mandelbrot B) 219, 348
Марков А 283
Маршалл (Marshall A W) 301
Мидлтон (Middleton D) 709
Миллс (Mills H D) 130
Моран (Moran P A P) 730
Мюнц (Muntz C) 300

Неве (Neveu J) 131, 297, 729
фон Неймаи (Neumann J) 62
Нейман (Neyman J) 228
Нелсон (Nelson E) 128 411
Ньюэл (Newell G F) 59

Орей (Orey C) 449
Орнштейн (Ornstein D) 257

Пайк (Pyke R) 229, 457, 539, 568
Петерсен (Petersen D P) 709
Петров В В 625
Пинкхэм (Pinkham R S) 86
Питман (Pittman E J) 635
Пойа (Polya G) 79, 195, 215, 228
Поллард (Pollard H) 424, 449, 520
Поляк (Pollak H O) 273
Полячек (Pollaczek F) 250
Порт (Port S C) 339
Прабу (Prabhu H U) 730
Прохоров Ю В 57, 406

Раит (Wright S) 399
Райт (Wright E M) 327
Рвачева Е Л 57
Реньи (Renyi A) 406
Риордан (Riordan J) 730
Рисс (Riesz M) 288
Рихтер (Richter W) 625
Роббинс (Robbins H E) 128, 323, 423
Розенблат (Rosenblatt M) 102, 346, 730

Рой (Roy S.) 423
Ройден (Royden H. L.) 283
Рэй (Ray D.) 395
Рэлей (Rayleigh) 50

Севидж (Savage J. L.) 157, 285
Серпинский (Sierpinski W.) 327
Скеллам (Skellam J. G.) 320
Скитович В. П. 104
Скороход А. В. 57, 730
Смирнов Н. В. 57, 62, 405
Смит (Smith W. L.) 423, 449, 568, 730
Снелл (Snell J. L.) 431
Спарре-Андерсен (Sparre-Andersen E.) 456, 481, 486, 489
Спицер (Spitzer F.) 197, 374, 386, 395, 569, 575, 580, 615
Стадден (Studden W.) 283, 730
Стейн (Stein C.) 678
Стивенс (Stevens W. L.) 44, 102
Стильтьес (Stieltjes) 283
Сьюэлл (Sewall) 399

Такач (Takacs L.) 542, 730
Тамаркин Я. Д. 273, 301
Таннер (Tanner J. C.) 237
Тейчер (Teicher H.) 346
Титчмарш (Titchmarsch E. C.) 688
Тортра (Tortrat A.) 676
Троттер (Trotter H. P.) 320, 411

Угахери (Ugaheri T.) 274
Уилкс (Wilks S. S.) 59
Уильямсон (Williamson R. E.) 431
Уиттекер (Whittaker J. M.) 709
Уокс (Wax N.) 730
Уолл (Woll J. W.) 411
Уоллес (Wallace D. L.) 621
Уолш (Walsh J. W.) 71

Фейер (Fejer L.) 706
Филлипс (Phillips R. S.) 288, 354, 520, 525, 730
де Финетти (Finetti B.) 61, 224, 283
Фишер (Fisher R. A.) 102, 337, 399
Фреше (Frechet H.) 207
Фукс (Fuchs W. H. Y.) 690

Халмош (Halmos P. R.) 131, 270
Хант (Hunt G.) 266
Харди (Hardy G. H.) 195, 327, 512
Харрис (Harris T. E.) 298
Хаусдорф (Hausdorff F.) 281
Хелли (Helly E.) 326
Хилле (Hille E.) 286, 288, 354, 476, 520, 575, 601, 730
Хинчин А. Я. 215, 224, 291, 474, 643, 664, 668, 730
Хобби (Hobby C.) 457
Хопф (Hopf E.) 471, 472
Хьюитт (Hewitt E.) 157, 285

Чандрасекар (Chandrasekhar S.) 50
Чернов (Chernoff H.) 346, 631
Чжоу (Chow S.) 423
Чжун (Chung K. L.) 134, 208, 256, 257, 288, 420, 449, 553, 690, 729

Шапиро (Shapiro J. M.) 647
Шварц (Schwartz L.) 699
Шеллинг (Schelling H.) 264, 285
Шеннон (Shannon C.) 708
Шепп (Shepp L.) 87, 197, 694, 728
Шоке (Choquet G.) 429, 670
Шохат (Shohat J. A.) 283, 301

Эйнштейн А. (Einstein A.) 395
Эйнштейн (Einstein A. Jr.) 228
Эссеен (Esséen G.) 607, 618, 620, 621
Эшер (Esscher F.) 625
Эрдёш (Erdös P.) 406, 424

Юл (Jule G. U.) 79

Яноши (Janossy L.) 385

ОГЛАВЛЕНИЕ

Предисловие к русскому изданию 5
Предисловие 8

Глава I. Показательные и равномерные плотности 13

§ 1. Введение 13
§ 2. Плотности. Свертки 16
§ 3. Показательная плотность 21
§ 4. Парадоксы, связанные с временем ожидания. Пуассоновский процесс 24
§ 5. Устойчивость неудач 29
§ 6. Времена ожидания и порядковые статистики 32
§ 7. Равномерное распределение 36
§ 8. Случайные разбиения 40
§ 9. Свертки и теоремы о покрытии 42
§ 10. Случайные направления 46
§ 11. Использование меры Лебега 51
§ 12. Эмпирические распределения 55
§ 13. Задачи 58

