Г. И. Ивченко, Ю. И. Медведев

Просто о сложном!

Urbi et orbi (лат.) — всем и каждому.

Это одновременно и расширенный учебник, и справочное пособие, и задачник.

Книга рассказывает о математической статистике, но одновременно и обучает ей.

«Если хотите увлечь вашим иконам — сделайте его привлекательным. Настолько привлекательным, чтобы книга открылась для ознакомления сухими листьями». Мы стремились следовать этому правилу.

ВВЕДЕНИЕ В МАТЕМАТИЧЕСКУЮ СТАТИСТИКУ



«Статистика знает всё»

URSS

Г. И. Ивченко, Ю. И. Медведев

ВВЕДЕНИЕ В МАТЕМАТИЧЕСКУЮ СТАТИСТИКУ



URSS
МОСКВА

ББК 22.172

Ивченко Григорий Иванович,
Медведев Юрий Иванович

Введение в математическую статистику: Учебник.
М.: Издательство ЛКИ, 2010. — 600 с.

Настоящая книга представляет собой своеобразный расширенный учебник по математической статистике. Данный учебник не ограничен рамками учебного стандарта или вузовской программы — он предназначен всем, кто интересуется математикой вообще и, в частности, хочет узнать, что такое современная математическая статистика, какие задачи и какими методами она решает, какие результаты в ней уже накоплены, какие проблемы в ней сегодня актуальны; наконец, каковы ее истоки, какой путь она прошла и какие ученые были ее творцами. По замыслу авторов, книга простым и доступным языком рассказывает о математической статистике и одновременно обучает ей. Вся теория объясняется и иллюстрируется на интересных и тщательно подобранных примерах. Книга может служить и задачником, так как содержит большой список упражнений для самостоятельного решения, а также справочным пособием по математической статистике, а в некоторых аспектах — и по теории вероятностей.

Книга будет интересна преподавателям, аспирантам и студентам естественных и технических вузов, в которых изучается математическая статистика, научным работникам, использующим в своей деятельности методы математической статистики, а также самому широкому кругу любителей математики.

Издательство ЛКИ.
117312, Москва, пр-т Шестидесятилетия Октября, 9.
Формат 60×90/16. Печ. л. 37,5. Зак. № 1748.

Отпечатано в ООО «ПК «Зауралье».
640022, Курганская обл., Курган, ул. К. Маркса, 106.

ISBN 978–5–382–01013–7

© Издательство ЛКИ, 2009

НАУЧНАЯ И УЧЕБНАЯ ЛИТЕРАТУРА
E-mail: URSS@URSS.ru
Каталог изданий в Интернете:
http://URSS.ru
Тел./факс: 7 (499) 135–42–16
URSS Тел./факс: 7 (499) 135–42–46

5735 ID 71268
9 785382 010137

Все права защищены. Никакая часть настоящей книги не может быть воспроизведена или передана в какой бы то ни было форме и какими бы то ни было средствами, будь то электронные или механические, включая фотокопирование и запись на магнитный носитель, а также размещение в Интернете, если на то нет письменного разрешения владельца.

Содержание

Предисловие 11

Введение 14

§ 1. Что такое математическая статистика, ее предмет и задачи 14

§ 2. Краткий исторический очерк развития математической статистики 25

Глава 1. (Вспомогательная, но очень важная!) **Основные распределения и их моделирование** . 30

Введение 30

§ 1.1. Основные дискретные модели математической статистики 31

1. Схема Бернулли и биномиальное распределение 31
2. Отрицательное биномиальное распределение 34
3. Распределение Пуассона 35
4. Гипергеометрическое распределение 37
5. Распределение Маркова—Пойа . 37
6. Полиномиальное распределение 39
7. Многомерное распределение Маркова—Пойа 40
8. Распределение степенного ряда . 42

§ 1.2. Основные абсолютно непрерывные модели 43

1. Нормальное распределение 44
2. Многомерное нормальное распределение 47
3. Гамма-распределение 51
4. Бета-распределение 53
5. Равномерное распределение 54
6. Распределение Стьюдента 55
7. Распределение Снедекора 56
8. Распределение Вейбулла 57
9. Распределение Парето 57
10. Распределение Дирихле 57
11. Преобразования случайных величин и векторов 58

§ 1.3. Моделирование выборок из конкретных распределений 62

1. Предварительные замечания 62
2. Моделирование распределений Бернулли $\mathrm{Bi}(1, p)$ и биномиального $\mathrm{Bi}(k, p)$ 64
3. Моделирование отрицательного биномиального распределения $\overline{\mathrm{Bi}}(r, p)$ 64
4. Моделирование полиномиальных испытаний 65
5. Моделирование пуассоновского распределения $\Pi(\lambda)$ 65

6. Моделирование непрерывных распределений 66
7. Моделирование показательного и связанных с ним распределений 67
8. Моделирование нормальных случайных чисел 67
9. Метод суперпозиций 68
10. Моделирование цепи Маркова 70
11. Метод Монте-Карло 73
Упражнения 74

Глава 2. Первичная обработка экспериментальных данных **89**

§ 2.1. Вариационный ряд выборки, эмпирическая функция распределения и гистограмма . 89
1. Порядковые статистики и вариационный ряд выборки 89
2. Эмпирическая функция распределения 90
3. Дальнейшие свойства э. ф. р. 96
4. Полигон частот, гистограмма 100
5. Ядерные оценки теоретической плотности 103
§ 2.2. Выборочные моменты: точная и асимптотическая теория 105
1. Выборочные моменты 105
2. Моменты выборочных среднего и дисперсии 108
3. Сходимость по вероятности выборочных моментов и функций от них 109
4. Асимптотическая нормальность выборочных моментов 114
5. Асимптотические доверительные интервалы для теоретических моментов 117
§ 2.3. Многомерные данные 118
1. Эмпирическая функция распределения 119
2. Моменты 120
3. Большие выборки 121
4. Добавление: нормальная модель 122
5. Другие корреляционные характеристики 125
§ 2.4. Выборочные квантили и порядковые статистики 128
1. Теоретические и эмпирические квантили 128
2. Распределение порядковых статистик 130
3. Асимптотическая нормальность средних членов вариационного ряда 133
4. Асимптотическое поведение крайних порядковых статистик 135
5. Асимптотическая теория для верхних экстремумов $X_{(n-m+1)}$, $m \geqslant 1$ 138
§ 2.5. Линейные и квадратичные статистики от нормальных выборок 140
1. Линейные и квадратичные статистики, условия их независимости 141
2. Распределения квадратичных статистик 143
3. Теорема Фишера 145
Упражнения 150

Глава 3. Общая теория оценивания неизвестных параметров распределений 159
§ 3.1. Статистические оценки и общие требования к ним 159
1. Понятие статистической оценки и ее среднеквадратичная ошибка 159
2. Несмещенные оценки 162
3. Оптимальные оценки 168
§ 3.2. Критерии оптимальности оценок, основанные на неравенстве Рао—Крамера. Эффективные оценки 173
1. Функция правдоподобия, вклад выборки, функция информации 173
2. Неравенство Рао—Крамера 175
3. Экспоненциальная модель 179
4. Критерий Бхаттачария оптимальности оценки 181
5. Критерии оптимальности в случае векторного параметра 182
§ 3.3. Достаточные статистики и оптимальные оценки 186
1. Достаточные статистики и критерий факторизации 186
2. Теорема Рао—Блекуэлла—Колмогорова 191
3. Экспоненциальные семейства и достаточные статистики 193
§ 3.4. Способы решения уравнения несмещенности 199
1. Модель степенного ряда, оценивание параметрических функций 200
2. Модели с выборочным пространством, зависящим от параметра θ 204
3. Метод подгонки 207
4. Гамма-модель с неизвестным параметром масштаба, оценивание параметрических функций 213
5. Другие применения гамма-модели для оценивания неизвестных параметров 217
§ 3.5. Оценки максимального правдоподобия . 219
1. Определение и примеры оценок максимального правдоподобия (о. м. п.) 219
2. Принцип инвариантности для о. м. п. 223
3. Метод накопления для приближенного вычисления о. м. п. 225
4. Асимптотические свойства о. м. п. 227
5. Асимптотические доверительные интервалы, основанные на о. м. п. 234
§ 3.6. Другие методы и принципы построения оценок 239
1. Метод моментов . 239
2. Метод минимума хи-квадрат 241
3. Модально несмещенные оценки 243
4. Медианно несмещенные оценки 245
5. Эквивариантные оценки 246
6. Байесовские и минимаксные оценки 250
7. Оценивание по цензурированным данным 258
§ 3.7 Объединение и улучшение оценок 264
1. Объединение оценок 264
2. Улучшение оценок 271
§ 3.8. Доверительное оценивание 276
1. Построение доверительного интервала с помощью центральной статистики 277

2. Построение доверительного интервала с использованием распределения точечной оценки параметра 281
3. Асимптотические доверительные интервалы 284
§ 3.9. Оценивание при выборе из конечной совокупности 288
1. Оценивание среднего совокупности 288
2. Оценивание состава совокупности 292
Упражнения 297

Глава 4. Проверка статистических гипотез 311

§ 4.1. Основные понятия и общие принципы теории проверки гипотез 311
§ 4.2. Проверка гипотезы о виде распределения 318
1. Критерий согласия Колмогорова 318
2. Критерий согласия хи-квадрат К. Пирсона 320
3. Критерий хи-квадрат для сложной гипотезы 327
4. Критерий квантилей 332
5. Критерий пустых ящиков 334
§ 4.3. Гипотеза и критерии однородности 340
1. Критерий однородности Смирнова 340
2. Критерий однородности хи-квадрат 341
3. Другие критерии в задаче о двух выборках 347
§ 4.4. Гипотеза независимости 350
1. Критерий независимости хи-квадрат 351
2. Критерий Спирмена 357
3. Критерий Кендалла 358
4. Случай m признаков 359
§ 4.5. Гипотеза случайности 361
1. Критерий инверсий 362
2. Проблема датчиков и обобщенный критерий хи-квадрат 364
Упражнения 368

Глава 5. Параметрические гипотезы 372

§ 5.1. Общие положения 372
§ 5.2. Выбор из двух простых гипотез. Критерий Неймана—Пирсона 374
1. Постановка задачи и предварительные соображения 374
2. Случай абсолютно непрерывных распределений 375
3. Критерий Неймана—Пирсона в случае дискретных распределений 386
§ 5.3. Сложные гипотезы 394
1. Р. н. м. критерии против сложных односторонних альтернатив. Модели с монотонным отношением правдоподобия 394
2. Двусторонние альтернативы, р. н. м. несмещенные критерии 398
3. Локальные наиболее мощные критерии 403
4. Проверка гипотез и доверительное оценивание 405
§ 5.4. Критерий отношения правдоподобия 409
1. Метод отношения правдоподобия для общих гипотез. 409
2. КОП для больших выборок 417
3. Дальнейшие асимптотические свойства КОП 421

4. Сложная нулевая гипотеза 423
5. Доверительные области максимального правдоподобия 426
§ 5.5. Проверка гипотез для конечных цепей Маркова 427
1. Гипотезы для конечных цепей Маркова 428
2. Критерий хи-квадрат для простой гипотезы 431
3. Критерий хи-квадрат для сложной гипотезы 432
4. Критерий отношения правдоподобия для общих параметрических гипотез 436
5. Критерий однородности 438
6. Оценивание порядка цепи 442
§ 5.6. Понятие о последовательном анализе. Критерий Вальда. 443
1. Определение критерия Вальда 444
2. О числе испытаний до момента остановки в критерии Вальда 445
3. О выборе границ в критерии Вальда 446
4. О среднем числе наблюдений в критерии Вальда 448
Упражнения 451

Глава 6. (Специальная) **Линейная регрессия и метод наименьших квадратов** **457**

§ 6.1. Модель линейной регрессии 457
§ 6.2. Оценивание параметров модели линейной регрессии 459
1. Метод наименьших квадратов 459
2. Оптимальность оценок наименьших квадратов 460
3. Оценивание остаточной дисперсии 462
4. Условные о. н. к. 463
5. Оптимальный выбор матрицы плана . 464
6. Примеры применения метода наименьших квадратов 465
7. Ортогональные многочлены Чебышева 470
§ 6.3. Нормальная регрессия 474
1. Модель нормальной регрессии 474
2. Оценки максимального правдоподобия (о. м. п.) параметров нормальной регрессии 475
3. Основная теорема теории нормальной регрессии 476
4. Доверительное оценивание параметров нормальной регрессии 477
5. Доверительная область для линейных комбинаций параметров $\underline{\beta}$ 479
6. Система совместных доверительных интервалов 480
7. Доверительный интервал для отклика 481
8. Проверка адекватности модели 482
§ 6.4. Общая линейная гипотеза нормальной регрессии 484
1. Понятие линейной гипотезы 484
2. F-критерий для линейной гипотезы 484
§ 6.5. Некоторые применения теории нормальной регрессии 490
1. Гипотеза о параллельности линий регрессии 490
2. Гипотеза однородности для нескольких нормальных выборок 493
3. Двойная классификация. Дисперсионный анализ 496
§ 6.6. Статистическая регрессия и прогнозирование 502

1. Задачи статистического прогноза 502
2. Условное математическое ожидание 503
3. Оптимальный предиктор и его свойства 505
4. Из истории регрессии 507
5. Линейный прогноз 508
6. Использование в прогнозе дополнительных переменных. Алгоритм обновления прогноза 511
7. Прогнозирование стационарных временных рядов 514
Упражнения 520

Глава 7. Элементы теории решений. Дискриминантный анализ 525

§ 7.1. Статистические решающие функции. Байесовское и минимаксное решения 525
1. Понятие решающей функции 525
2. Функция риска и допустимые решающие правила 526
3. Байесовское решение 527
4. Минимаксное решение 528
5. Оценивание параметров и проверка гипотез с позиций теории решений 531
§ 7.2. Классификация наблюдений 532
1. Задача классификации 532
2. Функция риска в задаче классификации 532
3. Байесовское решение 533
4. Минимаксное решение 535
§ 7.3. Классификация наблюдений в случае двух нормальных классов 535
1. Байесовский подход 536
2. Минимаксный подход 537
§ 7.4. Классификация нормальных наблюдений. Общий случай 538
1. Байесовский подход 538
2. Минимаксный подход 540
3. Классификация наблюдений в случае неизвестных параметров 540
Упражнения 542

Глава 8. Факторный анализ 546

§ 8.1. Постановка задач факторного анализа 546
§ 8.2. Неоднозначность решения в факторном анализе 548
§ 8.3. Вывод уравнений максимального правдоподобия 549
§ 8.4. Итерационный метод нахождения факторных нагрузок 553
§ 8.5. Проверка гипотезы о числе факторов 555
§ 8.6. Центроидный метод 558
§ 8.7 Оценка значений факторов 562

Глава 9. Компонентный анализ 564

§ 9.1 Постановка задач компонентного анализа 564
§ 9.2 Решение основных уравнений. Главные компоненты 566
§ 9.3. Метод Хотеллинга 568

Приложение **573**

1. Нормальное распределение 573
2. Распределение Пуассона 574
3. Биномиальное распределение 575
4. Распределение $\chi^2(n)$ 576
5. Распределение Стьюдента $S(n)$ 577
6. Распределение Снедекора $S(n_1, n_2)$ 578
7. Критерий Колмогорова 579
8. Критерий Смирнова 580
9. Равномерно распределенные случайные числа 581
10. Нормально распределенные $\mathcal{N}(0, 1)$ случайные числа 582

Литература **585**

Путеводитель по моделям в примерах и задачах **586**

Именной указатель **598**

Математическая статистика
является наукой о методах количественного анализа
массовых явлений, учитывающей одновременно
и качественное своеобразие этих явлений.

Б. В. Гнеденко

Предисловие

Книга человечнее, честнее, долговечней и ответственней любого другого источника информации.

В. А. Садовничий [1]

Уважаемый Читатель! Эта книга адресована тебе, поэтому мы хотим вначале кратко пояснить ее суть и особенности, объяснить те идеи, которыми мы руководствовались при ее написании. Прежде всего, это — книга *для всех*, т. е. не для определенного контингента обучающихся по той или иной специальности, того или иного профиля или направления, она не ограничена рамками определенного учебного стандарта или программы, наконец, заранее установленными ограничениями по объему. Она для тех, кто по существу, а не для «галочки» в ведомости, интересуется математикой вообще и, в частности, хочет узнать, что такое современная математическая статистика, какие задачи и какими методами она решает, какие результаты в ней уже накоплены, какие проблемы в ней сегодня актуальны, наконец, каковы ее истоки, какой путь она прошла, и кто был ее творцами (ведь *история науки так же важна, как и сама наука*). Итак, по замыслу, эта книга *рассказывает* о математической статистике, но одновременно и *обучает* ей, т. е. она может служить *расширенным* учебником, как покрывающим стандартные вузовские программы, предусматривающие чтение (в том или ином объеме) курса математической статистики, так и обеспечивающим возможность самообразования. Помимо этого, книга также может служить справочным пособием по математической статистике, а в некоторых аспектах — и по теории вероятностей. Из этой целевой установки вытекают все ее структурные и содержательные особенности.

Наш девиз: просто о сложном! Мы стремились писать так, чтобы излагаемое было доступно каждому, кто знаком с начальными элементами математического анализа, линейной алгебры и теории вероятностей. Причем помнить их либо обращаться за нужными сведениями в соответствующие учебники не обязательно, так как все необходимое напоминается (разъясняется) по ходу изложения собственно статистической теории. Никаких сложных математических конструкций и громоздких доказательств! Вся теория объясняется и иллюстрируется на интересных и тщательно подобранных примерах: «при изучении наук примеры полезнее правил» (И. Ньютон [2]). Для тех, кто имеет

---

[1] Садовничий Виктор Антонович (р. 1939) — российский математик, академик РАН, ректор МГУ им. М. В. Ломоносова (с 1992 г.).

[2] Ньютон Исаак (1643–1727) — английский математик, физик, механик, астроном.

целью глубоко освоить теорию статистического вывода, приводится большой список упражнений для самостоятельного решения, т. е. книга объединяет одновременно и учебник, и задачник. Умение решать задачи означает как знание теории, так и способность правильно этим знанием распорядиться, т. е. привести теорию в действие, ведь «если действовать не будешь, ни к чему ума палата» (Шота Руставели[3]). Упражнения, как правило, сформулированы так, что в самой формулировке уже заложен ответ, путь же к нему — это и есть тот творческий элемент, на котором можно узнать свои способности (*знать путь и пройти его — не одно и то же*). «Подсказкой» здесь (при затруднениях) могут служить соответствующие примеры из текста: на них можно освоить необходимые логику и аналитическую технику.

Теперь несколько слов для преподавателей (*как говорили древние*: *docendo discimus* (*лат.*) — *уча*, *учимся сами*). Одной из основных задач при преподавании курса математической статистики является выработка у студентов навыков обработки реальных данных по соответствующим алгоритмам с привлечением компьютерной техники при решении конкретных учебных задач. В настоящее время глубокое изучение теории математической статистики в отрыве от практики ее компьютерной реализации вообще невозможно. В то же время сам характер этой дисциплины доставляет большой простор преподавателю для организации практической работы студентов с привлечением компьютеров. Ключевым моментом при этом является моделирование на компьютере выборок из соответствующих распределений. Фактически на основе любой «теоретической» задачи, в которой речь идет о том или ином статистическом алгоритме анализа данных, преподаватель может поставить (и притом в неограниченном числе вариантов, варьируя конкретные значения параметров модели) соответствующую «практическую» задачу, формулируя в качестве предварительного этапа задание смоделировать исходные данные с использованием либо готовых таблиц случайных чисел, либо получаемых с помощью специальных программ. В дальнейшем, обрабатывая эти «экспериментальные» данные по соответствующему статистическому алгоритму, студент получает возможность сравнить предсказание теории с известными ему исходными параметрами, при которых моделировалась выборка. Подобный «игровой» элемент в организации практической работы по курсу повышает интерес студентов к изучению теории. Вопросам моделирования конкретных распределений уделяется в книге существенное внимание.

Наконец, мы приводим в конце книги достаточно обширный список литературы как общего характера, так и специальных монографий, в которых более полно и глубоко освещены различные аспекты статистической теории и ее приложений; эта литература может быть использована для самостоятельного и углубленного изучения отдельных вопросов.

В приложении приведены краткие статистические таблицы, необходимые для разбора примеров и решения задач, а также «Путеводитель по моделям

[3] Руставели Шота — грузинский поэт любви и красоты XII в.

в примерах и задачах», назначение которого — помочь быстро найти в тексте нужное место, где доказывается либо формулируется тот или иной конкретный факт (результат), относящийся к конкретной модели. Знак ■ в тексте означает окончание доказательства. Знаком • отмечены окончания примеров. Нумерация формул, теорем, рисунков и т. д. в каждом параграфе самостоятельная, при ссылках на материал другого параграфа указывается дополнительно его номер.

Наконец, еще одно пояснение об используемых обозначениях. По тексту нам будут встречаться как скалярные, так и векторные величины, обозначаемые в обоих случаях одним и тем же символом, например, θ — скаляр и $\theta = (\theta_1, \ldots, \theta_n)$ — вектор. Эта их размерностная спецификация всегда будет ясна из контекста. Однако в ряде случаев (там, где это диктуется содержательной стороной вопроса), чтобы подчеркнуть многомерный характер некоторой величины, мы будем дополнительно снабжать соответствующий символ чертой снизу, например, $\underline{x} = (x_1, \ldots, x_n)$ при $n > 1$, $\underline{\xi} = (\xi_1, \ldots, \xi_k)$ при $k > 1$ и т. д.

Восточная мудрость гласит: «Если хотите увлечь вашим знанием — сделайте его привлекательным. Настолько привлекательным, чтобы книги вчерашнего дня показались сухими листьями». Мы стремились следовать этому правилу. Насколько это нам удалось — судить тебе, наш Читатель.

Мы выражаем сердечную благодарность нашим коллегам и добрым товарищам: академику Ю. В. Прохорову, члену-корреспонденту РАН Б. А. Севастьянову и действительному члену Академии криптографии РФ В. П. Чистякову за постоянную поддержку нашей работы по математической статистике.

Авторы

Введение

...не проворным достается успешный бег, не храбрым — победа, не мудрым — хлеб, и не у разумных — богатство, и не искусным — благорасположение, но время и случай для всех их.

Екклесиаст (Библия. Ветхий Завет. 9:11)

Во всех изысканиях разума самое трудное — это начало.

Платон [1]

В любом начале волшебство таится,
Оно нам в помощь, в нем защита наша.

Г. Гессе [2] Игра в бисер

§ 1. Что такое математическая статистика, ее предмет и задачи

Итак, мы начинаем. Начинать же всегда надо с определения предмета обсуждения. Наш предмет — современная математическая статистика. Как ее можно определить кратко и, в то же время, понятно? Прежде всего — это математическая теория, а любую математическую теорию можно охарактеризовать указанием области тех практических задач, к решению которых она применяется. Поэтому мы начнем с того, что приведем несколько характерных примеров таких задач, которые решаются методами математической статистики.

Пример 1. Имеется урна, содержащая шары двух цветов, скажем, белые и черные. Пусть a — число всех шаров, a_1 — число белых из них и $a_2 = a - a_1$ — число черных шаров. Оба эти числа a_1 и a_2 (или хотя бы только a_1) нам могут быть неизвестны, и в этом случае может возникнуть потребность оценить неизвестную долю $p = a_1/a$ белых шаров в урне. Как поступить в этом случае? Можно, например, извлечь из урны все шары и посчитать число белых среди них — тогда мы узнаем точно числа a и a_1, следовательно, и $p = a_1/a$. Но такой «тривиальный» эксперимент по разным причинам может быть неосуществим (большой объем содержимого урны — трудно ее исчерпать, «дорого» это стоит и т. п.). Поэтому в таких ситуациях обычно поступают следующим

[1] Платон (427–347 до н. э.) — древнегреческий философ.

[2] Гессе Герман (1877–1962) — немецкий писатель.

образом: извлекают из урны некоторое число n шаров, подсчитывают число наблюдавшихся в выборке белых шаров (обозначим его m) и в качестве *оценки* (приближенного значения) для неизвестной доли $p = a_1/a$ используют величину $\widehat{p} = m/n$ — долю белых шаров среди n извлеченных. Приведенное описание такого эксперимента еще не совсем полно, так как организовать выбор n шаров можно разными способами. Можно, например, извлекать шары по одному последовательно, возвращая каждый раз вынутый шар обратно (в этом случае говорят *о выборе с возвращением*), а можно вынутый шар в урну не возвращать (*выбор без возвращения*) — в этом случае это то же самое, что извлечь сразу комплект из n шаров (при $n \leqslant a$); можно придумать и другие процедуры извлечения шаров из урны. Наконец, следует еще уточнить сам механизм отбора извлекаемых шаров, что делается уже с привлечением понятий теории вероятностей и прежде всего понятия *случайного* выбора (но об этом позже, см. пример 1 в § 1.1). Но во всех случаях возникает принципиальный вопрос: будет ли «эмпирическая» оценка $\widehat{p}$ отражать реальную картину, т. е. будет ли она близка (в каком-то смысле) к неизвестному нам истинному значению p? Более того, если в наших возможностях по-разному организовать эксперимент (в этом случае говорят об «активном» эксперименте), то какая из соответствующих оценок будет «лучше» (в каком-то смысле)? Здесь речь идет, следовательно, о сравнении различных экспериментов (способов действия) с позиций «качества» соответствующих оценок. На практике часто приходится учитывать еще и затраты на проведение эксперимента, так что экономические аспекты также приходится включать при этом в круг рассмотрения. Ответ на все эти непростые вопросы, а тем самым и рекомендации, как наиболее рационально (грамотно) нам действовать в подобных ситуациях, дает математическая статистика. •

Статистические задачи оценивания

Общее замечание. Использованный в этом примере язык урновой схемы («урна — шары») является примером абстрактной (математической) модели, подходящей для описания многих реальных ситуаций (в биологии, демографии, социологии, медицине, психологии, лингвистике, экономике и т. д.). Под шарами в урне можно понимать любую конечную совокупность объектов (элементов), которые (как в рассмотренном примере) могут быть разбиты по некоторому альтернативному признаку (пол, возраст, состояние здоровья, доход, сорт, емкость и т. п.) на два непересекающихся класса, и тогда задача состоит в том, чтобы с помощью некоторого статистического эксперимента оценить долю определенного класса во всей совокупности. Например, совокупность — это граждане страны, избирающие президента и разделившиеся на два класса по признаку «за — против», или партия каких-то изделий, которые могут быть исправными или дефектными и т. д. В более общем случае исследуемые объекты могут быть разбиты на большее, чем 2, число классов (тогда мы будем говорить об урне с шарами, скажем, N различных цветов), и тогда проблема состоит в оценивании долей всех классов, составляющих исследуемую совокупность. Во всех таких случаях язык урновых схем является универсальным экономным способом описания исследуемой ситуации и им широко пользуются в современной математической статистике и теории вероятностей.

Пример 2. Пусть требуется измерить некоторую величину a (например, диаметр некоторой звезды). Поскольку на результаты измерений влияют многие случайные факторы (точность настройки измерительного прибора, погрешность округления при считывании данных, психофизическое состояние оператора и т. д.), то результат i-го измерения X_i можно записать в виде $X_i = a+\varepsilon_i$, где ε_i — случайная погрешность измерения. Спрашивается, как по результатам n измерений $X_1, \ldots, X_n$ оценить величину a, точное значение которой нам неизвестно? Обычно в таких случаях в качестве оценки a используют среднее арифметическое наблюдений

$$\widehat{a} = \overline{X} = \frac{1}{n}(X_1 + \quad + X_n),$$

и математическая статистика позволяет обосновать разумность такого правила оценивания. •

Пример 3. Пусть из некоторых соображений известно, что переменная y линейно зависит от переменных $x_1, \ldots, x_r$: $y = A_0 + A_1x_1 + \quad + A_rx_r$ (в этих случаях говорят, что y «регрессирует» на $x_1, \quad, x_r$), причем коэффициенты $A_0, A_1, \quad, A_r$ неизвестны. При различных наборах $(x_{i1}, \quad, x_{ir})$, $i = 1, 2, \ldots, n$, измерены значения $y_i = A_0 + A_1x_{i1} + \ldots + A_rx_{ir} + \varepsilon_i$, где ε_i — ошибка измерения y при наборе $(x_{i1}, \quad, x_{ir})$. По значениям $(y_i, x_{i1}, \ldots, x_{ir})$ требуется оценить коэффициенты $A_0, A_1, \quad, A_r$. Задачи такого типа называются *регрессионными*, они часто встречаются, например, в экономических исследованиях. Так, y может означать спрос на некоторый товар, который зависит от его цены (x_1), потребительского дохода (x_2) и цен на конкурирующие и дополняющие товары $(x_3, x_4, \ldots)$, и эта зависимость в первом приближении может быть регрессионного типа. Регрессионные задачи составляют содержание отдельного раздела математической статистики — *регрессионного анализа*. Эти задачи будут рассмотрены в гл. 6. •

Во всех предыдущих примерах речь шла об *оценивании* тех или иных параметров в соответствующих математических моделях по результатам наблюдений или измерений, т. е. на основе *эмпирических* (опытных) данных. Такого типа задачи, называемые общим термином «задачи оценки (или оценивания) неизвестных параметров», составляют один из двух основных разделов математической статистики. Добавим к сказанному еще следующее.

Иногда нас интересует не точное значение неизвестного параметра, например, параметра a в примере 2, а требуется указать такой интервал $a_1 \leqslant a \leqslant a_2$, в котором с высокой гарантией (в статистике говорят «с высокой вероятностью», т. е. с вероятностью, близкой к 1) находится точное значение параметра. Такой интервал (a_1, a_2), границы которого также определяются по соответствующим эмпирическим данным, называют *доверительным интервалом*. Такие задачи также относятся к задачам оценивания, но при этом говорят об *интервальном* (или *доверительном*) *оценивании*. Общая теория оценивания излагается в гл. 3.

Элементы теории вероятностей

Перейдем теперь к примерам задач другого типа, рассматриваемых в математической статистике. Но для этого нам предварительно необходимо хотя бы кратко напомнить некоторые элементы теории вероятностей. В качестве основного пособия по теории вероятностей мы рекомендуем современный учебник для технических вузов В. П. Чистякова «Курс теории вероятностей» [27].

В любой статистической задаче мы имеем дело с некоторым набором исходных данных, которые получены в результате некоторого эксперимента (опыта): «опыт — вот учитель жизни вечный» (Гёте [3]). Термин *эксперимент* употребляется в статистике для обозначения любого процесса получения данных, все возможные исходы которого могут быть указаны заранее и действительный исход которого является одним из возможных. Обозначим эти данные символом X (это может быть число или вектор некоторой размерности) и в дальнейшем будем называть их *выборкой* (синоним термина *статистические данные*). Реализации выборки X обозначаются соответствующей строчной буквой x, а совокупность всех возможных реализаций обозначается $\mathfrak{X} = \{x\}$ и называется *выборочным пространством*. В математической статистике рассматриваются такие эксперименты, исход которых случаен, так что выборка X является случайной величиной с некоторым распределением (функцией распределения) F_X на выборочном пространстве $\mathfrak{X}$.

Оговоримся сразу, что в дальнейшем мы предполагаем, что случайная величина X (или ее распределение F_X) является либо дискретной (и тогда выборочное пространство состоит из конечного или счетного числа точек), либо абсолютно непрерывной (тогда $\mathfrak{X}$ является евклидовым пространством соответствующей размеренности или его частью), эти два случая обычно и встречаются в приложениях. В первом случае распределение выборки задается вероятностями отдельных ее реализаций

$$f(x) = \mathbf{P}\{X = x\}, \quad x \in \mathfrak{X}, \quad \sum_{x \in \mathfrak{X}} f(x) = 1,$$

во втором — плотностью распределения $f(x)$, т. е. такой неотрицательной функцией на $\mathfrak{X}$, что для любого параллелепипеда $B \subset \mathfrak{X}$, т. е. множества вида $B = \{x \quad a_i \leqslant x_i \leqslant b_i,\ i = 1, \quad , r\}$, если $x = (x_1, \quad , x_r)$, выполняется соотношение

$$\mathbf{P}\{X \in B\} = \int\limits_B f(x)\,dx, \qquad \int\limits_{\mathfrak{X}} f(x)\,dx = 1$$

(эта формула сохранятся для любых квадрируемых или кубируемых множеств B). (Использование символа $\mathbf{P}$ для обозначения вероятности события является в настоящее время общепринятым, поскольку во многих языках начальной буквой слова «вероятность» является P — *probability* в английском,

[3] Гёте Иоганн Вольфганг (1749–1832) — немецкий философ, поэт, ученый, естествоиспытатель.

probabilité во французском, *probabilitá* в итальянском, *probabilidad* в испанском, *probabilitas* в латинском языке и т. д.). Обычно из контекста задачи бывает ясно, о каком случае — дискретном или абсолютно непрерывном — идет речь, поэтому мы для единообразия используем одно обозначение $f(x)$ как для вероятности $\mathbf{P}\{X = x\}$ в дискретном случае, так и для плотности распределения — в абсолютно непрерывном, и для краткости в обоих случаях будем называть функцию $f(x)$ *плотностью*. Если функция $f(x)$ из каких-то соображений задана, то мы будем говорить, что задана *вероятностная модель* эксперимента. В задачах теории вероятностей предполагается, что вероятностная модель изучаемого явления полностью определена, и теория вероятностей занимается разработкой методов нахождения вероятностей различных сложных событий (исходя из известных вероятностей более простых событий) в рамках данной вероятностной модели. По этим вероятностям мы можем в дальнейшем строить научно обоснованные прогнозы и тем самым рационально планировать свои действия (*если вы хотите насмешить Господа Бога, расскажите ему о своих планах на завтра*).

Статистическая модель

Математическая статистика выделяется из теории вероятностей в самостоятельную науку, хотя основные методы и приемы рассуждений в ней остаются теми же. Причиной этого является именно специфичность задач, решаемых ею, в известной мере они являются обратными к задачам теории вероятностей. Дело в том, что на практике при изучении конкретного эксперимента распределение F_X наблюдаемой случайной величины X редко бывает известно полностью. Часто можно априори (т. е. до опыта) утверждать лишь, что F_X является элементом некоторого заданного класса (семейства) распределений $\mathcal{F} = \{F\}$ на выборочном пространстве $\mathfrak{X}$, так что $F_X \in \mathcal{F}$ — семейству априори *допустимых* распределений выборки. Этот класс $\mathcal{F}$ может в одних случаях включать в себя все распределения, которые можно задать на $\mathfrak{X}$ (ситуация полной неопределенности), в других же случаях он представляет собой некоторое более узкое семейство распределений, заданное в той или иной явной форме (что может отражать наличие определенной априорной информации о характере изучаемого явления). Если задан класс $\mathcal{F}$ допустимых распределений выборки, то говорят, что задана *статистическая модель* эксперимента.

Итак, статистическая модель описывает такие ситуации, когда имеется та или иная неопределенность в задании распределения F_X выборки (данных) X, и задача математической статистики (в общем плане) состоит в том, чтобы уменьшать эту неопределенность (насколько это возможно), используя информацию, доставляемую наблюдаемым исходом эксперимента (наблюдаемой реализацией выборки X). Говорят, что математическая статистика — это наука о *статистических выводах*: она уточняет (выявляет) структуру статистических моделей по результатам проводимых наблюдений. Методы, с помощью которых это делается, называют *статистическими методами*.

Статистические гипотезы

Сказанного уже достаточно, чтобы описать задачи второго типа, решаемые математической статистикой, — это задачи *проверки статистических гипотез*. Любое

утверждение о виде или свойствах распределения F_X выборки X называется *статистической гипотезой* (обычно говорят просто *гипотезой*). В общем виде задача проверки гипотез ставится так: выделяется некоторое подмножество $\mathcal{F}_0 \subset \mathcal{F}$ и требуется по данным X проверить, справедливо ли утверждение $H_0 \quad F_X \in \mathcal{F}_0$ или же оно ложно. В этом случае H_0 называют *основной* (*нулевой*) гипотезой. Распределения $F \in \mathcal{F}_1 = \mathcal{F} \backslash \mathcal{F}_0$, т. е. распределения из дополнительного (к $\mathcal{F}_0$) подмножества допустимых распределений выборки, называются *альтернативными*, а утверждение $H_1 \quad F_X \in \mathcal{F}_1$ — *альтернативной гипотезой*. Тогда говорят, что задача состоит в проверке гипотезы H_0 против альтернативы H_1 (кратко: задача (H_0, H_1)) или внутри общей гипотезы $H \quad F_X \in \mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$.

Статистический критерий

Правило, согласно которому мы, наблюдая X, принимаем решение принять гипотезу H_0 как истинную либо отклонить ее как ложную (принять альтернативную гипотезу H_1), называется *статистическим критерием* (или просто *критерием*). Математическая статистика и занимается разработкой рациональных принципов конструирования таких правил (критериев) — это и есть второй тип статистических задач.

> *Теория — это жернова, а опытные данные — зерно, которое засыпается в эти жернова и перерабатывается ими.*
>
> **Кельвин** [4)]

Добавим к сказанному еще следующее. На практике часто встречаются ситуации, когда эксперимент состоит в проведении повторных независимых наблюдений над некоторой случайной величиной ξ. Если проводится n наблюдений, то выборка $X = (X_1, \ldots, X_n)$ представляет собой n независимых копий величины ξ, т. е. представляет собой n-мерный случайный вектор, компоненты которого независимы и распределены так же, как ξ (можно также сказать, что X_i — это результат i-го наблюдения над ξ). В этом случае функция распределения выборки имеет вид $F_X(\underline{x}) = F_\xi(x_1) \ldots F_\xi(x_n)$, $\underline{x} = (x_1, \ldots, x_n)$, т. е. определяется распределением F_ξ наблюдаемой случайной величины ξ, и говорят, что $X = (X_1, \ldots, X_n)$ есть *случайная выборка объема n из распределения ξ* (или из $\mathcal{L}(\xi)$, $\mathcal{L} \equiv$ law (закон)). Статистическая модель для повторных независимых наблюдений будет в дальнейшем обозначаться кратко в виде $\mathcal{F} = \{F_\xi\}$, т. е. указанием лишь класса допустимых распределений исходной случайной величины ξ (обычно индекс ξ у F_ξ опускается, если это не приводит к недоразумениям). Случайная выборка является математической моделью независимых измерений, проводимых в одинаковых условиях, и именно такие модели мы будем рассматривать в основном в по-

[4)] Кельвин (Томсон Уильям) (1824–1907) — английский физик, почетный член Петербургской академии наук (1872), предложил абсолютную шкалу температур (шкала Кельвина).

следующем, хотя в ряде задач предположения о независимости и одинаковой распределенности компонент X_i не выполняются — на это каждый раз будет обращаться внимание.

Теперь приведем для иллюстрации несколько примеров конкретных статистических гипотез, чтобы продемонстрировать их разнообразие.

Задачи проверки гипотез

Пример 4 (Гипотеза о виде распределения). Пусть произведено n независимых наблюдений над некоторой случайной величиной ξ с неизвестной функцией распределения $F_\xi(x)$. Гипотезой, подлежащей проверке, может быть утверждение типа H_0 $F_\xi(x) = F(x)$, где функция $F(x)$ полностью задана, либо типа H_0 $F_\xi(x) \in \mathcal{F}_0$, где $\mathcal{F}_0$ — заданное семейство функций распределения. При этом обычно семейство $\mathcal{F}_0$ задается в *параметрическом* виде: $\mathcal{F}_0 = \{F(x;\theta),\ \theta \in \Theta\}$, т. е. распределения из класса $\mathcal{F}_0$ заданы с точностью до значений некоторого параметра θ с областью возможных значений Θ. Говорят, что в этом случае известен тип распределения наблюдаемой случайной величины, а неизвестен только параметр, от которого зависит распределение. Параметр θ может быть как скалярным, так и векторным, область (множество) Θ его допустимых значений называется *параметрическим множеством*. Если множество Θ состоит из одной точки, то соответствующее распределение полностью известно, и в этом случае гипотеза H_0 называется *простой*, в противном случае гипотеза H_0 — *сложная*. Эта терминология общепринята в математической статистике, и мы ею будем ниже пользоваться.

Сложная гипотеза

Например, наблюдается неотрицательная целочисленная случайная величина ξ, скажем, число вызовов, поступивших на телефонную станцию за фиксированное время t, и требуется по результатам наблюдений в n непересекающихся интервалах времени (каждый длиной t) проверить гипотезу H_0 $\mathcal{L}(\xi) \in \Pi(\theta)$, где $\Pi(\theta)$ — семейство всех пуассоновских распределений (здесь $\theta > 0$ — любое положительное число, т. е. Θ $\{\theta$ $\theta > 0\}$). Другими словами, в данном случае речь идет о гипотезе, что число вызовов $\xi = \xi(t)$ имеет некоторое пуассоновское распределение (список-справочник основных вероятностных распределений приводится ниже, в гл. 1). •

В качестве еще одной из возможных конкретизаций сформулированной гипотезы рассмотрим ситуацию, описанную выше в примере 2. В теории измерений обычно считают, что общая ошибка измерения ε_i складывается из большого числа частных ошибок, каждая из которых невелика. На основании центральной предельной теоремы теории вероятностей можно предположить, что результаты измерений X_i имеют нормальное распределение. Такое предположение также является статистической гипотезой о виде распределения наблюдаемых случайных величин. Поскольку нормальное распределение определяется двумя параметрами: своими средним значением и дисперсией, а гипотеза не фиксирует их значения (они могут быть произвольными), то и в этом случае проверяемая гипотеза — сложная.

Простая гипотеза

Наконец, приведем конкретный пример простой гипотезы о виде распределения. Рассмотрим опыт, заключающийся в бросании игральной кости. Здесь наблюдаемая случайная величина ξ есть число выпавших очков; ее возможными значениями являются числа $1, 2, \ldots, 6$, и если кость симметрична, то каждая ее грань выпадает с одной и той же вероятностью $1/6$, т. е. $\mathbf{P}\{\xi = k\} = 1/6$, $k = 1, 2, \quad, 6$. Тем самым предположение (гипотеза) H_0 о симметричности кости однозначно фиксирует распределение наблюдаемой величины, следовательно, в данном случае H_0 — простая гипотеза о виде распределения. Спрашивается, как по результатам n бросаний кости построить критерий, подтверждающий или же опровергающий гипотезу о симметричности кости? Такой критерий в этой и аналогичной задачах, известный как *критерий* χ^2 (хи-квадрат), мы построим ниже, а сейчас приведем связанный с данной моделью один забавный эпизод, изложенный Ж. Бертраном (1822–1900) в предисловии к его курсу «Исчисление вероятностей» (Париж, 1899): «Однажды в Неаполе преподобный Галиани увидел человека из Базиликаты, который, встряхивая 3 игральные кости в чашке, держал пари, что выбросит 3 шестерки... Вы скажете, такая удача возможна. Однако человеку из Базиликаты это удалось во второй раз, и пари повторилось. Он клал кости назад в чашку 3, 4, 5 раз и каждый раз выбрасывал 3 шестерки. „Черт возьми, — вскричал преподобный, — кости налиты свинцом!“ И так оно и было».

Здесь в шутливой форме дан типичный пример статистического вывода: если бы кости были симметричны, то наблюдаемое событие (5 раз подряд выпали 3 шестерки) имело бы ничтожно малую вероятность $(1/6^3)^5 \approx 4{,}71 \cdot 10^{-11}$ и, поскольку в эксперименте наблюдалось такое событие, вполне логично сделать вывод, что априорная модель (симметричность костей) ложна.

Пример 5 (Гипотеза однородности). Одной из важных прикладных задач математической статистики является задача проверки однородности исходных статистических данных, т. е. гипотезы H_0 о том, что закон распределения наблюдений оставался неизменным в течение всего эксперимента. Формализация такой проблемы имеет следующий вид. Пусть имеются две независимые выборки $\boldsymbol{X} = (X_1, \quad, X_n)$ и $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$, описывающие один и тот же процесс, явление и т. д., но полученные, скажем, в разное время или в разных условиях. Требуется ответить на вопрос, являются ли эти выборки однородными, т. е. полученными из одного и того же распределения, или же закон распределения наблюдений от выборки к выборке менялся? Если они однородны, то их можно объединить и работать с одной большой выборкой объема $n+m$ (заключающей в себе бóльшую информацию, нежели каждая выборка в отдельности), получая в итоге более обоснованные (более точные) статистические выводы о законе распределения наблюдений. Такая задача может возникнуть, например, при контроле качества некоторой продукции, когда по контрольным выборкам из различных партий надо установить, не менялось ли ее качество от смены к смене или в результате изменения технологического процесса и т. д.: если менялось (выборки неоднородны), то

производство надо остановить и подкорректировать технологический процесс. Еще одна иллюстрация: поступающие в вуз абитуриенты разбиты на два потока. Как по результатам экзамена по одному и тому же предмету установить, однородны ли оба потока? В таких ситуациях требуется проверить *гипотезу однородности* H_0 $F_1(x) \equiv F_2(x)$, где $F_1(x)$ и $F_2(x)$ — функции распределения выборок X и Y соответственно (эти функции нам неизвестны или же принадлежат некоторому заданному параметрическому семейству). В общем случае можно рассматривать произвольное конечное число независимых выборок. В современной статистике разработано большое число различных критериев проверки гипотезы однородности, объединяемых общим термином *критерии однородности*, и с этими критериями и их особенностями мы познакомимся в гл. 4 (§ 4.3). •

Пример 6 (Гипотеза независимости). В эксперименте наблюдается двумерная случайная величина $\xi = (\xi_1, \xi_2)$ с неизвестной функцией распределения $F_\xi(x, y)$ и есть основания предполагать, что компоненты ξ_1 и ξ_2 независимы. В этом случае надо проверить *гипотезу независимости* H_0 $F_\xi(x, y) = F_{\xi_1}(x)F_{\xi_2}(y)$, где $F_{\xi_1}(x)$ и $F_{\xi_2}(y)$ — некоторые одномерные (иногда говорят *маргинальные*) функции распределения. Гипотезы такого вида часто возникают в прикладных исследованиях, когда результаты наблюдений классифицируются по двум признакам, — в этом случае H_0 есть предположение о независимости этих признаков. Так каждый сдающий вступительные экзамены в вуз абитуриент может быть классифицирован по двум признакам: $\xi_1 \in \{1$ (принят), 0 (не принят)$\}$ и $\xi_2 \in \{1$ (мужчина), 0 (женщина)$\}$, — в этом случае гипотеза H_0 о независимости этих признаков означает отсутствие дискриминации по половому признаку при приеме в вуз. Еще одна иллюстрация: проведена частичная вакцинация населения против, скажем, гриппа; здесь признак ξ_1 означает наличие (1), либо отсутствие (0) прививки для конкретного индивидуума, а признак ξ_2 — состояние его здоровья: здоров (1) — болен (0), гипотеза же независимости H_0 интерпретируется как неэффективность вакцины. Гипотеза независимости может иметь и параметрический вид, т. е. когда априори постулируется, что распределение $\mathcal{L}(\xi_1, \xi_2)$ принадлежит некоторому параметрическому семейству $\mathcal{F} = \{F(x, y; \theta), \theta \in \Theta\}$. Если, например, $\mathcal{F}$ является семейством всех двумерных нормальных распределений, то гипотеза независимости в данном случае эквивалентна гипотезе о некоррелированности признаков ξ_1 и ξ_2 (поскольку независимость и некоррелированность для нормальных случайных величин эквивалентны!). В общем случае можно рассматривать k-мерную ($k \geqslant 2$) случайную величину $\xi = (\xi_1, \ldots, \xi_k)$ и проверять гипотезу независимости ее компонент. *Критерии независимости* будут рассмотрены в § 4.4 книги. •

Пример 7 (Гипотеза случайности). В предыдущих примерах априори предполагалось (постулировалось), что исходные данные представляют собой случайную выборку из некоторого распределения, т. е. компоненты вектора данных $X = (X_1, \ldots, X_n)$ независимы и одинаково распределены. Как правило, это

предположение бывает оправдано, так как вытекает из самого характера задачи, и не подвергается сомнению. Но иногда это исходное предположение само нуждается в проверке, т. е. оно рассматривается как статистическая гипотеза H_0, называемая *гипотезой случайности*. Формализуется такая гипотеза следующим образом. Пусть $F_X(x_1, \quad , x_n)$ обозначает функцию распределения выборки $X = (X_1, \quad , X_n)$, тогда подлежащая проверке гипотеза означает утверждение $H_0 \quad F_X(x_1, \ldots, x_n) = F(x_1) \ldots F(x_n)$, где $F(x)$ — некоторая одномерная функция распределения (она может быть полностью задана, либо задано семейство, которому она принадлежит, либо никак не специфицируется). Типичным примером ситуации, когда возникает необходимость проверки гипотезы случайности является работа генератора (датчика) случайных чисел. Под случайными числами мы понимаем последовательность $\{U_1, U_2, \ldots\}$ независимых и равномерно распределенных на отрезке $[0, 1]$ случайных величин. Такие числа широко используются в различных областях: в статистике — для моделирования случайных выборок из различных распределений (метод статистического моделирования будет описан в § 1.3), в криптографии — при получении ключей для зашифрования информации, в численном анализе и т. д. В практических задачах последовательность $\{U_1, U_2, \ldots\}$ строят либо с использованием готовых таблиц случайных чисел, либо генерируют с помощью специальных датчиков, встроенных непосредственно в ЭВМ, либо получают программным способом по некоторому вспомогательному алгоритму (в последнем случае получаются так называемые *псевдослучайные* числа, т. е. «очень похожие» на случайные). Во всех случаях (особенно в последнем) требуется осуществлять контроль за «качеством» вырабатываемой последовательности $\{U_i\}$ (т. е. чтобы эти числа были практически неотличимы от независимых равномерно распределенных чисел), что в математическом плане сводится к проверке гипотезы случайности (см. п. 2 § 4.5). •

Псевдослучайные числа

Приведенные примеры не охватывают всех возникающих в приложениях задач проверки статистических гипотез. Отметим еще один важный класс гипотез, которые называются *параметрическими*, — это гипотезы об истинном значении неизвестного параметра, определяющего заданное параметрическое семейство распределений (см. выше пример 4). Если класс априори допустимых распределений наблюдаемой случайной величины ξ задан в параметрическом виде, т. е. $\mathcal{L}(\xi) \in \mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$, то в общем случае параметрическая гипотеза имеет вид $H_0 \quad \theta \in \Theta_0 \subset \Theta$, т. е. она определяется указанием конкретного подмножества значений параметра θ. Например, если $\mathcal{F}$ есть класс всех нормальных распределений со средним θ_1 и дисперсией θ_2^2 (здесь $\theta = (\theta_1, \theta_2)$), то утверждение $H_0 \quad \theta_1 = \theta_{10}, \theta_2 = \theta_{20}$, где θ_{10} и θ_{20} — заданные числа, есть простая гипотеза о нормальном распределении наблюдений. Гипотеза же, выраженная равенством $\theta_1 = \theta_{10}$ и оставляющая значение дисперсии θ_2^2 неопределенным, — сложная. Общая теория проверки параметрических гипотез будет обсуждаться в гл. 5.

Параметрические гипотезы



О современной статистике

Итак, математическая статистика решает две основные проблемы:

1) оценивание различных параметров (характеристик) распределений наблюдаемых случайных величин и
2) проверка различных гипотез о свойствах этих величин (их распределений)

и тем самым позволяет в итоге подобрать подходящую вероятностную модель изучаемого эксперимента.

Помимо этого, современная статистическая практика связана с методами сбора, обработки и интерпретации данных, планированием эффективных и информативных экспериментов, принятием решений в условиях неопределенности и прогнозированием. Некоторые классы специфических задач выделяются при этом в отдельные разделы (теории) — так появляются термины «регрессионный анализ», «дисперсионный анализ», «дискриминантный анализ», «факторный анализ», «последовательный анализ», «компонентный анализ» и др. С некоторыми из них мы также познакомимся в дальнейшем.

Наконец, подчеркнем следующее. Практическая реализация современных статистических методов невозможна без интенсивного использования компьютерной техники. Именно компьютерная реализация статистических алгоритмов обработки информации делает их сегодня одним из наиболее эффективных инструментов принятия рациональных решений в самых разных областях человеческой деятельности (экономике, финансах, страховом деле, медицине, социологи и т.д.). Поэтому современный специалист-статистик должен как хорошо знать основы и методы теории статистического вывода, так и владеть навыками их практической реализации на компьютерах. Целям подготовки именно таких специалистов и служит настоящее пособие.

Итак, что же такое статистика? Некоторые понимают и исподьзуют ее, как пьяный фонарный столб, т. е. не для освещения, а как опору, чтобы не упасть. Широко известен, например, афоризм, что существует три вида лжи: ложь, наглая ложь и статистика. Но адресован он тем влиятельным лицам, которые в угоду своим политическим, финансовым и иным целям манипулируют статистическими данными, как им заблагорассудится, и не имеет ничего общего с наукой «Математическая статистика». На самом же деле, говоря словами английского статистика В. Томпсона, *статистика — это математическая теория того, как узнать нечто о мире через опыт.*

«Давайте присядем на это бревно у дороги, — говорю я, — и забудем бессердечность и сквернословие этих поэтов. Настоящую красоту нужно искать в великолепных рядах установленных фактов и общепринятых правил. В этом самом бревне, на котором мы сидим, миссис Сэмпсон, — говорю я, — скрыта статистика более прекрасная, чем любая поэма. Кольца показывают, что ему было шестьдесят лет. На глубине двух тысяч футов оно за три тысячи лет превратилось бы в уголь. Самая глубокая в мире угольная шахта находится в Киллингворте, близ Ньюкастла. В ящик длиной четыре фута, шириной три фута и высотой два фута

восемь дюймов войдет тонна угля. Если артерия порезана, сожмите ее выше раны. В ноге человека — тридцать костей. Лондонский Тауэр горел в 1841 году». «Продолжайте, мистер Пратт, — говорит миссис Сэмпсон. — Эти мысли так оригинальны и приятны. Я думаю, ничего нет прекраснее этой статистики».

О. Генри. Справочник Гименея

«Вы еще не рассказали мне, — заметила леди Наттэл, — чем занимается ваш жених».

«Он статистик», — ответила Лэймия с неприятным ощущением обороняющегося.

Леди Наттэл была, очевидно, застигнута врасплох. Ей не приходило в голову, что статистики вступают в нормальные общественные отношения. Этот род людей, как она предполагала, увековечен на некоторый второстепенный образ действий, наподобие мулов.

«Но, тетя Сара, это очень интересная профессия», — сказала с жаром Лэймия.

«Я не сомневаюсь в этом, — ответила ее тетя, которая, очевидно, очень сомневалась. — Выразить что-нибудь значительное в одних лишь цифрах, понятно, настолько невозможно, что должен существовать бесконечный простор для хорошо оплаченных советов относительно того, как это сделать. Но не думаете ли вы, что жизнь со статистиком была бы до некоторой степени, скажем, скучной?»

Лэймия молчала. Ей не хотелось говорить о поразительной глубине эмоциональных возможностей, которую она открыла под цифровой внешностью занятий Эдварда.

«Дело не в цифрах, — сказала она наконец, — а в том, что вы с ними делаете».

К. Мэндервил. Крушение Лэймии Гёрдлнек

§ 2. Краткий исторический очерк развития математической статистики

Первыми крупными работами, относящимися к математической статистике, были исследования Я. Бернулли[5] А. Муавра[6] и П. Лапласа[7] К. Гаусс[8] разработал теорию ошибок наблюдений и, в частности, обосновал один из ее основных принципов — метод наименьших квадратов. Научное обоснование закономерностей случайного рассеивания связано с именами русских мате-

[5] Бернулли Якоб (1700–1782) — швейцарский физик и математик.

[6] Муавр Абрахам де (1667–1754) — французский математик.

[7] Лаплас Пьер Симон (1749–1827) — крупнейший французский математик, физик, астроном, почетный член Петербургской АН (1802).

[8] Гаусс Карл Фридрих (1777–1855) — великий немецкий ученый — математик, физик, астроном, почетный член Петербургской АН (1824).

матиков П. Л. Чебышева [9], А. А. Маркова [10] и А. М. Ляпунова [11], значительно углубивших результаты известных к тому времени классических исследований.

Зарождение математической статистики как отдельной математической дисциплины относится к 20–30-м гг. XX в.

Ряд важнейших современных понятий и методов, оказавших большое влияние на развитие современной теории математической статистики, предложил английский математик и статистик Р. Фишер [12]: метод максимального правдоподобия, дисперсионный анализ, понятия состоятельности, достаточности, эффективности и др. Систематическое развитие теории проверки статистических гипотез началось с работ английского математика-статистика К. Пирсона [13] по критерию хи-квадрат; он также внес большой вклад в разработку теории корреляции и ее применение к проблемам наследственности и эволюции. Основной вклад в построение теории проверки статистических гипотез внесли Ю. Нейман [14] и Э. Пирсон [15], сын К. Пирсона, которые ввели понятия ошибок первого и второго рода и показали общность и значение метода отношения правдоподобия для построения критериев. Общая теория статистических решающих функций, охватывающая как теорию проверки статистических гипотез, так и теорию оценивания, создана венгерским статистиком А. Вальдом [16]

Большую роль в развитии математической статистики сыграли работы Г. Крамера [17], Э. Лемана [18], К. Рао [19], М. Кендалла [20], А. Стьюарта [21], С. Уилкса [22], а также отечественных ученых А. Н. Колмогорова [23], Б. В. Гнеденко [24],

---

[9] Чебышев Пафнутий Львович (1821–1894) — выдающийся русский математик и механик.

[10] Марков Андрей Андреевич (1856–1922) — выдающийся русский математик, академик.

[11] Ляпунов Александр Михайлович (1857–1918) — выдающийся русский математик и механик.

[12] Фишер Рональд Эйлмер (1890–1962) — крупнейший английский статистик, автор и исследователь многих современных понятий в математической статистике.

[13] Пирсон Карл (1857–1936) — английский математик-статистик, биолог и философ.

[14] Нейман Юрий (1894–1981) — американский статистик.

[15] Пирсон Эгон Шарп (1895–1980) — английский статистик.

[16] Вальд Абрахам (1902–1950) — американский статистик венгерского происхождения.

[17] Крамер Гаральд (1893–1985) — шведский статистик.

[18] Леман Эрих (р. 1917) — американский статистик.

[19] Рао Кальямпуди Радхакришна (р. 1920) — крупнейший индийский математик, статистик.

[20] Кендалл Морис Джордж (1907–1983) — английский статистик.

[21] Стьюарт Алан — английский статистик.

[22] Уилкс Самюэль Стэнли (1906–1964) — американский статистик.

[23] Колмогоров Андрей Николаевич (1903–1987) — великий советский ученый, один из крупнейших математиков XX в., академик АН СССР, создатель современной аксиоматики теории вероятностей.

[24] Гнеденко Борис Владимирович (1912–1995) — академик АН УССР, выдающийся математик, специалист по предельным теоремам теории вероятностей, истории математики, популяризатор науки, замечательный педагог.

Ю. В. Линника[25] Е. Е. Слуцкого[26] Н. В. Смирнова[27] В. И. Романовского[28] Л. Н. Большева[29], Ю. В. Прохорова[30], А. А. Боровкова[31], Б. А. Севастьянова[32], А. Н. Ширяева[33] и др.

Начиная с 50-х гг. XX в. в связи с появлением компьютеров лицо статистики изменилось. Если раньше отсутствие вычислительных возможностей вынуждало статистиков использовать слишком упрощенные модели, даже если эти модели были не всегда реалистичными, то теперь любое статистическое решение, которое можно просчитать на компьютере за сравнительно небольшой отрезок времени, стало «доступным». Таким образом, сложные и трудоемкие статистические методы вошли в повседневную практику. В это же время статистика, в определенном плане, стала эмпирической наукой: за несколько минут компьютеры могут порождать миллионы данных, используя которые можно «проверять» новейшие методы.

Математическая статистика — это непрерывно и интенсивно развивающаяся наука. В последние годы появилось большое количество глубоких результатов, охватывающих, по существу, все основные направления современной теории математической статистики. Достойный вклад в развитие новых методов этой теории вносит отечественная математическая школа.

В наши дни математическая статистика переживает бурный этап постоянного расширения областей практического применения ее методов, благодаря резкому росту арсенала средств, используемых исследователем. Внедрение в статистическую практику вычислительной техники позволило при решении задач математической статистики использовать методы статистического моделирования (Монте-Карло). Широкому распространению методов математической статистики способствовало создание больших универсальных пакетов прикладных статистических программ, обеспечивающих быстрый и эффективный компьютерный анализ больших массивов статистической информа-

---

[25] Линник Юрий Владимирович (1915–1972) — советский математик в области теории вероятности, теории чисел, статистики, академик АН СССР.

[26] Слуцкий Евгений Евгеньевич (1880–1948) — выдающийся российский и советский математик, статистик и экономист.

[27] Смирнов Николай Васильевич (1900–1966) — член-корреспондент АН СССР, крупнейший специалист по предельным теоремам теории вероятностей и математической статистики.

[28] Романовский Всеволод Иванович (1879–1954) — русский, советский, узбекский математик, академик АН Узбекской ССР.

[29] Большев Логин Николаевич (1922–1978) — советский математик, член-корреспондент АН СССР.

[30] Прохоров Юрий Васильевич (р. 1929) — советский математик, академик АН СССР. Основные труды по теории вероятностей, особенно по асимптотическим методам.

[31] Боровков Александр Алексеевич (р. 1931) — академик АН СССР, специалист в области теории вероятностей и математической статистики.

[32] Севастьянов Борис Александрович (р. 1923) — советский математик, член-корреспондент АН СССР.

[33] Ширяев Альберт Николаевич (р. 1934) — специалист в области теории вероятностей и математической статистики, член-корреспондент РАН.

ции. Именно в направлении соединения статистической теории и практики ее компьютерной реализации заключается существо современного этапа развития методов статистического анализа данных. Ибо «Бесполезно вычислять результат в такой форме, в какой потребитель не готов его использовать» (Хэмминг).

Методы математической статистики, соединенные с возможностями современной компьютерной техники, широко и эффективно используются в приложениях, являясь основой статистических методов исследования и контроля массового производства, статистических методов в области физики, биологии, медицины, социологии, метеорологии; их также применяют при обработке результатов статистического моделирования в различных задачах науки и техники.

В настоящее время статистические модели и методы играют все большую роль при решении экономических, финансовых, социальных проблем, а также различных проблем связи, телекоммуникации, защиты информации, криптографии и таких долго считавшихся далекими от математики наук, как психология, лингвистика, криминалистика и т. д. Такие области науки, как гидрология, климатология, звездная астрономия, антропология, также, оказывается, не могут обойтись без методов математической статистики.

Без методов этой науки сегодня невозможна разработка научно обоснованных прогнозов развития общества, включая прогнозы в политике, экономике, социологии.

В заключение, для подтверждения тезиса о том, что *история науки так же важна, как и сама наука*, приведем слова выдающегося немецкого математика, физика и философа, одного из создателей дифференциального и интегрального исчисления Г. В. Лейбница [34]. «Весьма полезно познать истинное происхождение замечательных открытий, особенно таких, которые были сделаны не случайно, а силою мысли. Это приносит пользу не только тем, что воздает каждому свое и побуждает других добиваться таких же похвал, сколько тем, что познание метода на выдающихся примерах ведет к развитию искусства открытия».

Все постоянство в зыбкости одной,
Все смутно пред глазами очевидца,
Я сложность вижу в истине простой,

Наукой правит случай — не закон...

Ф. Вийон [35]. Баллада поэтического состязания в Блуа (перевод Ю. А. Кожевникова)

[34] Лейбниц Готфрид Вильгельм фон (1646–1716) — выдающийся немецкий математик и философ.

[35] Вийон Франсуа (1431 – после 1463) — французский поэт и вор.

Всему начало — случай.

Старинная поговорка

...Помни, что случай плешив [36)]

Б. Грасиан [37)]**. Критикон**

[36)] В античном изобразительном искусстве представляли аллегорическую фигуру Случая в виде стоящего на колесе плешивого человека с одной лишь прядью волос, за которую надо его схватить, пока не умчался.

[37)] Грасиан Бальтасар (1601–1658) — испанский писатель.

Глава 1

(Вспомогательная, но очень важная!)

Основные распределения и их моделирование

Математические модели — объект и орудие труда математиков.

К. А. Рыбников [1)]

Введение

Выше мы определили статистическую модель и ее отношение к вероятностной модели: в обоих случаях модель задается плотностью распределения $f(x)$ наблюдаемой случайной величины ξ, а различие заключается в том, что́ мы априори предполагаем известным об этой плотности. В вероятностных задачах эта плотность полностью известна, в статистических же — она задается (предполагается) с той или иной степенью неопределенности, и статистик «борется» с этой неопределенностью. Часто мы предполагаем известным вид плотности $f(x)$, а неопределенность выражается через те или иные параметры, от которых зависит функция $f(x)$ (выше такие модели названы *параметрическими*). В теории вероятностей наиболее часто встречающиеся законы распределения (модели) имеют общепринятое наименование и обозначение — этот «язык» переносится и на соответствующие статистические модели. Проиллюстрируем это на одном конкретном примере.

Нормальная модель

Пусть из каких-то соображений (например, основываясь на центральной предельной теореме теории вероятностей) постулируется, что $\mathcal{L}(\xi)$ — нормальное или гауссовское распределение. Если обозначить через μ (мю) и σ^2 (сигма-квадрат) среднее и дисперсию нормального закона, то сказанное в символьном виде записывается так: $\mathcal{L}(\xi) = \mathcal{N}(\mu, \sigma^2)$ ($\mathcal{N}$ большое круглое является общепринятым обозначением нормального закона распределения). В терминах плотности f эта удобная лаконичная запись «расшифровывается» так:

$$f(x) = f(x|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty \qquad (1)$$

[1)] Рыбников Константин Алексеевич (1913–2004) — советский и российский математик, историк математики.

(здесь $-\infty<\mu<\infty$, $\sigma>0$ и использовано стандартное обозначение $\exp\{t\}=e^t$). В вероятностных задачах, связанных с нормальным распределением, — в этом случае говорят о *нормальной модели* — параметры μ и σ^2 считаются известными, и тем самым нормальная плотность $f(x|\mu,\sigma^2)$ полностью определена. Если же один из этих параметров (или оба) априори неизвестен, то мы имеем дело со *статистической нормальной моделью*. Неизвестный параметр модели принято обозначать символом θ (тэта), а область его возможных значений Θ (тэта большое), таким образом, в рассматриваемом случае мы имеем три варианта нормальной статистической модели:

1) $\mathcal{N}(\theta,\sigma^2)$, $\Theta=\{\theta\ \ -\infty<\theta<\infty\}$, — известна дисперсия и неизвестно среднее;
2) $\mathcal{N}(\mu,\theta^2)$, $\Theta=\{\theta\ \ 0<\theta<\infty\}$, — известно среднее и неизвестна дисперсия;
3) $\mathcal{N}(\theta_1,\theta_2^2)$, $\Theta=\{\theta=(\theta_1,\theta_2)\ \ -\infty<\theta_1<\infty,\ 0<\theta_2<\infty\}$, — оба параметра неизвестны.

Поскольку в дальнейшем мы будем иметь дело с конкретными моделями, которые широко используются в практических приложениях, мы сначала приведем для напоминания и удобства изложения и последующих ссылок список-справочник основных вероятностных распределений, сопровождая их некоторыми относящимися к теме комментариями. Это очень важный материал, поскольку он является базой для решения всех примеров и последующих задач.

§ 1.1. Основные дискретные модели математической статистики

Любая дорога начинается с первого шага.

В этом параграфе мы дадим определения и опишем необходимые нам в последующем основные свойства наиболее известных дискретных вероятностных распределений (как одномерных, так и многомерных), а также проиллюстрируем их использование в вероятностных и статистических задачах. Замечательным фактом является то, что существует несколько распределений большой общности, встречающихся в самых разнообразных задачах теории вероятностей и математической статистики. Прежде всего это биномиальное распределение, распределение Пуассона и нормальное распределение — с первого мы и начнем.

1. Схема Бернулли и биномиальное распределение

— Мадам, какова вероятность утром на улице встретить динозавра?
— Одна вторая: или встречу, или нет.

Говорят, что случайная величина ξ имеет *распределение Бернулли* с параметром p $(0<p<1)$, если она принимает лишь два значения, обозначаемые обычно

0 и 1, и при этом

$$\mathbf{P}\{\xi = 1\} = p = 1 - \mathbf{P}\{\xi = 0\}.$$

В терминах плотности $f(x)$ это можно записать в виде

$$f(x) = f(x|p) = p^x q^{1-x}, \quad x = 0, 1, \quad q = 1 - p. \tag{1}$$

Бернуллиевская модель

В символьных обозначениях сказанное выражается кратко так: $\mathcal{L}(\xi) = \mathrm{Bi}(1, p)$ и называется *бернуллиевской моделью*. Фундаментальная роль этой модели в теории вероятностей и математической статистике ясна: она является подходящей математической моделью для любого эксперимента с двумя исходами («успех»-«неуспех»), т. е. простейшего статистического эксперимента. Среднее и дисперсия такой случайной величины есть

$$\mathbf{E}\xi = p \quad \text{и} \quad \mathbf{D}\xi = pq \tag{2}$$

(символ $\mathbf{E}$ — от английского *expectation* (математическое ожидание ≡ среднее значение), $\mathbf{D}$ — от *dispersion* (дисперсия)). Если все случайные величины последовательности $\{X_1, X_2, \ldots\}$ (конечной или бесконечной) независимы и имеют одно и то же распределение $\mathrm{Bi}(1, p)$, то мы имеем *последовательность испытаний Бернулли* (или, кратко, *бернуллиевскую последовательность*), называемую так по имени Якоба Бернулли (1654–1705) — выдающегося швейцарского математика, впервые изучавшего такую схему (его основополагающие труды в области теории вероятностей изложены в посмертно изданном сочинении «Искусство предположений» (1713)).

Биномиальная модель

Пусть $\{X_1, \ldots, X_n\}$ — бернуллиевская последовательность с параметром p. Тогда сумма $X = X_1 + \ldots + X_n$ имеет *биномиальное распределение с параметрами n и p*, что кратко записывается в виде:

$$\mathcal{L}(x) = \mathrm{Bi}(n, p).$$

Эта случайная величина принимает, очевидно, лишь значения $0, 1, \ldots, n$ и при этом

$$\mathbf{P}\{X = x\} = f(x|n, p) = C_n^x p^x q^{n-x} \quad x = 0, 1, \ldots, n. \tag{3}$$

Термин «биномиальное распределение» связан с тем, что вероятности (3) являются членами известного «бинома Ньютона»:

$$1 = (p + q)^n = \sum_{x=0}^{n} C_n^x p^x q^{n-x}$$

Таким образом, *биномиальная модель* $\mathrm{Bi}(n, p)$ описывает распределение числа «успехов» в n испытаниях Бернулли с неизменной вероятностью «успеха» p. Здесь

$$\mathbf{E}X = np \quad \text{и} \quad \mathbf{D}X = npq. \tag{4}$$

Свойство воспроизводимости

Полезно знать также следующее свойство биномиального распределения: если случайные величины $X_1, \quad , X_k$ [redacted] независимы и

$$\mathcal{L}(X_j) = \mathrm{Bi}(n_j, p), \quad j = 1, \quad , k,$$

то

$$\mathcal{L}(X_1 + \quad + X_k) = \mathrm{Bi}(n_1 + \quad + n_k, p).$$

Это свойство биномиального распределения $\mathrm{Bi}(n, p)$ называется *воспроизводимостью* по параметру n.

Если параметр p нам неизвестен (а так на практике чаще всего и бывает!), то мы имеем *биномиальную статистическую модель*

$$\mathrm{Bi}(n, \theta), \quad \Theta = \{\theta \quad 0 < \theta < 1\}$$

(при $n = 1$ — *бернуллиевскую статистическую модель* $\mathrm{Bi}(1, \theta)$).

Примеры биномиольной модели

Пример 1. Рассмотрим ситуацию, описанную в примере 1 Введения. Если в данном эксперименте реализуется *схема выборо с возвращением*, и если условия эксперимента обеспечивают независимость извлечения каждого очередного шара от результатов предыдущих извлечений (практически это достигается тщательным перемешиванием содержимого урны перед каждым очередным извлечением), то мы имеем бернуллиевскую модель $\mathrm{Bi}(1, \theta)$, где под «успехом» понимается извлечение белого шара, для которой надо оценить параметр $\theta = a_1/a$ по результатам n испытаний. В данном случае наблюдаемая случайная величина X есть число белых шаров в выборке из n шаров, которая имеет биномиальное распределение $\mathrm{Bi}(n, \theta)$. •

Пример 2. Рассмотрим подробно описанный в примере 4 Введения эпизод с преподобным Галиани. Здесь мы имеем эксперимент, состоящий в пятикратном бросании трех игральных костей. Интересующая нас величина, характеризующая один опыт, — число выпавших шестерок на трех костях. Обозначим ее ξ. В свою очередь, она равна сумме трех бернуллиевских величин ξ_1, ξ_2 и ξ_3, где $\xi_i \quad 1$, если на i-й кости выпала шестерка, и 0 — в противном случае ($i = 1, 2, 3$). Если исходить из того, что кости одинаковы (это обычно не подвергается сомнению, хотя, на самом деле, — это тоже гипотеза), то случайные величины ξ_i, $i = 1, 2, 3$, имеют одно и то же распределение Бернулли $\mathrm{Bi}(1, \theta)$, и они, по условию опыта, независимы. Тем самым их сумма $\xi = \xi_1 + \xi_2 + \xi_3$ имеет распределение $\mathrm{Bi}(3, \theta)$, и над этой случайной величиной произведено пять независимых испытаний, т. е. мы имеем случайную выборку $\boldsymbol{X} = (X_1, \quad , X_5)$ из распределения $\mathrm{Bi}(3, \theta)$. Таким образом, мы построили биномиальную модель $\mathrm{Bi}(3, \theta)$ для описания нашего эксперимента, и в рамках этой модели требуется проверить гипотезу симметричности костей $H_0 \quad \theta = 1/6$ (т. е. гипотезу о том, что каждая грань выпадает с одинаковой вероятностью $1/6$). Как можно формализовать рассуждения преподобного? Общее число выпавших шестерок $X = X_1 + \ldots + X_5$ имеет биномиальное

распределение $\mathrm{Bi}(15, \theta)$, и по формуле (3)

$$\mathbf{P}\{X = 15\} = f(15|15, \theta) = \theta^{15}$$

Следовательно, при справедливости гипотезы H_0 вероятность наблюдаемого события, т. е. $X = 15$, равна 6^{-15} что ничтожно мало (!). Наверное, любой разумный человек на месте преподобного усомнился бы в истинности гипотезы H_0 в данном случае (при таком наблюдаемом исходе эксперимента). •

Замечание. Обратим внимание на следующее. В данном случае, если отбросить несущественные для математики эмоциональные нюансы (пари), а оставить лишь физическую суть эксперимента (в его ходе было брошено 3×5 15 костей), то можно сразу говорить о случайной выборке объема $n = 15$, из распределения Бернулли $\mathrm{Bi}(1, \theta)$ $X = (X_1, \ldots, X_{15})$, для которой исследуется случайная величина $X = X_1 +$ $+ X_{15}$ — общее число выпавших шестерок. Конечно, снова $\mathcal{L}(X) = \mathrm{Bi}(15, \theta)$.

2. Отрицательное биномиальное распределение

С бесконечной последовательностью испытаний Бернулли $\{X_1, X_2, \ldots\}$ связано еще одно важное дискретное распределение, которое обозначается $\overline{\mathrm{Bi}}(r, p)$ и называется *отрицательным биномиальным распределением с параметрами r и p* (здесь r — натуральное число). Это есть распределение числа «успехов» (1), предшествующих r-му «неуспеху» (0), и оно задается вероятностями

$$f(x|r, p) = C^x_{r+x-1} p^x q^r \qquad x = 0, 1, 2, \ldots \tag{5}$$

Заметим, что выражение для $f(x|r, p)$ совпадает с x-м членом разложения функции $q^r(1-p)^{-r}$ в ряд по степеням p, т. е. отрицательного бинома (отсюда и название):

$$1 = q^r(1-p)^{-r} = q^r \sum_{x=0}^{\infty} C^x_{r+x-1} p^x$$

Геометрическая модель

В частном случае $r = 1$ распределение $\overline{\mathrm{Bi}}(1, p)$ называется *геометрическим*: это есть распределение числа единиц, предшествующих первому нулю в бернуллиевской последовательности. Формула (5) в этом случае принимает вид

$$f(x|1, p) = p^x q, \quad x = 0, 1, 2, \ldots \tag{6}$$

Если случайная величина ξ имеет распределение $\overline{\mathrm{Bi}}(r, p)$, то

$$\mathbf{E}\xi = \frac{rp}{q} \quad \text{и} \quad \mathbf{D}\xi = \frac{rp}{q^2}. \tag{7}$$

Свойство воспроизводимости

Полезно знать также *свойство воспроизводимости* распределения $\overline{\mathrm{Bi}}(r, p)$ по параметру r: если случайные величины $X_1, \ldots, X_k$ независимы и

$$\mathcal{L}(X_j) = \overline{\mathrm{Bi}}(r_j, p), \quad j = 1, \ldots, k,$$

то

$$\mathcal{L}(X_1 + \quad + X_k) = \overline{\mathrm{Bi}}(r_1 + \quad + r_k, p).$$

Отсюда следует, что случайная величина с распределением $\overline{\mathrm{Bi}}(r, p)$ может быть реализована (построена) как сумма r независимых случайных величин с одинаковым геометрическим распределением $\overline{\mathrm{Bi}}(1, p)$, — этот факт используется для моделирования случайных величин с отрицательным биномиальным распределением (см. далее § 1.3).

Если параметр p неизвестен, то имеем *отрицательную биномиальную статистическую модель* $\overline{\mathrm{Bi}}(r, \theta)$, $\Theta = \{\theta \quad 0 < \theta < 1\}$ (при $r = 1$ — *геометрическую статистическую модель* $\overline{\mathrm{Bi}}(1, \theta)$). Эти модели часто используются, например, при разработке математических методов контроля качества промышленной продукции.

Замечание. В формуле (5) биномиальный коэффициент

$$C_{r+x-1}^{x} = \frac{r(r+1)\ldots(r+x-1)}{x!}$$

определен при любом действительном r, а при $r > 0$ этот коэффициент положителен, поэтому эта формула определяет распределение вероятностей при всех $r > 0$ и $0 < p < 1$, а не только при натуральных r. При натуральных r распределение $\overline{\mathrm{Bi}}(r, p)$ называют также *распределением Паскаля*[2)]

Распределение Паскаля

3. Распределение Пуассона

Случайная величина ξ имеет *распределение Пуассона*[3)] *с параметром* λ ($\lambda > 0$), если

$$\mathbf{P}\{\xi = x\} = f(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \tag{8}$$

что кратко записывается в виде $\mathcal{L}(\xi) = \Pi(\lambda)$; при этом $\lambda = \mathbf{E}\xi = \mathbf{D}\xi$.

Пуассоновская модель $\Pi(\lambda)$ обычно описывает *схему редких событий*: при некоторых предположениях о характере процесса появления случайных событий число событий, происшедших за фиксированный промежуток времени или в фиксированной области пространства, часто подчиняется пуассоновскому распределению. Примерами могут служить число частиц радиактивного распада, зарегистрированных счетчиком в течение некоторого времени t, число вызовов, поступивших на телефонную станцию за время t, число дефектов в куске ткани или в ленте фиксированной длины, число изюминок в кексе и т. д. Наконец, распределение Пуассона дает хорошую аппроксимацию биномиального распределения для больших значений n и малых значений p:

[2)] Паскаль Блез (1623–1662) — французский математик, механик, физик и философ, один из основателей теории вероятностей.

[3)] Пуассон Симеон Дени (1781–1840) — французский математик, механик и физик, иностранный почетный член Петербургской АН (с 1826).

$\mathrm{Bi}(n, p) \approx \Pi(np)$, если np невелико. Это свойство позволяет значительно упростить вычисления в биномиальной модели при указанных условиях.

Свойство воспроизводимости

Распределение Пуассона *воспроизводимо* по своему параметру, т. е. если $X_1, \quad , X_k$ — независимые случайные величины и $\mathcal{L}(X_j) = \Pi(\lambda_j)$, $j = 1, \quad , k$, то

$$\mathcal{L}(X_1 + \quad + X_k) = \Pi(\lambda_1 + \quad + \lambda_k).$$

Если параметр λ неизвестен, то имеем *пуассоновскую статистическую модель* $\Pi(\theta)$, $\Theta = \{\theta \quad \theta > 0\}$



Понятие распределения Пуассона впервые появилось в книге французского ученого Симеона Дениса Пуассона (1781–1840) «Исследования о вероятности судебных приговоров по уголовным и гражданским делам», опубликованной в 1837 г. Пуассон обнаружил, что если в n испытаниях Бернулли с вероятностью «успеха» p (см. п. 1) делать p все меньше и меньше и в то же время увеличивать n так, что произведение $np = \lambda$ остается постоянным, то вероятность (3) того, что будет ровно x успехов, стремится к $f(x|\lambda)$ (см. (8)). Следовательно, распределение Пуассона (8) является приближением для биномиального распределения (3). Возможность широкого применения и большая важность пуассоновского распределения не были поняты современниками Пуассона, более того, о нем почти полностью забыли. Однако после 1894 г. это распределение использовали при изучении странного феномена. За 20 лет между 1875 и 1894 гг. в 14 различных кавалерийских корпусах германской армии была собрана статистика трагических случаев, когда солдат был убит ударом копыта. Согласно 280 (= 20 лет × 14 корпусов) наблюдениям, 196 солдат погибли таким образом, т. е. в среднем за год $\lambda = 196/280 = 0{,}7$ Если бы число трагических исходов подчинялось распределению Пуассона с параметром $\lambda = 0{,}7$, то можно было бы ожидать, что при 280 наблюдениях в 139 случаях смертей нет, в 97 случаях — 1 смерть, в 34 случаях — 2 смерти и т. д. А что показала статистика? В действительности данные были соответственно 140, 91, 32 и т. д.; практика и теория оказались в столь хорошем согласии, что вряд ли можно было ожидать большего. Это сравнение появилось в 1897 г. в монографии Л. Борткевича «Закон малых чисел». Пуассоновское распределение стало широко применяться лишь в XX в. (по книге [23, с. 34]).

Распределение Эрланга

С распределением Пуассона связано еще одно интересное дискретное распределение — *распределение Эрланга*[4], задаваемое вероятностями

$$e_K = c\frac{\lambda^k}{k!}, \quad k = 0, 1, \quad , N, \quad c = \left(\sum_{k=0}^{N} \frac{\lambda^k}{k!}\right)$$

т. е. это — усеченное сверху в точке N пуассоновское распределение. В 1906 г. Эрланг установил, что в некоторых случаях лучшее приближение (для биномиального распределения) можно получить с использованием именно такого распределения.

[4] Эрланг Агнер Крарул (1878–1929) — датский математик.

Пример 3 (Дни рождения [25, с. 171–172]). Какова вероятность p_x того, что в группе из 500 человек ровно x родились 1 января? Если эти 500 человек выбраны случайно, то можно применить схему из 500 испытаний Бернулли с вероятностью «успеха» $p = 1/365$. Для пуассоновского приближения положим $\lambda = 500/365 = 1{,}3699$ Теоретические биномиальные вероятности (3) и их пуассоновские приближения (8) приведены в следующей таблице.

x	0	1	2	3	4	5	6
$\mathbf{P}_x = f(x\|500, 1/365)$	0,253	0,3484	0,2388	0,1089	0,0372	0,0101	0,0023
$\mathbf{P}_x \approx f(x\|500/365)$	0,2541	0,3481	0,2385	0,1089	0,0373	0,0102	0,0023

•

4. Гипергеометрическое распределение

Случайная величина ξ имеет *гипергеометрическое распределение с параметрами* a_1, a_2 и n, где a_1, a_2 и n — натуральные числа, причем

$$n \leqslant a = a_1 + a_2 \quad \mathcal{L}(\xi) = H(a_1, a_2, n),$$

если

$$\mathbf{P}\{\xi = x\} = f(x|a_1, a_2, n) = \frac{C_{a_1}^x C_{a_2}^{n-x}}{C_a^n}, \quad x = 0, 1, \quad , n. \tag{9}$$

Такое распределение возникает в следующей схеме. Рассмотрим ситуацию, описанную в примере 1 Введения и предположим, что n шаров извлекаются *наугад без возвращения*. Тогда, если ξ обозначает число белых шаров в выборке, то эта случайная величина имеет распределение $H(a_1, a_2, n)$ (сравни с примером 1).

Заметим, что вероятность (9) отличны от нуля лишь при $x \leqslant a_1$ и $n - x \leqslant a_2$, т. е. когда целое x лежит в интервале $\max(0, n - a_2) \leqslant x \leqslant \min(n, a_1)$.

Если $\mathcal{L}(\xi) = H(a_1, a_2, n)$, то ее первые два момента есть

$$\mathbf{E}\xi = \frac{na_1}{a} \quad \text{и} \quad \mathbf{D}\xi = \frac{na_1a_2(a-n)}{a^2(a-1)}. \tag{10}$$

5. Распределение Маркова—Пойа

Обобщение понятия зачастую бывает полезно для постижения его сущности.

А. Н. Колмогоров

Это распределение является обобщением одновременно и биномиального, и гипергеометрического распределений и возникает оно в следующей урновой схеме. Вновь вернемся к примеру 1 Введения и рассмотрим следующий эксперимент. Из урны, содержащей первоначально a_1 белых и $a_2 = a - a_1$ черных шаров, выбирается наугад (т. е. равновероятно) один шар, фиксируется его цвет и шар возвращается в урну с одновременным добавлением c

новых шаров того же цвета. Затем из урны (содержащей теперь $a + c$ шаров) снова производится случайное извлечение одного шара и повторяется тот же процесс. Здесь c может быть любым целым числом и, в частности, при $c = 0$ (новые шары не добавляются) мы имеем случайный выбор с возвращением, рассмотренный в примере 1, а при $c = -1$ (извлеченный шар в урну не возвращается) — схему случайного выбора без возвращения, рассмотренную в предыдущем п. 4 (в последнем случае процесс извлечения шаров кончается через a шагов из-за отсутствия шаров в урне). При $c > 0$ эта схема выбора обладает эффектом последействия: если извлекается шар какого-то цвета, то шанс (вероятность) извлечь шар такого же цвета при следующем испытании возрастает. Такая урновая схема может служить приближенной моделью явлений, подобных эпидемиям, когда осуществление некоторых событий увеличивает шанс их повторения.

Пусть ξ обозначает число белых шаров, наблюдавшихся при n извлечениях. Распределение вероятностей этой случайной величины называется *распределением Маркова—Пойа* (обозначается $\mathcal{L}(\xi) = \text{МП}(n; a_1, a_2, c)$), и оно имеет вид

$$\mathbf{P}\{\xi = x\} = f(x|a_1, a_2, c; n) = C_n^x \prod_{j=0}^{x-1}(a_1 + jc) \prod_{j=0}^{n-x-1}(a_2 + jc) \Big/ \prod_{j=0}^{n-1}(a + jc),$$
$$x = 0, 1, \tag{11}$$

Если воспользоваться биномиальными коэффициентами и учесть, что для любого действительного a и натурального x

$$C_a^x = \frac{a(a-1)\quad(a-x+1)}{x!} \quad \text{и} \quad C_a^0 = 1,$$

то формулу (11) при $c \neq 0$ (тогда числитель, и знаменатель в (11) можно разделить на c^n) можно записать также в любой из следующих двух форм:

$$f(x|a_1, a_2, c; n) = \frac{C^x_{a_1/c+x-1} C^{n-x}_{a_2/c+n-x-1}}{C^n_{a/c+n-1}} = \frac{C^x_{-a_1/c} C^{n-x}_{-a_2/c}}{C^n_{-a/c}}. \tag{12}$$

Подчеркнем, что из (11) при $c = 0$ мы получаем биномиальное распределение $\text{Bi}(n, p)$, $p = a_1/a$ (как в примере 1 для схемы выбора с возвращением), а из второго представления в (12) при $c = -1$ — гипергеометрическое распределение $H(a_1, a_2, n)$ (как в п. 4 для схемы выбора без возвращения), так что распределение Маркова—Пойа включает в себя как частные случаи оба эти распределения.

Если $\mathcal{L}(\xi) = \text{МП}(n; a_1, a_2, c)$, то среднее и дисперсия имеют вид

$$\mathbf{E}\xi = \frac{na_1}{a} \quad \text{и} \quad \mathbf{D}\xi = \frac{na_1a_2(a + cn)}{a^2(a + c)}. \tag{13}$$



Интересна история открытия этого распределеиия. Впервые оно появляется в работах выдающегося русского математика, академика А. А. Маркова (1856–1922), который провел исчерпывающий анализ этого распределения и опубликовал свои результаты в Известиях Петербургской Академии наук в 1917 г. Но, по-видимому, эта публикация «осталась незамеченной» за рубежом, так как, спустя 6 лет, в 1923 г. появляется работа Ф. Эггенбергера и Д. Пойа, где вводится такое же распределение и доказываются для него некоторые частные результаты. С этих пор даиное распределение стало называться в зарубежной литературе именем одного из его авторов — Пойа (выдающийся американский математик венгерского происхождения (1887–1985)), хотя более исторически оправдано название «распределение Маркова—Пойа».

6. Полиномиальное распределение

Полиномиальное распределение $M(n; \underline{p})$ с параметрами n и $\underline{p}$, где n — натуральное число, и $\underline{p}=(p_1,\ldots,p_N)$, $0<p_i<1$, $i=1,\ldots,N$, $p_1+\ldots+p_N=1$, — это распределение случайного вектора $\underline{\nu}=(\nu_1,\ldots,\nu_N)$ с целочисленными неотрицательными компонентами, удовлетворяющими условию

$$\nu_1 + \quad + \nu_N = n,$$

которое задается вероятностями

$$\mathbf{P}\{\underline{\nu}=\underline{x}\} = f(\underline{x}|n,\underline{p}) = \frac{n!}{x_1!\quad x_N!}p_1^{x_1}\quad p_N^{x_N} \tag{14}$$

где $\underline{x}=(x_1,\ \ldots,x_N)$ — произвольный вектор с целыми неотрицательными координатами, причем $x_1+\quad+x_N=n$. Кратко это записывается так: $\mathcal{L}(\underline{\nu})=M(n;\underline{p})$. Название «полиномиальное распределение» связано с тем, что вероятность (14) представляет собой общий член в разложении полинома $(p_1+\quad+p_N)^n$ по степеням $p_1,\quad,p_N$:

$$1=(p_1+\quad+p_N)^n=\sum_{x_1+\ldots+x_N=n}\frac{n!}{x_1!\ \ldots x_N!}p_1^{x_1}\quad p_N^{x_N}$$

Здесь

$$\mathbf{E}\nu_i = np_i,\quad \mathbf{D}\nu_i = np_i(1-p_i)\quad\text{и}$$
$$\operatorname{cov}(\nu_i,\nu_j) = -np_ip_j,\quad i,j=1,\quad,N,\quad i\neq j. \tag{15}$$

Полиномиальные испытания

Такое распределение возникает в схеме *полиномиальных испытаний*, т. е. независимых испытаний с N возможными исходами, вероятности которых не меняются от испытания к испытанию и равны $p_1,\quad,p_N$ соответственно: если произведено n испытаний и ν_i — число реализаций в них i-го исхода, $i=1,\quad,N$, то $\mathcal{L}(\underline{\nu})=M(n;\underline{p})$. Если $N=2$ (схема с двумя исходами), то мы имеем схему Бернулли (см. п. 1), для которой $\mathcal{L}(\nu_1)=\mathrm{Bi}(n,p_1)$; с другой стороны, так как $\nu_2=n-\nu_1$ и $p_2=1-p_1$, то двумерный вектор $(\nu_1,n-\nu_1)$ имеет полиномиальное распределение $M(n;p_1,1-p_1)$, эквивалентное, следовательно, биномиальному распределению $\mathrm{Bi}(n,p_1)$. Отсюда же вытекают следующие свойства полиномиального распределения:

1) распределение каждой компоненты ν_j вектора $\underline{\nu}$ является биномиальным $\mathrm{Bi}(n, p_j)$, $j = 1, \ldots, N$, и

Свойство воспроизводимости

2) если вектора $\underline{\nu}_1, \ldots, \underline{\nu}_k$ независимы и $\mathcal{L}(\underline{\nu}_i) = M(n_i; \underline{p})$, $i = 1, \ldots, k$, то

$$\mathcal{L}(\underline{\nu}_1 + \ldots + \underline{\nu}_k) = M(n_1 + \ldots + n_k; \underline{p})$$

(*воспроизводимость* по параметру n).