Глава II. Специальные плотности. Рандомизация 64

§ 1. Обозначения и определения 64
§ 2. Гамма-распределения 66
§ 3. Распределения математической статистики, связанные с гамма-распределениями 67
§ 4. Некоторые распространенные плотности 69
§ 5. Рандомизация и смеси 74
§ 6. Дискретные распределения 76
§ 7. *Бесселевы функции и случайные блуждания* 79
§ 8. Распределения на окружности 83
§ 9. Задачи 86

Глава III. Многомерные плотности. Нормальные плотности и процессы 89

§ 1. Плотности 89
§ 2. *Условные распределения* 95
§ 3. Возвращение к показательному и равномерному распределениям 98
§ 4. Характеризация нормального распределения 102
§ 5. Матричные обозначения. Матрица ковариаций 106
§ 6. Нормальные плотности и распределения 108
§ 7. Стационарные нормальные процессы 114
§ 8. Марковские нормальные плотности 122
§ 9. Задачи 128

Глава IV. **Вероятностные меры и пространства** 132

§ 1. Бэровские функции 132
§ 2. Функции интервалов и интегралы в $\mathscr{R}^r$ 135
§ 3. Вероятностные меры и пространства 142
§ 4. Случайные величины. Математические ожидания 145
§ 5. Теорема о продолжении 149
§ 6. Произведения пространств. Последовательности независимых случайных величин 153
§ 7. Нулевые множества. Пополнение 158

Глава V. **Вероятностные распределения в $\mathscr{R}^r$** 160

§ 1. Распределения и математические ожидания 161
§ 2. Предварительные сведения 170
§ 3. Плотности 174
§ 3а. Сингулярные распределения 177
§ 4. Свертки 179
§ 5. Симметризация 186
§ 6. Интегрирование по частям. Существование моментов 189
§ 7. Неравенство Чебышева 191
§ 8. Дальнейшие неравенства. Выпуклые функции 192
§ 9. Простые условные распределения. Смеси 196
§ 10. Условные распределения 200
§ 10а. Условные математические ожидания 203
§ 11. Задачи 206

Глава VI. **Некоторые важные распределения и процессы** 210

§ 1. Устойчивые распределения в $\mathscr{R}^1$ 210
§ 2. Примеры 216
§ 3. Безгранично делимые распределения в $\mathscr{R}^1$ 220
§ 4. Процессы с независимыми приращениями 224
§ 5. Обобщенные пуассоновские процессы и задачи о разорении 228
§ 6. Процессы восстановления 230
§ 7. Примеры и задачи 234
§ 8. Случайные блуждания 240
§ 9. Процессы массового обслуживания 244
§ 10. Возвратные и невозвратные случайные блуждания 252
§ 11. Общие марковские цепи 258
§ 12. Мартингалы 265
§ 13. Задачи 272

Глава VII. **Законы больших чисел. Применения в анализе** 275

§ 1. Основная лемма. Обозначения 275
§ 2. Полиномы Бернштейна. Абсолютно монотонные функции . . . 278

§ 3. Проблемы моментов 280
§ 4. Применение к симметрично зависимым случайным величинам 283
§ 5. Обобщенная формула Тейлора и полугруппы 286
§ 6. Формулы обращения для преобразования Лапласа 288
§ 7. Законы больших чисел для одинаково распределенных случайных величин 290
§ 8. Усиленный закон больших чисел для мартингалов 295
§ 9. Задачи 300

Глава VIII. **Основные предельные теоремы** 302

§ 1. Сходимость мер 302
§ 2. Специальные свойства 307
§ 3. Распределения как операторы 311
§ 4. Центральная предельная теорема 315
§ 5. Бесконечные свертки 324
§ 6. Теоремы о выборе 325
§ 7. Эргодические теоремы для цепей Маркова 330
§ 8. Правильно меняющиеся функции 334
§ 9. Асимптотические свойства правильно меняющихся функций 339
§ 10. Задачи 344

Глава IX. **Безгранично делимые распределения и полугруппы** . . . 349

§ 1. Общее знакомство с темой 349
§ 2. Полугруппы со сверткой 352
§ 3. Подготовительные леммы 356
§ 4. Случай конечных дисперсий 358
§ 5. Основная теорема 361
§ 6. Пример: устойчивые полугруппы 366
§ 7. Схемы серий 369
§ 8. Области притяжения 373
§ 9. Различные распределения. Теорема о трех рядах 378
§ 10. Задачи 381