Отметим еще, что часто вместо термина «полиномиальное» используется его синоним «мультиномиальное».

В приложениях обычно вектор вероятностей $\underline{p} = (p_1, \ldots, p_N)$ неизвестен — в этом случае мы имеем *полиномиольную статистическую модель* $M(n; \theta)$, $\theta = (\theta_1, \ldots, \theta_N)$, с параметрическим множеством $\Theta = \{\theta \quad 0 < \theta_i < 1, i = 1, \ldots, N, \theta_1 + \ldots + \theta_N = 1\}$.

Пример 4 (Обобщение примера 1 Введения). Имеется урна с шарами N различных цветов, которые мы условно обозначим $\mathcal{A}_1, \ldots, \mathcal{A}_N$. Пусть a_j — число шаров цвета $\mathcal{A}_j$, $j = 1, \ldots, N$, и $a = a_1 + \ldots + a_N$ — общее число шаров в урне. Рассмотрим следующий эксперимент: из урны по схеме выбора с возвращением наугад (т. е. равновероятно) извлекается n шаров (т. е. каждый раз любой шар может быть извлечен с одинаковой вероятностью $1/a$ и независимо от результатов предыдущих извлечений). Обозначим ν_j число наблюдавшихся $\mathcal{A}_j$-шаров, $j = 1, \ldots, N$. Тогда вектор $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ будет иметь полиномиальное распределение $M(n; \underline{p})$ с параметром $\underline{p} = (a_1/a, \ldots, a_N/a)$, т. е. вектором долей цветов $\mathcal{A}_1, \ldots, \mathcal{A}_N$ Если этот вектор нам неизвестен, то подходящей оценкой для него будет вектор относительных наблюдавшихся частот $\underline{\nu}/n = (\nu_1/n, \ldots, \nu_N/n)$, так как в силу (15) $\mathbf{E}\underline{\nu}/n = \underline{p}$ (в среднем оценка $\underline{\nu}/n$ совпадает со значением параметра $\underline{p}$). В последующем мы дадим строгое обоснование такого правила оценивания (см. пример 10 § 3.2). •

7. Многомерное распределение Маркова—Пойа

Вновь обратимся к ситуации, описанной в примере 4, и рассмотрим теперь такой эксперимент: последовательно в каждый момент времени $n = 1, 2, \ldots$ из урны наград извлекается один шар, фиксируется его цвет и шар возвращается обратно в урну с добавлением c новых шаров того же цвета, что и извлеченный шар (тем самым состав урны каждый раз увеличивается на c шаров, если $c > 0$, и уменьшается, если $c < 0$; при $c = 0$ он остается неизменным, т. е. в этом случае мы имеем схему выбора с возвращением, рассмотренную в примере 4). Вновь обозначим ν_j число $\mathcal{A}_j$-шаров, наблюдавшихся при n извлечениях, $j = 1, \ldots, N$. Тогда вектор $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ будет иметь распределение

$$\mathbf{P}\{\underline{\nu} = \underline{x}\} = \frac{n!}{x_1! \ldots x_N!} \frac{\prod_{j=1}^{N} a_j(a_j + c) \ldots (a_j + (x_j - 1)c)}{a(a + c) \ldots (a + (n - 1)c)}, \qquad (16)$$

где $\underline{x} = (x_1, \ldots, x_N)$ — произвольный вектор с целыми неотрицательными координатами, причем $x_1 + \quad + x_N = n$; множество таких векторов будем обозначать $\mathcal{K}_{Nn}$. Это распределение называется *мкогомерным* (*N-мерным*) *распределением Маркова—Пойа* с параметрами n, $\underline{a}$ и c и обозначается символом МП$(n; \underline{a}, c)$, где $\underline{a} = (a_1, \ldots, a_N)$ — вектор первоначального состава урны.

Формула (16) является многомерным (N-мерным) аналогом формулы (11) и сводится к ней при $N = 2$. Если $c \neq 0$, то разделив числитель и знаменатель в (16) на c^n и используя, как и в п. 5, биномиальные коэффициенты, этой формуле можно придать еще две формы — аналоги форм (12):

$$\mathbf{P}\{\underline{\nu} = \underline{x}\} = f(\underline{x}|\underline{a}, c; n) = \prod_{j=1}^{N} C_{a_j/c+x_j-1}^{x_j} \Big/ C_{a/c+n-1}^{n} = \prod_{j=1}^{N} C_{-a_j/c}^{x_j} \Big/ C_{-a/c}^{n},$$

$$\underline{x} \in \mathcal{K}_{Nn}. \tag{17}$$

Связь с другими распределениями

Отметим три важных частных случая этих соотношений. Если в формуле (16) положить $c = 0$, то мы придем к (14) с $\underline{p} = \underline{a}/a$, т. е. распределение Маркова—Пойа МП$(n; \underline{a}, 0)$ сводится к полиномиальному распределению $M(n; \underline{a}/a)$. Далее, если во втором представлении (17) положить параметр $c = -1$ (что соответствует схеме выбора без возвращения), то при $n \leqslant a$ (иначе биномиальный коэффициент C_a^n обратится в 0) получим распределение

$$\mathbf{P}\{\underline{\nu} = \underline{x}\} = \prod_{j=1}^{N} C_{a_j}^{x_j} \Big/ C_a^n, \quad \underline{x} \in \mathcal{K}_{Nn}, \tag{18}$$

которое является многомерным (N-мерным) аналогом гипергеометрического распределения (9) и сводится к нему при $N = 2$. Поэтому распределение, задаваемое вероятностями (18), называется *многомерным* (*N-мерным*) *гипергеометрическим распределением* с параметрами $\underline{a}$ и n. Оно обозначается символически $\mathcal{L}(\underline{\nu}) = H(\underline{a}, n)$ и характеризует собой схему случайной выборки без возвращения объема $n \leqslant a$ из конечной совокупности, состоящей из $a = a_1 + \quad + a_N$ элементов, разбитых по некоторому признаку на N непересекающихся классов размерами соответственно $a_1, \ldots, a_N$ (в данном случае ν_j интерпретируется как число наблюдавшихся в выборке элементов j-го класса, $j = 1, \quad, N$).

Наконец, если в первом представлении (17) все a_j положить равными c, то биномиальные коэффициенты в числителе станут равны 1, и мы придем к формуле

$$\mathbf{P}\{\underline{\nu} = \underline{x}\} = \left(C_{N+n-1}^{n}\right)^{-1}$$

задающей равномерное распределение на множестве

$\mathcal{K}_{Nn} = \{\underline{x} = (x_1, \ldots, x_N) \quad x_i = 0, 1, 2, \ldots, i = 1, \quad, N; x_1 + \quad + x_N = n\}$:

число элементов (векторов $\underline{x}$) этого множества равно C_{N+n-1}^{n} — числу решений уравнения $x_1 + \quad + x_N = n$ в целых неотрицательных числах, и всем им приписывается одинаковая вероятность.

Таким образом, многомерное распределение Марков—Пойа включает в себя в качестве частных случаев полиномиальное распределение, многомерное гипергеометрическое распределение и равномерное распределение на множестве $\mathcal{K}_{Nn}$.

Моменты МП-распределения

Если $\mathcal{L}(\underline{\nu}) = \text{МП}(n;\ \underline{a},\ c)$, то первые и вторые моменты случайного вектора $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ имеют вид

$$\mathbf{E}\nu_i = n\frac{a_i}{a},\quad \mathbf{D}\nu_i = n\frac{a_i(a-a_i)(a+cn)}{a^2(a+c)},\quad i = 1,\ \ ,N,\quad \text{и}$$
$$\operatorname{cov}(\nu_i, \nu_j) = -n\frac{a_i a_j(a+cn)}{a^2(a+c)},\quad i,j = 1,\ \ ,N,\quad i \neq j. \tag{19}$$

Полное изложение теории этого распределения имеется в работе авторов «Об урновой схеме Маркова—Пойа: от 1917 г. до наших дней» (Обозрение прикладной и промышленной математики. 1996. Т. 3. Вып. 4. С. 484–511).

8. Распределение степенного ряда

Целый класс дискретных распределений может быть построен по следующей схеме. Рассмотрим произвольный степенной ряд

$$f(\theta) = \sum_{x=l}^{\infty} a(x)\theta^x \qquad l \geqslant 0,$$

с неотрицательными коэффициентами ($a(x) \geqslant 0$ для всех $x \geqslant l$) и ненулевым радиусом сходимости R. Тогда при любом $\theta \in \Theta = (0, R)$ можно определить дискретное вероятностное распределение с плотностью

$$f(x|\theta) = a(x)\frac{\theta^x}{f(\theta)},\quad x = l, l+1, \tag{20}$$

(эти числа неотрицательны и в сумме дают 1). Такого типа распределение и называется *распределением степенного ряда*. Класс таких распределений весьма обширен и включает в себя многие стандартные распределения. В частности, в него входят распределение Пуассона $\Pi(\theta)$ (см. (8)): для него $f(\theta) = e^{\theta}$, $R = \infty$; отрицательное биномиальное распределение $\overline{\text{Bi}}(r, \theta)$ (см. (5)): для него $f(\theta) = (1-\theta)^{-r}$, $R = 1$; *логарифмическое распределение*, задаваемое вероятностями

Примеры роспределений степенного ряда

$$f(x|\theta) = \frac{\theta^x}{x}\Big/\ln\frac{1}{(1-\theta)},\quad x = 1, 2, \tag{21}$$

для него

$$f(\theta) = \ln\frac{1}{1-\theta} = \sum_{x=1}^{\infty}\frac{\theta^x}{x},\quad R = 1$$

и т.д., а также соответствующие усеченные слева распределения.

Примечание. Усеченным называется распределение, у которого некоторые значения запрещены. Например, *усеченным в нуле распределением Пуассона* является распределение, задаваемое вероятностями (сравни с (8))

$$f(x|\theta) = ce^{-\theta}\frac{\theta^x}{x!}, \quad x = 1, 2, \ldots \tag{22}$$

где константа c определяется из условия нормировки

$$\sum_{x=1}^{\infty} f(x|\theta) = 1,$$

и в данном случае она равна $(1 - e^{-\theta})^{-1}$

Моменты

Если случайная величина ξ имеет распределение степенного ряда (20), то ее первые два момента находятся по формулам

$$\mathbf{E}\xi \equiv \mu(\theta) = \theta\frac{f'(\theta)}{f(\theta)} \quad \text{и} \quad \mathbf{D}\xi \equiv \sigma^2(\theta) = \theta\mu'(\theta) \tag{23}$$

(убедитесь, что формулы для моментов в п. 2 и 3 являются частными случаями этих выражений).

Свойство воспроизводимости

Важным свойством распределения степенного ряда является следующее: если случайные величины $X_1, \ldots, X_n$ независимы и имеют одно и то же распределение (20), то их сумма $X = X_1 + \ldots + X_n$ также имеет распределение степенного ряда, порождаемое функцией $f^n(\theta)$, т. е.

$$\mathbf{P}\{X = x\} = f_n(x|\theta) = a_n(x)\frac{\theta^x}{f^n(\theta)}, \quad x = nl, nl + 1, \ldots \tag{24}$$

где коэффициенты $a_n(x)$ определяются разложением

$$f^n(\theta) = \sum_{x=nl}^{\infty} a_n(x)\theta^x$$

§ 1.2. Основные абсолютно непрерывные модели

Дорогу осилит идущий.

В этом параграфе мы продолжаем знакомство с основными вероятностными распределениями и их свойствами, сосредоточившись теперь на случае абсолютно непрерывных распределений. Принцип изложения здесь тот же, что и в предыдущем параграфе о дискретных распределениях. Напомним, что теперь функция $f(x)$ имеет смысл обычной плотности распределения, т. е. это неотрицательная функция, удовлетворяющая условию нормировки

$$\int_{\mathcal{E}} f(x)\,dx = 1,$$

где $\mathcal{E}$ — множество, на котором задано распределение наблюдаемой (изучаемой) случайной величины ξ.

1. Нормальное распределение

Это распределение уже встречалось нам ранее: его формальное определение дано во Введении к данной главе (см. (1)). Роль нормального распределения (нормальной модели) в теории вероятностей и математической статистике чрезвычайно велика и ее трудно переоценить. Нормальная модель возникает в таких статистических экспериментах, когда на исход эксперимента оказывает влияние большое число независимо действующих случайных факторов, каждый из которых лишь незначительно влияет на конечный результат (см. примеры 2 и 4 Введения). Для таких ситуаций имеется знаменитая теорема, называемая *центральной предельной теоремой* теории вероятностей (термин «центральная» отражает ее значимость, так ее впервые назвал Дьердь Пойа), согласно которой если $\{X_1, X_2, \ldots\}$ — последовательность независимых одинаково распределенных случайных величин с положительной дисперсией, то предельное при $n \to \infty$ распределение нормированной надлежащим образом суммы $S_n = X_1 + \ldots + X_n$ является нормальным. Именно: при $n \to \infty$

Центральная предельная теорема

$$\mathcal{L}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right) \to \mathcal{N}(0, 1), \tag{1}$$

где $\mu = \mathbf{E}X_1$, $\sigma^2 = \mathbf{D}X_1$. Более того, известно, что при некоторых условиях предельное распределение суммы S_n остается нормальным, даже если случайные величины $X_1, X_2, \ldots$ зависимы и неодинаково распределены. Возможно, именно поэтому нормальный закон распределения возникает как подходящая приближенная модель во многих областях исследований. Поэтому в конкретной практической задаче при построении соответствующей вероятностной или статистической модели часто предполагают, что распределение каждой наблюдаемой величины приближенно нормально. В силу этого статистические методы, разработанные для нормальной модели, имеют весьма широкую сферу применений и, следовательно, хорошо знать их — очень важно. (Пуанкаре[5]) как-то заметил с сарказмом, что все верят в универсальность нормального распределения: физики верят, потому, что думают, что математики доказали его логическую необходимость, а математики верят, так как считают, что физики проверили это лабораторными экспериментами [23, с. 49].)

Предельные теоремы

Общее замечание. В (1) нам впервые встречается запись типа $\mathcal{L}(\eta_n) \to \mathcal{L}(\eta)$, которую мы должны пояснить. Во многих задачах математической статистики (и теории вероятностей) рассматриваются последовательности случайных величин $\{\eta_n,\ n = 1, 2, \ldots\}$, сходящиеся в том или ином смысле к некоторому пределу η (эта) (случайной величине или константе),

[5] Пуанкаре Анри (1854–1912) — выдающийся французский математик, физик, астроном и философ.

когда $n \to \infty$, — в таких случаях говорят о соответствующих *предельных теоремах*. Мы будем использовать в дальнейшем лишь два простейших вида сходимости, без которых нам совершенно невозможно обойтись: *сходимость по вероятности* и *сходимость по распределению*, или (синоним) *слабая сходимость*. Напомним их определения. Последовательность $\{\eta_n\}$ называется сходящейся по вероятности к η, если

$$\lim_{n\to\infty} \mathrm{P}\{|\eta_n - \eta| > \varepsilon\} = 0 \quad \forall\, \varepsilon > 0,$$

что кратко записывается так: $\eta_n \xrightarrow{\mathrm{P}} \eta$. Очень важным при использовании этой сходимости является следующее свойство: если $\eta_n \xrightarrow{\mathrm{P}} \eta$ и функция φ непрерывна, то

$$\varphi(\eta_n) \xrightarrow{\mathrm{P}} \varphi(\eta). \tag{2}$$

Под сходимостью по распределению (слабой сходимостью) [что кратко записывается так: $\mathcal{L}(\eta_n) \to \mathcal{L}(\eta)$ или $\eta_n \xrightarrow{\mathcal{L}} \eta$] понимается сходимость соответствующих функций распределения

$$F_n(x) = \mathrm{P}\{\eta_n \leqslant x\}$$

к предельной функции распределения

$$F(x) = \mathrm{P}\{\eta \leqslant x\}$$

в каждой точке непрервности последней: $F_n(x) \to F(x)$, если в точке x функция $F(x)$ непрерывна. Известно, что из сходимости по вероятности следует слабая сходимость:

$$\eta_n \xrightarrow{\mathrm{P}} \eta \;\Rightarrow\; \mathcal{L}(\eta_n) \to \mathcal{L}(\eta).$$

Примером сходимости по вероятности является знаменитый *закон больших чисел*: если случайные величины $\eta_1, \eta_2, \ldots$ независимы, одинаково распределены и имеют математическое ожидание $\mathrm{E}\eta_1 = \mu$, то при $n \to \infty$

$$\frac{1}{n}(\eta_1 + \quad + \eta_n) \xrightarrow{\mathrm{P}} \mu; \tag{3}$$

примером же слабой сходимости является центральная предельная теорема (1). Эти предельные теоремы (или, как говорят, *асимптотический подход*) мы часто будем использовать в дальнейшем в различных статистических задачах, но это не есть «прихоть» математика, а «суровая» необходимость, ибо

> *Асимптотический подход — зачастую единственная возможность, открытая статистику в суровой действительности.*
>
> **Дж. Русас**

Моменты и свойство воспроизводимости

Вернемся к нормальному распределению и продолжим обсуждение его свойств. Уже отмечалось (но выделим это особо ввиду важности данного свойства), что если $\mathcal{L}(\xi) = \mathcal{N}(\mu, \sigma^2)$, то

$$\mathrm{E}\xi = \mu \quad \text{и} \quad \mathrm{D}\xi = \sigma^2 \tag{4}$$

Далее, можно показать, что всякая линейная комбинация нормально распределенных случайных величин также подчиняется нормальному закону.

Пусть, в частности, случайные величины $X_1, \quad , X_k$ независимы и

$$\mathcal{L}(X_i) = \mathcal{N}(\mu_i, \sigma_i^2), \quad i = 1, \quad , k.$$

Если $a_1, \ldots, a_k$ и b — постоянные, причем $a_i \neq 0$ хотя бы для одного значения i, то

$$\mathcal{L}(a_1 X_1 + \quad + a_k X_k + b) =$$
$$= \mathcal{N}(a_1\mu_1 + \quad + a_k\mu_k + b,\ a_1^2\sigma_1^2 + \quad + a_k^2\sigma_k^2). \tag{5}$$

В частности

$$\mathcal{L}(X_1 + \quad + X_k) = \mathcal{N}(\mu_1 + \quad + \mu_k,\ \sigma_1^2 + \quad + \sigma_k^2), \tag{6}$$

т. е. нормальное распределение $\mathcal{N}(\mu, \sigma^2)$ является *воспроизводящим* по параметрам μ и σ^2 одновременно.

Стандартное нормальное распределение

Нормальное распределение со средним 0 и дисперсией 1 называется *стандартным нормальным распределением*, его плотность и функция распределения обозначаются через φ и Φ (фи малое и фи большое) соответственно, так что

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{и} \quad \Phi(x) = \int_{-\infty}^{x} \varphi(t)\, dt, \quad -\infty < x < \infty. \tag{7}$$

Эти функции протабулированы и для них имеются обширные статистические таблицы.

Наконец, отметим, что если $\mathcal{L}(\xi) = \mathcal{N}(\mu, \sigma^2)$, $\sigma^2 > 0$, то по свойству (5), *нормированная случайная величина* $\xi' = (\xi - \mu)/\sigma$ имеет стандартное нормальное распределение $\mathcal{N}(0, 1)$.

Логнормальное распределение

С нормальным распределением тесно связано одно интересное распределение, называемое *логнормальным*. По определению, это распределение случайной величины $\eta = e^{\xi}$, где $\mathcal{L}(\xi) = \mathcal{N}(\mu, \sigma^2)$. Для первых двух моментов η справедливы соотношения

$$\mathbf{E}\eta = e^{\mu + \sigma^2/2}, \quad \mathbf{D}\eta = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} \tag{8}$$



Из истории: нормальное распределение, ставшее краеугольным камнем науки о случайном, было открыто французским математиком Абрахамом де Муавром (1667–1754) в 1733 г., о чем он сообщил некоторым своим друзьям. В 3-м издании своей книги «Доктрина шансов» (1756) он так писал о своем открытии мирового значения: «...Возьму на себя смелость утверждать, что это труднейшая проблема о случайном... (Де Муавр, как это ни странно, не включил свое открытие во 2-е издание книги в 1738 г.; первое издание было осуществлено в Англии в 1718 г., куда он переехал из Франции после отмены Нантского эдикта, который предоставлял гугенотам свободу вероисповедания) [23]. Некоторое время спустя нормальное распределение было снова открыто Гауссом в 1809 г. и Лапласом в 1812 г. в связи с работами по теории ошибок наблюдений. Лаплас дал первую (совершенную) формулировку центральной предельной теоремы и привел большое количество важных приложений этого распределения.

2. Многомерное нормальное распределение

Случайный вектор $\underline{\xi} = (\xi_1, \ldots, \xi_k)$ размерности $k \geqslant 2$ имеет *невырожденное* (или *собственное*) *многомерное* (k-*мерное*) *нормальное распределение с вектором средних* $\underline{\mu} = (\mu_1, \quad, \mu_k)$ *и матрицей вторых моментов* (*ковариаций*)

$$\Sigma = \left\|\operatorname{cov}(\xi_i, \xi_j)\right\|_1^k$$

(обозначается $\mathcal{L}(\underline{\xi}) = \mathcal{N}(\underline{\mu}, \Sigma)$), если его плотность $f(\underline{x}|\underline{\mu}, \Sigma)$ при всех $\underline{x} = (x_1, \ldots, x_k) \in R^k$ имеет вид

$$f(\underline{x}|\underline{\mu}, \Sigma) = (2\pi)^{-k/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\underline{x} - \underline{\mu})' \Sigma^{-1}(\underline{x} - \underline{\mu})\right\}. \qquad (9)$$

Элементы алгебры

Здесь приняты следующие обозначения, которые будут систематически использоваться и в дальнейшем: R^k — это k-мерное евклидово пространство, $|A| \equiv \det A$ — определитель матрицы $A \quad \|a_{ij}\|_1^k$ размера $k \times k$, A^{-1} — обратная матрица A, которая существует, если матрица A *невырождена* (т. е. $|A| \neq 0$), при матричных операциях векторы понимаются как вектор-столбцы, знак $'$ (штрих) означает транспонирование, наконец, выражение $\underline{x}'A\underline{x}$ есть лаконичная запись на языке матричной алгебры квадратичной формы $\sum_{i,j=1}^{k} a_{ij}x_ix_j$ для вектора $\underline{x} = (x_1, \quad, x_k)$ с матрицей коэффициентов A.

В (9) вектор $\underline{\mu} = \mathbf{E}\underline{\xi} \equiv (\mathbf{E}\xi_1, \ldots, \mathbf{E}\xi_k)$ может быть любой точкой R^k, а ковариационная матрица Σ — произвольной *симметрической* (т. е. $\Sigma' = \Sigma$) и *положительно определенной* (т. е. $\underline{x}'\Sigma\underline{x} > 0$, $\forall\, \underline{x} \neq \underline{0}$) матрицей (ее называют также *дисперсионной матрицей* и пишут $\mathbf{D}\underline{\xi} = \Sigma$). Напомним также, что симметрическая положительно определенная матрица обязательно невырождена, и обратная к ней матрица также является симметрической положительно определенной.

Таким образом, общее число параметров, которыми определяется невырожденное k-мерное распределение $\mathcal{N}(\underline{\mu}, \Sigma)$, с учетом симметричности матрицы Σ равно $k + k(k+1)/2$. Важнейшими свойствами этого распределения являются следующие:

Важнейшие свойства $\mathcal{N}(\underline{\mu}, \Sigma)$

1) *некоррелированность компонент вектора* $\underline{\xi}$ *эквивалентна их независимости* и

2) сохранение свойства нормальности при линейных преобразованиях: *если* $\underline{\eta} = L\underline{\xi}$, *где* L — *произвольная* $(m \times k)$-*матрица*, то $\mathcal{L}(\underline{\eta}) = \mathcal{N}(L\underline{\mu}, L\Sigma L')$, т. е. новый случайный вектор $\underline{\eta}$ имеет m-мерное нормальное распределение, параметры которого вычисляются по очень простому правилу: вектор средних $\mathbf{E}\underline{\eta} = L\underline{\mu}$, а дисперсионная матрица $\mathbf{D}\underline{\eta} = L\Sigma L'$.

Поскольку эти свойства многомерного нормального распределения чрезвычайно важны, и поскольку не все читатели достаточно знакомы с многомерными распределениями, мы обсудим их несколько более подробно. Мы покажем, что на самом деле они очень просто устанавливаются, если воспользоваться аппаратом *характеристических функций* — важнейшим инструментом доказательства различных теорем теории вероятностей. Но сначала мы кратко напомним их определение и некоторые из их основных свойств.

Характеристические функции

Характеристическая функция произвольной случайной величины (вектора) $\underline{X} = (X_1, \quad , X_k)$, $k \geqslant 1$, определяется формулой

$$\Psi_{\underline{X}}(\underline{t}) = \mathrm{E}e^{i\underline{t}'\underline{X}},$$

где i — мнимая единица, $\underline{t} = (t_1, \quad , t_k)$, $-\infty < t_j < \infty$, $j = 1, \quad , k$, и

$$\underline{t}'\underline{X} = t_1 X_1 + \quad + t_k X_k$$

— скалярное произведение векторов $\underline{t}$ и $\underline{X}$.

Характеристическая функция всегда существует, и для дискретной случайной величины с плотностью $f(\underline{x}) = \mathrm{P}\{\underline{X} = \underline{x}\}$, $\underline{x} \in \mathfrak{X}$, она вычисляется по формуле

$$\Psi_{\underline{X}}(\underline{t}) = \sum_{\underline{x}} e^{i\underline{t}'\underline{x}} f(\underline{x}), \tag{10}$$

а для абсолютно непрерывной случайной величины с плотностью $f(\underline{x})$ она есть интеграл Римана

$$\Psi_{\underline{X}}(\underline{t}) = \int_{\mathfrak{X}} e^{i\underline{t}'\underline{x}} f(\underline{x})\, d\underline{x}, \quad d\underline{x} = dx_1 \quad . dx_k. \tag{11}$$

Соответствие, установленное формулами (10) и (11) между характеристической функцией $\Psi_{\underline{X}}(\underline{t})$ и распределением случайной величины $\underline{X}$ (ее плотностью $f(\underline{x})$), является взаимно однозначным, т. е. по $\Psi_{\underline{X}}(\underline{t})$ плотность $f(\underline{x})$ восстанавливается (определяется) однозначно. Например, для одномерной дискретной случайной величины X с плотностью $f(x)$, $x = 0, 1, 2, \quad$ *формула обращения* есть

$$f(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itx} \Psi_X(t)\, dt, \tag{12}$$

если же случайная величина X имеет плотность $f(x)$ и функция $\Psi_X(t)$ абсолютно интегрируема, то формула обращения имеет вид

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \Psi_X(t)\, dt. \tag{13}$$

Таким образом, различные свойства изучаемых случайных величин можно формулировать как в терминах их плотностей, так и в терминах их характеристических функций.

Свойства характеристических функций

Перечислим теперь необходимые нам в последующем свойства характеристических функций. Таковых всего два.

1. *Линейное преобразование случайной величины.* Если случайный вектор $\underline{Y} = (Y_1, \ldots, Y_m)$ получается из вектора $\underline{X} = (X_1, \ldots, X_k)$ с помощью линейного преобразования $\underline{Y} = L\underline{X} + \underline{a}$, где L — произвольная $(m \times k)$-матрица и $\underline{a} = (a_1, \ldots, a_m)$ — произвольный вектор (L и $\underline{a}$ постоянны), то

$$\Psi_{\underline{Y}}(\underline{t}) = e^{i\underline{t}'\underline{a}} \Psi_{\underline{X}}(L'\underline{t}). \tag{14}$$

2. *Критерий независимости.* Если компоненты вектора $\underline{X}$ независимы, то

$$\Psi_{\underline{X}}(\underline{t}) = \Psi_{X_1}(t_1) \ldots \Psi_{X_k}(t_k). \tag{15}$$

Верно и обратное заключение: если имеет место разложение (15), то компоненты $X_1, \ldots, X_k$ независимы. Таким образом, представление (15) является удобным для работы критерием независимости компонент вектора $\underline{X}$.

После этого краткого экскурса в область теории вероятностей вернемся к обсуждаемому многомерному нормальному распределению $\mathcal{N}(\underline{\mu}, \Sigma)$ и дадим его новое (другое) определение в терминах характеристической функции.

Характеристическая функция $\mathcal{N}(\underline{\mu}, \Sigma)$

Случайный вектор $\underline{\xi} = (\xi_1, \ldots, \xi_k)$ имеет *k-мерное нормальное распределение $\mathcal{N}(\underline{\mu}, \Sigma)$, если его характеристическая функция $\Psi_{\underline{\xi}}(\underline{t}) = \mathbf{E} \exp\{i\underline{t}'\underline{\xi}\}$ имеет вид*

$$\Psi_{\underline{\xi}}(\underline{t}) = \exp\left\{i\underline{t}'\underline{\mu} - \frac{1}{2}\underline{t}'\Sigma\underline{t}\right\}, \quad \underline{t} \in R^k \tag{16}$$

Это определение шире, чем определение (9) невырожденного распределения, так как здесь не требуется существования обратной матрицы Σ^{-1}. Тем самым определение (16) остается в силе и в случае вырожденной матрицы Σ (т. е. при $|\Sigma| = 0$), — в этом случае распределение $\mathcal{N}(\underline{\mu}, \Sigma)$ называется *вырожденным* (или *несобственным*), и для него уже не существует k-мерной плотности типа (9). Мы не будем углубляться далее в эти тонкие вопросы, поскольку в последующем мы будем иметь дело лишь с невырожденными распределениями. Подчеркнем лишь, что в любом случае остаются справедливыми формулы

$$\mathbf{E}\underline{\xi} = \underline{\mu} \quad \text{и} \quad \mathbf{D}\underline{\xi} = \Sigma = \left\|\operatorname{cov}(\xi_i, \xi_j)\right\|_1^k \tag{17}$$

Если компоненты вектора $\underline{\xi}$ некоррелированы, т. е. $\operatorname{cov}(\xi_i, \xi_j) = 0$ при $i \neq j$, то дисперсионная матрица Σ становится диагональной:

$$\Sigma = \begin{Vmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_k^2 \end{Vmatrix}, \quad \text{где} \quad \sigma_j^2 = \mathbf{D}\xi_j, \quad j = 1, \ldots, k,$$

и формула (16) принимает вид

$$\Psi_{\underline{\xi}}(\underline{t}) = \exp\left\{i(t_1\mu_1 + \quad + t_k\mu_k) - \frac{1}{2}(\sigma_1^2 t_1^2 + \quad + \sigma_k^2 t_k^2)\right\} =$$

$$= \prod_{j=1}^{k} \exp\left\{it_j\mu_j - \frac{\sigma_j^2}{2}t_j^2\right\},$$

т. е. переменные $t_1, \quad , t_k$ разделяются, и тем самым характеристическая функция $\Psi_{\underline{\xi}}(\underline{t})$ распадается в произведение одномерных характеристических функций компонент вектора $\underline{\xi}$. По свойству (15) это означает независимость компонент вектора $\underline{\xi}$. Итак, из некоррелированности нормальных случайных величин следует их независимость, а поскольку из независимости случайных величин (с любым распределением!) следует их некоррелированность, то тем самым сформулированное ранее свойство 1) об эквивалентности этих двух свойств для нормальных величин доказано.

Еще проще устанавливается свойство 2) о сохранении нормальности при линейных преобразованиях. Действительно, если $\underline{\eta} = L\underline{\xi}$, то характеристическая функция вектора $\underline{\eta}$ равна в силу (16)

$$\Psi_{\underline{\eta}}(\underline{t}) = \mathbf{E}e^{i\underline{t}'\underline{\eta}} = \mathbf{E}e^{i\underline{t}'L\underline{\xi}} = \mathbf{E}e^{i(L'\underline{t})'\underline{\xi}} = \Psi_{\underline{\xi}}(L'\underline{t}) = \exp\left\{i\underline{t}'(L\underline{\mu}) - \frac{1}{2}\underline{t}'(L\Sigma L')\underline{t}\right\},$$

что есть характеристическая функция нормального закона $\mathcal{N}(L\underline{\mu}, L\Sigma L')$.

Следствием этого свойства является следующий фундаментальный результат, который доказывается при помощи выбора соответствующей матрицы L: *совместное распределение любого подмножества компонент вектора $\underline{\xi}$ является нормальным, и соответствующие подвектор вектора $\underline{\mu}$ и подматрица матрицы Σ будут вектором средних и дисперсионной матрицей этого распределения.*

Стандартное многомерное нормальное распределение

Многомерное нормальное распределение с нулевым вектором средних ($\underline{\mu} = \underline{0}$) и единичной дисперсионной матрицей

$$\Sigma = \mathbb{1} = \begin{Vmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{Vmatrix}$$

называется *стандартным*; о соответствующем случайном векторе $\underline{\xi}$ также говорят как о *стандартном нормальном векторе*. В этом случае компоненты этого вектора являются независимыми стандартными нормальными случайными величинами, каждая из которых имеет плотность, указанную в (7).

Специальное линейное преобразование нормального вектора

В заключение отметим еще одно интересное свойство нормального распределения. В алгебре доказывается, что любая симметрическая положительно определенная матрица Σ может быть приведена

с помощью ортогонального преобразования, заданного ортогональной матрицей $\mathcal{U}$ (т. е. $\mathcal{U}^{-1} = \mathcal{U}'$), к диагональному виду:

$$\mathcal{U}'\Sigma\mathcal{U} = \Lambda = \begin{Vmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{Vmatrix},$$

где $\lambda_j > 0$, $j = 1, \ldots, k$, — собственные (или характеристические) числа Σ.

Пусть теперь $\mathcal{L}(\underline{\xi}) = \mathcal{N}(\underline{\mu}, \Sigma)$ и $\underline{\eta} = \mathcal{U}'\underline{\xi}$. Тогда $\mathcal{L}(\underline{\eta}) = \mathcal{N}(\mathcal{U}'\underline{\mu}, \Lambda)$, т. е. вектор $\underline{\eta}$ будет иметь уже некоррелированные (следовательно, и независимые) компоненты. Далее, так как все $\lambda_j > 0$, то определена диагональная матрица

$$\Lambda^{-1/2} = \begin{Vmatrix} \lambda_1^{-1/2} & & 0 \\ & \ddots & \\ 0 & & \lambda_k^{-1/2} \end{Vmatrix}$$

Рассмотрим теперь случайный вектор

$$\underline{Z} = \Lambda^{-1/2}(\underline{\eta} - \mathcal{U}'\underline{\mu}) = \Lambda^{-1/2}\mathcal{U}'(\underline{\xi} - \underline{\mu}).$$

Это будет также нормальный вектор с нулевым вектором средних ($\mathbf{E}\underline{Z} = \underline{0}$) и дисперсионной матрицей $\Lambda^{-1/2}\Lambda\Lambda^{-1/2} = \mathbf{1}$, т. е. $\underline{Z}$ — стандартный нормальный вектор. Таким образом, *всегда можно указать линейное преобразование, переводящее невырожденный* (!) *нормальный вектор в стандартный нормальный вектор*: для построения этого преобразования $\mathcal{U}$ надо решить линейную систему $\Sigma\underline{u}_j = \lambda_j\underline{u}_j$, $j = 1, \ldots, k$, для определения собственных векторов $\underline{u}_1, \ldots, \underline{u}_k$ и собственных чисел $\lambda_1, \ldots, \lambda_k$ дисперсионной матрицы Σ, и тогда $\mathcal{U} = \|\underline{u}_1 \ldots \underline{u}_k\|$.

Мы ограничиваемся этим минимумом сведений о многомерном нормальном распределении, ибо сказано: «Не пытайся объять необъятное — это невозможно!», и переходим к следующему (после нормального) по значимости и широте применений непрерывному распределению, каковым является гамма-распределение.

3. Гамма-распределение

Случайная величина ξ имеет *гамма-распределение с параметрами* a *и* λ ($a > 0$, $\lambda > 0$): $\mathcal{L}(\xi) = \Gamma(a, \lambda)$, если ее плотность имеет вид

$$f(x|a, \lambda) = \begin{cases} \dfrac{x^{\lambda-1}e^{-x/a}}{\Gamma(\lambda)a^{\lambda}} & \text{при } x \geqslant 0, \\ 0 & \text{при } x < 0. \end{cases} \tag{18}$$

Здесь $\Gamma(\lambda)$ — гамма-функция, определяемая соотношением

$$\Gamma(\lambda) = \int_0^{\infty} t^{\lambda-1}e^{-t}\,dt, \quad \lambda > 0. \tag{19}$$

Полезно помнить следующие важные свойства этой функции:

$$\Gamma(\lambda+1)=\lambda\Gamma(\lambda),\quad \lambda>0,\quad \Gamma(1)=1,$$

следовательно, $\Gamma(n+1)=n!$ для всех натуральных n; далее $\Gamma(1/2)=\sqrt{\pi}$, следовательно, для всех натуральных n значение $\Gamma(n/2)$ также может быть вычислено в явном виде; для больших значений λ имеет место асимптотическая *формула Стирлинга*[6]:

Формула Стирлинга

$$\Gamma(\lambda+1)\sim\sqrt{2\pi\lambda}\lambda^{\lambda}e^{-\lambda}$$

(напомним, что «$\sim$» есть символ асимптотической эквивалентности: $a\sim b$ при $a, b\to\infty$, если $a/b\to 1$; формула Стирлинга, играющая исключительно важную роль в математике, была доказана в 1730 г., она утверждает, что $n!$ приблизительно равняется $\sqrt{2\pi n}(n/e)^n$; независимо от Стирлинга и раньше него формулу для факториалов $n!\approx c\sqrt{n}(n/e)^n$ открыл Муавр, Стирлинг же нашел вид константы $c=\sqrt{2\pi}$).

В представлении (18) a называется *параметром масштаба*, а λ — *параметром формы*. Частные случаи гамма-распределения $\Gamma(a,\lambda)$ при некоторых специальных значениях этих параметров имеют свои названия. Так, при $\lambda=1$ распределение $\Gamma(a,1)$ называется *показательным* (или *экспоненциальным*) *распределением с параметром масштаба* a, а $\Gamma(1,1)$ — *стандартным показательным распределением*: оно задается плотностью $f(x|1,1)=e^{-x}\quad x\geqslant 0$.

Показательное распределение

Если n — натуральное число, то гамма-распределение с параметрами $a=2$ и $\lambda=n/2$ называется *χ^2-распределением с n степенями свободы* и обозначается $\Gamma(2,n/2)=\chi^2(n)$.

Распределение χ^2

Наконец, при натуральных λ $\Gamma(a,\lambda)$ называется *распределением Эрланга порядка* λ. Это распределение суммы λ независимых случайных величин с одинаковым показательным распределением $\Gamma(a,1)$.

Приведем также вид соответствующей характеристической функции

Характеристическая функция и моменты

$$\Psi_{\xi}(t)=\mathbf{E}e^{it\xi}$$

и моментов:

$$\Psi_{\xi}(t)=\frac{1}{(1-iat)^{\lambda}},\quad t\in R^1 \tag{20}$$

и $\mathbf{E}\xi^b=a^b\Gamma(\lambda+b)/\Gamma(\lambda)$ при $b>-\lambda$, в частности,

$$\mathbf{E}\xi=a\lambda,\quad \mathbf{D}\xi=a^2\lambda. \tag{21}$$

[6] Стирлинг Джеймс (1692–1770) — шотландский математик.

Свойство воспроизводимости

Важным свойством гамма-распределения является его *воспроизводимость* по параметру формы λ: если случайные величины $X_1, \quad , X_k$ независимы и при этом $\mathcal{L}(X_j) = \Gamma(a, \lambda_j)$, $j = 1, \quad , k$, то

$$\mathcal{L}(X_1 + \quad + X_n) = \Gamma(a_1, \lambda_1 + \ldots + \lambda_k).$$

Наконец, отметим следующее полезное свойство: если $\mathcal{L}(\xi) = \Gamma(a, \lambda)$, то для любой постоянной $c > 0$ $\mathcal{L}(c\xi) = \Gamma(ca, \lambda)$, в частности, если случайная величина ξ имеет стандартное показательное распределение $\Gamma(1, 1)$, то $\mathcal{L}(a\xi) = \Gamma(a, 1)$ — это свойство используется в алгоритмах моделирования (см. § 1.3).

Модель гамма-рапределения широко используется в задачах теории надежности и теории массового обслуживания. Особую роль играет при этом показательное распределение $\Gamma(a, 1)$. Этому распределению часто подчиняются случайные длины интервалов между последовательными моментами наступления событий в, так называемых, пуассоновских потоках, времена «жизни» различных технических устройств и т. д.

В задачах математической статистики гамма-распределение играет большую роль благодаря тесной связи с нормальным распределением, в частности, там, где рассматриваются квадратичные формы от нормальных случайных величин. Наиболее ярким примером является $\chi^2(n) = \Gamma(2, n/2)$: это распределение суммы квадратов $X_1^2 + \quad + X_n^2$, где $X_1, \quad , X_n$ — независимые стандартные нормальные случайные величины (см. упр. 38).

Другой важный результат состоит в следующем. Пусть $(X_1, \ldots, X_n)$ — случайная выборка из нормального распределения $\mathcal{N}(\mu, \sigma^2)$. Обозначим

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{и} \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Теорема Фишера

Тогда величины $\overline{X}$ и S^2 независимы и при этом

$$\mathcal{L}(\overline{X}) = \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{а} \quad \mathcal{L}\left(\frac{nS^2}{\sigma^2}\right) = \chi^2(n-1). \tag{22}$$

Этот результат известен как знаменитая «теорема Фишера» (1925), и на нем основывается огромное число важных результатов в статистике нормальных моделей.

4. Бета-распределение

Бета-распределение $\mathrm{Be}(a, b)$, $a > 0$, $b > 0$, задается плотностью

$$f(x|a, b) = \begin{cases} \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} & \text{при} \quad 0 \leqslant x \leqslant 1, \\ 0 & \text{в остальных случаях.} \end{cases} \tag{23}$$

Если $\mathcal{L}(\xi) = \mathrm{Be}(a, b)$, то

$$\mathbf{E}\xi = \frac{a}{a+b} \quad \text{и} \quad \mathbf{D}\xi = \frac{ab}{(a+b)^2(a+b+1)}. \tag{24}$$

В случае, когда $a = b = 1$, бета-распределение $\mathrm{Be}(1, 1)$ является равномерным распределением на интервале $(0, 1)$ (см. следующий пункт).

Связь с гамма-распределением

Один из наиболее важных случаев возникновения бета-распределения выражается следующим утверждением: *если случайные величины X_1 и X_2 независимы и имеют гамма-распределения $\Gamma(a, \lambda_1)$ и $\Gamma(a, \lambda_2)$ соответственно, то случайная величина $Y = X_1/(X_1 + X_2)$ имеет распределение $\mathrm{Be}(\lambda_1, \lambda_2)$.*

Отметим также, что нормирующий множитель в (23)

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$

называется *бета-функцией* и обозначается

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

(отсюда и название распределения $\mathrm{Be}(a, b)$).

5. Равномерное распределение

Равномерное распределение $\mathcal{U}(a, b)$, $-\infty < a < b < \infty$, имеет постоянную плотность на интервале $[a, b]$:

$$f(x|a, b) = \begin{cases} \dfrac{1}{b-a} & \text{при} \quad a \leqslant x \leqslant b, \\ 0 & \text{в остальных случаях.} \end{cases} \tag{25}$$

Если $\mathcal{L}(\xi) = \mathcal{U}(a, b)$, то

$$\mathcal{L}\left(\frac{\xi - a}{b-a}\right) = \mathcal{U}(0, 1) = \mathrm{Be}(1, 1)$$

и

$$\mathbf{E}\xi = \frac{a+b}{2}, \quad \mathbf{D}\xi = \frac{(b-a)^2}{12}. \tag{26}$$

Равномерное распределение $\mathcal{U}(a, b)$ описывает процесс «выбора точки наудачу» в интервале $[a, b]$. Так, если $[a, b]$ — интервал между последовательными отправлениями автобуса от остановки, то время ожидания пассажира, не знающего расписания и пришедшего на остановку, есть случайная величина с распределением $\mathcal{U}(a, b)$. Распределение $\mathcal{U}(0, 1)$ играет особую роль в методах моделирования с помощью компьютеров случайных величин с заранее заданными распределениями (об этом подробно будет идти речь в § 1.3).

Такие методы широко используют для приближенных вычислений интегралов, решения дифференциальных и интегральных уравнений и т. д. (см. также пример 7 Введения).

6. Распределение Стьюдента

Стьюдент (Student) — псевдоним английского статистика Уильяма Д. Госсета, впервые использовавшего это распределение в 1908 г. С 1899 г. он работал в Дублине на пивоваренном заводе Гиннесса, и его начальник настоял на том, чтобы Госсет писал под псевдонимом. По определению, *распределением Стьюдента* (синоним: *t-распределение*) *с n степенями свободы* $S(n)$ называется распределение случайной величины (*стьюдентова отношения*)

$$t_n = \frac{\xi}{\sqrt{\chi_n^2/n}}, \tag{27}$$

где случайные величины ξ и χ_n^2 независимы и при этом $\mathcal{L}(\xi) = \mathcal{N}(0, 1)$, а $\mathcal{L}(\chi_n^2) = \chi^2(n)$. Плотность $f(x|n)$ этого распределения имеет вид

$$f(x|n) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \qquad -\infty < x < \infty. \tag{28}$$

Статистический пример

Приведем типичный статистический пример, где возникает это распределение. Пусть $(X_1, \quad, X_n)$ — случайная выборка из распределения $\mathcal{N}(\theta_1, \theta_2^2)$ (т. е. рассматривается нормальная модель с неизвестными средним значением и дисперсией). Тогда в силу теоремы Фишера (см. (22)) стьюдентово отношение (27) примет вид

$$t_{n-1} = \frac{\sqrt{n}}{\theta_2} \frac{\overline{X} - \theta_1}{\sqrt{\frac{nS^2}{n-1}\theta_2^2}} = \sqrt{n(n-1)} \frac{\overline{X} - \theta_1}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2}}, \tag{29}$$

и при этом $\mathcal{L}(t_{n-1}) = S(n-1)$ (независимо от значения неизвестных параметров!).

Если случайная величина X имеет t-распределение $S(n)$, то при $n > 2$

$$\mathbf{E}X = 0 \quad \text{и} \quad \mathbf{D}X = \frac{n}{n-2}. \tag{30}$$

Распределение Коши

В случае, когда $n = 1$, t-распределение называется *распределением Коши*[7] — это распределение отношения

[7] Коши Огюстен Луи (1798–1857) — французский математик, иностранный почетный член Петербургской АН (1831). Один из основоположников теории аналитических функций.

двух независимых стандартных нормальных величин, оно имеет плотность

$$f(x|1) = \frac{1}{\pi}\,\frac{1}{1+x^2}, \quad -\infty < x < \infty, \tag{31}$$

его характеристическая функция есть $\Psi(t) = e^{-|t|}$ и для него не существуют моменты, в том числе и математическое ожидание.

Распределение Коши интересно своими связями с другими распределениями, возникает в некоторых физических задачах, связанных с блужданием частиц, имеет ряд оригинальных аналитических свойств. Например, по закону Коши распределена функция $\operatorname{tg}\eta$, где случайная величина η равномерно распределена на интервале $[-\pi/2, \pi/2]$.

7. Распределение Снедекора

По определению, *распределением Снедекора с n и m степенями свободы* (обозначается $S(n, m)$) называется распределение случайной величины

$$F_{n,m} = \frac{\chi_n^2}{n}\,\frac{\chi_m^2}{m} = \frac{m}{n}\,\frac{\chi_n^2}{\chi_m^2}, \tag{32}$$

F-распределение

где случайные величины χ_n^2 и χ_m^2 независимы и имеют χ^2-распределения с n и m степенями свободы соответственно. Часто это распределение называется также *F-распределением* или *распределением Фишера*.

Плотность $f(x|n, m)$ распределения $S(n, m)$ имеет вид

$$f(x|n,m) = \left(\frac{n}{m}\right)^{n/2} \frac{\Gamma\left(\dfrac{n+m}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)}\,\frac{x^{n/2-1}}{(1+nx/m)^{(n+m)/2}}, \quad x > 0, \tag{33}$$

и $f(x|n, m) = 0$ при $x \leqslant 0$.

Статистический пример

Это распределение играет в статистике важную роль, так как отношение некоторых независимых квадратичных форм от нормальных случайных величин подчиняется F-распределению. Вот один из соответствующих примеров.

Пусть $(X_1, \ldots, X_n)$ — случайная выборка из нормального распределения $\mathcal{N}(\mu_1, \sigma^2)$, а $(Y_1, \ldots, Y_m)$ — независимая с первой случайная выборка из распределения $\mathcal{N}(\mu_2, \sigma^2)$. Тогда при $n \geqslant 2$, $m \geqslant 2$ и любых μ_1, μ_2 и σ^2 отношение

$$F_{n-1,m-1} = \frac{\dfrac{1}{(n-1)}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2}{\dfrac{1}{(m-1)}\displaystyle\sum_{i=1}^{m}(Y_i - \overline{Y})^2} \tag{34}$$

будет иметь распределение $S(n-1, m-1)$, что следует из теоремы Фишера (см. (22)).

Среднее и дисперсия случайной величины $F_{n,m}$ имеют вид (при $m > 4$)

$$\mathbf{E}F_{n,m} = \frac{m}{m-2} \quad \text{и} \quad \mathbf{D}F_{n,m} = \frac{2m^2(n+m-2)}{n(m-4)(m-2)^2}. \tag{35}$$

8. Распределение Вейбулла

Распределение Вейбулла $W(a, \alpha, b)$ зависит в общем случае от трех параметров: *параметра сдвига (положения)* $a \in R^1$, *параметра формы* $\alpha > 0$ и *параметра масштаба* $b > 0$ и задается функцией распределения

$$F_\xi(x) = 1 - \exp\left\{-\left(\frac{x-a}{b}\right)^\alpha\right\}, \quad x \geqslant a \tag{36}$$

($F_\xi(x) = 0$ при $x < a$). Здесь

$$\mathbf{E}\xi = a + b\,\Gamma\left(1 + \frac{1}{\alpha}\right) \quad \text{и} \quad \mathbf{D}\xi = b^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right] \tag{37}$$

Частный случай $W(a, 1, b)$ известен как *двухпараметрическое показательное (экспоненциальное) распределение*, а $W(a, 2, b)$ — как *распределение Релея*.

9. Распределение Парето

Случайная величина ξ имеет *распределение Парето*[8] *с параметрами* x_0 *и* α ($x_0, \alpha > 0$), если ее плотность такова:

$$f(x|x_0, \alpha) = \begin{cases} \dfrac{\alpha x_0^\alpha}{x^{\alpha+1}} & \text{при} \quad x \geqslant x_0, \\ 0 & \text{в остальных случаях.} \end{cases} \tag{38}$$

Это распределение применяется в качестве подходящей модели для широкого класса эмпирических наблюдений, относящихся, например, к распределению дохода, размера городов или частоты слов в языке.

Если в (38) $\alpha > 2$, то справедливы следующие формулы для первых двух моментов:

$$\mathbf{E}\xi = \frac{\alpha x_0}{\alpha - 1} \quad \text{и} \quad \mathbf{D}\xi = \frac{\alpha x_0^2}{(\alpha-1)^2(\alpha-2)}. \tag{39}$$

10. Распределение Дирихле

Распределение Дирихле[9] $D(\underline{\alpha})$, где $\underline{\alpha} = (\alpha_1, \ldots, \alpha_k)$, $\alpha_i > 0$, $i = 1, \quad, k$, — *параметрический вектор*, задается плотностью

$$f(\underline{x}|\underline{\alpha}) = \frac{\Gamma(\alpha_1 + \quad + \alpha_k)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_k)} x_1^{\alpha_1 - 1} \quad x_k^{\alpha_k - 1} \quad \underline{x} = (x_1, \quad, x_k) \in \mathcal{E}, \tag{40}$$

[8] Парето Вильфредо (1848–1923) — итальянский экономист и социолог.

[9] Дирихле Петер Густав (1805–1859) — немецкий математик.

где $\mathcal{E}$ — единичный симплекс:

$$\mathcal{E} = \{\underline{x} \;\; 0 \leqslant x_i \leqslant 1,\ i = 1, \quad , k,\ x_1 + \quad + x_k = 1\}.$$

Сравнение формулы (40) с формулой (23) для плотности бета-распределения показывает, что распределение Дирихле при $k = 2$ сводится к бета-распределению, поэтому распределение $D(\underline{\alpha})$ иногда называют *многомерным бета-распределением*.

Если $\mathcal{L}(\underline{\xi} = (\xi_1, \ldots, \xi_k)) = D(\underline{\alpha})$, то первые и вторые моменты случайного вектора $\underline{\xi}$ имеют вид (далее обозначено $\alpha = \alpha_1 + \quad + \alpha_k$):

$$\begin{gathered} \mathbf{E}\xi_i = \frac{\alpha_i}{\alpha}, \quad \mathbf{D}\xi_i = \frac{\alpha_i(\alpha - \alpha_i)}{\alpha^2(\alpha + 1)}, \\ \operatorname{cov}(\xi_i, \xi_j) = -\frac{\alpha_i \alpha_j}{\alpha^2(\alpha + 1)}, \quad i \neq j, \end{gathered} \tag{41}$$

при этом $\mathcal{L}(\xi_i) = \mathrm{Be}(\alpha_i, \alpha - \alpha_i)$ (см. п. 4), $i = 1, \quad , k$.

Замечание. Обратим внимание на то, что в данном случае координаты вектора $\underline{\xi}$ связаны линейным соотношением $\xi_1 + \ldots + \xi_k = 1$, из которого одна из них, скажем ξ_k, может быть выражена через остальные: $\xi_k = 1 - \xi_1 - \ldots - \xi_{k-1}$. Если, в соответствии с этим, в (40) переменную x_k записать как $1 - x_1 - \ldots - x_{k-1}$, то $f(\underline{x}|\underline{\alpha})$ будет обычной $(k-1)$-мерной плотностью распределения случайного вектора $(\xi_1, \ldots, \xi_{k-1})$, которая положительна в тех точках симплекса $S \subset R^{k-1}$, где $0 \leqslant x_i \leqslant 1$, $i = 1, \ldots, k-1$, и $0 \leqslant x_1 + \quad + x_{k-1} \leqslant 1$. Именно при такой интерпретации формулы (40) мы получаем при $k = 2$ представление (23). Далее, так как интеграл от плотности $f(\underline{x}|\underline{\alpha})$ по симплексу S должен быть равен 1, то мы получаем формулу

$$\int_S x_1^{\alpha_1 - 1} \ldots x_{k-1}^{\alpha_{k-1} - 1} (1 - x_1 - \ldots - x_{k-1})^{\alpha_k - 1}\, dx_1 \ldots dx_{k-1} = \frac{\Gamma(\alpha_1) \ldots \Gamma(\alpha_k)}{\Gamma(\alpha_1 + \ldots + \alpha_k)}, \tag{42}$$

Формула Дирихле

известную в анализе как *формула (интеграл) Дирихле* (этот интеграл был впервые исследован П. Дирихле в 1839 г. и этим объясняется название распределения $D(\underline{\alpha})$).

11. Преобразования случайных величин и векторов

Мы завершаем разговор об основных вероятностных распределениях техническим, но совершенно необходимым для последующего разделом, в котором мы напомним некоторые формулы из теории вероятностей, которые часто используются для нахождения явного вида распределений при преобразованиях случайных величин. Выше мы уже встречались с примерами таких преобразований. Так, в п. 2 мы установили, как вычислить распределение линейно преобразованного нормального случайного вектора. Но есть и задачи (с ними мы встретимся тоже), в которых необходимо иметь дело с преобразованиями не линейными, а более сложными, как говорят, функциональными. Поэтому, хотя соответствующие формулы гораздо сложнее, но без них нам не обойтись, и мы их здесь приведем для последующих ссылок на них.

Сдвиг-масштабное преобразование

Начнем с простейшего случая *сдвиг-масштабного преобразования* случайной величины ξ. Пусть ξ имеет непрерывную функцию распределения $F_\xi(x) = \mathbf{P}\{\xi \leqslant x\}$ Рассмотрим произвольное линейное преобразование $\eta = a\xi + b$, где $a > 0$, $b \in R^1$ Тогда очевидно, что функция распределения преобразованной случайной величины η равна

$$F_\eta(x) = F_\xi\left(\frac{x-b}{a}\right).$$

Если, как и выше, через $f(x)$ обозначать плотность исходной величины ξ $\left(f(x) \quad F'_\xi(x)\right)$, а через $g(x)$ — плотность преобразованной величины η $\left(g(x) = F'_\eta(x)\right)$, то в терминах плотностей предыдущее запишется так:

$$g(x) = \frac{1}{a} f\left(\frac{x-b}{a}\right). \tag{43}$$

В этом случае b называют *параметром сдвига* (*положения*), a — *параметром масштаба*. В статистических задачах часто считается, что исходная плотность f нам известна, а параметры a и b (или один из них) неизвестны, такую модель (43) называют *сдвигомасштабной*: в этом случае наблюдаемая случайная величина η предполагается полученной с помощью некоторого линейного преобразования над «стандартной» случайной величиной ξ, распределение которой полностью известно. Таким образом, сдвигомасштабная модель является специальной формой общей параметрической модели. Примерами сдвигомасштабных моделей являются нормальная модель $\mathcal{N}(\mu, \sigma^2)$ (здесь $a = \sigma$, $b = \mu$, а стандартным является распределение $\mathcal{N}(0, 1)$, задаваемое формулами (7)), показательная модель $\Gamma(a, 1)$ (здесь имеется только один параметр масштаба a, а стандартным является распределение $\Gamma(1, 1)$, определенное в п. 3), равномерное распределение $\mathcal{U}(a, b)$ (здесь «стандартом» является $\mathcal{U}(0, 1)$) и т. д. Сдвигомасштабные преобразования часто используются в задачах теории вероятностей и математической статистики.

Общее функциональное преобразование

Перейдем теперь к общему случаю произвольного функционального преобразования. Пусть случайный вектор $\underline{X} = (X_1, \quad, X_k)$ имеет абсолютно непрерывное распределение с плотностью $f(\underline{x})$, $\underline{x} = (x_1, \ldots, x_k) \in$ $\in S \subseteq R^k$ и $\underline{h} = (h_1, \ldots, h_k)$ $S \to R^k$ — произвольное взаимно однозначное и гладкое (т. е. все частные производные $\partial h_i(\underline{x})/\partial x_j$ непрерывны) преобразование, якобиан которого

$$J(\underline{x}) = \det \begin{Vmatrix} \dfrac{\partial h_1(\underline{x})}{\partial x_1} & \dfrac{\partial h_k(\underline{x})}{\partial x_1} \\ \cdot & \cdot \\ \dfrac{\partial h_1(\underline{x})}{\partial x_k} & \dfrac{\partial h_k(\underline{x})}{\partial x_k} \end{Vmatrix}$$

не обращается на S в нуль. Тогда плотность распределения преобразованного случайного вектора $\underline{Y} = \underline{h}(\underline{X}) = \big(h_1(\underline{X}), \quad , h_k(\underline{X})\big)$ имеет следующий вид:

$$g(\underline{y}) = \frac{f\big(\underline{h}^{-1}(\underline{y})\big)}{\big|J\big(\underline{h}^{-1}(\underline{y})\big)\big|}, \quad \underline{y} = (y_1, \quad , y_k) \in \underline{h}(S), \tag{44}$$

где $\underline{h}^{-1}$ — обратное к $\underline{h}$ преобразование: $\underline{h}^{-1}\big(\underline{h}(\underline{x})\big) \equiv \underline{x}$.

Частные случаи

Эта «страшная» (в силу своей общности) формула на самом деле чаще всего применяется в более «прозрачных» частных случаях, два из которых мы выделим особо. Если $k = 1$, то речь идет о преобразовании случайной величины: $Y = h(X)$, где $h(x)$ — взаимно однозначная гладкая функция с необращающейся в нуль производной. В этом случае формула (44) для плотности Y принимает вид

$$g(y) = \frac{f\big(h^{-1}(y)\big)}{\big|h'\big(h^{-1}(y)\big)\big|}. \tag{45}$$

Линейное преобразование

Для случая линейного преобразования: $\underline{Y} = A\underline{X} + \underline{b}$ при $|A| \equiv a \neq 0$ формула (44) конкретизируется к виду

$$g(\underline{y}) = \frac{f\big(A^{-1}(\underline{y} - \underline{b})\big)}{|a|} \tag{46}$$

(для одномерного случая $k = 1$ формула (46) сводится к (43)).

Эти два случая ($k = 1$ и линейное преобразование) чаще всего и встречаются в приложениях. В заключение приведем два иллюстративных (но полезных для дальнейшего) примера использования этих формул.

Пример 1. Пусть $\mathcal{L}(\xi) = \mathcal{U}(0, 1)$, т. е. ξ имеет стандартное равномерное распределение, и $\eta = -\ln \xi$. Здесь $f(x) = 1$, $x \in (0, 1)$, $h(x) = -\ln x$, $h'(x) = -1/x$, $h^{-1}(y) = e^{-y}$ $y > 0$, поэтому по формуле (45) плотность преобразованной случайной величины η есть e^{-y}, $y > 0$. Таким образом, в данном случае $\mathcal{L}(\eta) = \Gamma(1, 1)$, т. е. η имеет стандартное показательное распределение. Этот факт используется при моделировании показательного распределения, о чем будет идти речь в следующем параграфе. •

Получение нормальных случайных величин из равномерных

Пример 2 (Немного более сложный, но поучительный и очень полезный!). Пусть случайные величины X_1 и X_2 независимы и $\mathcal{L}(X_i) = \mathcal{U}(0, 1)$, $i = 1, 2$. Рассмотрим их преобразование

$$Y_1 = \sqrt{-2 \ln X_1} \sin(2\pi X_2), \quad Y_2 = \sqrt{-2 \ln X_1} \cos(2\pi X_2) \tag{47}$$

и найдем распределение пары $Y = (Y_1, Y_2)$. Здесь

$$y_1 = h_1(x_1, x_2) = \sqrt{-2 \ln x_1} \sin(2\pi x_2),$$
$$y_2 = h_2(x_1, x_2) = \sqrt{-2 \ln x_1} \cos(2\pi x_2),$$
$$(x_1, x_2) \in (0, 1)^2,$$

и все условия применимости формулы (44) выполнены. Имеем

$$\frac{\partial h_1(x_1, x_2)}{\partial x_1} = -\frac{\sin(2\pi x_2)}{x_1\sqrt{-2\ln x_1}},$$

$$\frac{\partial h_2(x_1, x_2)}{\partial x_1} = -\frac{\cos(2\pi x_2)}{x_1\sqrt{-2\ln x_1}},$$

$$\frac{\partial h_1(x_1, x_2)}{\partial x_2} = 2\pi\sqrt{-2\ln x_1}\cos(2\pi x_2),$$

$$\frac{\partial h_2(x_1, x_2)}{\partial x_2} = -2\pi\sqrt{-2\ln x_1}\sin(2\pi x_2),$$

откуда находим якобиан $J(x_1, x_2)$ нашего преобразования: это определитель

$$J(x_1, x_2) = \begin{vmatrix} -\dfrac{\sin(2\pi x_2)}{x_1\sqrt{-2\ln x_1}} & -\dfrac{\cos(2\pi x_2)}{x_1\sqrt{-2\ln x_1}} \\ \cdot & \cdot\cdot\cdot\cdot\cdot \\ 2\pi\sqrt{-2\ln x_1}\cos(2\pi x_2) & -2\pi\sqrt{-2\ln x_1}\sin(2\pi x_2) \end{vmatrix} = \frac{2\pi}{x_1}.$$

Поскольку совместная плотность пары (X_1, X_2) в данном случае есть произведение равномерных плотностей на интервалах соответственно $0 < x_1 < 1$ и $0 < x_2 < 1$, то $f(x_1, x_2) = 1$, $(x_1, x_2) \in (0, 1)^2$, и в силу (44) плотность $g(y_1, y_2)$ преобразованной пары (Y_1, Y_2) есть $x_1/2\pi$ Осталось выразить x_1 через (y_1, y_2). В данном случае это очень просто, так как у нас

$$y_1^2 + y_2^2 = -2\ln x_1.$$

Отсюда

$$x_1 = \exp\left\{-\frac{1}{2}(y_1^2 + y_2^2)\right\},$$

и окончательно получаем, что искомая плотность имеет вид

$$g(y_1, y_2) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(y_1^2 + y_2^2)\right\}, \quad (y_1, y_2) \in R^2$$

Но это есть двумерная стандартная нормальная плотность (см. п. 2), т. е.

$$Y = (Y_1, Y_2)$$

является стандартным нормальным случайным вектором (его компоненты независимы и $\mathcal{L}(Y_i) = \mathcal{N}(0, 1)$, $i = 1, 2$).

Таким образом с помощью «хитрого» преобразования (47) мы можем получать из независимых стандартных равномерно распределенных случайных величин (т. е. в определенном плане — простейших случайных величин) независимые же стандартные нормальные случайные величины. Этот факт также используется в алгоритмах статистического моделирования (см. следующий параграф). •

Замечание. В связи с последним примером обратим внимание заинтересованного читателя на следующий нюанс. В данном случае нам не потребовалось вычислять полностью обратное преобразование $\underline{h}^{-1}(\underline{y})$, т. е. находить обе обратные функции $x_1 = x_1(y_1, y_2)$ и $x_2 = x_2(y_1, y_2)$: поскольку плотность f здесь постоянна, а якобиан $J(x_1, x_2)$ зависит лишь от x_1, то достаточно знать лишь первую обратную функцию $x_1 = x_1(y_1, y_2)$. Подобное обстоятельство встречается достаточно часто в задачах, связанных с использованием формулы (44). Поэтому в конкретных задачах не надо торопиться вычислять $\underline{h}^{-1}(\underline{y})$: сначала надо вычислить якобиан $J(\underline{x})$ и получить явное выражение для $f(\underline{x})/|J(\underline{x})|$, а затем, в зависимости от вида этого выражения, уже решать задачу его представления через новую переменную $\underline{y}$. В общем, как говорили наши умные предки: «Festina lente» — торопись медленно (Октавиан Август (63 г. до н. э. – 14 г. н. э.), римский император).

Дальнейшие свойства рассмотренных конкретных вероятностных распределений (как точные, так и асимптотические), а также связи между ними обсуждаются в упражнениях в конце главы.

§ 1.3. Моделирование выборок из конкретных распределений

Точность результата не может быть выше точности исходных данных.

Для изучения и практической иллюстрации эффективности различных статистических процедур необходимо иметь наборы конкретных экспериментальных данных, которые можно было бы рассматривать в качестве реализаций некоторой случайной величины ξ с известным законом распределения $\mathcal{L}(\xi)$. Обрабатывая эти данные по соответствующему статистическому алгоритму, мы в этом случае получаем возможность сравнить предсказание теории с известными нам параметрами этого распределения и тем самым «подвергнуть теорию практической проверке». Такой «игровой» элемент при организации практической работы по статистической обработке данных повышает интерес и к изучению самой теории статистического вывода. Для получения таких исходных данных (выборок) широко используется *метод статистического моделирования*, с помощью которого, используя современные компьютеры, можно быстро «смоделировать» выборку $\boldsymbol{X} = (X_1, \ldots, X_n)$ из любого заданного распределения $\mathcal{L}(\xi)$. Описанию соответствующих алгоритмов моделирования и посвящен настоящий параграф.

1. Предварительные замечания

Обычно для получения реализации последовательности независимых случайных чисел с нужным нам распределением используют реализацию последовательности независимых и равномерно распределенных на интервале $[0, 1]$ случайных чисел

$$U_0, U_1, U_2, \ldots . \tag{1}$$

Датчики случайных чисел

В свою очередь, для этого необходимо иметь практически неограниченный источник, вырабатывающий последовательность (1), — *датчик случайных чисел*. В современных компьютерах, как правило, имеются встроенные физические датчики случайных чисел, принцип работы которых основан на использовании в качестве источников случайности быстро флуктуирующих шумовых процессов, вырабатываемых радиоэлементами. Отметим, что получаемые при этом числа — дискретные по своей природе, так как принимают значения в пределах разрядности компьютера (так, при двоичном представлении чисел с N разрядами в ячейках случайное число может принимать 2^N возможных значений).

Псевдослучайные числа

Однако, чаще всего в качестве чисел (1) используют, так называемые, *псевдослучайные числа*, т. е. детерминированные числа, получаемые по некоторому алгоритму и обладающие в той или иной мере свойствами случайных чисел, так сказать, «похожие» на них. Имеются различные методы получения псевдослучайных чисел, наиболее известным из которых является так называемый *линейный конгруэнтный метод*, в котором полагают $U_n = Z_n/m$, где $\{Z_n\}$ — последовательность, определяемая рекуррентным соотношением

$$Z_{n+1} = aZ_n + c \,(\mathrm{mod}\, m), \quad n = 0, 1, 2,$$

Z_0 — начальное значение, a, c, m — натуральные числа (операция приведения чисел по модулю (mod) означает, что если $r = sm + l$, $0 \leqslant l < m$, то $r \,(\mathrm{mod}\, m) = l$). Числа $\{U_n\}$, получаемые таким способом, конечно, не являются случайными (они детерминированные), но при специальном выборе констант a, c, m и Z_0 свойства этих чисел «очень похожи» на свойства последовательности (1) и потому их используют в качестве исходного материала (т. е. в качестве последовательности (1)) для решения модельных задач, иллюстрирующих различные выводы теории. Приведем условия, которым следуют при практическом выборе значений определяющих линейный конгруэнтный метод констант:

1) c и m — взаимно простые числа;
2) $b = a - 1$ кратно p для любого простого p, являющегося делителем m;
3) b кратно 4, если m кратно 4.

Но «правильный» выбор этих констант еще не «гарантирует» хороших свойств псевдослучайных чисел. Даже в датчиках, рекомендованных для широкого использования, нередко обнаруживаются существенные недостатки. Поэтому «качество» псевдослучайных чисел требуется контролировать при помощи различных статистических критериев (см. пример 7 Введения), общая теория которых излагается в гл. 4 (см. п. 2 § 4.5).

После этих замечаний перейдем непосредственно к алгоритмам моделирования основных вероятностных распределений, описанных в предыдущих параграфах, считая, что в нашем распоряжении имеется неограниченная

последовательность (1) «простейших» случайных величин с распределением $\mathcal{U}(0, 1)$ (см. п. 5 § 1.2).

2. Моделирование распределений Бернулли $\mathrm{Bi}(1, p)$ и биномиального $\mathrm{Bi}(k, p)$ (см. п. 1 § 1.1)

Чтобы получить реализацию бернуллиевской последовательности $\{X_1, X_2, \ldots\}$, где

$$\mathbf{P}\{X_n = 1\} = 1 - \mathbf{P}\{X_n = 0\} = p$$

при заданном $p \in (0, 1)$, очевидно, достаточно положить

$$X_n = I(\mathcal{U}_n \leqslant p), \quad n = 1, 2,$$

(здесь и далее везде $I(A)$ обозначает индикатор события A, т. е. $I(A) = 1$, если A имеет место, и $I(A) = 0$ в противном случае). Действительно, независимость случайных величин $\{X_n\}$ обеспечивается независимостью исходных величин $\{\mathcal{U}_n\}$ и

$$\mathbf{P}\{X_n = 1\} = \mathbf{P}\{\mathcal{U}_n \leqslant p\} = p \quad \text{для любого} \quad n = 1, 2,$$

Имея последовательность Бернулли $\{X_n\}$ легко смоделировать выборку из биномиального распределения $\mathrm{Bi}(k, p)$. Именно, поскольку

$$\mathcal{L}(X_1 + \quad + X_k) = \mathrm{Bi}(k, p),$$

то положив

$$X'_j = X_{(j-1)k+1} + \quad + X_{jk}, \quad j = 1, \quad , n,$$

получим реализацию выборки $(X'_1, \quad , X'_n)$ объема n из распределения $\mathrm{Bi}(k, p)$.

3. Моделирование отрицательного биномиального распределения $\overline{\mathrm{Bi}}(r, p)$ (см. п. 2 § 1.1)

Поскольку число единиц до первого нуля в последовательности Бернулли $\{X_n\}$ имеет геометрическое распределение $\overline{\mathrm{Bi}}(1, p)$, то, положив $Y = \min\{j \geqslant 0 \quad X_{j+1} = 0\}$, будем иметь $\mathcal{L}(Y) = \overline{\mathrm{Bi}}(1, p)$. Отсюда следует, что алгоритм моделирования выборки $(Y_1, \quad , Y_n)$ из распределения $\overline{\mathrm{Bi}}(1, p)$ можно задать формулами

$$Y_i = \min\{j \geqslant 0 \quad X_{Y_{i-1}+j+1} = 0\}, \quad i = 1, \quad , n, \quad Y_0 = 0.$$

Далее, поскольку

$$\mathcal{L}(Y_1 + \quad + Y_r) = \overline{\mathrm{Bi}}(r, p)$$

(свойство воспроизводимости отрицательного биномиального распределения), то смоделировать выборку $(Y'_1, \ . \ , Y'_n)$ из распределения $\overline{\mathrm{Bi}}(r, p)$ можно (как и выше для биномиального распределения) по формулам

$$Y'_j = Y_{(j-1)r+1} + \quad + Y_{jr}, \quad j = 1, \quad , n,$$

т. е. просуммировав числа $\{Y_i\}$ последовательными группами по r подряд идущих чисел.

4. Моделирование полиномиальных испытаний (см. п. 6 § 1.1)

Пусть требуется смоделировать n независимых наблюдений над дискретной случайной величиной ξ с распределением

$$\mathbf{P}\{\xi = l\} = p_l, \quad l = 1, \quad , N \quad (p_1 + \quad + p_N = 1).$$

Разобьем отрезок $[0, 1]$ на N интервалов $\mathcal{E}_1, \quad , \mathcal{E}_N$, где

$$\mathcal{E}_1 = [0, p_1], \quad \mathcal{E}_l = (p_1 + \quad + p_{l-1}, \ p_1 + \quad + p_l], \quad l = 2, \quad , N$$

(так что $\mathcal{E}_l$ имеет длину p_l, $l = 1, \quad , N$), и положим

$$X_i = \sum_{l=1}^{N} lI(\mathcal{U}_i \in \mathcal{E}_l), \quad i = 1, \quad , n. \tag{2}$$

Так как X_i определяется только через $\mathcal{U}_i$, то случайные величины $\{X_i\}$ независимы и одинаково распределены, а

$$\mathbf{P}\{X_i = l\} = \mathbf{P}\{\mathcal{U}_i \in \mathcal{E}_l\} = |\mathcal{E}_l| = p_l,$$

т. е. они распределены так же, как ξ. Таким образом, $(X_1, \quad , X_n)$ — выборка из $\mathcal{L}(\xi)$.

Замечание. В случае произвольных исходов $a_1, \quad , a_N$ случайную величину X_i следует определить по формуле

$$X_i = \sum_{l=1}^{N} a_l I(\mathcal{U}_i \in \mathcal{E}_l).$$

5. Моделирование пуассоновского распределения $\Pi(\lambda)$ (см. п. 3 § 1.1)

Установим предварительно следующий факт.

Лемма. Определим случайную величину ξ равенством

$$\xi = \max\left\{j \ \prod_{i=1}^{j} \mathcal{U}_i \geqslant e^{-\lambda}\right\}, \quad \lambda > 0.$$

Тогда $\mathcal{L}(\xi) = \Pi(\lambda)$.

Доказательство. Перепишем определение ξ в эквивалентном виде

$$\xi = \max\left\{j \quad \sum_{i=1}^{j} \ln \mathcal{U}_i \leqslant \lambda\right\}$$

(при $j = 0$ пустая сумма, как это обычно принято, считается равной 0, а пустое произведение полагается равным 1). Тогда события

$$\{\xi < m\} \quad \text{и} \quad \left\{-\sum_{i=1}^{m} \ln \mathcal{U}_i > \lambda\right\},$$

очевидно, эквивалентны. Далее, в примере 1 в § 1.2 было показано, что

$$\mathcal{L}(-\ln \mathcal{U}_1) = \Gamma(1, 1)$$

— стандартное показательное распределение, а по воспроизводящему свойству гамма-распределения (см. п. 3 § 1.2)

$$\mathcal{L}\left(\sum_{i=1}^{m} \ln \mathcal{U}_1\right) = \Gamma(1, m)$$

— *распределение Эрланга порядка* m, поэтому (см. (18) § 1.2)

$$\mathbf{P}\{\xi < m\} = \mathbf{P}\left\{-\sum_{i=1}^{m} \ln \mathcal{U}_i > \lambda\right\} = \frac{1}{(m-1)!}\int_{\lambda}^{\infty} x^{m-1}e^{-x}\,dx = \sum_{r=0}^{m-1} e^{-\lambda}\frac{\lambda^r}{r!},$$

$$m \geqslant 1.$$

Но это и означает (см. (8) § 1.1), что случайная величина ξ имеет распределение $\Pi(\lambda)$. ■

Из этой леммы следует, что выборку $(X_1, \quad, X_n)$ из распределения $\Pi(\lambda)$ можно смоделировать по формулам

$$X_k = \max\left\{j \;\Big|\; \prod_{i=1}^{j} \mathcal{U}_{X_{k-1}+i} \geqslant e^{-\lambda}\right\}, \quad k = 1, 2, \quad, n, \quad X_0 = 0. \qquad (3)$$

6. Моделирование непрерывных распределений

Для моделирования случайных величин с непрерывным распределением в ряде случаев можно использовать следующий простой факт: если $F(x)$ — непрерывная и строго монотонно возрастающая функция распределения и $F^{-1}(y)$ — обратная к ней функция $(F^{-1}(F(x)) \equiv x)$, то для случайной величины $\eta = F^{-1}(\mathcal{U})$, где $\mathcal{L}(\mathcal{U}) = \mathcal{U}(0, 1)$, имеем

$$\mathbf{P}\{\eta \leqslant x\} = \mathbf{P}\{\mathcal{U} \leqslant F(x)\} = F(x).$$

Таким образом, случайная величина $\eta = F^{-1}(\mathcal{U})$ имеет функцию распределения $F(x)$, и если обратная функция $F^{-1}(y)$ может быть явно вычислена, то $F^{-1}(\mathcal{U}_i)$, $i = 1, \ldots, n$, — выборка из распределения F.

Замечание. Если функция $F(x)$ не является строго монотонной, то вывод остается в силе, если под обратной функцией понимать

$$F^{-1}(y) = \sup\{x \,.\, F(x) \leqslant y\}.$$

Заметим еще, что если $(X_i,\ i = 1, \ldots, n)$ — выборка из непрерывного распределения $F(x)$, то $(aX_i + b,\ i = 1, \quad, n)$, где $a > 0$, $b \in R^1$, будет выборкой из распределения сдвигомасштабного типа $F\big((x-b)/a\big)$ (см. п. 11 § 1.2

соотношение (43)). Так, например, $(a\mathcal{U}_i + b,\ i = 1, \ldots, n)$ — выборка из равномерного распределения $\mathcal{U}(b, a+b)$, а если $\mathcal{L}(X_i) = \mathcal{N}(0, 1)$, т. е. $\{X_i\}$ — стандартные нормальные числа, то $\mathcal{L}(\sigma X_i + \mu) = \mathcal{N}(\mu, \sigma^2)$; в частности, чтобы смоделировать выборку из распределения Коши (см. (31) § 1.2), надо положить

$$X_i = \operatorname{tg}\left(\pi\mathcal{U}_i - \frac{\pi}{2}\right) = -\operatorname{ctg}(\pi\mathcal{U}_i), \quad i = 1, \quad , n.$$

7. Моделирование показательного и связанных с ним распределений (см. п. 3 § 1.2)

Чтобы смоделировать выборку $(X_1, \ldots, X_n)$ из показательного распределения $\Gamma(a, 1)$, надо положить $X_i \quad -a\ln\mathcal{U}_i$, $i = 1, \quad , n$. Суммируя полученные числа по m штук: $Y_1 = X_1 + \ldots + X_m$, $Y_2 = X_{m+1} + \ldots + X_{2m}$, , получим выборку из распределения Эрланга $\Gamma(a, m)$.

8. Моделирование нормальных случайных чисел (см. п. 1 § 1.2)

Большое число статистических задач связано с анализом выборок из нормального распределения $\mathcal{N}(\mu, \sigma^2)$, при этом (см. замечание в п. 6) достаточно уметь моделировать выборки из стандартного нормального распределения $\mathcal{N}(0, 1)$ (см. (7) § 1.2). В данном случае явной формулы для обратной функции $\Phi^{-1}(y)$ к стандартной нормальной функции распределения не существует, поэтому описанный в п. 6 алгоритм моделирования непригоден. Один из методов получения стандартных нормальных случайных чисел из равномерных (1) состоит в построении соответствующего явно вычислимого функционального преобразования. Пример такого преобразования нам уже встречался в примере 2 в § 1.2 (см. там соотношения (47)). Из полученного в этом примере результата следует, что каждая пара $(\mathcal{U}_{2k-1}, \mathcal{U}_{2k})$, $k = 1, 2, \ldots$, чисел последовательности (1) с помощью преобразования

$$\begin{aligned} X_{2k-1} &= \sqrt{-2\ln\mathcal{U}_{2k-1}}\sin(2\pi\mathcal{U}_{2k}), \\ X_{2k} &= \sqrt{-2\ln\mathcal{U}_{2k-1}}\cos(2\pi\mathcal{U}_{2k}) \end{aligned} \tag{4}$$

порождает пару (X_{2k-1}, X_{2k}) независимых стандартных нормальных чисел.

Другой путь получения нормальных чисел из последовательности (1) равномерных чисел основан на использовании центральной предельной теоремы теории вероятностей (см. соотношение (1) § 1.2). В соответствии с этой теоремой можно положить (см. формулы (26) § 1.2, в соответствии с которыми $\mathbf{E}\mathcal{U}_i = 1/2$, $\mathbf{D}\mathcal{U}_i = 1/12$)

$$X_{NK} = \frac{\mathcal{U}_{(k-1)N+1} + \quad + \mathcal{U}_{kN} - N/2}{\sqrt{N/12}}, \quad k = 1, 2, \tag{5}$$

Этот алгоритм, очевидно, легче реализуем, чем предыдущий, однако он приводит лишь к приближенно нормальным числам, так как лишь в пределе при

$N \to \infty$ распределение случайных величин X_{NK} будет стандартным нормальным $\mathcal{N}(0, 1)$. Но практически удовлетворительное приближение к распределению $\mathcal{N}(0, 1)$ получается уже при $N = 12$ — это значение параметра N в алгоритме (5) обычно и используют для конкретных вычислений.

9. Метод суперпозиций

В ряде случаев моделируемое распределение может быть представлено в виде *смеси* более простых распределений, и тогда для моделирования такого распределения может быть использован *метод суперпозиций*, описанию которого посвящен данный пункт. Но сначала дадим определение смеси.

Смеси распределений

Определение Пусть $p_k \geqslant 0$, $\sum\limits_k p_k = 1$, $F_k(x)$ — некоторые функции распределения. Тогда функция распределения

$$F(x) = \sum_k p_k F_k(x)$$

называется смесью функций $F_k(x)$ с весами p_k.

Аналогично определяется смесь распределений в терминах плотностей.

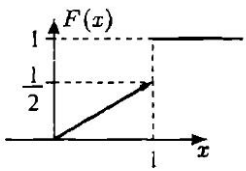


Рис. 1

Пример 1. Функция распределения

$$F(x) = \begin{cases} x/2, & 0 \leqslant x < 1, \\ 1, & x \geqslant 1, \end{cases}$$

(см. рис. 1) является смесью с весами $1/2$ функций распределения равномерного $\mathcal{U}(0, 1)$ и вырожденного в точке $x = 1$ распределений. •

Пример смеси

Пример 2. Пусть распределение на $[0, 1]$ задано плотностью вида

$$f(x) = \sum_{k=0}^{\infty} a_k x^k \quad \text{с} \quad a_k \geqslant 0.$$

Тогда так как

$$\int_0^1 f(x)\,dx = \sum_{k=0}^{\infty} \frac{a_k}{k+1} = 1,$$

то положив $p_k = a_k/(k+1)$, будем иметь

$$f(x) = \sum_{k=0}^{\infty} p_k (k+1) x^k.$$

Но

$$(k+1)x^k = f(x|k+1, 1)$$

— плотность бета-распределения (см. (23) § 1.2), следовательно,

$$f(x) = \sum_{k=0}^{\infty} p_k f(x|k+1, 1),$$

т. е. такое распределение представляет собой смесь бета-распределений $\mathrm{Be}(k+1, 1)$ с весами $p_k = a_k/(k+1)$. •

Метод суперпозиции

Теперь, чтобы смоделировать случайную величину, функция распределения или плотность которой являются смесью, можно использовать **метод суперпозиции**:

1) разыграть значение дискретной случайной величины, принимающей значение k с вероятностью p_k, $k = 0, 1, .$ (как в п. 4), допустим, что при этом было получено значение k_0;
2) смоделировать случайную величину с функцией распределения F_{k_0} (плотностью f_{k_0}) некоторым способом (скажем, методом обратной функции, см. п. 6).

Полиномы Бернштейна

Для произвольной **непрерывной** плотности $f(x)$, заданной на отрезке $[0, 1]$, можно использовать способ **приближенного** моделирования методом суперпозиции, основанный на аппроксимации *полиномами Бернштейна* [10]:

$$f_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k}$$

Основой для этого служит известная в анализе

Теорема Вейерштрасса [11] Если функция $f(x)$ непрерывна на $[0, 1]$, то $f_n(x) \to f(x)$ равномерно по x при $n \to \infty$.

Применение теоремы Вейерштрасса

По этой теореме непрерывную на отрезке $[0, 1]$ плотность $f(x)$ можно равномерно приблизить плотностями — *нормированными полиномами Бернштейна*

$$\widetilde{f}_n(x) = \frac{f_n(x)}{d_n}, \quad \text{где} \quad d_n = \int_0^1 f_n(x)\,dx,$$

[10] Бернштейн Сергей Натанович (1880–1968) — советский математик, академик АН СССР, выдающийся специалист по теории дифференциальных уравнений, теории приближения функций многочленами и теории вероятностей.

[11] Вейерштрасс Карл Теодор Вильгельм (1815–1897) — немецкий математик, выдающийся специалист по математическому анализу, теории функций и линейной алгебре. Почетный член Петербургской АН (1864).

так как из равномерной сходимости $f_n(x)$ к $f(x)$ следует сходимость

$$d_n \to \int_0^1 f(x)\,dx = 1 \quad \text{при} \quad n \to \infty.$$

Для моделирования случайной величины $\widetilde{\xi}_n$ с плотностью $\widetilde{f}_n(x)$ методом суперпозиции представим $\widetilde{f}_n(x)$ в виде

$$\widetilde{f}_n(x) = \sum_{k=0}^{n} \frac{f(k/n)}{d_n(n+1)} \left[(n+1)C_n^k x^k (1-x)^{n-k}\right]$$

Здесь в квадратных скобках стоит плотность $f(x|k+1,\ n-k+1)$ бета-распределения $\mathrm{Be}(k+1,\ n-k+1)$ (см. п. 4 § 1.2), поэтому

$$1 = \int_0^1 \widetilde{f}_n(x)\,dx = \sum_{k=0}^{n} \frac{f(k/n)}{d_n(n+1)},$$

откуда

$$d_n = \frac{1}{k+1} \sum_{k=0}^{n} f\left(\frac{k}{n}\right).$$

Следовательно, положив

$$p_k = f\left(\frac{k}{n}\right) \Big/ \sum_{k=0}^{n} f\left(\frac{k}{n}\right),$$

получим представление

$$\widetilde{f}_n(x) = \sum_{k=0}^{n} p_k f(x|k+1,\ n-k+1),$$

которое позволяет применить метод суперпозиции для моделирования случайной величины $\widetilde{\xi}_n$

10. Моделирование цепи Маркова

Методом статистического моделирования можно моделировать и зависимые испытания с различным характером зависимости. Мы рассмотрим для иллюстрации простейший тип стохастической зависимости — *марковскую* зависимость. Напомним, что последовательность случайных величин $\{\nu_t, t = 0, 1, 2, \ \}$, принимающих значения из конечного множества $\mathcal{E} = \{1, 2, \ , N\}$, называемого *пространством состояний*, является *цепью Маркова*, если для любого $t = 0, 1, 2, \ldots$ распределение вероятностей случайной величины ν_{t+1} зависит лишь от того, какое значение приняла предыдущая величина ν_t, и не зависит от более ранних величин ν_τ, $\tau < t$. Формально это можно выразить через условные вероятности так:

$$\mathrm{P}\{\nu_{t+1} = i_{t+1} | \nu_\tau = i_\tau,\ \tau \leqslant t\} = \mathrm{P}\{\nu_{t+1} = i_{t+1} | \nu_t = i_t\}, \quad i_t, i_{t+1} \in \mathcal{E}. \quad (6)$$

Если при этом вероятности в правой части (6) не зависят от «времени» t, то цепь Маркова называется *однородной*. Обычно рассматривают только однородные цепи, которые являются подходящей моделью многих случайных явлений; такую модель ввел в 1907 г. А. А. Марков, именем которого она и называется.

Если для однородной цепи Маркова $\{\nu_t\}$ обозначить вероятность перехода из состояния i в состояние j через

$$p_{ij} = \mathbf{P}\{\nu_{t+1} = j | \nu_t = i\}, \quad i, j \in \mathcal{E}, \tag{7}$$

то эти числа образуют *матрицу переходных вероятностей* $\mathbf{P} = \|p_{ij}\|_1^N$, элементы которой обладают следующими свойствами:

$$p_{ij} \geqslant 0, \quad \sum_{j=1}^{N} p_{ij} = 1, \quad i = 1, \ldots, N;$$

Стохастическая [redacted] матрица

такая матрица называется *стохастической*. Чтобы полностью определить цепь Маркова, надо еще задать *вектор начальных вероятностей* $\underline{p} = (p_1, \ldots, p_N)$, где $p_i = \mathbf{P}\{\nu_0 = i\}$, $i = 1, \ldots, N$. Зная вектор $\underline{p}$ и матрицу $\mathbf{P}$ можно легко вычислить совместное распределение величин $\nu_0, \nu_1, \ldots, \nu_T$ при любом T:

$$\mathbf{P}\{\nu_0 = i_0, \ \nu_1 = i_1, \ \nu_2 = i_2, \quad \ldots \quad \nu_T = i_T\} = p_{i_0} p_{i_0 i_1} p_{i_1 i_2} \quad \ldots \quad p_{i_{T-1}} \tag{8}$$

Моделирование [redacted] цепи

Пусть теперь требуется смоделировать реализацию однородной цепи Маркова $\{\nu_t, \ t = 0, 1, \quad \ldots, T\}$ с произвольной стохастической матрицей переходных вероятностей $\mathbf{P} = \|p_{ij}\|_1^N$ и вектором начальных вероятностей $\underline{p} = (p_1, \ldots, p_N)$. Для этого мы используем модифицированный алгоритм моделирования полиномиальных испытаний (см. п. 4). Именно, построим сначала систему разбиений отрезка $[0, 1]$ на N интервалов в соответствии с вектором $\underline{p}$ и строками матрицы $\mathbf{P}$:

$$\mathcal{E}^{(0)} = \{\mathcal{E}_1^{(0)}, \ldots, \mathcal{E}_N^{(0)}\}, \quad \mathcal{E}^{(i)} = \{\mathcal{E}_1^{(i)}, \ldots, \mathcal{E}_N^{(i)}\}, \quad i = 1, \ldots, N,$$

где $|\mathcal{E}_j^{(0)}| = p_j$, $|\mathcal{E}_j^{(i)}| = p_{ij}$, $j = 1, \quad \ldots, N$, и определим случайные величины

$$\zeta^{(0)}(\mathcal{U}) = \sum_{j=1}^{N} j I(\mathcal{U} \in \mathcal{E}_j^{(0)}), \quad \zeta^{(i)}(\mathcal{U}) = \sum_{j=1}^{N} j I(\mathcal{U} \in \mathcal{E}_j^{(i)}), \tag{9}$$

$$i = 1, \quad \ldots, N,$$

где случайная величина $\mathcal{U}$ равномерно распределена на отрезке $[0, 1]$.