Глава X. **Марковские процессы и полугруппы** 383

§ 1. Псевдопуассоновский тип 384
§ 2. Вариант: линейные приращения 387
§ 3. Скачкообразные процессы 389
§ 4. Диффузионные процессы в $\mathscr{R}^1$ 394
§ 5. Прямое уравнение. Граничные условия 400
§ 6. Диффузия в многомерном случае 407
§ 7. Подчиненные процессы 408
§ 8. Марковские процессы и полугруппы 413
§ 9. «Показательная формула» в теории полугрупп 417
§ 10. Производящие операторы. Обратное уравнение 420

Глава XI Теория восстановления 423

§ 1. Теорема восстановления 423
§ 2. Уравнение $\zeta = F * \zeta$ 429
§ 3. Устойчивые процессы восстановления 431
§ 4. Уточнения 436
§ 5. Центральная предельная теорема 438
§ 6. Обрывающиеся (невозвратные) процессы 440
§ 7. Применения 444
§ 8 Существование пределов в случайных процессах 446
§ 9. Теория восстановления на всей прямой 448
§ 10 Задачи 453

Глава XII. Случайные блуждания в $\mathscr{R}^1$ 456

§ 1. Обозначения и соглашения 457
§ 2 Двойственность 461
§ 3. Распределение лестничных высот Факторизация Винера—Хопфа 466
§ 4 Примеры 472
§ 5. Применения 477
§ 6. Одна комбинаторная лемма 480
§ 7. Распределение лестничных моментов 481
§ 8 Закон арксинуса 484
§ 9. Различные дополнения 489
§ 10 Задачи 491

Глава XIII. Преобразование Лапласа. Тауберовы теоремы. Резольвенты 495

§ 1. Определения. Теорема непрерывности 495
§ 2. Элементарные свойства 500
§ 3. Примеры 502
§ 4 Вполне монотонные функции. Формулы обращения 504
§ 5 Тауберовы теоремы 508
§ 6. Устойчивые распределения 514
§ 7. Безгранично-делимые распределения 516
§ 8 Многомерный случай 519
§ 9. Преобразования Лапласа для полугрупп 520
§ 10. Теорема Хилле—Иосида 526
§ 11 Задачи 530

Глава XIV. Применение преобразования Лапласа 534

§ 1. Уравнение восстановления: теория 534
§ 2. Уравнение типа уравнения восстановления: примеры 536
§ 3. Предельные теоремы, включающие распределения арксинуса . . 539
§ 4 Периоды занятости и соответствующие ветвящиеся процессы 542

§ 5. Диффузионные процессы 544
§ 6. Процессы размножения и гибели. Случайные блуждания . . . 549
§ 7. Дифференциальные уравнения Колмогорова 553
§ 8. Пример: чистый процесс размножения 559
§ 9. Вычисление $P(\infty)$ и времен первого прохождения 562
§ 10. Задачи 566

Глава XV. **Характеристические функции** 569

§ 1. Определение. Основные свойства 569
§ 2. Специальные плотности. Смеси 573
§ 3. Единственность. Формулы обращения 579
§ 4. Свойства регулярности 584
§ 5. Центральная предельная теорема для одинаково распределенных слагаемых 588
§ 6. Условие Линдеберга 592
§ 7. Характеристические функции многомерных распределений . . . 596
§ 8. Две характеризации нормального распределения 600
§ 9. Задачи 603

Глава XVI. **Асимптотические разложения, связанные с центральной предельной теоремой** 607

§ 1. Обозначения 608
§ 2. Асимптотические разложения для плотностей 609
§ 3. Сглаживание 613
§ 4. Асимптотические разложения для распределений 616
§ 5. Теорема Берри—Эссеена 620
§ 6. Большие отклонения 622
§ 7. Различно распределенные слагаемые 626
§ 8. Задачи 630

Глава XVII. **Безгранично делимые распределения** 632

§ 1. Теорема о сходимости 632
§ 2. Безгранично делимые распределения 638
§ 3. Примеры. Специальные свойства 644
§ 4. Устойчивые характеристические функции 648
§ 5. Области притяжения 652
§ 6. Устойчивые плотности 657
§ 7. Схема серий 659
§ 8. Класс L 663
§ 9. Частичное притяжение. «Универсальные законы» 666
§ 10. Бесконечные свертки 669
§ 11. Многомерный случай 670
§ 12. Задачи 671

Глава XVIII. **Применение методов Фурье к случайным блужданиям** . 675

§ 1. Основное тождество 675
§ 2 Конечные интервалы Вальдовская аппроксимация 678
§ 3 Факторизация Винера—Хопфа 681
§ 4. Обсуждение результатов Применения 684
§ 5. Уточнения 687
§ 6 Возвращения в нуль 689
§ 7. Критерии возвратности 690
§ 8 Задачи 693

Глава XIX **Гармонический анализ** 695

§ 1 Равенство Парсеваля 695
§ 2 Положительно определенные функции 697
§ 3 Стационарные процессы 700
§ 4. Ряды Фурье 703
§ 5 Формула суммирования Пуассона 707
§ 6. Положительно определенные последовательности 710
§ 7. L^2-теория 713
§ 8 Случайные процессы и стохастические интегралы 719
§ 9. Задачи 726

Предметный указатель 736
Именной указатель 744