Так построенные случайные величины $\zeta^{(\cdot)}$ (дзета) будут иметь распределения на множестве $\mathcal{E} = \{1, 2, \ldots, N\}$, задаваемые соответственно вектором $\underline{p}$ и соответствующими строками матрицы $\mathbf{P}$:

$$\mathbf{P}\{\zeta^{(0)} = j\} = p_j, \quad \mathbf{P}\{\zeta^{(i)} = j\} = p_{ij}, \quad j = 1, \ldots, N, \quad i = 1, \ldots, N.$$

После этого уже можно моделировать цепь с использованием случайных чисел (1) по следующему алгоритму:

$$\nu_0 = \zeta^{(0)}(\mathcal{U}_0), \quad \nu_1 = \zeta^{(\nu_0)}(\mathcal{U}_1),$$
$$\nu_2 = \zeta^{(\nu_1)}(\mathcal{U}_2), \qquad \nu_T = \zeta^{(\nu_{T-1})}(\mathcal{U}_T). \tag{10}$$

Действительно, здесь

$$\mathbf{P}\{\nu_0 = j\} = \mathbf{P}\{\zeta^{(0)} = j\} = p_j$$

и

$$\mathbf{P}\{\nu_{t+1} = j | \nu_t = i\} = \mathbf{P}\{\zeta^{(i)} = j\} = p_{ij},$$

что и требовалось.

Примеры моделирования цепи

Пример 3 (Случайное блуждание с поглощением). Частица, стартуя в момент $t=0$ из точки k, $0<k<N$, блуждает по целым точкам отрезка $[0, N]$. Пусть S_t — положение частицы в момент $t = 0, 1, 2 \quad (S_0 = k)$. Если в момент t частица находилась в точке l $(S_t = l)$ и при этом $1 \leqslant l \leqslant N \quad 1$, то в следующий момент $t+1$ она может оказаться в точке $l+1$ с вероятностью $p(\mathbf{P}\{S_{t+1} = l+1 | S_t = l\} = p)$ и в точке $l-1$ с вероятностью $q = 1-p(\mathbf{P}\{S_{t+1} = = l-1 | S_t = l\} = q)$; в точках 0 и N частица поглощается и блуждание прекращается $(\mathbf{P}\{S_{t+1} = 0 | S_t = 0\} = \mathbf{P}\{S_{t+1} = N | S_t = N\} \quad 1)$. Здесь мы имеем однородную цепь Маркова $\{S_t, t = 0, 1, 2, \ldots\}$ с пространством состояний $\{0, 1, 2, \quad , N\}$, вектором начальных вероятностей $\underline{p} = (0 \quad 010 \quad 0)$, где 1 стоит на $(k+1)$-м месте, и матрицей переходных вероятностей $\|p_{ij}\|_0^N$, в которой $p_{00} = p_{NN} = 1$, $p_{i,i+1} = p$, $p_{i,i-1} = q$, если $1 \leqslant i \leqslant N-1$, и $p_{ij} = 0$ при $|i-j| > 1$.

Таким образом, блуждание частицы по целым точкам отрезка $[0, N]$ с поглощением на концах описывается указанной специальной цепью Маркова. Чтобы смоделировать такое блуждание, надо положить $S_0 = k$, $S_{t+1} = S_t + X_{t+1}$ до тех пор, пока $0 < S_t < N$, где $X_{t+1} = 1$, если $\mathcal{U}_{t+1} \leqslant p$, и $X_{t+1} = -1$, если $\mathcal{U}_{t+1} > p$; если же S_t впервые станет равным 0 либо N, то процесс прекращается. •

Пример 4 (Случайное блуждание с отражением). Пусть переходные вероятности $p_{i,i+1}$, $p_{i,i-1}$ для $1 \leqslant i \leqslant N-1$ и $p_{i,j}$ для $|i-j| > 1$ остаются теми же самыми. Если определить еще $p_{00} = q$, $p_{01} = p$, $p_{NN} = p$, $p_{N,N-1} = q$, то полученная цепь Маркова описывает блуждание частицы по целым точкам отрезка $[0, N]$ с отражением на концах. Моделирование такого блуждания отличается от предыдущего случая только в конечных точках: при $S_t = 0$ полагаем $S_{t+1} = S_t + X_{t+1} = X_{t+1}$, где

$$X_{t+1} = \begin{cases} 1, & \text{если} \quad \mathcal{U}_{t+1} \leqslant p, \\ 0, & \text{если} \quad \mathcal{U}_{t+1} > p, \end{cases}$$

а при $S_t = N$ полагаем $S_{t+1} = S_t + X_{t+1} = N + X_{t+1}$, где

$$X_{t+1} = \begin{cases} 0, & \text{если } U_{t+1} \leqslant p, \\ & \text{если } U_{t+1} > p. \end{cases}$$

•

11. Метод Монте-Карло

Отметим еще один круг задач, при решении которых можно эффективно использовать реализации случайных испытаний. Речь идет о, так называемом, *методе Монте-Карло*, или *методе статистических испытаний* (его название происходит от города Монте-Карло в княжестве Монако, знаменитого своими игорными домами). Этот вычислительный метод часто используется при отыскании значений различных величин (определяемых, например, некоторыми уравнениями или интегралами), когда этим величинам можно придать вероятностную интерпретацию. Сущность метода Монте-Карло состоит в следующем: исходя из смысла вычисляемой величины a, подбирают такую случайную величину ξ, чтобы ее среднее значение совпадало с a: $\mathbf{E}\xi = a$. Далее моделируют выборку $(X_1, \quad , X_n)$ из распределения $\mathcal{L}(\xi)$ и в качестве оценки для a используют выборочное среднее

$$\overline{X} = \frac{1}{n}(X_1 + \quad + X_n).$$

Обоснованием этого метода служит следующее

Утверждение. Пусть $X_1, \quad , X_n$ — независимые наблюдения над случайной величиной ξ, $\mathbf{E}\xi = a$, $\mathbf{D}\xi > 0$ и $\mathbf{E}\xi^4 < \infty$. Тогда распределение случайной величины $T_n = \sqrt{n}(\overline{X} - a)/S$, где $S^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, сходится при $n \to \infty$ к стандартному нормальному распределению $\mathcal{N}(0, 1)$ (см. упр. 44 к гл. 2).

Отсюда следует, что если γ — заданное число, близкое к 1 $(0 < \gamma < 1)$, и величина c_γ удовлетворяет равенству $\Phi(c_\gamma) = (1 + \gamma)/2$, где $\Phi(x)$ — стандартная нормальная функция распределения (см. (7) § 1.2), то

$$\mathbf{P}\{|T_n| < c_\gamma\} = \mathbf{P}\left\{\frac{\sqrt{n}\,|\overline{X} - a|}{S} < c_\gamma\right\} \to \Phi(c_\gamma) - \Phi(-c_\gamma) = 2\Phi(c_\gamma) - 1 = \gamma \tag{11}$$

(здесь использовано равенство $\Phi(-x) = 1 - \Phi(x)$, $\forall\, x \in R^1$).

Соотношение (11) означает, что ошибка в определении a этим методом приблизительно с вероятностью γ не превосходит $c_\gamma S/\sqrt{n}$, когда n достаточно велико.

Пусть, например, требуется вычислить интеграл

$$a = \int\!\!\cdots\!\!\int_{v_r} f(t_1, \ldots, t_r)\, dt_1 \ldots dt_r, \tag{12}$$

где область интегрирования есть r-мерный куб $v_r = \{(t_1, \quad, t_r) \ \ 0 \leqslant t_i \leqslant 1, i = 1, \quad, r\}$ Очевидно, что положив здесь $\xi = f(\mathcal{U}_1, \ldots, \mathcal{U}_r)$, где $\{\mathcal{U}_i\}$ — числа из последовательности (1), получим $a = \mathrm{E}\xi$. Таким образом, в качестве приближенного значения интеграла (12) в данном случае можно взять величину

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad X_k = f\left(\mathcal{U}_{(k-1)r+1}, \quad, \mathcal{U}_{kr}\right), \quad k = 1, \quad, n.$$

Например, этим методом можно получить приближенное значение числа

$$e = \int_0^1 e^x\,dx + 1: \qquad \widehat{e} = \frac{1}{n}\sum_{i=1}^{n} e^{\mathcal{U}_i} + 1,$$

взяв, скажем, $n = 100$ равномерных чисел $\{\mathcal{U}_i\}$ Здесь точность метода легко оценить, сравнив значение оценки $\widehat{e}$ с известным нам точным значением $e = 2{,}71828\ldots$

Упражнения

Свои способности человек может узнать, только попытавшись приложить их.

Сенека [12)]

1 Пусть ξ имеет распределение Бернулли $\mathrm{Bi}(1, p)$. Показать, что $\mathrm{E}\xi$ и $\mathrm{D}\xi$ имеют вид (2) § 1.1, а характеристическая функция $\psi(t) = \mathrm{E}e^{it\xi} = pe^{it} + q$. Убедиться в справедливости следующих формул:

$$\mathrm{E}\xi = \frac{1}{i}\psi'(0), \quad \mathrm{D}\xi = -\psi''(0) + (\psi'(0))^2 \tag{1}$$

2 Пусть X имеет биномиальное распределение $\mathrm{Bi}(n, p)$. Показать, что ее среднее и дисперсия имеют вид (4) § 1.1, а для ее характеристической функции $\psi(t)$ справедливо представление $\psi(t) = (pe^{it} + q)^n$ Проверить справедливость формул (1).

3 Доказать свойство воспроизводимости распределения $\mathrm{Bi}(n, p)$ по параметру n: если $X_1, \ldots, X_k$ — независимые биномиальные случайные величины и $\mathcal{L}(X_j) = \mathrm{Bi}(n_j, p)$, $j = 1, \ldots, k$, то $\mathcal{L}(X_1 + \ldots + X_k) = \mathrm{Bi}(n_1 + \ldots + n_k, p)$.

◀ *Указание.* Воспользоваться критерием независимости для характеристических функций (см. (15) § 1.2) и результатом упражнения 2. ▶

4 Доказать, что биномиальные вероятности $f(x|n, p)$ (см. (3) § 1.1) при изменении x от 0 до n сначала монотонно возрастают, затем монотонно убывают, достигая наибольшего значения при $x = m$, где m определяется неравенствами $(n+1)p - 1 < m \leqslant (n+1)p$; если же $m = (n+1)p$, то наибольшее значение достигается дважды: $f(m-1|n, p) = f(m|n, p)$.

[12)] Сенека (4 г. до н. э. – 65 г. н. э.) — древнеримский философ, воспитатель Нерона; по его приказу покончил жизнь самоубийством.

5 Убедиться в том, что формула (5) § 1.1 для плотности отрицательной биномиальной величины ξ может быть записана также в виде

$$f(x|r,p) = (-1)^x C_{-r}^x p^x q^r \qquad x = 0, 1, 2, \ldots$$

◀ *Указание.* Воспользоваться замечанием о биномиальных коэффициентах в конце п. 2 § 1.1. ▶

6 Пусть $\mathcal{L}(\xi) = \overline{\mathrm{Bi}}(r,p)$. Показать, что ее характеристическая функция имеет вид

$$\psi(t) = \mathrm{E}e^{it\xi} = \left(\frac{q}{1 - pe^{it}}\right)^r$$

и, воспользовавшись формулами (1), убедиться в справедливости формул (7) § 1.1.

7 Доказать свойство воспроизводимости распределения $\overline{\mathrm{Bi}}(r,p)$ по параметру r: если $X_1, \ldots, X_k$ — независимые случайные величины и

$$\mathcal{L}(X_j) = \overline{\mathrm{Bi}}(r_j, p), \quad j = 1, \ldots, k,$$

то

$$\mathcal{L}(X_1 + \ldots + X_k) = \overline{\mathrm{Bi}}(r_1 + \ldots + r_k, p).$$

◀ *Указание.* Воспользоваться критерием независимости для характеристических функций. ▶

8 Пусть в бернуллиевской последовательности $\{X_j,\ j \geqslant 1\}$ с параметром p случайная величина η_r означает число испытаний до появления r-го нуля включительно. Выписать плотность величины η_r и найти ее среднее и дисперсию.

◀ *Указание.* $\eta_r = \xi + r$, где $\mathcal{L}(\xi) = \overline{\mathrm{Bi}}(r,p)$. ▶

9 Пусть случайная величина ξ имеет распределение Пуассона $\Pi(\lambda)$. Показать, что ее характеристическая функция есть $\exp\{\lambda(e^{it} - 1)\}$ и по формулам (1) вычислить первые два момента ξ.

10 Доказать свойство воспроизводимости распределения $\Pi(\lambda)$: если $X_1, \ldots, X_k$ — независимые пуассоновские случайные величины с параметрами $\lambda_1, \ldots, \lambda_k$ соответственно, то

$$\mathcal{L}(X_1 + \ldots + X_k) = \Pi(\lambda_1 + \quad + \lambda_k).$$

◀ *Указание.* Воспользоваться критерием независимости для характеристических функций. ▶

11 Пусть случайные величины X_1 и X_2 независимы и $\mathcal{L}(X_i) = \Pi(\lambda_i)$, $i = 1, 2$. Доказать следующий факт для условного распределения:

$$\mathcal{L}(X_1 | X_1 + X_2 = n) = \mathrm{Bi}(n,p) \quad \text{при} \quad p = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

12 Доказать следующую предельную теорему (*теорему Пуассона*) о сходимости биномиального распределения к пуассоновскому: если $n \to \infty$, а $p \to 0$ так, что $np \to \lambda > 0$, то $\mathrm{Bi}(n,p) \to \Pi(\lambda)$, т. е. (см. (3) и (8) § 1.1) $f(x|n,p) \to f(x|\lambda)$, $x = 0, 1, 2, .$

◀ *Указание.* Воспользоваться следующим свойством *о непрерывном соответствии между характеристическими функциями и распределениями случайных величин*: пусть имеется последовательность случайных величин $\{\eta_n\}$ и последовательность $\{\psi_n(t)\}$ соответствующих характеристических функций, тогда, если для любого t $\psi_n(t) \to \psi(t)$ при

$n \to \infty$ и $\psi(t)$ есть характеристическая функция некоторой случайной величины η, то $\eta_n \xrightarrow{\mathcal{L}} \eta$. Верно и обратное, т. е. из слабой сходимости $\eta_n \xrightarrow{\mathcal{L}} \eta$ следует сходимость соответствующих характеристических функций.
В рассматриваемом случае воспользоваться упр. 2 и 9 и пределом $(1+\varepsilon_n z)^n \to e^{\alpha z}$ если $n \to \infty$, $\varepsilon_n \to 0$, а $n\varepsilon_n \to \alpha > 0$. ►

13 Доказать следующее утверждение о сходимости отрицательного биномиального распределения к пуассоновскому: $\overline{\mathrm{Bi}}(r,p) \to \Pi(\lambda)$, если $r \to \infty$, а $p \to 0$ так, что $rp \to \lambda > 0$.

◄ *Указание.* См. указание к упр. 12, использовать упр. 6 и 9. ►

14 Пусть независимые случайные величины $X_1, X_2, \ldots$ имеют логарифмическое распределение (21) § 1.1, а независимая от них величина Y имеет распределение Пуассона $\Pi(r \ln(1/(1-\theta)))$. Определим случайную величину Z как сумму случайного числа Y слагаемых из последовательности $\{X_i\}$ $Z = X_1 + X_2 + \ldots + X_Y$, $Z = 0$ при $Y = 0$. Доказать, что $\mathcal{L}(Z) = \overline{\mathrm{Bi}}(r,\theta)$.

◄ *Указание.* Установить сначала, что характеристическая функция суммы $X_1 + \ldots + X_n$ есть $f^n(\theta e^{it})/f^n(\theta)$, где $f(\theta) = \ln(1/(1-\theta))$ (см. (21) и (24) § 1.1), и применить формулу полной вероятности; в итоге получить, что $\mathrm{E}e^{itZ} = \left((1-\theta)/(1-\theta e^{it})\right)^r$ (см. упр. 6). ►

15 Установить справедливость формул (10) § 1.1 для $\mathrm{E}\xi$ и $\mathrm{D}\xi$ гипергеометрической случайной величины.

◄ *Указание.* Воспользовавшись формулой свертки:

$$\sum_x C_A^x C_B^{n-x} = C_{A+B}^n$$

и равенством (далее используется обозначение $(x)_r = x(x-1)\ldots(x-r+1)$, $(x)_0 = 1$) $(x)_r C_A^x = (A)_r C_{A-r}^{x-r}$, получить общее представление для r-го факториального момента

$$\mathrm{E}(\xi)_r = \sum_x (x)_r \frac{C_{a_1}^x C_{a_2}^{n-x}}{C_a^n} = (a_1)_r \frac{C_{a-r}^{n-r}}{C_a^n} = \frac{(a_1)_r (n)_r}{(a)_r}, \quad r = 1, 2, \ldots$$

Далее использовать формулы $\mathrm{E}\xi = \mathrm{E}(\xi)_1$, $\mathrm{D}\xi = \mathrm{E}(\xi)_2 + \mathrm{E}\xi - (\mathrm{E}\xi)^2$ ►

16 Пусть в (9) § 1.1 $a_1 = ap$, $a_2 = aq$, $q = 1-p$, $0 < p < 1$. Убедиться в том, что при $a \to \infty$ $f(x|ap, aq, n) \to C_n^x p^x q^{n-x}$, т. е. гипергеометрическое распределение $H(ap, aq, n)$ сходится к биномиальному распределению $\mathrm{Bi}(n,p)$, когда $a \to \infty$, а n и p фиксированы.

◄ *Указание.* Записать формулу для $f(x|ap, aq, n)$ в виде

$$f(x|ap, aq, n) = C_n^x \frac{p(p-1/a)\ldots(p-(x-1)/a)\cdot q(q-1/a)\ \ (q-(n-x-1)/a)}{(1-1/a)\ \ (1-(n-1)/a)}$$

и перейти здесь к пределу при $a \to \infty$. ►

17 *Оценка числа рыб в озере.* Пусть a — (неизвестное) число рыб в озере. Предположим, что из озера вылавливают a_1 рыб, помечают их и выпускают обратно. При повторном отлове n рыб среди них оказалось x помеченных. Считается, что результаты двух отловов можно рассматривать как случайные бесповторные выборки из совокупности всех рыб в озере. Тогда вероятность того, что второй улов содержит ровно x помеченных рыб, есть как раз гипергеометрическая вероятность $f(x|a_1, a_2, n)$. На практике

числа a_1, n и x наблюдаются. Спрашивается, как по этой информации оценить неизвестное число a рыб в озере? Можно, например, найти такое значение a, при котором вероятность $f(x|a_1, a-a_1, n)$ достигает своего наибольшего значения, поскольку для такого a наше наблюдение имеет максимальную вероятность. Это значение $\hat{a}$ называется *оценкой максимального правдоподобия* для a. Рассматривая отношение

$$\frac{f(x|a_1, a-a_1, n)}{f(x|a_1, a-1-a_1, n)} = \frac{(a-a_1)(a-n)}{a(a-a_1-n+x)},$$

убедиться в том, что оценка $\hat{a}$ есть наибольшее целое число, не превосходящее na_1/x. Так, например, при $a_1 = 1000$, $n = 1000$ и $x = 100$ величина $\hat{a} = 10\,000$.

18 Пусть целочисленный случайный вектор $\underline{\nu} = (\nu_1, \quad, \nu_N)$ имеет полиномиальное распределение $M(n, \underline{p})$. Показать, что его производящая функция

$$\varphi(\underline{t}) = \mathrm{E}\prod_{j=1}^{N} t_j^{\nu_j}$$

есть

$$\varphi(\underline{t}) = (p_1 t_1 + \quad + p_N t_N)^n \qquad \underline{t} = (t_1, \ldots, t_N),$$

а производящая функция подвектора $\underline{\nu_k} = (\nu_1, \ldots, \nu_k)$, $k < N$ имеет вид

$$\varphi(t_1, \quad + t_n, 1, \ldots, 1) = \left(1 + \sum_{j=1}^{k} p_j(t_j - 1)\right)^n$$

Получить отсюда, что

$$\mathcal{L}(\nu_1) = \mathrm{Bi}(n, p_1), \quad \mathcal{L}(\nu_2|\nu_1 = x_1) = \mathrm{Bi}\left(n - x_1, \frac{p_2}{1-p_1}\right)$$

и вообще

$$\mathcal{L}(\nu_j|\nu_l = x_l,\ l = 1, \ldots, j-1) = \mathrm{Bi}\left(n - x_1 - \ldots - x_{j-1}, \frac{p_j}{1 - p_1 - \quad - p_{j-1}}\right),$$
$$2 \leqslant j \leqslant N$$

19 (*Продолжение*). Вывести формулы для смешанных факториальных моментов произвольных порядков компонент вектора $\underline{\nu}$:

$$\mathrm{E}\prod_{j=1}^{N}(\nu_j)_{r_j} = \frac{\partial^{r_1+\ldots+r_N}}{\partial t_1^{r_1} \quad \partial t_N^{r_N}}\varphi(1, \ldots, 1) = (n)_{r_1+\ldots+r_N}\prod_{j=1}^{N} p_j^{r_j}$$

В частности,

$$\mathrm{E}\nu_j = np_j, \quad \mathrm{E}\nu_j(\nu_j - 1) = n(n-1)p_j^2, \quad \mathrm{E}\nu_i\nu_j = n(n-1)p_ip_j, \quad i \neq j.$$

Получить отсюда формулы (15) § 1.1.

20 (*Продолжение*). Обозначим

$$\sigma_{ij} = \begin{cases} p_i(1-p_i) & \text{при } i = j, \\ -p_ip_j & \text{при } i \neq j, \end{cases}$$

и $\Sigma_k = \|\sigma_{ij}\|_1^k$, $k \leqslant N$. Убедиться в том, что $|\Sigma_k| = \det\Sigma_k = p_1 \ldots p_k(1 - p_1 - \ldots - p_k)$, следовательно, $|\Sigma_N| = 0$ (отражение связи $\nu_1 + \quad + \nu_N = n$) и $|\Sigma_k| > 0$ при $k < N$. Проверить, что $\Sigma_{N-1}^{-1} = \|g_{ij}\|_1^{N-1}$, где $g_{ii} = 1/p_i + 1/p_N$, $g_{ij} = 1/p_N$, $i \neq j$.

21 Доказать воспроизводимость полиномиального распределения $M(n, \underline{p})$ по параметру n: если векторы $X_1, \ldots, X_k$ независимы и $\mathcal{L}(X_i) = M(n_i, \underline{p})$, $i = 1, \ldots, k$, то

$$\mathcal{L}(X_1 + \quad + X_k) = M(n_1 + \quad + n_k, \underline{p}).$$

◀ *Указание.* Воспользоваться критерием независимости для характеристических (или, что то же самое, для производящих) функций. ▶

Общее замечание. Ранее (в § 1.2) мы напомнили определение и некоторые свойства характеристических функций, которые используются для изучения свойств любых случайных величин. Но при работе с дискретными случайными величинами часто более удобно использование ***производящих функций***, определяемых соотношением

Производящие функции

$$\varphi_\xi(z) = \mathbf{E}z^\xi = \sum_x z^x \mathbf{P}\{\xi = x\}, \quad |z| \leqslant 1$$

(для многомерной случайной величины $\underline{\xi} = (\xi_1 \ldots \xi_k)$ производящая функция определяется как $\varphi_{\underline{\xi}}(z_1, \ldots, z_k) = \mathbf{E}z_1^{\xi_1} \ldots z_k^{\xi_k}$). Характеристическая функция $\psi_\xi(t) = \mathbf{E}e^{it\xi}$ выражается через производящую функцию равенством $\psi_\xi(t) = \varphi_\xi(e^{it})$, и потому свойства производящих функций аналогичны свойствам характеристических функций. Например, критерий независимости звучит так: компоненты (дискретного) случайного вектора $X = (X_1, \ldots, X_k)$ независимы тогда и только тогда, когда его производящая функция имеет вид

$$\varphi_X(z_1, \ldots, z_k) = \varphi_{X_1}(z_1) \ldots \varphi_{X_k}(z_k).$$

22 Доказать, что между полиномиальным и пуассоновским распределениями существует следующая связь: если случайные величины $\eta_1, \ldots, \eta_N$ независимы и $\mathcal{L}(\eta_j) = \Pi(\lambda_j)$, $j = 1, \ldots, N$, то для их совместного условного распределения справедливо представление

$$\mathcal{L}(\eta_1, \ldots, \eta_N | \eta_1 + \ldots + \eta_N = n) = M(n, p_1, \ldots, p_N),$$

где

$$p_j = \frac{\lambda_j}{\lambda}, \quad j = 1, \ldots, N, \quad \lambda = \lambda_1 + \ldots + \lambda_N.$$

◀ *Указание.* Вычислить условную вероятность

$$\mathbf{P}\{\eta_1 = x_1, \quad \eta_N = x_N | \eta_1 + \quad + \eta_N = n\}$$

и убедиться, что она имеет вид (14) § 1.1. ▶

23 Пусть в схеме полиномиальных испытаний число испытаний есть случайная величина Y с распределением Пуассона $\Pi(\lambda)$. Показать, что частоты исходов $\nu_1, \ldots, \nu_N$ независимы и $\mathcal{L}(\nu_j) = \Pi(\lambda p_j)$, $j = 1, \ldots, N$.

◀ *Указание.* Используя упр. 18 и формулу полной вероятности, вычислить совместную производящую функцию частот $\nu_1, \ldots, \nu_N$; воспользоваться критерием независимости для производящих функций. ▶

24 Показать, что производящая функция распределения степенного ряда (20) § 1.1 есть $\varphi(z) = f(\theta z)/f(\theta)$ и с ее помощью доказать формулы (23) § 1.1 для моментов.

◀ *Указание.* Через производящую функцию моменты находятся по формулам:

$$\mathbf{E}\xi = \varphi'_\xi(1), \quad \mathbf{D}\xi = \varphi''_\xi(1) + \varphi'_\xi(1) - \left(\varphi'_\xi(1)\right)^2. \ ▶$$

25 (*Продолжение*). Найти производящую функцию суммы $X_1 + \quad + X_n$ независимых слагаемых, имеющих одно и то же распределение степенного ряда, и с ее помощью установить формулу (24) § 1.1.

26 Исходя из вида характеристической функции нормальной случайной величины ξ с $\mathcal{L}(\xi) = \mathcal{N}(\mu, \sigma^2)$:

$$\psi(t) = \exp\left\{it\mu - \frac{\sigma^2}{2}t^2\right\}$$

и используя формулы (1), убедиться в справедливости формул (4) § 1.2.

27 С помощью характеристических функций установить свойство (5) § 1.2 нормальных случайных величин.

28 Показать, что если X_1 и X_2 — независимые одинаково распределенные нормальные случайные величины, то их сумма $X_1 + X_2$ и разность $X_1 - X_2$ независимы.

◀ *Указание.* Вычислить $\operatorname{cov}(X_1 + X_2, X_1 - X_2)$ и убедиться в том, что она равна 0. Вообще, если $X_1, \ldots, X_n$ — случайная выборка из нормального распределения $\mathcal{N}(\mu, \sigma^2)$ и $\overline{X} = (1/n)(X_1 + \ldots + X_n)$, то $\overline{X}$ и вектор $(X_1 - \overline{X}, \ldots, X_n - \overline{X})$ независимы, так как все ковариации $\operatorname{cov}(\overline{X}, X_i - \overline{X}) = 0$, $i = 1, \ldots, n$. Отсюда, между прочим, следует независимость $\overline{X}$ и S^2 (см. (22) § 1.2), так как S^2 есть функция лишь от разностей $X_i - \overline{X}$, $i = 1, \ldots, n$. ▶

29 Пусть случайный вектор $\underline{\xi} = (\xi_1, \ldots, \xi_k)$ имеет нормальное распределение

$$\mathcal{N}(\underline{\mu}, \Sigma = \|\sigma_{ij}\|_1^k).$$

Доказать, что $\mathbf{E}\underline{\xi} = \underline{\mu}$, $\mathbf{D}\underline{\xi} = \Sigma$, т.е. $\mathbf{E}\xi_j = \mu_j$, $\operatorname{cov}(\xi_i, \xi_j) = \sigma_{ij}$, $\quad, j = 1, \ldots, k$.

◀ *Указание.* Использовать общую формулу для моментов

$$\frac{\partial^{r_1+\ldots+r_k}}{\partial t_1^{r_1} \quad \partial t_k^{r_k}} \Psi_{\underline{\xi}}(\underline{0}) = i^{r_1+\ldots+r_k}\mathbf{E}\left(\xi_1^{r_1} \ldots \xi_k^{r_k}\right). \blacktriangleright$$

30 (*Продолжение*). Показать, что подвектор $\underline{\xi}_r = (\xi_1, \ldots, \xi_r)$, $1 \leqslant r < k$, имеет нормальное распределение $\mathcal{N}\left((\mu_1, \quad, \mu_r), \Sigma_r = \|\sigma_{ij}\|_1^r\right)$.

◀ *Указание.* Характеристическая функция подвектора $\underline{\xi}_r$ есть

$$\Psi_{\underline{\xi}}(t_1, \ldots, t_r, 0, \ldots, 0),$$

где $\Psi_{\underline{\xi}}(t_1, \ldots, t_k)$ дана в (16) § 1.2. ▶

31 Пусть $X = (X_1, \quad, X_k)$ — стандартный нормальный вектор и U — произвольная $k \times k$-ортогональная матрица (т.е. $U' = U^{-1}$). Показать, что $Y = UX$ также стандартный нормальный вектор.

32 (*Продолжение*). Пусть теперь L — произвольная $m \times k$-матрица и $\underline{a}$ — произвольный m-вектор. Показать, что m-вектор $Y = LX + \underline{a}$ имеет нормальное распределение $\mathcal{N}(\underline{a}, LL')$.

33 Пусть вектор $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ имеет полиномиальное распределение $M(n, \underline{p})$, где $0 < p_j < 1$, $j = 1, \ldots, N$. Обозначим $\nu_j^* = (\nu_j - np_j)/\sqrt{n}$, $j = 1, \ldots, N$, $\underline{\nu^*} = (\nu_1^*, \ldots, \nu_N^*)$ (заметим, что $\nu_1^* + \ldots + \nu_N^* = 0$). Доказать *центральную предельную теорему* для полиномиального распределения: если $n \to \infty$, то $\mathcal{L}(\underline{\nu^*}) \to \mathcal{N}(\underline{0}, \Sigma_N)$, где дисперсионная матрица Σ_N указана в упр. 20.

◀ *Указание.* Воспользоваться свойством о непрерывном соответствии между характеристическими функциями и распределеннями случайных величин (см. указание к упр. 12), которое справедливо также и для векторных случайных величин, и установить сходимость (используя упр. 18)

$$\lim_{n\to\infty} \ln \Psi_{\underline{\nu}^*}(\underline{t}) = -\frac{1}{2}\sum_{j=1}^{N} p_j t_j^2 + \frac{1}{2}\left(\sum_{j=1}^{N} p_j t_j\right)^2 = -\frac{1}{2}\underline{t}'\Sigma_N \underline{t}. \ \blacktriangleright$$

Замечание. Поскольку матрица Σ_N вырождена (см. упр. 20), то предельное нормальное распределение является несобственным, однако для любого подвектора $\underline{\nu}^*_k = (\nu_1^*, \quad, \nu_k^*)$, $k < N$, будем иметь уже сходимость к собственному нормальному распределению: $\mathcal{L}(\underline{\nu}^*_k) \to \mathcal{N}(\underline{0}, \Sigma_k)$. В частности, при $k = 1$ получаем знаменитую *теорему Муавра—Лапласа* о сходимости биномиального распределения к нормальному: $\mathcal{L}(\nu_1^*) \to \mathcal{N}(0, p_1(1-p_1))$ или

$$\mathcal{L}\left(\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}}\right) \to \mathcal{N}(0,1)$$

(поскольку $\mathcal{L}(\nu_1) = \mathrm{Bi}(n, p_1)$, см. упр. 18).

34 (*Продолжение*). Пусть теперь при $n \to \infty$ вероятности $p_j \to 0$, $j = 1, \quad, N-1$, и при этом $np_j \to \lambda_j$, $0 < \lambda_j < \infty$, $j = 1, \quad, N-1$. Это означает, что при большом числе испытаний n исходы с номерами $1, 2, \quad, N-1$ будут встречаться редко (в среднем $\lambda_1, \ldots, \lambda_{N-1}$ раз соответственно). Доказать, что компоненты $\nu_1, \ldots, \nu_{N-1}$ будут асимптотически независимы и при этом $\mathcal{L}(\nu_j) \to \Pi(\lambda_j)$, $j = 1, \ldots, N-1$.

◀ *Указание.* Установить сходимость для характеристических функций

$$\Psi_{\nu_1,\ldots,\nu_{N-1}}(t_1, \quad, t_{N-1}) \to \prod_{j=1}^{N-1} \exp\left\{\lambda_j(e^{it_j} - 1)\right\}.$$

(см. упр. 9 и 18). ▶

35 (*Продолжение*). Доказать, что распределение случайной величины $\eta = n - \nu_N$ сходится к распределению Пуассона $\Pi(\lambda_1 + \quad + \lambda_{N-1})$.

◀ *Указание.* Заметить, что $\eta = \nu_1 + \quad + \nu_{N-1}$ и воспользоваться упр. 10 (или предыдущим результатом). ▶

36 Установнть справедливость формул (21) § 1.2 для моментов гамма-распределения, предварительно получив представление (20) § 1.2 для его характеристической функции и воспользовавшись соотношениями (1) в упр. 1.

37 Пусть $X_1, \ldots, X_k$ — независимые случайные величины и $\mathcal{L}(X_j) = \Gamma(a, \lambda_j)$, $j = 1, \quad, k$. Показать, что для любой постоянной $c > 0$

$$\mathcal{L}\big(c(X_1 + \ldots + X_k)\big) = \Gamma(ca, \lambda_1 + \quad + \lambda_k).$$

◀ *Указание.* Воспользоваться свойствами характеристических функций. ▶

38 Показать, что если $X_1, \ldots, X_n$ — независимые стандартные нормальные величины, то $\mathcal{L}(X_1^2 + \ldots + X_n^2) = \chi^2(n) = \Gamma(2, n/2)$. Основываясь на центральной предельной теореме (см. (1) § 1.2), установить, что при $n \to \infty$

$$\mathcal{L}\left(\frac{X_1^2 + \quad + X_n^2 - n}{\sqrt{2n}}\right) \to \mathcal{N}(0,1).$$

◄ *Указание.* Сначала установить, что $\mathcal{L}(X_1^2) = \Gamma(2, 1/2)$, далее воспользоваться свойством воспроизводимости гамма-распределения. ►

39 Пусть $X_1, \quad , X_k$ — независимые показательные случайные величины с распределениями $\Gamma(a_i, 1)$, $i = 1, \quad , k$, соответственно. Показать, что случайная величина $Y = \min\{X_1, \ldots, X_k\}$ имеет распределение $\Gamma\big((1/a_1 + \quad + 1/a_k)^{-1}, 1\big)$

◄ *Указание.* Вычислить $\mathrm{P}\{Y > y\}$. ►

40 Пусть k-мерный нормальный вектор $\underline{\xi}$ имеет собственное распределение $\mathcal{N}(\underline{\mu}, \Sigma)$. Доказать, что квадратичная форма

$$Q(\underline{\xi}) = (\underline{\xi} - \underline{\mu})'\Sigma^{-1}(\underline{\xi} - \underline{\mu})$$

имеет распределение $\chi^2(k)$.

◄ *Указание.* Рассмотреть линейное преобразование $\underline{Z} = \Lambda^{-1/2}\mathcal{U}'(\underline{\xi} - \underline{\mu})$, переводящее вектор $\underline{\xi}$ в стандартный нормальный вектор (см. п. 2 § 1.2), тогда

$$Q(\underline{\xi}) = \underline{Z}'\underline{Z} = Z_1^2 + \quad + Z_k^2,$$

далее использовать упр. 38. ►

41 Пусть о распределении случайных величин ξ и Λ (лямбда большое) известно, что $\mathcal{L}(\Lambda) = \Gamma(a, r)$, где r — целое, а условное распределение $\mathcal{L}(\xi|\Lambda = \lambda) = \Pi(\lambda)$. Показать, что $\mathcal{L}(\xi) = \overline{\mathrm{Bi}}(r, p)$ при $p = a/(a + 1)$.

◄ *Указание.* Вычислить безусловные вероятности

$$\mathrm{P}\{\xi = k\} = \int_0^\infty e^{-\lambda}\frac{\lambda^k}{k!}\,\frac{\lambda^{r-1}}{\Gamma(r)a^r}e^{-\lambda/a}\,d\lambda, \quad k = 0, 1,$$

и убедиться, что они имеют вид (5) § 1.1. ►

42 Пусть случайная величина ξ имеет бета-распределение $\mathrm{Be}(a, b)$. Вычислить $\mathrm{E}\xi^r(1 - \xi)^s$, $r, s \geqslant 0$, и получить отсюда формулы для $\mathrm{E}\xi$ и $\mathrm{D}\xi$ (24) § 1.2.

◄ *Указание.* Воспользоваться определением бета-функции $B(a, b)$ и свойствами гамма-функции $\Gamma(\lambda)$ (см. п. 3 § 1.2). ►

43 Пусть случайные величины X_1 и X_2 независимы и $\mathcal{L}(X_i) = \Gamma(a, \lambda_i)$, $i = 1, 2$. Доказать, что случайные величины $Y_1 = X_1 + X_2$ и $Y_2 = X_1/(X_1 + X_2)$ независимы и при этом $\mathcal{L}(Y_1) = \Gamma(a, \lambda_1 + \lambda_2)$, а $\mathcal{L}(Y_2) = \mathrm{Be}(\lambda_1, \lambda_2)$.

◄ *Указание.* Применить формулу (44) § 1.2 при вычислении совместной плотности преобразованного вектора $Y = (Y_1, Y_2)$ и убедиться в том, что она есть произведение маргинальных (одномерных) плотностей указанных гамма- и бета-распределений. ►

44 Пусть о распределении случайных величин ξ и η известно, что

$$\mathcal{L}(\eta) = \mathrm{Be}\left(\frac{a_1}{c}, \frac{a_2}{c}\right),$$

где a_1, a_2 и c — натуральные числа, а условное распределение $\mathcal{L}(\xi|\eta = p) = \mathrm{Bi}(n, p)$. Показать, что безусловное распределение ξ есть распределение Маркова—Пойа $\mathrm{МП}(n; a_1, a_2, c)$ (см. (11) § 1.1); получить отсюда формулы (13) § 1.1.

◀ *Указание.* Вычислить безусловные вероятности

$$\mathbf{P}\{\xi = x\} = C_n^x \int_0^1 p^x (1-p)^{n-x} \frac{\Gamma(a/c)}{\Gamma(a_1/c)\Gamma(a_2/c)} p^{a_1/c-1}(1-p)^{a_2/c-1}\,dp,$$

$$a = a_1 + a_2,$$

и, воспользовавшись свойствами гамма-функции, убедиться в том, что они имеют вид (11) § 1.1. При вычислении моментов учесть, что r-й факториальный момент биномиального распределения $\mathbf{Bi}(n, p)$ есть $(n)_r p^r$ (см. упр. 19). ▶

45 Вывести формулы (26) § 1.2 для моментов равномерного распределения $\mathcal{U}(a, b)$, а также получить вид его характеристической функции

$$\Psi(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

46 Пусть $(X_1, \ldots, X_n)$ — случайная выборка из распределения $\mathcal{U}(a, b)$ и $Y = \min\{X_1, \ldots, X_n\}$, $Z = \max\{X_1, \ldots, X_n\}$ Вывести вид совместной плотности случайных величин Y и Z:

$$f(y, z) = \begin{cases} \dfrac{n(n-1)}{(b-a)^n}(z-y)^{n-2} & \text{при } a \leqslant y < z \leqslant b, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Получить следующие формулы для моментов:

$$\mathbf{E}Y = \frac{na+b}{n+1}, \quad \mathbf{E}Z = \frac{a+nb}{n+1},$$

$$\mathbf{D}Y = \mathbf{D}Z = \frac{n(b-a)^2}{(n+1)^2(n+2)}, \quad \operatorname{cov}(Y, Z) = \frac{(b-a)^2}{(n+1)^2(n+2)}.$$

47 Пусть случайная величина ξ имеет равномерное распределение $\mathcal{U}(-\pi/2, \pi/2)$. Показать, что случайная величина $\eta = \operatorname{tg}\xi$ имеет распределение Коши, задаваемое плотностью (31) § 1.2.

◀ *Указание.* Сначала выписать функцию распределения $F_\eta(x) = \mathbf{P}\{\eta \leqslant x\}$, дифференцированием которой получить плотность. ▶

48

а) Пусть X_1 и X_2 — независимые стандартные нормальные случайные величины. Показать, что их отношение X_1/X_2 распределено по закону Коши.

◀ *Указание.* Плотность отношения $\zeta = \xi/\eta$ двух независимых случайных величин, плотности которых f_ξ и f_η известны, вычисляется по формуле

$$f_\zeta(y) = \int_{-\infty}^{\infty} f_\xi(xy) f_\eta(x)|x|\,dx. \quad \blacktriangleright \qquad (2)$$

б) Распределение Коши с параметром сдвига a (обозначается $\mathcal{K}(a)$) определяется плотностью

$$\frac{1}{\pi} \cdot \frac{1}{1+(x-a)^2}, \quad x \in R^1.$$

(так что формула (31) § 1.2 соответствует случаю $a = 0$). Убедиться в том, что характеристическая функция закона $\mathcal{K}(a)$ есть $\exp\{iat - |t|\}$ и доказать следующий факт: если случайные величины $X_1, \ldots, X_n$ независимы и $\mathcal{L}(X_j) = \mathcal{K}(a_j)$, $j = 1, \ldots, n$, то $\mathcal{L}(\overline{X}) = \mathcal{K}(\bar{a})$, где черта означает среднее арифметическое.

49 Проверить, что стьюдентово отношение t_n (см. (27) § 1.2) имеет моменты $\mathrm{E}t_n^k$ лишь при $k < n$, при этом моменты нечетного порядка равны 0, а

$$\mathrm{E}t_n^{2r} = \frac{1 \cdot 3 \ldots (2r-1)n^r}{(n-2)(n-4)\ldots(n-2r)} \quad \text{при} \quad 2r < n.$$

Получить отсюда формулы (30) § 1.2.

◀ *Указание.* Учесть, что если $\mathcal{L}(\xi) = \mathcal{N}(0, 1)$, то $\mathrm{E}\xi^{2r+1} = 0$, $\mathrm{E}\xi^{2r} = 1 \cdot 3 \ldots (2r-1)$, а также формулы для моментов гамма-распределения (см. п. 3 § 1.2). ▶

50 (*Продолжение*). Установить, что

$$\mathcal{L}\left(\frac{1}{1 + t_n^2/n}\right) = \mathrm{Be}\left(\frac{n}{2}, \frac{1}{2}\right).$$

◀ *Указание.* Воспользоваться упр. 43. ▶

51

а) Показать, что если X имеет распределение Снедекора $S(n, m)$, то $Y = 1/X$ имеет распределение $S(m, n)$.

б) Пусть t_n имеет распределение Стьюдента $S(n)$. Показать, что $\mathcal{L}(t_n^2) = S(1, n)$.

52 Пусть $F(x; n, m)$ — функция распределения закона $S(n, m)$, а $B(x; a, b)$ — функция бета-распределения $\mathrm{Be}(a, b)$. Установить равенство

$$F(x; n, m) = B\left(\frac{nx}{m + nx}; \frac{n}{2}, \frac{m}{2}\right), \quad x > 0.$$

Получить отсюда выражение (33) § 1.2 для плотности распределения $S(n, m)$. Найти моменты этого распределения (35) § 1.2.

◀ *Указание.* Записать соотношение (32) § 1.2 в виде

$$F_{n,m} = \frac{m}{n} \frac{Y}{1 - Y}, \quad \text{где} \quad Y = \frac{\chi_n^2}{\chi_n^2 + \chi_m^2},$$

и воспользоваться упр. 43, в силу которого $\mathcal{L}(Y) = \mathrm{Be}(n/2, m/2)$. Моменты вычисляются по формулам

$$\mathrm{E}F_{n,m}^r = \left(\frac{m}{n}\right)^r \mathrm{E}(\chi_n^2)^r \mathrm{E}(\chi_m^2)^{-r}$$

используя далее формулы для моментов гамма-распределения. ▶

53 (*Продолжение*). Установить, что

$$\mathcal{L}\left(\frac{nF_{n,m}}{m + nF_{n,m}}\right) = \mathrm{Be}\left(\frac{n}{2}, \frac{m}{2}\right)$$

54 Пусть $X = (X_1, \quad, X_l, X_{l+1}, \ldots, X_{l+m})$ — случайная выборка из показательного распределения $\Gamma(a, 1)$ и

$$Y = \frac{m}{l} \frac{X_1 + \quad + X_l}{X_{l+1} + \ldots + X_{l+m}}.$$

Доказать, что $\mathcal{L}(Y) = S(2l, 2m)$.

◄ *Указание.* Воспользоваться упр. 37, в силу которого

$$\mathcal{L}\left(\frac{2}{a}(X_1 + \ldots + X_l)\right) = \Gamma(2, l) = \chi^2(2l)$$

и аналогично

$$\mathcal{L}\left(\frac{2}{a}(X_{l+1} + \quad + X_{l+m})\right) = \chi^2(2m). \ ►$$

55 Пусть случайная величина ξ имеет распределение Вейбулла $W(a, \alpha, b)$ (см. (36) § 1.2). Доказать формулы (37) § 1.2 для $\mathbf{E}\xi$ и $\mathbf{D}\xi$.

56 (*Продолжение*). Пусть $(X_1, \ldots, X_n)$ — случайная выборка из распределения $W(a, \alpha, b)$ и $Y = \min\{X_1, \ldots, X_n\}$. Показать, что

$$\mathcal{L}\left(n^{1/\alpha}\frac{Y - a}{b}\right) = W(1, \alpha, 0)$$

и получить отсюда следующие формулы для моментов Y

$$\mathbf{E}Y = a + b\,\Gamma\left(1 + \frac{1}{\alpha}\right)n^{-1/\alpha},$$

$$\mathbf{D}Y = b^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]n^{-2/\alpha}$$

◄ *Указание.* Вычислить $\mathbf{P}\{Y > y\}$. ►

57 Пусть ξ имеет распределение Парето с параметрами x_0 и α (см. (38) § 1.2). Для каких натуральных значений k существуют моменты $\mathbf{E}\xi^k$? Показать, что при $\alpha > 2$ справедливы формулы (39) § 1.2 для $\mathbf{E}\xi$ и $\mathbf{D}\xi$.

58 (*Продолжение*). Доказать, что случайная величина $\eta = \ln(\xi/x_0)$ имеет показательное распределение $\Gamma(1/\alpha, 1)$ (см. п. 3 § 1.2).

◄ *Указание.* Применить формулу (45) § 1.2 либо непосредственно вычислить функцию распределения $F_\eta(t) = \mathbf{P}\{\eta \leqslant t\}$ ►

59 Пусть о распределении случайных величин ξ и η известно, что ξ имеет распределение Парето с параметрами x_0 и α, а условное распределение η при условии, что $\xi = \theta$, является равномерным $\mathcal{U}(0, \theta)$. Доказать, что условное распределение $\mathcal{L}(\xi|\eta = x)$ есть распределение Парето с параметрами $\max(x_0, x)$ и $\alpha + 1$.

◀ *Указание.* Если $f_\xi(\theta)$ и $f_\eta(x)$ есть маргинальные плотности величин ξ и η, а $f_{\eta|\xi}(x|\theta)$ — плотность условного распределения $\mathcal{L}(\eta|\xi = \theta)$, то плотность $f_{\xi|\eta}(\theta|x)$ условного распределения $\mathcal{L}(\xi|\eta = x)$ вычисляется по *формуле Байеса* [13]

Условные распределения и формула Байеса

$$f_{\xi|\eta}(\theta|x) = \frac{f_\xi(\theta) f_{\eta|\xi}(x|\theta)}{f_\eta(x)}; \qquad (3)$$

при этом

$$f_\eta(x) = \int f_{\eta|\xi}(x|\theta) f_\xi(\theta)\, d\theta$$

является «нормирующим» множителем, не зависящим от переменной θ условной плотности $f_{\xi|\eta}(\theta|x)$, и потому вычислять его явное значение в конкретной задаче не обязательно (плотность любого распределения достаточно вычислять с точностью до нормирующего множителя: $f(x) = cp(x) \cong p(x)$).
В нашем случае записать

$$f_\xi(\theta) \cong \theta^{-a-1} I\,(\theta > x_0), \quad f_{\eta|\xi}(x|\theta) = \theta^{-1} I\,(\theta > x),$$

где (напомним) $I(\cdot)$ — индикатор. ►

60 Предположим, что $X_1, \ldots, X_k$ — независимые случайные величины и X_i имеет гамма распределение $\Gamma(a, \lambda_i)$, $i = 1, \ldots, k$. Пусть случайный вектор $Y = (Y_1, \ldots, Y_k)$ определяется соотношениями

$$Y_i = \frac{X_i}{X_1 + \quad + X_k}, \qquad , k.$$

Доказать, что вектор Y и случайная величина $X = X_1 + \quad + X_k$ независимы и при этом Y имеет распределение Дирихле $D(\underline{\lambda} = (\lambda_1, \quad , \lambda_n))$ (распределение X указано в упр. 37).

◀ *Указание.* Это упражнение является обобщением упр. 43 и доказывается аналогично. ►

61 Доказать формулы (41) § 1.2 для моментов распределения Дирихле $D(\underline{a})$.

◀ *Указание.* Использовать формулу (42) § 1.2 для интеграла Дирихле и свойства гамма-функции. ►

62 Пусть о распределении случайных векторов $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ и $\underline{\xi} = (\xi_1, \ldots, \xi_N)$ известно, что $\mathcal{L}(\underline{\xi}) = \mathrm{D}(\underline{a}/c)$, где $\underline{a} = (a_1, \ldots, a_N)$ и все a_i и c — натуральные числа, а условное распределение $\mathcal{L}(\underline{\nu}|\underline{\xi} = \underline{p})$ есть полиномиальное распределение $M(n; \underline{p})$ (см. п. 6 § 1.1). Показать, что безусловное распределение вектора $\underline{\nu}$ есть N-мерное распределение Маркова—Пойа $\mathrm{M\Pi}(n; \underline{a}, c)$ (см. п. 7 § 1.1); получить отсюда формулы (19) § 1.1.

◀ *Указание.* Это упражнение является многомерным обобщением упражнения 44 и доказывается аналогично с использованием формулы (42) § 1.2 для интеграла Дирихле и свойств гамма-функции: в данном случае надо вычислить безусловные вероятности

$$\mathrm{P}\{\underline{\nu} = \underline{x}\} = \frac{n!}{x_1! \ldots x_N!} \int_S \prod_{j=1}^N p_j^{x_j} \frac{\Gamma(a/c)}{\Gamma(a_1/c) \ldots \Gamma(a_N/c)} \prod_{j=1}^N p_j^{a_j/c-1}\, dp_1 \quad \ldots dp_{N-1}$$

[13] Байес Томас (1702–1761) — английский математик-вероятностник.

(здесь область интегрирования S определена в (42) § 1.2, $p_N = 1 - p_1 - \quad - p_{N-1}$, $a = a_1 + \quad + a_N$ и $\underline{x} \quad (x_1, \quad , x_N)$ — произвольный целочисленный вектор с $x_1 + \ldots + x_N = n$), и убедиться в том, что они имеют вид (17) § 1.2. При вычислении моментов учесть формулу для факториальных моментов полиномиального распределения (см. упр. 19). ►

Замечание. В этом упражнении речь идет о модели *смеси вероятностных распределений*: в данном случае рассматривается смесь полиномиальных распределений $M(n; \underline{p})$, где вектор $\underline{p}$ предполагается случайным вектором, имеющим распределение Дирихле, по этому «смешивающему» распределению. Таким образом, установленный результат можно сформулировать в этих терминах следующим образом: *многомерное распределение Маркова—Пойа является смесью полиномиальных распределений по распределению Дирихле.*

Соответствующий «одномерный» результат упражнения 44 теперь звучит так: распределение Маркова—Пойа МП$(n; a_1, a_2, c)$ является смесью биномиальных распределений Bi(n, p) по бета-распределению: $\mathcal{L}(p) = \mathrm{Be}(a_1/c, a_2/c)$.

63 Пусть о распределении случайных векторов $\underline{\nu} = (\nu_1, \quad , \nu_N)$ и $\underline{\xi} = (\xi_1, \quad , \xi_N)$ известно, что $\underline{\xi}$ имеет распределение Дирихле $\mathrm{D}(\underline{\alpha})$, а условное распределение $\mathcal{L}(\underline{\nu}|\underline{\xi} = \theta)$ есть полиномиальное распределение $M(n; \theta)$. Показать, что условное распределение

$$\mathcal{L}(\underline{\xi}|\underline{\nu} = \underline{x}) = \mathrm{D}(\underline{\alpha} + \underline{x}).$$

◄ **Указание (См. указание к упр. 59).** Если $x_1, \quad , x_N$ — целые неотрицательные числа, удовлетворяющие условию $x_1 + \ldots + x_N = n$, то в нашем случае условная плотность

$$f_{\underline{\nu}|\underline{\xi}}(\underline{x}|\theta) \cong \theta_1^{x_1} \quad \theta_N^{x_N}$$

а $f_{\underline{\xi}}(\theta) \cong \theta_1^{a_1 - 1} \quad \theta_N^{a_N - 1}$, следовательно,

$$f_{\underline{\xi}|\underline{\nu}}(\theta|\underline{x}) \cong f_{\underline{\xi}}(\theta) f_{\underline{\nu}|\underline{\xi}}(\underline{x}|\theta) \cong \theta_1^{a_1 + x_1 - 1} \ldots \theta_N^{a_N + x_N - 1} \quad ►$$

Замечание. В статистических приложениях часто предполагается, что наблюдаемая выборка (данные) X имеет распределение из некоторого параметрического семейства распределений $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$ (в данном случае $X = \underline{\nu}$ и $\mathcal{F} = M(n; \theta)$ — семейство полиномиальных распределений), а о параметре этого семейства предполагается, что он случаен, и его *априорное* распределение $\mathcal{L}(\theta) \in \mathcal{F}^*$ — заданному семейству априорных распределений θ (в нашем случае $\mathcal{F}^* = \mathrm{D}(\underline{\alpha})$). Семейство $\mathcal{F}^*$ называется *сопряженным* к семейству $\mathcal{F}$ (обозначается $\mathcal{F}^* \lhd \mathcal{F}$), если при $X = x$ *апостериорное* распределение параметра $\mathcal{L}(\theta|X = x) \in \mathcal{F}^*$ т. е. находится в том же классе распределений, что и априорное распределение параметра. В этих терминах полученный результат звучит так: $\mathrm{D}(\underline{\alpha}) \lhd M(n; \theta)$, т. е. *распределение Дирихле сопряжено к полиномиальному распределению.*

Сопряженные семейства распределений

Аналогичный результат мы имели и в упр. 59: распределение Парето сопряжено к равномерному распределению $\mathcal{U}(0, \theta)$.

64 Убедиться в справедливости также следующих утверждений (ниже $\underline{X} = (X_1, \ldots, X_n)$ — случайная выборка и $x = x_1 + \quad + x_n$):

1) $\mathrm{Be}(a, b) \lhd \mathrm{Bi}(m, \theta)$, при этом

$$\mathcal{L}(\theta|\underline{X} = \underline{x}) = \mathrm{Be}(a + x, b + nm - x);$$

2) $\mathrm{Be}(a, b) \triangleleft \widetilde{\mathrm{Bi}}(r, \theta)$, при этом

$$\mathcal{L}(\theta|\underline{X} = \underline{x}) = \mathrm{Be}(a + x, b + nr);$$

3) $\Gamma(a, \lambda) \triangleleft \Pi(\theta)$, при этом

$$\mathcal{L}(\theta|\underline{X} = \underline{x}) = \Gamma\left(\frac{a}{na + 1}, \lambda + x\right),$$

4) $\Gamma(a, \lambda) \triangleleft \Gamma(\theta^{-1}, 1)$, при этом

$$\mathcal{L}(\theta|\underline{X} = \underline{x}) = \Gamma\left(\frac{a}{ax + 1}, \lambda + n\right);$$

5) $\mathcal{N}(\mu, \sigma^2) \triangleleft \mathcal{N}(\theta, b^2)$, при этом

$$\mathcal{L}(\theta|\underline{X} = \underline{x}) = \mathcal{N}(\mu_1, \sigma_1^2), \quad \text{где}$$

$$\mu_1 = \left(\frac{\mu}{\sigma^2} + \frac{x}{b^2}\right)\sigma_1^2, \quad \sigma_1^2 = \left(\frac{1}{\sigma^2} + \frac{n}{b^2}\right)^{-1}$$

6) распределение Парето с параметрами x_0 и α сопряжено к равномерному распределению $\mathcal{U}(0, \theta)$, при этом апостериорное распределение θ при наблюдении $\boldsymbol{X} = \underline{x}$ есть распределение Парето с параметрами $\max(x_0, x_1, \quad , x_n)$ и $\alpha + n$.

◀ *Указание.* При нахождении апостериорных плотностей (3) следовать указанию к упр. 59. ▶

65 Пусть случайный вектор $\underline{\xi} = (\xi_1, \xi_2)$ имеет двумерное собственное нормальное распределение

$$\mathcal{N}\left((\mu_1, \mu_2), \left\| \begin{matrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{matrix} \right\|\right)$$

где $-1 < \rho = \mathrm{corr}(\xi_1, \xi_2) < 1$. Доказать, что условное распределение

$$\mathcal{L}(\xi_2|\xi_1 = x) = \mathcal{N}(M(x), \sigma^2),$$

где условное среднее

$$M(x) = \mathbf{E}(\xi_2|\xi_1 = x) = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho(x - \mu_1)$$

— *функция регрессии* ξ_2 на ξ_1, являющаяся в данном случае линейной по x, а условная дисперсия

$$\sigma^2 = \mathbf{D}(\xi_2|\xi_1 = x) = (1 - \rho^2)\sigma_2^2$$

не зависит от x. Существуют ли другие распределения $\mathcal{L}(\underline{\xi})$, отличные от нормального, обладающие такими же свойствами условного распределения $\mathcal{L}(\xi_2|\xi_1 = x)$?

◀ *Указание.* Вычислить условную плотность величины ξ_2 при условии $\xi_1 = x$ по формуле $f_{\xi_2|\xi_1}(y|x) = f_{\xi_1\xi_2}(x, y)/f_{\xi_1}(x)$ и убедиться в том, что она равна

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(y - M(x))^2\right\}. \blacktriangleright$$

66 (Обобщение на многомерный случай). Пусть $\underline{X} = (X_1, \ldots, X_{r-1})$, $Y = X_r$, $r \geqslant 2$, и совместное распределение

$$\mathcal{L}(\underline{X}, Y) = \mathcal{N}\big((\mu_1, \ldots, \mu_r), \Sigma = \|\sigma_{ij}\|_1^r\big),$$

где матрица Σ не вырождена и $\Sigma^{-1} = \|\sigma^{ij}\|_1^r$. Доказать, что

$$\mathcal{L}(Y|\underline{X} = \underline{x}) = \mathcal{N}\big(M(\underline{x}), 1/\sigma^{rr}\big),$$

где

$$M(\underline{x}) = \mathbf{E}(Y|\underline{X} = \underline{x}) = \mu_r - \sum_{i=1}^{r-1}(x_i - \mu_i)\frac{\sigma^{ir}}{\sigma^{rr}}, \quad \underline{x} = (x_1, \ldots, x_{r-1}).$$

◀ *Указание.* Записав совместную плотность в виде (см. (9) § 1.2)

$$f_{\underline{X}Y}(x_1, \ldots, x_{r-1}, x_r) = C \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{r}\sigma^{ij}(x_i - \mu_i)(x_j - \mu_j)\right\} =$$

$$= C \exp\left\{-\frac{1}{2}\left[\sigma^{rr}(x_r - \mu_r)^2 + 2(x_r - \mu_r)\sum_{i=1}^{r-1}\sigma^{ir}(x_i - \mu_i)\right] + \quad\right\},$$

получить, что переменная x_r войдет в выражение условной плотности $f_{Y|\underline{X}}(x_r|\underline{x})$ в виде

$$\exp\left\{-\frac{\sigma^{rr}}{2}\left(x_r - \mu_r + \sum_{i=1}^{r-1}\sigma^{ir}\frac{x_i - \mu_i}{\sigma^{rr}}\right)^2\right\}. \blacktriangleright$$

67 Оценить методом Монте-Карло величину площади $a = \pi/4$ под дугой единичной окружности $x^2 + y^2 = 1$, $x, y \geqslant 0$.

Глава 2

Первичная обработка экспериментальных данных

Дело не в цифрах, а в том, что вы с ними делаете.

Данная глава представляет собой общее введение в теорию выборочного (статистического) метода. Здесь рассматривается ситуация, когда исходные данные (выборка) получаются в результате проведения повторных независимых наблюдений (измерений) над некоторой случайной величиной ξ, закон распределения которой в течение всего эксперимента остается неизменным. Как уже отмечалось во Введении, в этом случае выборка представляет собой n-мерный вектор $X = (X_1, \ldots, X_n)$, где n — число наблюдений (объем выборки), а компоненты X_i — независимые копии наблюдаемой величины ξ, т. е. имеющие такое же распределение, как и ξ. В этом случае говорят, что $X = (X_1, \ldots, X_n)$ есть случайная выборка (или просто выборка) из распределения ξ (или из $\mathcal{L}(\xi)$). В теории выборочного метода (или кратко: выборочной теории) изучаются различные свойства случайной выборки и функций от нее. В этой главе мы введем основные понятия выборочной теории, приведем фундаментальные теоремы математической статистики, исследуем в точной и асимптотической (т. е. при большом объеме выборки) постановках свойства основных характеристик случайной выборки.

§ 2.1. Вариационный ряд выборки, эмпирическая функция распределения и гистограмма

Верно определяйте слова, и вы освободите мир от половины недоразумений.

Р. Декарт [1)]

1. Порядковые статистики и вариационный ряд выборки

Итак, пусть $X = (X_1, \ldots, X_n)$ — выборка из некоторого распределения $\mathcal{L}(\xi)$. Произвольной реализации $\underline{x} = (x_1, \ldots, x_n)$ этой выборки можно поставить в соответствие упорядоченную последовательность

$$x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}, \tag{1}$$

[1)] Декарт Рене (1596–1650) — французский математик, физик.

располагая $x_1, \ldots, x_n$ в порядке их возрастания, так что $x_{(1)} = \min\{x_1, \ldots, x_n\}$, $x_{(2)}$ — второе по величине значение, $x_{(n)} = \max\{x_1, \ldots, x_n\}$.

Обозначим через $X_{(k)}$ случайную величину, которая для каждой реализации $\underline{x}$ выборки $\boldsymbol{X}$ принимает значение $x_{(k)}$, $k = 1, \ldots, n$. Так по выборке $\boldsymbol{X}$ определяют новую последовательность случайных величин $X_{(1)}, \ldots, X_{(n)}$, называемых *порядковыми статистиками* выборки; при этом $X_{(k)}$ — *k-я порядковая статистика*, а $X_{(1)}$ и $X_{(n)}$ — *экстремальные* (соответственно *минимальное и максимальное*) *значения выборки*, их разность $\rho = X_{(n)} - X_{(1)}$ называют *размахом* выборки. Из определения порядковых статистик следует, что они упорядочены по возрастанию их значений, т. е. они образуют возрастающую последовательность

$$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}, \tag{2}$$

Вариационный ряд

которая называется *вариационным рядом* выборки $\boldsymbol{X}$. Симметричные относительно концов элементы последовательности (2) $X_{(m)}$ и $X_{(n-m+1)}$ иногда называют соответственно *m-м минимальным и m-м максимальным* значениями выборки, или m-ми *нижним и верхним экстремумами* ($m = 1, 2, \ldots$). Важной характеристикой вариационного ряда выборки является его *медиана* (или середина), которая равна $X_{(k+1)}$, если $n = 2k + 1$, и $\left(X_{(k)} + X_{(k+1)}\right)/2$ при $n = 2k$. В дальнейшем под термином «основные характеристики выборки» будем понимать совокупность ее экстремальных значений, размаха и медианы.

Итак, вариационный ряд — это расположенные в порядке возрастания их величин элементы выборки. Подчеркнем, что для заданной реализации $\underline{x} = (x_1, \ldots, x_n)$ выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$ реализацией последовательности (2) является последовательность (1). Вариационный ряд является одним из стандартных способов представления выборки.

2. Эмпирическая функция распределения

Введем теперь фундаментальное понятие математической статистики, каковым является *эмпирическая функция распределения*. Для этого определим сначала для каждого действительного x случайную величину $\mu_n(x)$, равную числу элементов выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$, значения которых не превосходят x, т. е.

$$\mu_n(x) = \sum_{i=1}^{n} I(X_i \leqslant x), \tag{3}$$

и положим $\widehat{F}_n(x) \quad \frac{1}{n}\mu_n(x)$ — это и есть эмпирическая функция распределения (далее используется сокращение э. ф. р.), соответствующая выборке $\boldsymbol{X}$. Функцию распределения $F(x)$ наблюдаемой случайной величины ξ в этом контексте называют *теоретической функцией распределения*. По своему определению э. ф. р. — случайная функция: для каждого $x \in R^1$ значение $\widehat{F}_n(x)$ — случайная величина, принимающая значения $0, 1/n, 2/n, \ldots,$

$(n-1)/n$, $n/n=1$, при этом

$$\mathbf{P}\left\{\widehat{F}_n(x)=\frac{k}{n}\right\}=\mathbf{P}\{\mu_n(x)=k\}.$$

С каждым X_i, можно связать два события: $\{X_i \leqslant x\}$ и $\{X_i > x\}$. Вероятности этих событий, очевидно, равны

$$p=\mathbf{P}\{X_i \leqslant x\}=F(x), \quad q=\mathbf{P}\{X_i > x\}=1-F(x).$$

Если событие $\{X_i \leqslant x\}$ назвать успехом, то $\mu_n(x)$ является числом успехов в n независимых испытаний Бернулли с вероятностью успеха p, поэтому $\mu_n(x)$ имеет биномиальное распределение $\mathrm{Bi}(n,p)$ с $p=F(x)$ (см. п. 1 § 1.1). Следовательно,

$$\mathbf{P}\left\{\widehat{F}_n(x)=\frac{k}{n}\right\}=C_n^k F^k(x)(1-F(x))^{n-k}, \quad k=0,1,\ldots,n. \tag{4}$$

Распределение э. ф. р.

Итак, э. ф. р. (как и вариационный ряд) — некоторая сводная характеристика выборки. Для каждой реализации $\underline{x}$ выборки X функция $\widehat{F}_n(x)$ однозначно определена и обладает всеми свойствами функции распределения: изменяется от 0 до 1, не убывает и непрерывна справа. При этом она кусочно постоянна и возрастает только в точках последовательности (1). Если все компоненты вектора $\underline{x}$ различны (в последовательности (1) все неравенства строгие), то функция $\widehat{F}_n(x)$ задается, очевидно, соотношениями

$$\widehat{F}_n(x)=\begin{cases} 0 & \text{при} \quad x < x_{(1)}, \\ k/n & \text{при} \quad x_{(k)} \leqslant x < x_{(k+1)}, \quad k=1,\ldots,n-1, \\ 1 & \text{при} \quad x \geqslant x_{(n)}, \end{cases}$$

т. е. в этом случае число скачков равно n и величины всех скачков равны $1/n$; типичный график функции $\widehat{F}_n(x)$ имеет вид, изображенный на рис. 1.

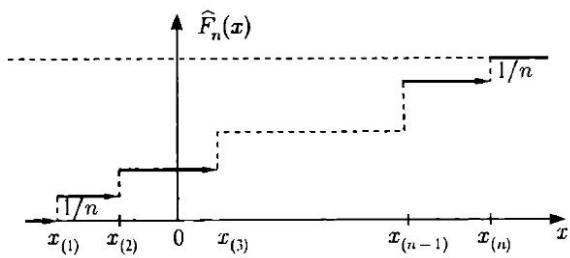


Рис. 1

Пример 1. Реализацией выборки $X \quad (X_1, \quad, X_6)$ являются следующие данные: $-1{,}5$; 2,6; 1,2; $-2{,}1$; 0,1; 0,9. Найти ее основные характеристики и построить эмпирическую функцию распределения.

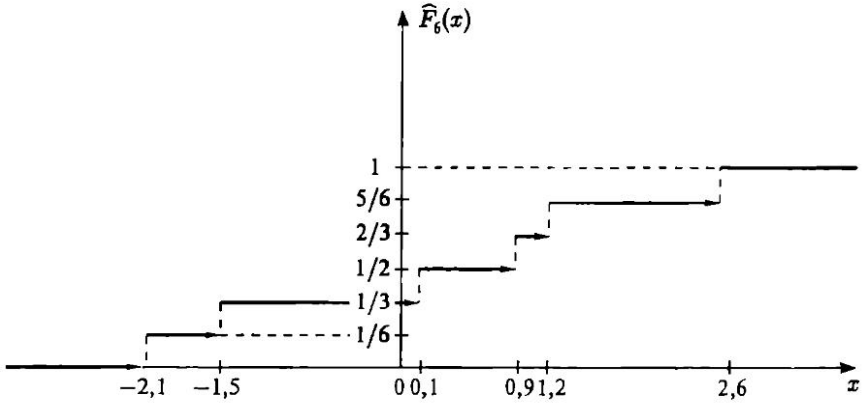


Рис. 2

Прежде всего строим вариационный ряд: $-2{,}1$; $-1{,}5$; $0{,}1$; $0{,}9$; $1{,}2$; $2{,}6$. Отсюда видно, что минимальное (максимальное) значение выборки равно $-2{,}1$ ($2{,}6$), размах равен $2{,}6-(-2{,}1)=4{,}7$, а медиана равна

$$\frac{x_{(3)}+x_{(4)}}{2}=\frac{0{,}1+0{,}9}{2}=0{,}5.$$

Эмпирическая функция распределения $\widehat{F}_6(x)$ является ступенчатой, возрастает скачками величиной $1/6$ в точках вариационного ряда, при этом $\widehat{F}_6(x)=0$ при $x<x_{(1)}=-2{,}1$ и $\widehat{F}_6(x)=1$ при $x\geqslant x_{(6)}=2{,}6$. Таким образом,

$$\widehat{F}_6(x)=\begin{cases}0 & \text{при} \quad x<-2{,}1,\\ 1/6 & \text{при} \quad -2{,}1\leqslant x<-1{,}5,\\ 1/3 & \text{при} \quad -1{,}5\leqslant x<0{,}1,\end{cases} \qquad \widehat{F}_6(x)=\begin{cases}1/2 & \text{при} \quad 0{,}1\leqslant x<0{,}9,\\ 2/3 & \text{при} \quad 0{,}9\leqslant x<1{,}2,\\ 5/6 & \text{при} \quad 1{,}2\leqslant x<2{,}6,\\ 1 & \text{при} \quad x\geqslant 2{,}6.\end{cases}$$

График этой функции изображен на рис. 2. •

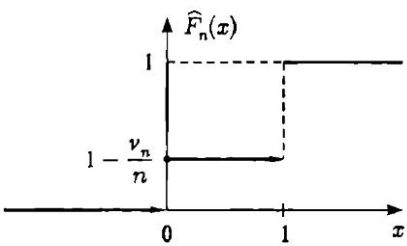


Рис. 3

Пример 2. Пусть $X=(X_1,\ \ldots,X_n)$ — выборка из распределения Бернулли $\mathrm{Bi}(1,p)$ (см. п. 1 § 1.1). Обозначим $\nu_n=X_1+\ \ldots\ +X_n$ число единиц в выборке. Тогда скачки э. ф. р. $\widehat{F}_n(x)$ будут иметь место лишь в точках 0 и 1 и очевидно (рис. 3)

$$\widehat{F}_n(x)=\begin{cases}0 & \text{при} \quad x<0,\\ 1-\dfrac{\nu_n}{n} & \text{при} \quad 0\leqslant x<1,\\ 1 & \text{при} \quad x\geqslant 1,\end{cases}$$ •

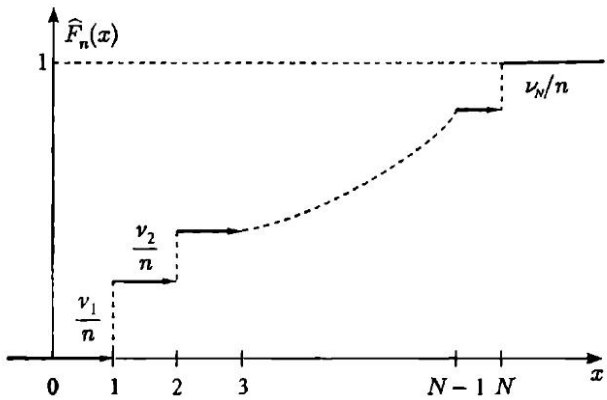


Рис. 4

Пример 3. Пусть $X=(X_1,\ldots,X_n)$ — выборка из дискретного распределения $\mathcal{L}(\xi)=\begin{pmatrix}1 & 2 & & N\\ p_1 & p_2 & & p_N\end{pmatrix}$. Обозначим $\nu_j=\sum_{i=1}^{n} I(X_i=j)$ число реализаций исхода «j», $j=0,1,\ \ldots,N$, тогда вектор $\underline{\nu}=(\nu_1,\ldots,\nu_N)$ имеет полиномиальное распределение $M(n;\underline{p}=(p_1,\ \ldots,p_N))$ (см. п. 6 § 1.1), а отдельная частота ν_j — биномиальное распределение $\mathrm{Bi}(n,p_j)$.

Э. ф. р. для дискретного распределения

В данном случае э. ф. р. $\widehat{F}_n(x)$ будет иметь скачки лишь в точках $1,2,\ \ldots,N$ и при этом величина скачка в точке j есть

$$\Delta\widehat{F}_n(j)=\widehat{F}_n(j)-\widehat{F}_n(j-0)=\frac{\nu_j}{n},\quad j=1,\ \ldots,N$$

Тем самым график $\widehat{F}_n(x)$ будет иметь вид, указанный на рис. 4.

Здесь $\mathbf{P}\{\Delta\widehat{F}_n(j)=0\}=\mathbf{P}(\nu_j=0)=(1-p_j)^n$ что мало при больших n, т. е. в большой выборке скачок в точке j наверняка будет иметь место. Более того, так как

$$\mathbf{P}\Big\{\bigcup_{j=1}^{N}\{\Delta\widehat{F}_n(j)=0\}\Big\}\leqslant\sum_{j=1}^{N}\mathbf{P}\{\Delta\widehat{F}_n(j)=0\}=\sum_{j=1}^{N}(1-p_j)^n\to 0,$$

при $n\to\infty$, то в больших выборках с вероятностью, близкой к 1, скачки э. ф. р. $\widehat{F}_n(x)$ будут иметь место во всех точках $1,2,\ \ldots,N$, а случайными будут лишь величины этих скачков. •

Если же теоретическая функция распределения $F_\xi(x)=F(x)$ непрерывна, то с вероятностью 1 все элементы выборки $X=(X_1,\ldots,X_n)$ будут различны, и график э. ф. р. $\widehat{F}_n(x)$ будет иметь вид, изображенный на рис. 1, т. е. случайными теперь будут точки скачков, величины же скачков не случайны и равны $1/n$.

Таким образом, для выборок из дискретных и непрерывных распределений характер соответствующих эмпирических функций распределения будет различным. Тем не менее в любом случае э. ф. р. $\widehat{F}_n(x)$ с увеличением объема выборки n сближается в каждой точке x с теоретической функцией распределения $F(x)$.

Сходимость по вероятности э. ф. р.

Теорема 1. Для любого x $(-\infty < x < \infty)$ и любого $\varepsilon > 0$ при $n \to \infty$

$$\mathbf{P}\{|\widehat{F}_n(x) - F(x)| < \varepsilon\} \longrightarrow 1,$$

т. е. $\widehat{F}_n(x) \xrightarrow{\mathbf{P}} F(x)$ (сходимость по вероятности см. п. 1 § 1.2).

Доказательство. Это утверждение является простым следствием упоминавшегося ранее закона больших чисел (см. (3) § 1.2), так как, обозначив, $\eta_j = I(X_j \leqslant x)$, $j = 1, \ldots, n$, мы получим последовательность независимых, одинаково распределенных случайных величин с математическим ожиданием $\mathbf{E}\eta_1 = \mathbf{P}\{X_1 \leqslant x\} = F(x)$, а $\widehat{F}_n(x) = \dfrac{1}{n}\sum\limits_{j=1}^{n}\eta_j$. Таким образом, $\widehat{F}_n(x) \xrightarrow{\mathbf{P}} F(x)$, что и утверждается в теореме. ■

Скорость сближения для э. ф. р.

Можно установить и скорость сближения э. ф. р. $\widehat{F}_n(x)$ с $F(x)$. Так как из (3) следует, что (см. (4) § 1.1)

$$\begin{aligned}\mathbf{E}\widehat{F}_n(x) &= \frac{1}{n}\mathbf{E}\mu_n(x) = F(x),\\ \mathbf{D}\widehat{F}_n(x) &= \frac{1}{n^2}\mathbf{D}\mu_n(x) = \frac{1}{n}F(x)(1 - F(x)),\end{aligned} \tag{5}$$

то согласно неравенству Чебышева (для любой случайной величины ξ $\mathbf{P}\{|\xi - \mathbf{E}\xi| > \varepsilon\} \leqslant \mathbf{D}\xi/\varepsilon^2$) имеем для любого $t > 0$

$$\mathbf{P}\{\sqrt{n}|\widehat{F}_n(x) - F(x)| > t\} \leqslant \frac{F(x)(1 - F(x))}{t^2} \leqslant \frac{1}{4t^2} \tag{6}$$

(последнее неравенство следует из того, что для любой функции распределения $F(x)$ имеет место оценка $F(x)(1 - F(x)) \leqslant 1/4$).

Полученная оценка (6) универсальна, так как она не зависит ни от теоретической функции распределения $F(x)$, ни от точки x, ни от объема выборки n. В частности, из нее следует, что

$$\mathbf{P}\left\{|\widehat{F}_n(x) - F(x)| > \frac{5}{\sqrt{n}}\right\} \leqslant 0{,}01.$$

Более того оценку можно получить, если число наблюдений n велико и в этой ситуации воспользоваться теоремой Муавра—Лапласа о сходимости биномиального распределения к нормальному (см. замечание к упр. 33 гл. 1): если $0 < F(x) < 1$, то при $n \to \infty$

Теорема Муавра—Лапласа для э. ф. р.

$$\mathcal{L}\left(\frac{\mu_n(x) - nF(x)}{\sqrt{nF(x)(1-F(x))}}\right) \longrightarrow \mathcal{N}(0,1).$$

В терминах э. ф. р. $\widehat{F}_n(x) = \mu_n(x)/n$ это соотношение можно переписать в следующем виде: обозначим $W_n(x) = \sqrt{n}\big(\widehat{F}_n(x) - F(x)\big)$ так называемый *эмпирический процесс*, тогда

Эмпирический процесс

$$\mathcal{L}\big(W_n(x)\big) \longrightarrow \mathcal{N}\big(0, F(x)(1-F(x))\big). \tag{7}$$

Отсюда можно записать следующую цепочку соотношений:

$$\begin{aligned}\mathbf{P}\big\{|W_n(x)| < t\big\} &= \mathbf{P}\left\{\widehat{F}_n(x) - \frac{t}{\sqrt{n}} < F(x) < \widehat{F}_n(x) + \frac{t}{\sqrt{n}}\right\} \approx \\ &\approx \Phi\left(\frac{t}{\sqrt{F(x)(1-F(x))}}\right) - \Phi\left(-\frac{t}{\sqrt{F(x)(1-F(x))}}\right) = \\ &= 2\Phi\left(\frac{t}{\sqrt{F(x)(1-F(x))}}\right) - 1 \geqslant 2\Phi(2t) - 1.\end{aligned}$$

Если мы хотим сделать правую часть равной заданному числу γ, близкому к 1 (например, $\gamma = 0{,}99$), то решив уравнение $2\Phi(2t) - 1 = \gamma$, получим

$$t_\gamma = \frac{1}{2}\Phi^{-1}\left(\frac{1+\gamma}{2}\right),$$

где Φ^{-1} — обратная функция к стандартной нормальной функции распределения (см. (7) § 1.2). В итоге мы получим соотношение

Доверительный интервал для теоретической функции распределения

$$\mathbf{P}\left\{\widehat{F}_n(x) - \frac{t_\gamma}{\sqrt{n}} < F(x) < \widehat{F}_n(x) + \frac{t_\gamma}{\sqrt{n}}\right\} \gtrsim \gamma, \tag{8}$$

которое означает, что для больших выборок с вероятностью, приблизительно равной γ, значение теоретической функции распределения будет отличаться от соответствующего значения эмпирической функции распределения не больше, чем на $t_\gamma/\sqrt{n}$. Таким образом, в (8) определен случайный интервал $\big(\widehat{F}_n(x) \mp t_\gamma/\sqrt{n}\big)$ с центром в точке $\widehat{F}_n(x)$, который с высокой вероятностью γ накрывает значение теоретической функции распределения $F(x)$, если в этой точке $F(x) \neq 0$. Такой интервал $\big(\widehat{F}_n(x) \mp t_\gamma/\sqrt{n}\big)$ называется *асимптотическим γ-доверительным интервалом* для $F(x)$. Ширина этого интервала при

больших n имеет порядок $1/\sqrt{n}$, т. е. это очень узкий интервал, с высокой точностью локализующий значение $F(x)$, если оно нам неизвестно. Например, при $t_\gamma = 2$ величина $\gamma = 1 - 6 \ \ 10^{-5}$

Из этих рассуждений можно сделать следующий принципиальный вывод: если объем выборки большой, то э. ф. р. $\widehat{F}_n(x)$ в каждой точке x, где $0 < F(x) < 1$, отличается от $F(x)$ на величину порядка $1/\sqrt{n}$, т. е. $\widehat{F}_n(x)$ сближается с $F(x)$ с ростом n и тем самым э. ф. р. может рассматриваться в качестве приближенного значения (*оценки*) теоретической функции распределения $F(x)$, если она нам неизвестна. Функцию $\widehat{F}_n(x)$ поэтому часто называют *статистическим аналогом* для $F(x)$: $\widehat{F}_n(x) \approx F(x)$.

Состоятельность э. ф. р.

Замечание. Указанное в теореме 1 свойство сходимости $\widehat{F}_n(x) \xrightarrow{P} F(x)$, $n \to \infty$, называется в статистике свойством *состоятельности* э. ф. р. $\widehat{F}_n(x)$ как оценки для $F(x)$. Таким образом, теорема 1 утверждает, что э. ф. р. $\widehat{F}_n(x)$ является состоятельной оценкой для теоретической функции распределения $F(x)$ в каждой точке x.

3. Дальнейшие свойства э. ф. р.

(Этот материал для тех, кто хочет заглянуть поглубже в обсуждаемую проблематику)

Выше мы обсуждали одномерные свойства э. ф. р. $\widehat{F}_n(x)$, т. е. ее поведение в фиксированной точке x, и для этого нам было достаточно знать лишь свойства биномиального распределения $\mathrm{Bi}(n, F(x))$. Более глубокие особенности э. ф. р. проявляются, если рассматривать ее поведение не в отдельной фиксированной точке x, а в произвольной конечной системе точек $x_1 < x_2 < \ldots < x_{N-1}$, т. е. если рассматривать случайный вектор $\left(\widehat{F}_n(x_1), \quad , \widehat{F}_n(x_{N-1})\right)$ Пусть $x_0 = -\infty$, $x_N = \infty$, точки $x_1, \ \ldots, x_{N-1}$ таковы, что

$$0 = F(x_0) < F(x_1) < \quad < F(x_{N-1}) < F(x_N) = 1.$$

Обозначим

$$p_i = \Delta_i F \equiv F(x_i) - F(x_{i-1}), \quad i = 1, \quad , N,$$

и введем случайные величины

$$\nu_i = \Delta_i \mu_n = \mu_n(x_i) - \mu_n(x_{i-1}) = \sum_{j=1}^{N} I(x_{i-1} < X_j \leqslant x_i), \quad i = 1, \quad , N$$

Заметим, что ν_i есть число реализаций события $A_i = \{x_{i-1} < \xi \leqslant x_i\}$ в n независимых испытаниях, а так как события $A_1, \quad , A_N$ образуют полную группу попарно несовместных исходов с вероятностями их реализаций в отдельном испытании $p_1, \quad , p_N$ соответственно, то случайный вектор $\underline{\nu} = (\nu_1, \quad , \nu_N)$ имеет полиномиальное распределение $M\left(n; \underline{p} = (p_1, \quad , p_N)\right)$ (см. п. 6 § 1.1). Но из определения э. ф. р. следует, что

$$\widehat{F}_n(x_i) = \frac{1}{n}\mu_n(x_i) = \frac{1}{n}(\nu_1 + \ldots + \nu_i), \quad i = 1, \ldots, N,$$

и тем самым многомерные свойства э. ф. р. $\widehat{F}_n(x)$ определяются полиномиальным распределением. Для полиномиального распределения известна центральная предельная теорема о его сходимости при $n \to \infty$ к многомерному нормальному распределению (см. упр. 33 к гл. 1), воспользовавшись которой мы сейчас легко опишем поведение любых многомерных распределений э. ф. р. для больших выборок. Для этого вычислим сначала приращения э. ф. р. $\widehat{F}_n(x)$ и эмпирического процесса $W_n(x) = \sqrt{n}\big(\widehat{F}_n(x) - F(x)\big)$ на интервалах (x_{i-1}, x_i):

Многомерные свойства э. ф. р.

$$\Delta_i \widehat{F}_n = \frac{1}{n}\Delta_i \mu_n = \frac{\nu_i}{n},$$

$$\Delta_i W_n = \sqrt{n}(\Delta_i \widehat{F}_n - \Delta_i F) = \sqrt{n}\left(\frac{\nu_i}{n} - p_i\right) = \frac{\nu_i - np_i}{\sqrt{n}} \equiv \nu_i^*$$

Следовательно,

$$\mathcal{L}(\Delta_1 W_n, \Delta_2 W_n, \ldots, \Delta_N W_n) = \mathcal{L}(\nu_1^*, \nu_2^*, \ldots, \nu_N^*).$$

Последнее же распределение по упр. 33, гл. 1, сходится при $n \to \infty$ к нормальному распределению $\mathcal{N}(\underline{0}, \Sigma_N)$, где дисперсионная матрица $\Sigma_N = \|\sigma_{ij}\|_1^N$ имеет элементы $\sigma_{ij} = p_i(\delta_{ij} - p_j)$, δ_{ij} — символ Кронекера ($\delta_{ii} = 1$, $\delta_{ij} = 0$, $i \neq j$).

Таким образом, совместное распределение любых приращений эмпирического процесса (т. е. центрированной и нормированной э. ф. р.) асимптотически (при $n \to \infty$) нормально:

$$\mathcal{L}(\Delta_1 W_n, \Delta_2 W_n, \ldots, \Delta_N W_n) \longrightarrow \mathcal{N}(\underline{0}, \Sigma_N) \tag{9}$$

(это значительное усиление свойства асимптотической нормальности (7) эмпирического процесса в фиксированной точке).

Результату (9) можно придать следующую интепретацию. Обозначим $W^0(x)$ случайную функцию (процесс), обладающую тем свойством, что

$$\mathcal{L}(\Delta_1 W^0, \Delta_2 W^0, \ldots, \Delta_N W^0) = \mathcal{N}(\underline{0}, \Sigma_N),$$

т. е. совместные распределения всех ее приращений нормальны (такая функция $W^0(x)$ называется *гауссовским процессом*, и она может быть конструктивно определена, но это уже выходит за рамки нашего пособия). Тогда (9) означает, что при больших n эмпирический процесс $W_n(x)$ приближенно устроен как гауссовский процесс

Предельные распределения функционалов от эмпирического процесса

$$W^0(x) \quad W_n(x) \approx W^0(x).$$

Поэтому следует ожидать, что и распределения «хороших» функционалов $g(\cdot)$ от эмпирического процесса будут слабо сходиться к соответствующим распределениям $g(W^0)$:

$$\mathcal{L}\big(g(W_n)\big) \longrightarrow \mathcal{L}\big(g(W^0)\big) \quad \text{при} \quad n \to \infty.$$

Если это так, и если распределение предельного функционала $g(W^0)$ можно вычислить в явном виде, то на этом пути можно получать предельные теоремы для распределений различных функционалов $g(W_n)$ от эмпирического процесса. Именно на этом пути получена знаменитая теорема А. Н. Колмогорова (1933) о предельном распределении максимального отклонения

$$\mathbf{D}_n = \mathbf{D}_n(X) = \sup_{-\infty<x<\infty} \left|\widehat{F}_n(x) - F(x)\right| \tag{10}$$

э. ф. р. от $F(x)$ на всей оси.

Теорема 2 (Теорема Колмогорова). Если функция $F(x)$ непрерывна, то при любом фиксированном $t > 0$

$$\lim_{n\to\infty} \mathbf{P}\{\sqrt{n}\mathbf{D}_n \leqslant t\} = K(t) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2t^2} \tag{11}$$

Распределение Колмогорова и доверительная зона для теоретической функции распределения

Распределение $K(t)$, определенное в (11), называется *распределением Колмогорова*, и оно не зависит от вида теоретической функции распределения $F(x)$.

При этом предельную функцию распределения $K(t)$, которая подробно протабулирована, можно с хорошим приближением использовать для практических расчетов уже при $n \geqslant 20$.

Теорему Колмогорова обычно применяют для того, чтобы определить границы, в которых с заданной вероятностью находится график теоретической функции распределения $F(x)$, если она неизвестна. Пусть для заданного $\gamma \in (0, 1)$ число t_γ определяется уравнением $K(t_\gamma) = \gamma$. Тогда из (11) имеем:

$$\mathbf{P}\{\sqrt{n}\mathbf{D}_n \leqslant t_\gamma\} = \mathbf{P}\{\widehat{F}_n(x) - t_\gamma/\sqrt{n} \leqslant F(x) \leqslant \\ \leqslant \widehat{F}_n(x) + t_\gamma\sqrt{n} \text{ для всех } x\} \xrightarrow[n\to\infty]{} K(t_\gamma) = \gamma$$

Таким образом, при больших n с вероятностью, близкой к γ, значения функции $F(x)$ для всех x удовлетворяют неравенствам

$$\widehat{F}_n(x) - t_\gamma/\sqrt{n} \leqslant F(x) \leqslant \widehat{F}_n(x) + t_\gamma/\sqrt{n}.$$

Так как $0 \leqslant F(x) \leqslant 1$, то эти неравенства можно несколько уточнить:

$$\max\{0, \widehat{F}_n(x) - t_\gamma/\sqrt{n}\} \leqslant F(x) \leqslant \min\{\widehat{F}_n(x) + t_\gamma/\sqrt{n}, 1\} \tag{12}$$

Область плоскости, определяемая верхней и нижней границами в (12) называется *асимптотической γ-доверительной зоной* для теоретической функции распределения. Для определения числовых значений t_γ при различных γ можно воспользоваться табулированными значениями функции $K(t)$, например, $t_{0,95} = 1{,}3581$, $t_{0,99} = 1{,}6276$.

Полезно знать также следующий результат, принадлежащий В. И. Гливенко[2] (1933).

Теорема 3 (Теорема Гливенко).

$$\mathbf{P}\Big\{\lim_{n\to\infty} \mathbf{D}_n = 0\Big\} = 1.$$

Другими словами, максимальное отклонение эмпирической функции распределения от теоретической функции распределения на всей оси с вероятностью 1 будет сколь угодно мало при достаточно большом объеме выборки.

Наконец, сформулируем еще один важный результат выборочной теории, принадлежащий Н. В. Смирнову (1939), который раскрывает другие свойства эмпирических функций распределения.

Теорема 4 (Теорема Смирнова). Пусть $\widehat{F}_{1n}(x)$ и $\widehat{F}_{2m}(x)$ — две э. ф. р., построенные на основе двух независимых выборок объемов n и m из одного и того же распределения $\mathcal{L}(\xi)$, и

Э. ф. р. для двух выборок

$$\mathbf{D}_{n,m} = \sup_{-\infty<x<\infty} \big|\widehat{F}_{1n}(x) - F_{2m}(x)\big|. \tag{13}$$

Тогда, если теоретическая функция распределения $F(x)$ непрерывна, то для любого фиксированного $t > 0$

$$\lim_{n,m\to\infty} \mathbf{P}\left\{\sqrt{\frac{nm}{n+m}}\mathbf{D}_{n,m} \leqslant t\right\} = K(t), \tag{14}$$

где функция $K(t)$ определена в (11).

Эту теорему обычно используют для проверки гипотезы однородности о том, что две выборки получены из одного и того же распределения (см. пример 5 Введения и п. 1 § 4.3).

Подводя итог обсуждению асимптотических свойств (свойств для больших выборок) эмпирической функции распределения, можно сделать следующее общее заключение. То, что при больших n $\widehat{F}_n(x) \approx F(x)$, находит отражение в следующем фундаментальном принципе математической статистики: для любой характеристики g наблюдаемой случайной величины ξ, которая есть некоторый функционал от теоретической функции распределения $F(x)$ $g = g(F)$, можно построить ее *статистический аналог (оценку)* $\widehat{g}_n = g\,(\widehat{F}_n)$, и тогда для больших n можно использовать $\widehat{g}_n$ в качестве приближенного значения (оценки) для g $\widehat{g}_n \approx g$.

Фундаментальный принцип математической стотистики

[2] Гливенко Валерий Иванович (1896–1940) — российский математик, специалист в области теории вероятностей и математического анализа.

4. Полигон частот, гистограмма

Помимо вариационного ряда и эмпирической функции распределения существуют и другие способы наглядного представления статистических данных, облегчающих работу с ними. Так, если наблюдаемая в эксперименте случайная величина ξ дискретна и принимает значения $a_1, a_2, \ldots$ то более наглядное представление о ее законе распределения дадут относительные частоты $\nu_r^* = \nu_r/n$, где ν_r — число элементов выборки $X = (X_1, \quad, X_n)$, принявших значение a_r: $\nu_r = \sum_{j=1}^{n} I(X_j = a_r)$, $r = 1, 2, \quad$ В этом случае, согласно закону больших чисел, при $n \to \infty$ $\nu_r^* \xrightarrow{\mathbf{P}} \mathbf{P}\{\xi = a_r\}$, $r = 1, 2, \quad$ т. е. ν_r^* сближается с ростом n с теоретической вероятностью $\mathbf{P}\{\xi = a_r\}$, и потому, по крайней мере для больших выборок, относительные частоты ν_r^* можно рассматривать в качестве приближенных значений (оценок) для неизвестных вероятностей $\mathbf{P}\{\xi = a_r\}$.

Познавательная ценность теории вероятностей обусловлена тем, что массовые случайные явления в своем совокупном действии создают строгие закономерности. Само понятие математической вероятности было бы бесплодно, если бы не находило своего осуществления в виде частоты появления какого-либо результата при многократном повторении однородных условий.

А. Н. Колмогоров

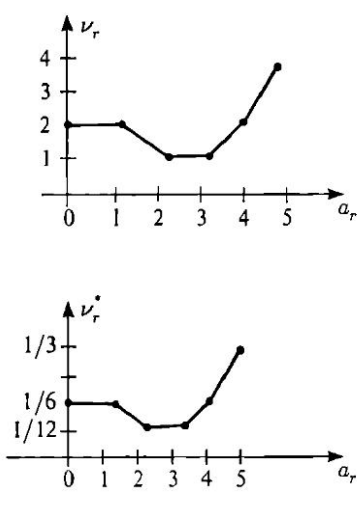


Рис. 5

Исходные данные дискретного типа обычно записывают в виде последовательности пар $\{(a_r, \nu_r)\}$, где $\nu_r > 0$, или в виде таблицы

a_1	a_2	. .
ν_1	ν_2	

,

называемой *статистическим рядом* (подчеркнем, что в этом случае $\sum_r \nu_r = n$).

Наглядным представлением таких данных является *полигон частот*, который представляет собой ломаную с вершинами в точках (a_r, ν_r), $r = 1, 2, \quad$ (см. рис. 5).

Можно рассматривать также статистический ряд $\{(a_r, \nu_r^*)\}$ (здесь $\sum_r \nu_r^* = 1$) и соответствующий полигон относительных частот.

Пример 4. Наблюденные значения целочисленной величины ξ для 12 опытов оказались равными 5, 1, 4, 0, 1, 4, 3, 5, 0, 5, 5, 2. Статистические ряды и полигоны частот этих данных таковы:

a_r	0	1	2	3	4	5	Σ
ν_r	2	2	1	1	2	4	12

a_r	0	1	2	3	4	5	Σ
ν_r^*	1/6	1/6	1/12	1/12	1/6	1/3	1

Метод группировки и гистограмма

Для непрерывной случайной величины ξ, обладающей непрерывной плотностью $f(x)$, также можно построить по соответствующей выборке $X = (X_1, \ldots, X_n)$ статистический аналог $\widehat{f}_n(x)$ для плотности $f(x)$, который называется *гистограммой*. Для этого используется *метод группировки*, в соответствии с которым область Δ возможных значений ξ разбивается на некоторое число N непересекающихся интервалов $\Delta_1, \ldots, \Delta_N$ (так что $\Delta = \bigcup_{r=1}^{N} \Delta_r$), подсчитывают числа $\nu_1, \quad , \nu_N$ наблюдений $X_1, \quad , X_n$, попавших в соответствующие интервалы:

$$\nu_r = \sum_{j=1}^{N} I(X_j \in \Delta_r), \quad r = 1, \quad , N \quad \left(\text{так что } \sum_{r=1}^{N} \nu_r = n\right),$$

и строят кусочно-постоянную функцию

$$\widehat{f}_n(x) = \frac{\nu_r}{n|\Delta_r|} \quad \text{при} \quad x \in \Delta_r, \quad r = 1, \quad , N \tag{15}$$

(здесь $|\Delta_r|$ — длина интервала Δ_r). То, что построенная по такому правилу гистограмма $\widehat{f}_n(x)$ действительно «похожа» на теоретическую плотность $f(x)$, следует из закона больших чисел, согласно которому при $n \to \infty$ относительная частота ν_r/n сближается с теоретической вероятностью

$$\mathbf{P}\{\xi \in \Delta_r\} = \int_{\Delta_r} f(x)\, dx.$$

Но этот интеграл по теореме о среднем равен $f(a_r)|\Delta_r|$, где a_r — некоторая внутренняя точка интервала Δ_r (при малом Δ_r в качестве a_r можно взять, например, середину интервала). Таким образом, при больших n и достаточно «мелком» разбиении $\{\Delta_r\}$ $\widehat{f}_n(x) \approx f(a_r)$ при $x \in \Delta_r$, т. е. гистограмма $\widehat{f}_n(x)$ будет достаточно хорошо приближать график плотности $f(x)$, следовательно, $\widehat{f}_n(x)$ можно рассматривать в качестве *статистического аналога* (оценки) для $f(x)$. •



Рис. 6

Полигон частот

Наряду с гистограммой, в качестве приближения для неизвестной теоретической плотности $f(x)$ можно использовать кусочно-линейный график, называемый *полигоном частот*, и который строится так: если построена гистограмма $\widehat{f}_n(x)$, то ординаты, соответствующие серединам интервалов группировки, последовательно соединяют отрезками прямых (см. рис. 6). Построенный таким образом кусочно-линейный график также является статистическим аналогом теоретической плотности. Группированные данные удобно представлять в виде соответствующего статистического ряда $\{(a_r, \nu_r)\}$, где a_r — середина r-го интервала группировки.

Метод группировки применяется и для дискретных данных, когда, например, множество возможных значений наблюдаемой случайной величины счетно (как для пуассоновской или геометрической величины). Относительные частоты ν_r/n в любом случае являются статистическими аналогами (оценками) для теоретических вероятностей $\mathbf{P}\{\xi \in \Delta_r\}$.

Метод группировки наблюдений — удобный и часто более экономный способ записи информации, содержащейся в выборке. Если, например, $n = 10^4$ а точность измерений наблюдаемых значений на интервале $(0, 1)$ сравнима с 0,1, то ясно, что практически бесполезно знать все 10^4 измерений, а достаточно указать лишь 10 частот $\nu_1, \quad , \nu_{10}$ попадания в интервалы

$$\Delta_r = \left(\frac{r-1}{10}, \frac{r}{10}\right], \quad r = 1, \qquad 10,$$

т. е. знать лишь гистограмму выборки. Однако, этот способ обладает и очевидными недостатками, связанными с некоторой неопределенностью в способе разбиения $\{\Delta_r\}$ и частичной потерей информации при огрублении данных (ведь фактически мы все наблюдения, попавшие в интервал Δ_r, заменяем на среднюю точку a_r этого интервала). Поэтому гистограмму используют практически лишь на предварительном этапе анализа статистических данных.

Пример 5. Наблюдения над непрерывной случайной величиной ξ, округленные с точностью 10^{-2}, оказались равными 0,59; 0,16; 0,44; 0,48; 0,90; 0,19; 0,65; 0,42; 0,35; 0,84; 0,28; 0,63; 0,54; 0,12; 0,18; 0,67; 0,94; 0,63; 0,33; 0,03. Построим гистограмму и полигон частот этих данных, взяв в качестве интервалов группировки

$$\Delta_r = \left(\frac{r-1}{10}, \frac{r}{10}\right), \quad r = 1, \ldots, 10.$$

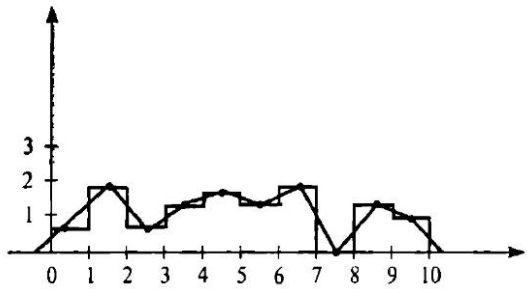


Рис. 7

Частоты интервалов здесь равны

r	1	2	3	4	5	6	7	8	9	10	Σ
ν_r	1	4	1	2	3	2	4	0	2	1	20

$|\Delta_r| = 10^{-1}$ $n|\Delta_r| = 2$, поэтому по формуле (15) значение гистограммы $\widehat{f}_{20}(x)$ в r-м интервале Δ_r равно $\nu_r/2$. Вид гистограммы и полигона частот указан на рис. 7 •

5. Ядерные оценки теоретической плотности

(Этот материал также для желающих заглянуть поглубже в проблему оценивания теоретической плотности)

Один из современных подходов в задаче оценивания неизвестной теоретической плотности $f(x)$ основан на использовании так называемых *ядерных оценок*, т. е. оценок вида

$$\widetilde{f}_n(x) = \frac{1}{na_n} \sum_{i=1}^{n} k\left(\frac{x - X_i}{a_n}\right) \tag{16}$$

при соответствующем выборе ядра $k(t)$ и последовательности чисел $a_n \to 0$ при $n \to \infty$. В качестве ядра $k(t)$ здесь выбирают гладкие ограниченные плотности (например, равномерную плотность на интервале $(-1/2, 1/2)$), что обеспечивает хорошее сближение $\widetilde{f}_n(x)$ с $f(x)$ при $n \to \infty$. Будем далее считать, что выполняется условие

$$d^2 = \int k^2(t)\, dt < \infty \tag{17}$$

и $a_n \to 0$ при $n \to \infty$ так, что $na_n \to \infty$ (например, $a_n = n^{-\alpha}$ $0 < \alpha < 1$).

Тогда можно показать, что при этих условиях для непрерывной и ограниченной плотности $f(x) > 0$ при $n \to \infty$

$$\mathcal{L}\left(\sqrt{na_n}\left(\widetilde{f}_n(x) - \mathbf{E}\widetilde{f}_n(x)\right)\right) \to \mathcal{N}\left(0, \sigma^2(x)\right), \tag{18}$$

где

$$\sigma^2(x) = d^2 f(x);$$

при этом

$$\mathbf{E}\widetilde{f}_n(x) = \mathbf{E}\frac{1}{a_n}k\left(\frac{x-\xi}{a_n}\right) = \frac{1}{a_n}\int k\left(\frac{x-t}{a_n}\right)f(t)\,dt =$$

$$\int k(z)f(x-a_n z)dz \to f(x),$$

так как

$$\int k(z)\,dz = 1.$$

Таким образом, (18) означает, что $\widetilde{f}_n(x)$ сближается с $f(x)$ при $n \to \infty$: отклонение $\widetilde{f}_n(x)$ от $\mathbf{E}\widetilde{f}_n(x)$ (которое, в свою очередь, близко к $f(x)$) имеет порядок $1/\sqrt{na_n}$, который мал при больших n.

Доказательство утверждения (18) состоит в применении центральной предельной теоремы к сумме (16) независимых и одинаково распределенных слагаемых в «схеме серий», т. е. когда слагаемые (а, следовательно, и их распределение) зависят от числа n (оно выходит за рамки нашего пособия, и мы его опускаем).

Вопрос об оптимальном выборе «свободных» параметров a_n и функции (ядра) $k(t)$ зависит здесь от свойств гладкости теоретической плотности $f(x)$. Так, если $f(x)$ имеет непрерывные производные порядка $2m > 2$, то можно выбором a_n и $k(t)$ добиться скорости сближения $\widetilde{f}_n(x)$ с $f(x)$ порядка $n^{-2m/(4m+1)}$ при $n \to \infty$.

Прямоугольное ядро

На практике часто используют «прямоугольные» ядра:

$$k(t) = \begin{cases} 1, & -\frac{1}{2} \leqslant t \leqslant \frac{1}{2}, \\ 0 & \text{в противном случае.} \end{cases}$$

В таком случае $\widetilde{f}_n(x) = \nu_n(x)/(na_n)$, где

$$\nu_n(x) = \sum_{i=1}^{n} I\left(x - \frac{a_n}{2} \leqslant X_i \leqslant x + \frac{a_n}{2}\right)$$

— число наблюдений в интервале $\left[x - \frac{a_n}{2},\ x + \frac{a_n}{2}\right]$

Имеются также результаты вида:

$$\mathbf{P}\left\{\lim_{n\to\infty} \sup_x \left|\widetilde{f}_n(x) - f(x)\right| = 0\right\} = 1$$

(при соответствующих условиях на $f(x)$, $k(t)$ и a_n) и вида

$$\mathbf{P}\left\{\sup_x \left|\widetilde{f}_n(x) - f(x)\right| > \varepsilon_n\right\} \sim \delta_n,$$

позволяющие рассчитывать асимптотические доверительные зоны для $f(x)$, основанные на ядерной оценке $\widetilde{f}_n(x)$ вида (16).

§ 2.2. Выборочные моменты: точная и асимптотическая теория

Пусть $X = (X_1, \quad , X_n)$ — выборка из распределения $\mathcal{L}(\xi)$ и $F(x)$ и $\widehat{F}_n(x)$ — соответственно теоретическая и эмпирическая функции распределения. На $\widehat{F}_n(x)$ можно смотреть как на функцию распределения некоторой дискретной случайной величины, принимающей n значений: $X_1, \quad , X_n$ — с вероятностями, равными $1/n$ (если какое-либо значение встретится в выборке k раз, то этому значению соответствует вероятность k/n). Об этом распределении говорят как об *эмпирическом* или *выборочном* распределении (отсюда и термин *эмпирическая функция распределения* для $\widehat{F}_n(x)$). Как для исходного распределения $\mathcal{L}(\xi)$ вводятся различные числовые характеристики (математическое ожидание, или среднее, дисперсия, моменты и т. д.), так и для эмпирического распределения, связанного с выборкой X, вводятся аналогичные характеристики, называемые *эмпирическими* (или *выборочными*): *выборочное среднее*, *выборочная дисперсия* и т. д. Таким образом, эмпирическая (или выборочная) характеристика является *статистическим аналогом* соответствующей теоретической характеристики, аналогично тому, как эмпирическая функция распределения $\widehat{F}_n(x)$ является статистическим аналогом теоретической функции распределения $F(x)$. В общем случае, если $g = \mathbf{E}g(\xi)$ есть некоторая теоретическая характеристика наблюдаемой случайной величины ξ, то ее статистический аналог, т. е. соответствующая эмпирическая (или выборочная) характеристика, вычисляется по формуле

$$\widehat{g}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i). \tag{1}$$

1. Выборочные моменты

Наиболее важными характеристиками случайной величины ξ являются ее моменты $\alpha_k = \mathbf{E}\xi^k$, а также центральные моменты $\mu_k = \mathrm{E}(\xi - \alpha_1)^k$ (когда они существуют). Их статистическими аналогами, вычисляемыми по соответствующей выборке $X = (X_1, \quad , X_n)$, являются *выборочные моменты* соответственно *обычные*

$$\widehat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{2}$$

и *центральные*

$$\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\alpha}_1)^k. \tag{3}$$

Выборочное среднее

Особенно важны моменты первого и второго порядков. При $k = 1$ величину $\widehat{\alpha}_1$ называют *выборочным средним* и обозначают стандартным символом $\overline{X}$ (черта означает среднее арифметическое):

$$\overline{X} = \widehat{\alpha}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{4}$$

Выборочная дисперсия

При $k = 2$ величину $\widehat{\mu}_2$ называют *выборочной дисперсией* и также обозначают стандартным символом $S^2 = S^2(X)$:

$$S^2 = \widehat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{5}$$

Таким образом, выборочные среднее и дисперсия являются статистическими аналогами теоретических среднего (математического ожидания) $\mathbf{E}\xi$ и дисперсии $\mathbf{D}\xi$, когда они существуют.

Между выборочными $\widehat{\alpha}_k$ и выборочными центральными $\widehat{r}_k$ моментами сохраняются те же соотношения, что и между теоретическими обычными α_k и центральными μ_k моментами. В общем случае при $k \geqslant 2$ из (3) имеем по формуле бинома Ньютона

$$\widehat{\mu}_k = \sum_{r=0}^{k}(-1)^r C_k^r \overline{X}^r \widehat{\alpha}_{k-r},$$

т.е. выборочные центральные моменты являются многочленами от обычных выборочных моментов. В частности,

$$\begin{aligned} S^2 &= \widehat{\alpha}_2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2, \\ \widehat{\mu}_3 &= \widehat{\alpha}_3 - 3\overline{X}\,\widehat{\alpha}_2 + 2\overline{X}^3, \\ \widehat{\mu}_4 &= \widehat{\alpha}_4 - 4\overline{X}\,\widehat{\alpha}_3 + 6\overline{X}^2\widehat{\alpha}_2 - 3\overline{X}^4 \end{aligned} \tag{6}$$

Если выборочные данные представлены в виде статистического ряда $\{(a_r, \nu_r)\}$ (когда наблюдаемая величина ξ дискретна или когда применяется метод группировки), то выборочные моменты вычисляются по формулам

$$\widehat{\alpha}_k = \frac{1}{n}\sum_r a_r^k \nu_r, \quad \widehat{\mu}_k = \frac{1}{n}\sum_r (a_r - \widehat{\alpha}_1)^k \nu_r;$$

в частности,

$$\begin{aligned} \widehat{\alpha}_1 &= \bar{a} = \frac{1}{n}\sum_r a_r \nu_r, \\ \widehat{\mu}_2 = S^2 &= \frac{1}{n}\sum_r (a_r - \bar{a})^2 \nu_r = \frac{1}{n}\sum_r a_r^2 \nu_r - \bar{a}^2. \end{aligned} \tag{7}$$

В этом случае для упрощения вычислений исходные данные часто удобно преобразовать, положив $b_r = (a_r - \widetilde{a})/c$, где сдвиг $\widetilde{a}$ есть значение a_r с максимальной частотой ν_r и единица масштаба $c > 0$ — произвольна. Средние $\bar{a}$ и $\bar{b}$, и дисперсии $S^2(a)$, $S^2(b)$ соответственно исходных и преобразованных данных связаны при этом соотношениями

$$\bar{a} = c\bar{b} + \widetilde{a}, \quad S^2(a) = c^2 S^2(b). \tag{8}$$

Пример 1. Вычислим выборочные среднее и дисперсию по исходным и группированным данным, приведенным в примере 5 § 2.1. По формулам (4) и (6) имеем

$$\overline{X} = \frac{9{,}37}{20} \approx 0{,}468, \quad S^2 \approx 0{,}067.$$

Далее, середины интервалов Δ_r в данном случае равны $a_r = r \cdot 0{,}1 - 0{,}05$, при этом $\widetilde{a} = a_7 = 0{,}65$; кроме того, здесь удобно выбрать в качестве единицы масштаба $c = 0{,}1$. Таким образом, статистический ряд преобразованных по формуле $b_r = (a_r - \widetilde{a})/c = (r \cdot 0{,}1 - 0{,}7)\,10 = r - 7$ данных имеет вид

b_r	-6	-5	-4	-3	-2	-1	0	2	3	Σ
ν_r	1	4	1	2	3	2	4	2	1	20

Теперь уже вычисляем (см. (7)–(8))

$$\sum_r b_r \nu_r = -37, \quad \bar{b} = -\frac{37}{20} = -1{,}85$$

и, следовательно, $\bar{a} = 0{,}65 - 0{,}185 = 0{,}465$.

Аналогично,

$$\sum_r b_r^2 \nu_r = 201, \quad S^2(b) = \frac{1}{n}\sum_r b_r^2 \nu_r - \bar{b}^2 = \frac{201}{20} - 1{,}85^2 \approx 6{,}628$$

и, следовательно, $S^2(a) \approx 0{,}066$. •

Пример 2. Проводились опыты с бросанием одновременно 12 игральных костей, при этом наблюдалось число ξ костей с числом очков 4, 5 или 6. Данные для $n = 4096$ опытов приведены в таблице, где ν_r обозначает число опытов, в которых наблюдалось значение $\xi = r$ $(r = 0, 1, \quad , 12)$

r	0	1	2	3	4	5	6	7	8	9	10	11	12	Σ
ν_r	0	7	60	198	430	731	948	847	536	257	71	11	0	4096

Назовем успехом выпадение на одной кости 4, 5 или 6 очков — вероятность этого равна $1/2$, тогда наблюдаемая случайная величина ξ равна числу успехов в 12 испытаниях Бернулли с вероятностью успеха $1/2$, т. е ξ имеет биномиальное распределение $\mathrm{Bi}(12, 1/2)$. По формулам (4) § 1.1 теоретическое среднее есть $\mathbf{E}\xi = 6$, а теоретическая дисперсия $\mathbf{D}\xi = 3$.

Вычислим теперь выборочное среднее и дисперсию для приведенных данных. Здесь мы имеем дело со статистическим рядом $\{(r, \nu_r)\}$, поэтому

вычисления производим по формулам (7)–(8), предварительно преобразовав данные, положив $b_r = r - 6$ (поскольку здесь $\widetilde{a} = 6$). Статистический ряд так преобразованных данных имеет вид

b_r	-5	-4	-3	-2	-1	0	1	2	3	4	5	Σ
ν_r	7	60	198	430	731	948	847	536	257	71	11	4096

Здесь

$$\sum_r b_r \nu_r = 569, \quad \bar{b} = \frac{569}{4096} \approx 0{,}14$$

и, следовательно, $\widehat{\alpha}_1 = \widetilde{a} + \bar{b} \approx 6{,}14$.

Аналогично,

$$\sum_r b_r^2 \nu_r = 12\,083 \quad \text{и} \quad \widehat{\mu}_2 = S^2(b) = \frac{1}{n}\sum_r b_r^2 \nu_r - \bar{b}^2 \approx 2{,}93.$$

В данном случае различие между теоретическими и выборочными моментами незначительно. •

Выборочные моменты — это функции от выборки X, т. е. они являются случайными величинами, поэтому важно знать их вероятностные свойства, т. е. уметь находить их распределения (которые иногда называют *выборочными распределениями*) и изучать различные характеристики этих распределений. Рассмотрим сначала некоторые характеристики выборочных среднего и дисперсии.

2. Моменты выборочных среднего и дисперсии

Так как X_i независимы и распределены так же, как ξ, то

$$\mathbf{E}\overline{X} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}X_i = \mathbf{E}\xi = \alpha_1, \quad \mathbf{D}\overline{X} = \frac{1}{n^2}\sum_{i=1}^{n} \mathbf{D}X_i = \frac{1}{n}\mathbf{D}\xi = \frac{\mu_2}{n}. \tag{9}$$

Рассмотрим теперь выборочную дисперсию S^2 (см. (5)). Положив

$$Y_i = X_i - \alpha_1, \quad i = 1, \quad , n,$$

запишем (5) в виде

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2 \tag{10}$$

Поскольку $\mathbf{E}Y_i = 0$, $\mathbf{E}Y_i^2 = \mu_2$ и $\mathbf{E}Y_iY_j = \mathbf{E}Y_i\mathbf{E}Y_j = 0$ $(i \neq j)$, то

$$\mathbf{E}\overline{Y}^2 = \frac{1}{n^2}\sum_{i,j=1}^{n} \mathbf{E}Y_iY_j = \frac{1}{n^2}\sum_{i=1}^{n} \mathbf{E}Y_i^2 = \frac{\mu_2}{n}.$$

Отсюда и из (10) находим

Среднее выборочной дисперсии

$$\mathbf{E}S^2 = \frac{n-1}{n}\mu_2. \qquad (11)$$

Перейдем к вычислению $\mathbf{D}S^2$ Из (10) находим

$$(S^2)^2 = \frac{1}{n^2}\left(\sum_{i=1}^{n} Y_i^2\right)^2 - \frac{2}{n}\overline{Y}^2\sum_{i=1}^{n} Y_i^2 + \overline{Y}^4 \qquad (12)$$

Так как случайные величины $Y_1, \ldots, Y_n$ независимы и $\mathbf{E}Y_i = 0$, то в правой части равенства

$$\mathbf{E}\overline{Y}^4 = \frac{1}{n^4}\sum_{i,k,l=1}^{n} \mathbf{E}Y_iY_jY_kY_l$$

отличны от нуля только n слагаемых $\mathbf{E}Y_i^4 = \mu_4$ и $C_n^2C_4^2 = 3n(n-1)$ слагаемых $\mathbf{E}Y_i^2Y_j^2 = \mu_2^2$, $i \neq j$. Поэтому

$$\mathbf{E}\overline{Y}^4 = \frac{1}{n^4}\left(n\mu_4 + 3n(n-1)\mu_2^2\right) = \frac{\mu_4 + 3(n-1)\mu_2^2}{n^3}.$$

Аналогично находим

$$\frac{1}{n^2}\mathbf{E}\left(\sum_{i=1}^{n} Y_i^2\right)^2 = \frac{\mu_4 + (n-1)\mu_2^2}{n} = \mathbf{E}\left(\overline{Y}^2\sum_{i=1}^{n} Y_i^2\right).$$

С учетом этих соотношений из (12) и (11) по формуле

$$\mathbf{D}S^2 = \mathbf{E}(S^2)^2 - (\mathbf{E}S^2)^2$$

получим

Дисперсия выборочной дисперсии

$$\mathbf{D}S^2 = \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3} =$$

$$= \frac{(n-1)^2}{n^3}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right). \qquad (13)$$

Аналогично можно находить моменты и более высоких порядков, хотя с увеличением порядка вид формул и их вывод усложняются.

3. Сходимость по вероятности выборочных моментов и функций от них

В этом пункте мы проанализируем свойства выборочных моментов, определенных равенствами (2) и (3), для больших выборок, т. е. их асимптотическое поведение при неограниченном возрастании объема выборки n. Чтобы подчеркнуть зависимость моментов $\widehat{\alpha}_k$ и $\widehat{\mu}_k$ от объема выборки, будем их

в дальнейшем снабжать индексом n, т. е. писать $\widehat{\alpha}_{nk}$ и $\widehat{\mu}_{nk}$. Первые два момента случайной величины $\widehat{\alpha}_{nk}$ соответственно равны (везде далее предполагаем, что все соответствующие моменты наблюдаемой случайной величины ξ существуют)

$$\begin{aligned} \mathbf{E}\widehat{\alpha}_{nk} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}X_i^k = \mathbf{E}\xi^k = \widehat{\alpha}_k, \\ \mathbf{D}\widehat{\alpha}_{nk} &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}X_i^k = \frac{1}{n}\mathbf{D}\xi^k = \frac{1}{n}\left(\mathbf{E}\xi^{2k} - (\mathbf{E}\xi^k)^2\right) = \frac{\alpha_{2k} - \alpha_k^2}{n}. \end{aligned} \tag{14}$$

На основании неравенства Чебышева (напоминание см. перед формулой (6) § 2.1) отсюда следует, что для любого $\varepsilon > 0$ при $n \to \infty$

$$\mathbf{P}\{|\widehat{\alpha}_{nk} - \alpha_k| < \varepsilon\} \longrightarrow 1, \tag{15}$$

т. е. выборочный момент $\widehat{\alpha}_{nk}$ сходится по вероятности при $n \to \infty$ к соответствующему теоретическому моменту α_k. Тем самым $\widehat{\alpha}_{nk}$ можно использовать в качестве приближенного значения (оценки) для α_k, когда число наблюдений n достаточно велико. Аналогичное утверждение справедливо и для выборочных центральных моментов: $\widehat{\mu}_{nk} \xrightarrow{\mathbf{P}} \mu_k$, и вообще для любых выборочных характеристик, которые имеют вид непрерывных функций от конечного числа величин $\widehat{\alpha}_{nk}$ (центральный момент $\widehat{\mu}_{nk}$, как показано выше в п. 1 (см. (6)), выражается через $\widehat{\alpha}_{n1}, \ldots, \widehat{\alpha}_{nk}$ в виде многочлена степени k). Этот вывод является следствием следующей общей теоремы о сходимости функций от случайных величин.

Теорема 1. Пусть случайные величины $\eta_1(n), \ldots, \eta_r(n)$ сходятся по вероятности при $n \to \infty$ к некоторым постоянным $c_1, \ldots, c_r$ соответственно. Тогда для любой непрерывной функции $\varphi(x_1, \ldots, x_r)$ случайная величина

$$\zeta(n) = \varphi(\eta_1(n), \ldots, \eta_r(n)) \xrightarrow{\mathbf{P}} \varphi(c_1, \ldots, c_r).$$

Доказательство. Поскольку φ непрерывна, то для любого $\varepsilon > 0$ найдется $\delta = \delta(\varepsilon)$ такое, что $|\varphi(x_1, \ldots, x_r) - \varphi(c_1, \ldots, c_r)| < \varepsilon$ при $|x_i - c_i| < \delta$, $i = 1, \ldots, r$. Введем события $B_i \quad \{|\eta_i(n) - c_i| < \delta\}$, $i = 1, \ldots, r$ Тогда событие $B = B_1 \ldots B_r$ влечет событие

$$C = \{|\zeta(n) - \varphi(c_1, \ldots, c_r)| < \varepsilon\}$$

Отсюда

$$\begin{aligned} \mathbf{P}\{C\} &\geqslant \mathbf{P}\{B\} = 1 - \mathbf{P}\{\overline{B}\} = \\ &= 1 - \mathbf{P}\{\overline{B}_1 \cup \ldots \cup \overline{B}_r\} \geqslant 1 - \sum_{i=1}^{r}\mathbf{P}\{\overline{B}_i\}. \end{aligned} \tag{16}$$

Далее, из сходимости $\eta_i(n) \xrightarrow{\mathbf{P}} c_i$ имеем, что для данного δ и любого $\gamma > 0$ найдется $n_i = n_i(\gamma)$ такое, что

$$\mathbf{P}\{\overline{B}_i\} = \mathbf{P}\{|\eta_i(n) - c_i| \geqslant \delta\} < \frac{\gamma}{r}$$

при $n \geqslant n_i$. Пусть $n_0 = \max\{n_1, \ldots, n_r\}$, тогда при $n \geqslant n_0$ выполняются одновременно все неравенства $\mathbf{P}\{\overline{B}_i\} < \gamma/r$, $i = 1, \quad , r$. Следовательно, из (16) получим

$$\mathbf{P}\{|\zeta(n) - \varphi(c_1, \ldots, c_r)| < \varepsilon\} \geqslant 1 - \gamma \quad \text{при} \quad n \geqslant n_0.$$

Но это и означает утверждение теоремы. ■

Замечание. Очевидно, что для справедливости теоремы 1 достаточно потребовать непрерывности функции φ только в некоторой окрестности точки $(c_1, \ldots, c_r)$.

В частности, теорема 1 дает, что выборочная дисперсия S^2 (см. (5)) сходится по вероятности к теоретической дисперсии

$$\mu_2 = \mathbf{D}\xi \quad S^2 \xrightarrow{\mathbf{P}} \mu_2.$$

Подправленная выборочная дисперсия

Однако, в отличие от формулы (14), где математическое ожидание $\widehat{\alpha}_{nk}$ совпало с соответствующим теоретическим моментом α_k, в (11) правая часть смещена относительно μ_2. Если вместо S^2 ввести величину

$$S_0^2 = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2, \tag{17}$$

то для нее получим

$$\mathbf{E}S_0^2 = \mu_2, \quad \mathbf{D}S_0^2 = \left(\frac{n}{n-1}\right)^2 \mathbf{D}S^2 \tag{18}$$

Из неравенства Чебышева и формул (18) и (13) следует, что для любого $\varepsilon > 0$ при $n \to \infty$

$$\mathbf{P}\{|S_0^2 - \mu_2| < \varepsilon\} \to 1, \tag{19}$$

т. е. $S_0^2 \xrightarrow{\mathbf{P}} \mu_2$. Таким образом, подправленная выборочная дисперсия S_0^2 в (17) уже будет иметь математическое ожидание, совпадающее с теоретической дисперсией μ_2, сохраняя при этом свойство сходимости по вероятности к μ_2 с ростом объема выборки (что вполне естественно, поскольку поправочный коэффициент $n/(n-1) \to 1$ при $n \to \infty$).

Теперь настало время ввести несколько новых понятий, которыми широко оперирует современная математическая статистика — это понятия *статистики, несмещенности и состоятельности.*

Определение 1 Любая случайная величина T, являющаяся функцией от выборки $X = (X_1, \ldots, X_n)$: $T = T(X)$, называется *статистикой.*

Примерами конкретных статистик являются выборочные моменты $\widehat{\alpha}_k$ и выборочные центральные моменты $\widehat{\mu}_k$, статистикой является выборочная характеристика $\widehat{g}_n$ в (1). Приведем еще два примера конкретных статистик, используемых в различных статистических задачах.

При исследовании свойств графика плотности распределения непрерывной случайной величины часто рассматривают такие характеристики, как *коэффициенты асимметрии* γ_1 и *эксцесса* γ_2, определяемые формулами

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}, \quad \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3. \tag{20}$$

Если график плотности симметричен относительно нуля, то $\gamma_1 = 0$ и по значению γ_1 судят о степени отклонения от симметрии. Аналогично, для нормального распределения $\gamma_2 = 0$, поэтому о кривых плотности с $\gamma_2 = 0$ говорят, что они имеют *нормальный эксцесс*. Если же $\gamma_2 > 0$ ($\gamma_2 < 0$), то говорят, что эксцесс кривой плотности положителен (отрицателен). Если задана выборка $X = (X_1, \quad , X_n)$ из абсолютно непрерывного распределения $\mathcal{L}(\xi)$, то *выборочные коэффициенты асимметрии* $\widehat{\gamma}_{n1}$ и *эксцесса* $\widehat{\gamma}_{n2}$ определяются как статистики

$$\widehat{\gamma}_{n1} = \frac{\widehat{\mu}_{n3}}{\widehat{\mu}_{n2}^{3/2}} = \frac{\widehat{\mu}_{n3}}{S^3} \quad \text{и} \quad \widehat{\gamma}_{n2} = \frac{\widehat{\mu}_{n4}}{S^4} - 3. \tag{21}$$

Эти статистики как непрерывные функции от выборочных моментов в силу теоремы 1 сходятся по вероятности при $n \to \infty$ к соответствующим теоретическим характеристикам (20).

Определение 2 Статистика $T = T(X)$, используемая в качестве приближенного значения для некоторой (неизвестной) теоретической характеристики g : $T \approx g$, называется *несмещенной в среднем* (или просто *несмещенной*) *оценкой* g (для g), если выполняется условие

$$\mathbf{E}\, T(X) = g. \tag{22}$$

Свойство несмещенности интуитивно привлекательно: оно означает, что по крайней мере «в среднем» используемая оценка приводит к желаемому результату. Поэтому в конкретных задачах оценивания часто ищут именно несмещенную оценку. С примерами несмещенных оценок мы уже выше встречались. Так, первое соотношение (5) § 2.1 означает, что эмпирическая функция распределения $\widehat{F}_n(x)$ является несмещенной оценкой теоретической функции распределения $F(x)$ в каждой точке $x \in R^1$; аналогично, ввиду (14) выборочный момент $\widehat{\alpha}_{nk}$ является несмещенной оценкой для теоретического момента α_k, а ввиду (18) статистика S_0^2, определенная в (17), является несмещенной оценкой теоретической дисперсии μ_2. Выборочная же дисперсия S^2 в качестве оценки для $\mu_2 = \mathbf{D}\xi$ свойством несмещенности не обладает: она имеет *смещение* $\mathbf{E}S^2 - \mu_2 = -\mu_2/n$ (см. (11)), которое стремится к 0 при $n \to \infty$ — такие оценки называются *асимптотически несмещенными*. Тем самым выборочная дисперсия S^2 является лишь асимптотически несмещенной оценкой μ_2. Точно

так же и другие выборочные центральные моменты $\widehat{\mu}_{nk}$ (см. (3)) являются асимптотически несмещенными оценками для теоретических центральных моментов μ_k, поскольку для них известно, что $\mathbf{E}\,\widehat{\mu}_{nk} - \mu_k \to 0$ при $n \to \infty$.

Определение 3 Статистика $T_n = T_n(X_1, \ldots, X_n)$ называется *состоятельной* оценкой для теоретической характеристики g, если $T_n \xrightarrow{\mathbf{P}} g$ при $n \to \infty$.

Таким образом, состоятельность в статистике — это эквивалентное понятие для сходимости по вероятности. Оно означает, что для выборок большого объема значительная разница между значениями оценки и оцениваемой характеристики маловероятна, и потому разумно (по крайней мере для больших выборок) принять статистику T_n за приближенное значение неизвестной характеристики g. Состоятельность является обязательным требованием для любого правила оценивания. Подчеркнем, что это свойство оценок имеет асимптотический характер и связано лишь с большими выборками (с увеличением информации, доставляемой выборкой, точность оценки должна возрастать!). В отличие от этого свойство несмещенности никак не связано с числом наблюдений. Отметим также, что все выборочные моменты (как обычные, так и центральные) являются состоятельными оценками соответствующих теоретических моментов, выборочные коэффициенты асимметрии и эксцесса (21) также являются состоятельными оценками соответствующих теоретических характеристик (20).

Критерий состоятельности

Наконец, подчеркнем, что для проверки состоятельности некоторой несмещенной оценки T_n для g достаточно убедиться, что ее дисперсия $\mathbf{D}T_n \to 0$ при $n \to \infty$, так как в этом случае по неравенству Чебышева

$$\mathbf{P}\{|T_n - g| > \varepsilon\} \leqslant \frac{\mathbf{D}T_n}{\varepsilon^2} \to 0, \quad \forall\, \varepsilon > 0. \tag{23}$$

На основании этого критерия и ввиду (5) § 2.1 можно заключить, что эмпирическая функция распределения $\widehat{F}_n(x)$ является состоятельной оценкой теоретической функции распределения $F(x)$ в любой точке $x \in R^1$.

Ввиду особой практической важности оценивания теоретических среднего $\alpha_1 = \mathbf{E}\xi$ и дисперсии $\mu_2 = \mathbf{D}\xi$ суммируем полученные выше результаты (9), (18) и (19) в форме отдельной теоремы.

Теорема 2. Выборочное среднее

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

и статистика

$$S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

являются несмещенными и состоятельными оценками соответственно для $\mathbf{E}\xi$ и $\mathbf{D}\xi$.

Пример 3. Рассмотрим ситуацию, описанную в примере 1 § 1.1. В данном случае речь идет об оценивании теоретического среднего $\theta = \mathbf{E}\xi$ (доли белых шаров в урне) в схеме Бернулли $\mathrm{Bi}(1, \theta)$ по выборке $\boldsymbol{X} = (X_1, \quad, X_n)$, где $X_j = 1$, если в j-м испытании извлечен белый шар, и $X_j = 0$ в противном случае. Здесь выборочное среднее

$$\overline{X} = \frac{1}{n}(X_1 + \quad + X_n) = \frac{X}{n},$$

где X — число наблюдавшихся белых шаров, является несмещенной и состоятельной оценкой $\theta = a_1/a$. •

4. Асимптотическая нормальность выборочных моментов

Продолжим исследование свойств выборочных моментов для больших выборок и обсудим теперь асимптотическое при $n \to \infty$ поведение их выборочных распределений. Но сначала введем некоторые дополнительные обозначения и определения для удобства дальнейшего изложения (см. Общее замечание в п. 1 § 1.2).

Если распределение случайной величины η_n сходится при $n \to \infty$ к распределению случайной величины η и при этом $\mathcal{L}(\eta) = \mathcal{N}(\mu, \sigma^2)$, то будем писать $\mathcal{L}(\eta_n) \to \mathcal{N}(\mu, \sigma^2)$. Далее иногда будем говорить, что случайная величина η_n *асимптотически нормальна с параметрами* μ_n и σ_n^2 или, короче, *асимптотчески нормальна* $\mathcal{N}(\mu_n, \sigma_n^2)$, и записывать это следующим образом:

$$\mathcal{L}(\eta_n) \sim \mathcal{N}(\mu_n, \sigma_n^2), \quad \text{если} \quad \mathcal{L}\left(\frac{\eta_n - \mu_n}{\sigma_n}\right) \to \mathcal{N}(0, 1).$$

Найдем сначала асимптотические распределения выборочных моментов $\widehat{\alpha}_{nk}$ (см. (2)). Величина $n\widehat{\alpha}_{nk} = \sum_{i=1}^{n} X_i^k$ является суммой независимых одинаково распределенных случайных величин. Если конечен момент $\alpha_{2k} = \mathbf{E}\xi^{2k}$ то к этой сумме можно применить центральную предельную теорему теории вероятностей (см. соотношение (1) § 1.2). Так как $\mathbf{E}X_i^k = \alpha_k$, $\mathbf{D}X_i^k = \alpha_{2k} - \alpha_k^2$ (см. (14)), то величина

$$\frac{n\widehat{\alpha}_{nk} - n\alpha_k}{\sqrt{n(\alpha_{2k} - \alpha_k^2)}} = \frac{\widehat{\alpha}_{nk} - \alpha_k}{\sqrt{(\alpha_{2k} - \alpha_k^2)/n}}$$

асимптотически нормальна $\mathcal{N}(0, 1)$. Таким образом, справедлива следующая теорема.

Теорема 3. Если конечен теоретический момент α_{2k}, то при $n \to \infty$ выборочный момент $\widehat{\alpha}_{nk}$ асимптотически нормален $\mathcal{N}\left(\alpha_k, \dfrac{\alpha_{2k} - \alpha_k^2}{n}\right)$

Полезность этой теоремы состоит в том, что она позволяет оценивать для больших выборок вероятности заданных отклонений значений выборочных моментов от теоретических. Действительно, из этой теоремы имеем, что при любом фиксированном $t > 0$ и $n \to \infty$

$$\mathbf{P}\left\{\sqrt{\frac{n}{\alpha_{2k}-\alpha_k^2}}\,|\widehat{\alpha}_{nk}-\alpha_k|<t\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-t}^{t}e^{-x^2/2}dx=\Phi(t)-\Phi(-t)=2\Phi(t)-1$$

(обозначения см. в (7) § 1.2).

В частности, из теоремы 3 следует, что если теоретическая дисперсия $\mu_2 = \mathbf{D}\xi$ существует, то при $n \to \infty$ выборочное среднее $\overline{X} = \widehat{\alpha}_{n1}$ асимптотически нормально $\mathcal{N}(\alpha_1, \mu_2/n)$. Отметим в этой связи, что если $\mathcal{L}(\xi) = \mathcal{N}(\alpha_1, \mu_2)$, то случайная величина $\overline{X}$, как сумма независимых нормальных случайных величин, также нормальна с параметрами α_1 и μ_2/n (см. соотношение (5) § 1.2), т. е. в данном случае $\mathcal{L}(\overline{X}) = \mathcal{N}(\alpha_1, \mu_2/n)$ при любом n.

Справедлив также более общий результат об асимптотической нормальности любого конечного числа выборочных моментов $\widehat{\alpha}_{nk}$. Но для доказательства этого надо знать многомерный вариант центральной предельной теоремы, который формулируется следующим образом.

Многомерная центральная предельная теорема. Пусть r-мерные случайные векторы $\underline{\eta}_k = (\eta_{k1}, \quad, \eta_{kr})$, $k = 1, 2, \quad$ независимы, одинаково распределены и имеют конечные моменты

$$a_i = \mathbf{E}\eta_{ki}, \quad b_{ij} = \operatorname{cov}(\eta_{ki}, \eta_{kj}), \quad i, j = 1, \quad, r.$$

Тогда при $n \to \infty$

$$\mathcal{L}(\zeta_{n1}, \quad, \zeta_{nr}) \longrightarrow \mathcal{N}\big(\underline{0}, B = \|b_{ij}\|_1^r\big),$$

где

$$\zeta_{ni} = \frac{(\eta_{1i} + \quad + \eta_{ni} - na_i)}{\sqrt{n}}, \quad i = 1, \quad, r$$

(определение многомерного нормального распределения дано в п. 2 § 1.2).

Рассмотрим теперь вектор $\underline{A}_n = (\widehat{\alpha}_{n1}, \quad, \widehat{\alpha}_{nr})$, составленный из произвольного числа r первых выборочных моментов. Обозначим

$$\underline{\alpha} = (\alpha_1, \quad, \alpha_r) = \mathbf{E}\underline{A}_n \quad \text{и} \quad \Sigma = \|\sigma_{ij}\|_1^r, \quad \text{где} \quad \sigma_{ij} = \alpha_{i+j} - \alpha_i\alpha_j.$$

Легко убедиться, что дисперсионная матрица $\mathbf{D}\underline{A}_n = (1/n)\Sigma$ (см. упр. 21). Введем, далее, векторы $\underline{\eta}_k = (X_k, X_k^2, \quad, X_k^r)$, $k = 1, \ldots, n$, — они являются независимыми и одинаково распределенными, так как $X_1, \ldots, X_n$ независимы и одинаково распределены, при этом $\mathbf{E}\underline{\eta}_k = \underline{\alpha}$, $\mathbf{D}\underline{\eta}_k = \Sigma$. Осталось заметить, что $n\underline{A}_n = \underline{\eta}_1 + \ldots + \underline{\eta}_n$, и применить к этой сумме многомерную

центральную предельную теорему:

$$\mathcal{L}\left(\frac{n\underline{A}_n - n\underline{\alpha}}{\sqrt{n}}\right) \longrightarrow \mathcal{N}(0, \Sigma)$$

или, что эквивалентно, $\mathcal{L}(\underline{A}_n) \sim \mathcal{N}\big(\underline{\alpha}, (1/n)\Sigma\big)$.

Таким образом, справедлива следующая теорема о совместном распределении выборочных моментов.

Теорема 4. Если конечен теоретический момент α_{2r}, то при $n \to \infty$ совместное распределение выборочных моментов $(\widehat{\alpha}_{n1}, \ldots, \widehat{\alpha}_{nr})$ асимптотически нормально $\mathcal{N}\big(\underline{\alpha} = (\alpha_1, \quad, \alpha_r), (1/n)\Sigma\big)$.

Эта теорема полезна больше не сама по себе, а своим важным следствием, которое мы приводим без доказательства. Именно, справедлива

Асимптотическая нормальность функций от выборочных моментов

Теорема 5. Для любой дифференцируемой функции φ от r переменных

$$\mathcal{L}\big(\sqrt{n}(\varphi(\underline{A}_n) - \varphi(\underline{\alpha}))\big) \longrightarrow \mathcal{N}(0, \sigma^2)$$

при условии, что $\sigma^2 = \underline{b}'\Sigma\underline{b} > 0$, где

$$\underline{b} = \operatorname{grad}\varphi(\underline{\alpha}) = \left(\frac{\partial\varphi(\underline{x})}{\partial x_1}, \qquad \frac{\partial\varphi(\underline{x})}{\partial x_r}\right)\Bigg|_{\underline{x}=\underline{\alpha}}$$

Более того, если $\sigma^2 = \sigma^2\big(\{\alpha.\}\big)$ — непрерывная функция от теоретических моментов $\{\alpha.\}$ и $\sigma_n^2 = \sigma^2\big(\{\widehat{\alpha}_{n\cdot}\}\big)$ (т. е. моменты $\alpha.$ заменены соответствующими выборочными моментами $\widehat{\alpha}_{n\cdot}$), то

$$\mathcal{L}\left(\sqrt{n}\,\frac{\varphi(\underline{A}_n) - \varphi(\underline{\alpha})}{\sigma_n}\right) \longrightarrow \mathcal{N}(0, 1). \tag{24}$$

Из теоремы 5, в частности, следует, что асимптотически нормальными являются и центральные выборочные моменты $\widehat{\mu}_{nk}$, поскольку они являются непрерывными функциями (многочленами) от обычных выборочных моментов. Например, соответствующее утверждение для выборочной дисперсии S^2 формулируется следующим образом (см. упр. 23).

Асимптотическая нормальность выборочной дисперсии

Теорема 6. Если конечен теоретический момент μ_4, то при $n \to \infty$ выборочная дисперсия S^2 асимптотически нормальна $\mathcal{N}\big(\mu_2, (\mu_4 - \mu_2^2)/n\big)$. Более того,

$$\mathcal{L}\left(\frac{\sqrt{n}(S^2 - \mu_2)}{\sqrt{\widehat{\mu}_{n4} - S^4}}\right) \longrightarrow \mathcal{N}(0, 1). \tag{25}$$

5. Асимптотические доверительные интервалы для теоретических моментов

Теоремы предыдущего пункта, в особенности теорема 5 в варианте (24), используются в статистике для построения доверительных интервалов для теоретических моментов. С примером доверительного интервала мы уже встречались в п. 2 § 2.1, где был построен асимптотический γ-доверительный интервал для теоретической функции распределения $F(x)$ в произвольной точке x, где $F(x) \neq 0$. Дадим теперь общее определение доверительного интервала. Пусть нас интересует некоторая теоретическая характеристика g наблюдаемой случайной величины ξ и имеется выборка $\boldsymbol{X} = (X_1, \quad , X_n)$ из $\mathcal{L}(\xi)$.

Определение 4 *γ-доверительным интервалом* для g называется такой случайный интервал $(T_1(\boldsymbol{X}), T_2(\boldsymbol{X}))$, $T_1(\boldsymbol{X}) < T_2(\boldsymbol{X})$, который содержит внутри себя (накрывает) неизвестное значение g с вероятностью, не меньшей γ:

$$\mathbf{P}\{T_1(\boldsymbol{X}) < g < T_2(\boldsymbol{X})\} \geqslant \gamma. \tag{26}$$

Здесь $T_1(\boldsymbol{X})$ и $T_2(\boldsymbol{X})$ — некоторые статистики (функции от выборки), называемые соответственно *нижней и верхней доверительными границами*, а γ — задаваемый заранее *доверительный уровень*, который обычно выбирается достаточно близким к 1 (например, $\gamma = 0{,}9$; $0{,}95$; $0{,}99$ и т. д.). Длина $T_2(\boldsymbol{X}) - T_1(\boldsymbol{X})$ такого интервала характеризует точность локализации оцениваемой характеристики g, а величина γ отражает «степень готовности мириться с возможностью ошибки», т. е. является показателем надежности доверительного интервала.

Если выполнено условие (26), то практически это означает, что в большом числе независимых применений такого правила оценивания мы можем ошибиться, утверждая, что $g \in (T_1(\boldsymbol{X}), T_2(\boldsymbol{X}))$, примерно в $(1-\gamma)100\,\%$ случаях.

Общая теория доверительного оценивания будет обсуждаться позже (см. гл. 3 § 3.8), а сейчас мы продемонстрируем такой подход в оценивании теоретических моментов. При этом, поскольку мы будем основываться на предельной теореме 5, то речь будет идти о построении *асимптотических доверительных интервалов*.

Итак, пусть нас интересует характеристика $g = \varphi(\underline{\alpha})$, $\underline{\alpha} = (\alpha, \quad , \alpha_r)$, где φ — заданная непрерывно дифференцируемая функция. Тогда на основании (24) мы можем записать следующую цепочку соотношений:

$$\mathbf{P}\left\{\frac{\sqrt{n}}{\sigma_n}\left|\varphi(\underline{A}_n) - g\right| < c_\gamma\right\} = \mathbf{P}\left\{g \in \left(\varphi(\underline{A}_n) \mp \frac{c_\gamma \sigma_n}{\sqrt{n}}\right)\right\} \longrightarrow$$
$$\longrightarrow \Phi(c_\gamma) - \Phi(-c_\gamma) = 2\,\Phi(c_\gamma) - 1 = \gamma, \tag{27}$$

если $c_\gamma = \Phi^{-1}\big((1+\gamma)/2\big)$

Но это и означает, что $(\varphi(\underline{A}_n) \mp c_\gamma \sigma_n/\sqrt{n})$ есть асимптотический γ-доверительный интервал для $g = \varphi(\underline{\alpha})$. Центр этого интервала находится в случайной точке $\varphi(\underline{A}_n)$, а его длина равна $2c_\gamma\sigma_n/\sqrt{n}$, т. е. при больших объемах

выборки это достаточно узкий интервал, обеспечивающий высокую точность локализации оцениваемой характеристики $g \quad \varphi(\underline{\alpha})$. Таким образом, мы имеем общее решение доверительного оценивания произвольных (непрерывно дифференцируемых) функций от любого конечного числа теоретических моментов. На практике нам обычно бывают нужны более частные случаи оценивания отдельных моментов — прежде всего среднего и дисперсии наблюдаемой случайной величины ξ. Поэтому выпишем еще в явном виде несколько конкретизаций этого общего решения.

Пример 4. Рассмотрим задачу построения доверительного интервала для произвольного теоретического момента $\alpha_k = \mathbf{E}\xi^k \quad k \geqslant 1$. В этом случае, вместо использования общего алгоритма (27), проще непосредственно использовать теорему 3, заменив в ее формулировке асимптотическую дисперсию $(\alpha_{2k} - \alpha_k^2)/n$ ее оценкой $(\widehat{\alpha}_{n,2k} - \widehat{\alpha}_{nk}^2)/n$. В итоге это даст искомый асимптотический γ-доверительный интервал для α_k вида

$$\left(\widehat{\alpha}_{nk} \mp c_\gamma \sqrt{\frac{\widehat{\alpha}_{n,2k} - \widehat{\alpha}_{nk}^2}{n}}\right). \tag{28}$$

Доверительный интервал для среднего

Полагая здесь $k = 1$, получим соответствующий интервал для теоретического среднего $\alpha_1 = \mathbf{E}\xi$:

$$\left(\overline{X} \mp \frac{c_\gamma S}{\sqrt{n}}\right). \tag{29}$$

•

Пример 5. Чтобы построить асимптотический γ-доверительный интервал для теоретической дисперсии $\mu_2 = \mathbf{D}\xi$, надо просто воспользоваться результатом (25): искомый интервал есть

Доверительный интервал для дисперсии

$$\left(S^2 \mp c_\gamma \sqrt{\frac{\widehat{\mu}_{n4} - S^4}{n}}\right) \tag{30}$$

•

§ 2.3. Многомерные данные

Высшая математика проще элементарной в том смысле, что исследовать непроходимую чащу легче с самолета, нежели пешком (но для этого надо иметь самолет и уметь им управлять!).

Иногда наблюдаемая в эксперименте случайная величина представляет собой вектор некоторой размерности $k \geqslant 2$: $\underline{\xi} = (\xi_1, \quad , \xi_k)$. В этом случае выборка из $\mathcal{L}(\underline{\xi})$ есть совокупность k-мерных векторов $\{X_l = (X_{l1}, \quad ., X_{lk}),$ $l = 1, \ldots, n\}$ и обработка таких многомерных данных уже более трудоемка,

так как помимо обычных выборочных моментов здесь надо вычислять эмпирические (выборочные) аналоги для смешанных теоретических моментов, среди которых важнейшим для приложений является второй смешанный момент или ковариация.

Чтобы не загромождать изложение техническими деталями, рассмотрим подробно лишь простейший случай $k = 2$. Итак, пусть наблюдается двумерная случайная величина $\underline{\xi} = (\xi_1, \xi_2)$ и $\{\boldsymbol{X}_l = (X_{l1}, X_{l2}),\ l = 1, \ldots, n\}$ — повторные независимые наблюдения над ней. Обозначим $F(x_1, x_2) = \mathbf{P}\{\xi_1 \leqslant x_1, \xi_2 \leqslant x_2)\}$ (теоретическую) функцию распределения, $\mathbf{E}\underline{\xi} = (\mathbf{E}\xi_1, \mathbf{E}\xi_2) = \underline{\mu} = (\mu_1, \mu_2)$ — вектор средних и $\mathbf{D}\xi_j = \sigma_j^2$, $j = 1, 2$, $\operatorname{cov}(\xi_1, \xi_2) = \sigma_{12}$ — вторые центральные моменты вектора $\underline{\xi}$. Как в данном случае по выборочным данным построить эмпирические (выборочные) аналоги этих характеристик?

1. Эмпирическая функция распределения

Эмпирическая функция распределения определяется по аналогии с одномерным случаем формулой

$$\widehat{F}_n(x_1, x_2) = \frac{1}{n} \sum_{l=1}^{n} I(X_{l1} \leqslant x_1, X_{l2} \leqslant x_2) \equiv \frac{1}{n} \sum_{l=1}^{n} \eta_l. \tag{1}$$

В (1) $\{\eta_l = I(X_{l1} \leqslant x_1, X_{l2} \leqslant x_2)\}$ — независимые одинаково распределенные случайные величины, принимающие значения 0 и 1, т. е. бернуллиевские величины, причем

$$\mathbf{E}\eta_l = \mathbf{P}\{X_{l1} \leqslant x_1, X_{l2} \leqslant x_2\} = F(x_1, x_2),$$
$$\mathbf{D}\eta_l = F(x_1, x_2)\big(1 - F(x_1, x_2)\big).$$

Отсюда и из (1) следует, что

$$\mathbf{E}\widehat{F}_n(x_1, x_2) = \frac{1}{n} \sum_{l=1}^{n} \mathbf{E}\eta_l = F(x_1, x_2), \tag{2}$$

т. е. э. ф. р. $\widehat{F}_n(x_1, x_2)$ является несмещенной оценкой для $F(x_1, x_2)$ при любых x_1, x_2.

Далее

$$\mathbf{D}\widehat{F}_n(x_1, x_2) = \frac{1}{n^2} \sum_{l=1}^{n} \mathbf{D}\eta_l = \frac{1}{n} F(x_1, x_2)\big(1 - F(x_1, x_2)\big), \tag{3}$$

что стремится к нулю при $n \to \infty$, т. е. $\widehat{F}_n(x_1, x_2)$ является состоятельной оценкой для $F(x_1, x_2)$.

Таким образом, здесь имеется полная аналогия с одномерным случаем (см. (5) § 2.1).

2. Моменты

Обратимся теперь к моментам. Обозначим $\boldsymbol{X}_j = (X_{1j}, \ldots, X_{nj})$, $j = 1, 2$, выборочные данные для j-й координаты вектора $\underline{\xi} = (\xi_1, \xi_2)$ и пусть $\overline{X}_j$, S_j^2 — соответствующие выборочные среднее и дисперсия, вычисляемые по формулам (4)–(5) § 2.2.

Тогда $\overline{X}_j$ является эмпирическим аналогом для теоретического среднего $\mu_j = \mathrm{E}\xi_j$, а S_j^2 — для дисперсии $\sigma_j^2 = \mathrm{D}\xi_j$

Выборочная ковариация

Выборочная же (*эмпирическая*) *ковариация* — аналог теоретической ковариации

$$\sigma_{12} = \operatorname{cov}(\xi_1, \xi_2) = \mathrm{E}(\xi_1 - \mu_1)(\xi_2 - \mu_2) = \mathrm{E}\xi_1\xi_2 - \mu_1\mu_2$$

— определяется формулами

$$S_{12} = \frac{1}{n}\sum_{l=1}^{n}(X_{l1} - \overline{X}_1)(X_{l2} - \overline{X}_2) = \frac{1}{n}\sum_{l=1}^{n} X_{l1}X_{l2} - \overline{X}_1\overline{X}_2. \qquad (4)$$

Выборочный коэффициент корреляции

Соответственно, *выборочный* (*эмпирический*) *коэффициент корреляции* — аналог теоретического коэффициента корреляции

$$\rho = \rho(\xi_1, \xi_2) = \operatorname{corr}(\xi_1, \xi_2) = \frac{\operatorname{cov}(\xi_1, \xi_2)}{\sqrt{\mathrm{D}\xi_1\mathrm{D}\xi_2}}$$

— определяется формулой

$$\widehat{\rho}_n = \frac{S_{12}}{S_1 S_2}. \qquad (5)$$

Исследуем некоторые свойства этих новых выборочных характеристик. В § 2.2 мы установили, что выборочная дисперсия не является несмещенной оценкой теоретической дисперсии, но легко построить статистику уже являющуюся несмещенной оценкой дисперсии (см. формулы (17)–(18) § 2.2). Аналогичный факт имеет место и для выборочной ковариации. Именно, справедлива

Лемма. Статистика

$$\frac{n}{n-1}S_{12} = \frac{1}{n-1}\sum_{l=1}^{n}(X_{l1} - \overline{X}_1)(X_{l2} - \overline{X}_2) \qquad (6)$$

является несмещенной оценкой теоретической ковариации $\sigma_{12} = \operatorname{cov}(\xi_1, \xi_2)$.

Доказательство. Введем случайную величину $\eta = \xi_1 + \xi_2$. Выборочные данные для нее есть $X_l = X_{l1} + X_{l2}$, $l = 1, \ldots, n$, а ее дисперсия $\mathrm{D}\eta = \sigma_1^2 + \sigma^2 + 2\sigma_{12}$. Построим несмещенную оценку для $\mathrm{D}\eta$: это есть статистика

$$\frac{1}{n-1}\sum_{l=1}^{n}(X_l - \overline{X})^2 = \frac{1}{n-1}\sum_{l=1}^{n}\big((X_{l1} - \overline{X}_1) + (X_{l2} - \overline{X}_2)\big)^2 =$$

$$= \frac{1}{n-1}\sum_{l=1}^{n}(X_{l1}-\overline{X}_1)^2 + \frac{1}{n-1}\sum_{l=1}^{n}(X_{l2}-\overline{X}_2)^2 + \frac{2}{n-1}\sum_{l=1}^{n}(X_{l1}-\overline{X}_1)(X_{l2}-\overline{X}).$$

В этом разложении

$$\frac{1}{(n-1)}\sum_{l=1}^{n}(X_{lj}-\overline{X}_j)^2$$

есть несмещенная оценка для σ_j^2, $j = 1, 2$, следовательно, последнее слагаемое — несмещенная оценка для $2\sigma_{12}$. ■

В общем случае для k-мерной наблюдаемой величины $\underline{\xi} = (\xi_1, \ldots, \xi_k)$ с $\mathbf{E}\underline{\xi} = \underline{\mu} = (\mu_1, \ldots, \mu_k)$ и дисперсионной матрицей $\mathbf{D}\underline{\xi} = \Sigma = \|\sigma_{ij} = \operatorname{cov}(\xi_i, \xi_j)\|_1^k$ выборочные моменты первого и второго порядков вычисляются по формулам

$$\overline{X}_j = \frac{1}{n}\sum_{l=1}^{n} X_{lj}, \quad S_j^2 = \frac{1}{n}\sum_{l=1}^{n}(X_{lj}-\overline{X}_j)^2$$

$$S_{ij} = \frac{1}{n}\sum_{l=1}^{n}(X_{li}-\overline{X}_i)(X_{lj}-\overline{X}_j), \quad i \neq j.$$

Таким образом, здесь мы имеем вектор выборочных средних $(\overline{X}_1, \quad, \overline{X}_k)$ и выборочную дисперсионную матрицу $\widehat{\Sigma} = \|S_{ij}\|_1^k$, где $S_{ii} = S_i^2$, $i = 1, \quad, k$.

Вектор выборочных средних будет несмещенной оценкой для вектора теоретических средних $\underline{\mu} = (\mu_1, \quad, \mu_k)$, т. е. $\mathbf{E}\overline{X}_i = \mu_i$, $i = 1, \ldots, k$, а подправленная выборочная дисперсионная матрица $n/(n-1)\widehat{\Sigma}$ — несмещенной оценкой для Σ: $\mathbf{E}(n/(n-1)\widehat{\Sigma}) = \Sigma$ (это равенство следует понимать поэлементно).

3. Большие выборки

Обсудим теперь асимптотические свойства введенных новых эмпирических характеристик, предполагая, что теоретические ковариации существуют. Тогда по закону больших чисел при $n \to \infty$

$$\frac{1}{n}\sum_{l=1}^{n} X_{li}X_{lj} \xrightarrow{\mathbf{P}} \mathbf{E}(\xi_i\xi_j),$$

но также и $\overline{X}_j \xrightarrow{\mathbf{P}} \mathbf{E}\xi_j$, $S_j^2 \xrightarrow{\mathbf{P}} \mathbf{D}\xi_j$. Следовательно, по теореме 1 § 2.2

Состоятельность выборочных ковариаций и коэффициентов корреляций

$$S_{ij} \xrightarrow{\mathbf{P}} \sigma_{ij} \quad \text{и} \quad \widehat{\rho}_{ij} = \frac{S_{ij}}{S_iS_j} \xrightarrow{\mathbf{P}} \rho_{ij} = \rho(\xi_i, \xi_j), \quad i \neq j. \tag{7}$$

Таким образом, выборочные ковариации и коэффициенты корреляций являются состоятельными оценками соответствующих теоретических ковариаций и коэффициентов корреляций (конечно, если все теоретические дисперсии $\sigma_i^2 > 0$).

Можно также показать, что $\mathbf{E}\widehat{\rho}_{ij} - \rho_{ij} \longrightarrow 0$ при $n \to \infty$, т. е. $\widehat{\rho}_{ij}$ являются лишь асимптотически несмещенными оценками для ρ_{ij}.

4. Добавление: нормальная модель

Предположим, что наблюдается двумерная нормальная случайная величина:

$$\mathcal{L}(\xi_1, \xi_2) = \mathcal{N}\left((\mu_1, \mu_2), \left\| \begin{matrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{matrix} \right\|\right), \quad |\rho| < 1$$

(см. п. 2 § 1.2), и требуется по соответствующей выборке рассчитать доверительный интервал для теоретического коэффициента корреляции ρ.

Это можно сделать на основании результата Р. Фишера (1915 г.), который показал, что если от выборочного коэффициента корреляции $\widehat{\rho}_n$, определенного в (5), перейти к новой статистике

$$Z_n = \frac{1}{2} \ln \frac{1 + \widehat{\rho}_n}{1 - \widehat{\rho}_n},$$

то уже для небольших объемов выборки n статистика Z_n будет распределена приблизительно нормально, со средним значением и дисперсией, заданными приближенными выражениями

$$\begin{gathered} \mathbf{E}Z_n \approx \zeta + \frac{\rho}{2(n-1)}, \quad \text{где} \\ \zeta = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}, \quad \text{и} \quad \mathbf{D}Z_n \approx \frac{1}{n-3}. \end{gathered} \tag{8}$$

Другими словами, $\mathcal{L}\left(\sqrt{n-3}(Z_n - \zeta)\right) \longrightarrow \mathcal{N}(0, 1)$ при $n \to \infty$.

Отсюда, так же, как и в п. 5 § 2.2, мы можем построить асимптотический γ-доверительный интервал для ζ: $\left(Z_n \mp c_\gamma/\sqrt{n-3}\right)$, $c_\gamma = \Phi^{-1}\left((1+\gamma)/2\right)$. Поскольку функция $\zeta = \zeta(\rho)$ в (8) монотонно возрастает, то разрешая уравнения

$$\zeta(\rho) = Z_n \mp \frac{c_\gamma}{\sqrt{n-3}}$$

Доверительный интервал для ρ

относительно ρ, мы получим асимптотический γ-доверительный интервал для самого коэффициента корреляции ρ:

$$\rho \in \Delta_\gamma = \left(\zeta^{-1}\left(Z_n - \frac{c_\gamma}{\sqrt{n-3}}\right), \zeta^{-1}\left(Z_n + \frac{c_\gamma}{\sqrt{n-3}}\right)\right) \tag{9}$$

(здесь ζ^{-1} — обратная функция к ζ). Поскольку в этом алгоритме число n предполагается большим, то в (9) можно ограничиться лишь первыми двумя членами разложения в ряд Тейлора[3] функции $\zeta^{-1}(Z_n + \varepsilon_n)$, $\varepsilon_n \longrightarrow 0$, и ис-

[3] Тейлор Брук (1685–1731) — английский математик и философ, в 1712 г. нашел общую формулу разложения функций в степенной ряд, носящую его имя.

пользовать приближенный интервал

$$\Delta_\gamma \approx \left(\zeta^{-1}(Z_n) \mp \frac{c_\gamma}{\sqrt{n-3}} \frac{d}{dt}\zeta^{-1}(t)\Big|_{t=Z_n}\right) = \\ = \frac{e^{2Z_n}-1}{e^{2Z_n}+1}\left(1 \mp \frac{4c_\gamma e^{2Z_n}}{(e^{4Z_n}-1)\sqrt{n-3}}\right). \tag{10}$$

При больших n это достаточно узкий интервал, окружающий точку $(e^{2Z_n}-1)/(e^{2Z_n}+1)$. В свою очередь, если значение статитстики $\widehat{\rho}_n$ мало, тогда $Z_n \approx \widehat{\rho}_n$, и интервал (10) может быть заменен приближенным интервалом $\Delta'_\gamma = \left(\widehat{\rho}_n \mp c_\gamma/\sqrt{n-3}\right)$, который уже является симметричным интервалом с центром в точке $\widehat{\rho}_n$. Построенный интервал Δ'_γ может быть использован также для проверки гипотезы H_0, состоящей в том, что $\rho = \rho_0$, где ρ_0 — некоторое малое заданное число. В частности, при $\rho_0 = 0$ гипотеза H_0 будет означать независимость компонент ξ_1 и ξ_2 наблюдаемой двумерной нормальной случайной величины $\underline{\xi} = (\xi_1, \xi_2)$, поскольку в данном случае некоррелированность ($\rho_0 = 0$) эквивалентна независимости (см. п. 2 § 1.2). Рассуждаем при этом мы так. Если гипотеза H_0 $\rho = \rho_0$ верна, то

$$\mathbf{P}\left\{\widehat{\rho}_n - \frac{c_\gamma}{\sqrt{n-3}} < \rho_0 < \widehat{\rho}_n + \frac{c_\gamma}{\sqrt{n-3}}\right\} \approx \gamma$$

или, что то же,

$$\mathbf{P}\left\{|\widehat{\rho}_n - \rho_0| \geqslant \frac{c_\gamma}{\sqrt{n-3}}\right\} \approx 1-\gamma \equiv \alpha.$$

При достаточно малом значении α (при γ близком к 1) событие $|\widehat{\rho}_n - \rho_0| \geqslant c_\gamma/\sqrt{n-3}$ практически невозможно, и если для данной выборки обнаруживается, что $|\widehat{\rho}_n - \rho_0| \geqslant c_\gamma/\sqrt{n-3}$, то мы должны отклонить гипотезу о равенстве $\rho = \rho_0$, как противоречащую данным. В противном случае гипотеза считается согласующейся с данными, и она принимается. При этом мы можем ошибиться, отклонив гипотезу H_0, если она верна, но вероятность этого приблизительно равна α, т. е. мала.

Гипотеза независимости

Обычно такую методику применяют для проверки *гипотезы независимости* $H_0: \rho = 0$ (понятие общей гипотезы независимости см. в примере 6 Введения), и в этом случае правило принятия решения формулируется следующим образом: задаемся малой вероятностью α ошибочно отклонить истинную гипотезу H_0 (α называется *уровнем значимости*) и определяем число $c_\gamma = \Phi^{-1}\left((1+\gamma)/2\right)$, $\gamma = 1-\alpha$. Далее по имеющейся выборке вычисляем значение статистики $\widehat{\rho}_n$ по формуле (5), и если окажется, что $|\widehat{\rho}_n| \geqslant c_\gamma/\sqrt{n-3}$, то гипотеза H_0 отвергается, в противном случае, т. е. при $|\widehat{\rho}_n| < c_\gamma/\sqrt{n-3}$, она принимается.

Таким образом, здесь мы встретились с конкретным примером проверки статистической гипотезы, общая теория которых будет детально обсуждаться в гл. 4.

Замечание. Плотность $f_{n,\rho}(r)$ статистики $\widehat{\rho}_n$ имеет вид [15, с. 436]: при $n > 2$

$$f_{n,\rho}(r) = \frac{n-2}{\pi}(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2}\int\limits_0^1 \frac{x^{n-2}}{(1-\rho r x)^{n-1}}\frac{dx}{\sqrt{1-x^2}}, \quad |r|<1,$$

т. е. зависит лишь от n и ρ, и при $\rho = 0$ (т. е. при независимости ξ_1 и ξ_2)

$$f_{n,0}(r) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}(1-r^2)^{(n-4)/2}, \quad |r|<1.$$

Если (при $\rho = 0$) рассмотреть преобразование

$$T_n = \sqrt{n-2}\,\frac{\widehat{\rho}_n}{\sqrt{1-\widehat{\rho}_n^2}},$$

то, используя формулу (45) § 1.2, нетрудно убедиться в том, что статитстика T_n имеет распределение Стьюдента $S(n-2)$ (это мы оставляем читателю в качестве простого упражнения на применение формулы преобразования (45) § 1.2). Поскольку функция $y = \sqrt{n-2}\,x/\sqrt{1-x^2}$ при $|x| < 1$ монотонно возрастает от $-\infty$ до ∞, то отсюда следует, что события $\{|T_n| > t_\alpha\}$ и $\{|\widehat{\rho}_n| > z_\alpha\}$ при $z_\alpha = t_\alpha/\sqrt{n-2+t_\alpha^2}$ эквивалентны. Поэтому, если при заданном α, $0 < \alpha < 1$, число t_α выбрано так, чтобы выполнялось условие $\mathbf{P}\{|T_n| > t_\alpha\} = \alpha$, то будет выполняться также условие $\mathbf{P}\{|\widehat{\rho}_n| > z_\alpha\} = \alpha$. На этом факте, аналогично предыдущему, можно предложить следующий алгоритм для проверки гипотезы независимости $H_0 \quad \rho \quad 0$, который уже «работает» при любом числе наблюдений $n > 2$. Пусть задан малый уровень значимости α. Тогда, поскольку плотность распределения Стьюдента (см. (28) § 1.2) симметрична относительно нуля, то, обозначая функцию распределения закона $S(n)$ через $S_n(x)$, из цепочки соотношений

$$\alpha = \mathbf{P}\{|T_n| > t_\alpha\} = 2\int\limits_{t_\alpha}^{\infty} f(x|n-2)\,dx = 2\big(1 - S_{n-2}(t_\alpha)\big)$$

получаем, что t_α определяется уравнением

$$S_{n-2}(t_\alpha) = 1 - \frac{\alpha}{2}, \quad \text{т. е.} \quad t_\alpha = S_{n-2}^{-1}\left(1-\frac{\alpha}{2}\right),$$

где $S_{n-2}^{-1}(\cdot)$ — обратная функция к $S_{n-2}(x)$.

Далее вычисляем по имеющимся данным значение $\widehat{\rho}_n$, и если окажется, что $|\widehat{\rho}_n| > z_\alpha = t_\alpha/\sqrt{n-2+t_\alpha^2}$, то мы отвергаем гипотезу H_0 как противоречащую данным, поскольку вероятность такого события при гипотезе H_0 (равная, по предыдущему, α) мала при малом α. В противном случае, т. е. при $|\widehat{\rho}_n| \leqslant z_\alpha$, гипотеза H_0 принимается. Вероятность ошибочно отклонить при этом истинную гипотезу H_0 в точности равна α.

5. Другие корреляционные характеристики

В теории статистической регрессии (о которой подробно будет идти речь в гл. 6 § 6.6) важную роль играют, помимо обычных корреляций (коэффициентов корреляций $\operatorname{corr}(\xi, \eta) \equiv \rho(\xi, \eta)$), другие, более сложные корреляционные характеристики совокупности изучаемых случайных величин. Здесь мы дадим их определения и построим их эмпирические (выборочные) аналоги.

Итак, пусть имеется совокупность случайных величин Y и $\underline{X} = (X_1, \ldots, X_p)$, связанных некоторой стохастической зависимостью, причем $\underline{X}$ доступно наблюдению (измерению), а значение Y непосредственно измерить невозможно (например, Y связано будущим). Тогда возникает задача предсказания (прогноза, оценки) величины Y по информации, доставляемой измерением величин $X_1, \ldots, X_p$, которые в этом контексте называются *предсказывающими переменными*. Функция от предсказывающих переменных $\varphi(\underline{X})$, которую используют в качестве оценки для $Y \cdot \varphi(\underline{X}) \approx Y$, называют *предиктором* величины Y по $\underline{X}$. Задачей разработки методов построения наилучших (в каком-то смысле) предикторов занимается *теория статистической регрессии*.

Для измерения точности предиктора φ обычно используется *среднеквадратичная ошибка* (с. к. о.)

$$\Delta(\varphi) = \mathbf{E}(\varphi(\underline{X}) - Y)^2$$

Предиктор, минимизирующий с. к. о. в заданном классе предикторов L называется *оптимальным предиктором*, или *прогнозом* (в классе L), и обозначается символом φ^*

Прогноз

$$\varphi^* = \arg\min_{\varphi \in L} \Delta(\varphi), \tag{11}$$

его с. к. о. есть

$$\Delta^* = \Delta(\varphi^*) = \mathbf{E}(\varphi^*(\underline{X}) - Y)^2 = \inf_{\varphi \in L} \mathbf{E}(\varphi(\underline{X}) - Y)^2 \tag{12}$$

Таким образом, в общем случае задача построения наилучшего (в среднеквадратичном смысле и в заданном классе предикторов L) прогноза для величины Y по $\underline{X}$ сводится к решению экстремальной проблемы (11)–(12).

Часто ограничиваются классом *линейных предикторов*, т. е. когда L есть класс линейных функций: $L = \{\varphi \quad \varphi(\underline{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p\}$ В этом случае оптимальный предиктор всегда существует (при условии существования вторых моментов и невырожденности дисперсионной матрицы $\Sigma = \mathbf{D}\underline{X}$) и имеет вид:

Линейный прогноз

$$\begin{aligned} Y_p^* = \varphi^*(\underline{X}) &= \beta_0^* + \beta_1^* X_1 + \ldots + \beta_p^* X_p = \\ &= \beta_0^* + \underline{\beta}^{*\prime} \underline{X} = \mathbf{E}Y + \underline{\beta}^{*\prime}(\underline{X} - \mathbf{E}\underline{X}), \end{aligned} \tag{13}$$

где

$$\underline{\beta}^* = \Sigma^{-1} \underline{a}, \quad \underline{a} = \operatorname{cov}(Y, \underline{X}) = \left(\operatorname{cov}(Y_1, X_1), \ldots, \operatorname{cov}(Y_1, X_p)\right)$$

и

$$\beta_0^* = \mathbf{E}Y - \underline{\beta}^{*\prime} \mathbf{E}\underline{X};$$

при этом его с. к. о.

$$\Delta^* = \sigma_Y^2 - \underline{a}'\Sigma^{-1}\underline{a}, \quad \sigma_Y^2 = \mathbf{D}Y. \tag{14}$$

Таким образом, как сам оптимальный предиктор, так и его с. к. о. выражаются в данном случае лишь через первые и вторые моменты системы случайных величин $(Y, \underline{X})$.

Множественный коэффициент корреляции

В изложенной теории линейного прогноза особую роль играет величина

$$\rho^2_{Y\underline{X}} = \frac{\underline{a}'\Sigma^{-1}\underline{a}}{\sigma_Y^2} = \rho^2\big(Y, \varphi^*(\underline{X})\big), \tag{15}$$

называемая *множественным коэффициентом корреляции*; она является естественной *мерой зависимости между* Y и $\underline{X}$ (*мерой точности прогноза*) и с ее помощью можно сравнивать различные совокупности предсказывающих переменных в конкретных задачах. Согласно (15) множественный коэффициент корреляции между Y и совокупностью $\underline{X} = (X_1, \quad , X_p)$ равен квадрату обычного коэффициента корреляции между Y и оптимальным линейным предиктором для Y по $\underline{X}$. Чем ближе $\rho^2_{Y\underline{X}}$ к 1, тем меньше с. к. о. Δ^* (см. (14)), а при $\rho^2_{Y\underline{X}} = 1$ с. к. о. $\Delta^* = 0$, т. е. прогноз абсолютно точен: в этом случае $Y = \varphi^*(\underline{X})$, т. е. Y функционально связана с $\underline{X}$.

В практических приложениях теоретические моменты системы случайных величин $(Y, \underline{X})$ обычно неизвестны, поэтому все сведения, необходимые для построения оптимального предиктора, получают в результате обработки статистических данных, которые представляют собой выборку из распределения $\mathcal{L}(Y, \underline{X})$. Оценив по такой вспомогательной выборке (по результатам прошлых измерений $\underline{X}$ и Y) первые и вторые моменты, как это описано выше в п. 2,

Эмпирический предиктор

и заменив этими оценками теоретические характеристики $\mu_0 = \mathbf{E}Y$, $\underline{\mu} = \mathbf{E}\underline{X}$, $\underline{a}$ и Σ, строят *эмпирический предиктор*

$$\widehat{\varphi}^*(\underline{X}) = \widehat{\beta}_0^* + \underline{\widehat{\beta}}^{*\prime}\underline{X}, \quad \underline{\widehat{\beta}}^* = \widehat{\Sigma}^{-1}\underline{\widehat{a}}, \quad \widehat{\beta}_0^* = \widehat{\mu}_0 - \underline{\widehat{\beta}}^{*\prime}\underline{\widehat{\mu}}, \tag{16}$$

который и используют для предсказания в других случаях (в будущих измерениях). Точность эмпирического предиктора (16) оценивается *эмпирическим множественным коэффициентом корреляции*

$$\widehat{\rho}^2_{Y\underline{X}} = \frac{\underline{\widehat{a}}'\widehat{\Sigma}^{-1}\underline{\widehat{a}}}{S^2(Y)} \tag{17}$$

(в формулах (16)–(17) $\underline{\widehat{a}} = (S_{01}, \quad , S_{0p})$, $\widehat{\Sigma} = \|S_{ij}\|_1^p$, индекс «0» соответствует Y). Множественный коэффициент корреляции $\rho^2_{Y\underline{X}}$ принято записывать в виде $\rho^2_{0(1\ldots p)}$.

Еще одна полезная корреляционная характеристика возникает при рассмотрении следующего вопроса. Корреляция между двумя величинами, скажем, в нашем случае, между Y и X_p, характеризуется коэффициентом кор-

реляции ρ_{0p} $\operatorname{corr}(Y, X_p)$, который иногда называется также *полным коэффициентом корреляции* Y и X_p. Если рассматривать Y и X_p совместно с остальными $p-1$ величинами $X_1, \ldots, X_{p-1}$, то можно считать, что изменение величин Y и X_p вызывается в какой-то мере изменением остальных величин. Спрашивается, как в такой ситуации измерить непосредственную связь Y и X_p, исключая, так сказать, влияние промежуточных переменных X_1, , X_{p-1}? Для этого служит *частный коэффициент корреляции* величин Y и X_p относительно X_1, . X_{p-1}, обозначаемый символом $\rho_{0p(1\ldots p-1)}$. Для его вычисления поступают следующим образом. Строят оптимальные линейные предикторы для Y и X_p, основанные на промежуточных переменных $X_1, \ldots, X_{p-1}$, по алгоритму (13), т. е.

Частный коэффициент корреляции

$$Y^* = \beta_0^* + \beta_1^* X_1 + \ldots + \beta_{p-1}^* X_{p-1}$$

и

$$X_p^* = \alpha_0^* + \alpha_1^* X_1 + \ldots + \alpha_{p-1}^* X_{p-1},$$

и рассматривают «остатки» $e_1 = Y - Y^*$ и $e_2 = X_p - X_p^*$ тем самым исключается влияние «общих факторов» $X_1, \ldots, X_{p-1}$ на рассматриваемые переменные Y и X_p). После этого вычисляется обычный коэффициент корреляции $\rho(e_1, e_2)$ между этими остатками, который и есть частный коэффициент корреляции $\rho_{0p(1\ldots p-1)}$ величин Y и X_p относительно X_1, , X_{p-1}. В теории линейного прогнозирования доказывается, что

$$\rho_{0p(1\ldots p-1)} = \frac{\rho^{0p}}{(\rho^{00}\rho^{pp})^{1/2}}, \tag{18}$$

где $\|\rho^{ij}\| = \|\rho_{ij}\|^{-1}$ и

$$\rho_{ij} = \rho(X_i, X_j), \quad i, j = 0, 1, \quad , p \quad (X_0 = Y).$$

Из (18) следует, что частный коэффициент корреляции определяется через обычные коэффициенты корреляции системы величин Y, X_1, , X_p, следовательно, его эмпирический аналог строится заменой в (18) теоретических корреляций ρ_{ij} на выборочные корреляции $\widehat{\rho}_{ij}$, определенные в (7):

$$\widehat{\rho}_{0p(1\ldots p-1)} = \frac{\widehat{\rho}^{0p}}{(\widehat{\rho}^{00}\widehat{\rho}^{pp})^{1/2}}. \tag{19}$$

Из формул (17) и (19) видно, что эмпирические множественный и частный коэффициенты корреляций представляют собой непрерывные функции от выборочных вторых моментов, поэтому по теореме 1 § 2.2 на основании соотношений (7) можно заключить, что они являются состоятельными оценками соответствующих теоретических характеристик.

Пример 1. Рассмотрим случай $p = 2$. Здесь предиктор $Y^* = \beta_0^* + \beta_1^* X_1$, где согласно (13) $\beta_1^* = \operatorname{cov}(Y, X_1)/\mathbf{D}X_1 = \rho_{01}\dfrac{\sigma_0}{\sigma_1}$ и $\beta_0^* = \mathbf{E}Y - \beta_1^* \mathbf{E}X_1$.

Аналогично, предиктор $X_2^* = \alpha_0^* + \alpha_1^* X_1$, где

$$\alpha_1^* = \frac{\operatorname{cov}(X_2, X_1)}{\mathbf{D}X_1} = \rho_{12}\frac{\sigma_2}{\sigma_1}, \quad \alpha_0^* = \mathbf{E}X_2 - \alpha_1^*\mathbf{E}X_1.$$

Следовательно, остатки имеют вид

$$e_1 = Y - Y^* = Y - \beta_1^* X_1 - \beta_0^*, \quad e_2 = X_2 - X_2^* = X_2 - \alpha_1^* X_1 - \alpha_0^*,$$

а их с. к. о. по (14) равны $\mathbf{D}e_1 = \sigma_0^2(1-\rho_{01}^2)$ и $\mathbf{D}e_2 = \sigma_2^2(1-\rho_{12}^2)$. Далее непосредственно вычисляем

$$\operatorname{cov}(e_1, e_2) = \operatorname{cov}(Y - \beta_1^* X_1, X_2 - \alpha_1^* X_1) = (\rho_{02} - \rho_{01}\rho_{12})\sigma_0\sigma_2,$$

$$\rho(e_1, e_2) = \frac{\operatorname{cov}(e_1, e_2)}{\sqrt{\mathbf{D}e_1\mathbf{D}e_2}} = \frac{\rho_{02} - \rho_{01}\rho_{12}}{\sqrt{(1-\rho_{01}^2)(1-\rho_{12}^2)}}. \tag{20}$$

С другой стороны, обращая корреляционную матрицу $\|\rho_{ij}\|_0^2$ и вычисляя элементы

$$\rho^{02} = \frac{\rho_{02} - \rho_{01}\rho_{12}}{|\rho_{ij}|}, \quad \rho^{00} = \frac{1-\rho_{12}^2}{|\rho_{ij}|}, \quad \rho^{22} = \frac{1-\rho_{01}^2}{|\rho_{ij}|},$$

по формуле (18) приходим к тому же результату (20).

Вычислим, наконец, оптимальный предиктор (13).

Здесь

$$\Sigma = \begin{Vmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{Vmatrix}, \quad \underline{a} = (\rho_{01}\sigma_0\sigma_1,\ \rho_{02}\sigma_0\sigma_2)$$

и в итоге находим

$$\beta_1^* = \frac{\sigma_0}{\sigma_1}\ \frac{\rho_{01} - \rho_{02}\rho_{12}}{1-\rho_{12}^2}, \quad \beta_2^* = \frac{\sigma_0}{\sigma_2}\ \frac{\rho_{02} - \rho_{01}\rho_{12}}{1-\rho_{12}^2}.$$

Таким образом,

$$\varphi^*(X_1, X_2) = \mathbf{E}Y + \beta_1^*(X_1 - \mathbf{E}X_1) + \beta_2^*(X_2 - \mathbf{E}X_2)$$

с указанными коэффициентами β_1^* и β_2^*. Множественный же коэффициент корреляции (15) здесь имеет вид

$$\rho_{0(12)}^2 = \frac{\rho_{01}^2 + \rho_{02}^2 - 2\rho_{01}\rho_{02}\rho_{12}}{1-\rho_{12}^2}. \tag{21}$$

•

§ 2.4. Выборочные квантили и порядковые статистики

1. Теоретические и эмпирические квантили

Важный класс характеристик наблюдаемой (скалярной) случайной величины ξ составляют *квантили*. По определению, p-квантиль ζ_p — это решение уравнения $F(\zeta_p) = p$, $0 < p < 1$, где $F(x)$ — функция распределения ξ. Если

функция $F(x)$ строго возрастает, то однозначно определена обратная к ней функция F^{-1}, и тогда $\zeta_p = F^{-1}(p)$. В общем случае, для произвольной непрерывной справа функции F обратная к ней функция определяется равенством

Квантильная функция

$$F^{-1}(p) = \inf\{x \quad F(x) \geqslant p\} \qquad (1)$$

и называется *квантильной функцией*.

Итак, в любом случае (для любой функции распределения) p-квантиль случайной величины ξ (или ее распределения) есть $\zeta_p = F^{-1}(p)$, т. е. значение квантильной функции (1) в точке $p \in (0, 1)$ (см. рис. 1)

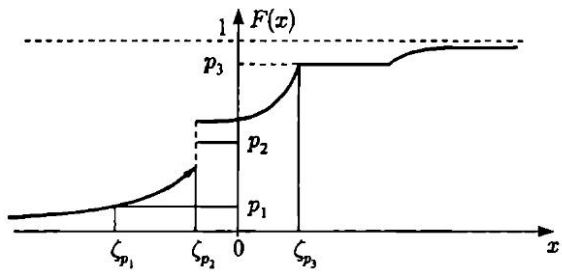


Рис. 1

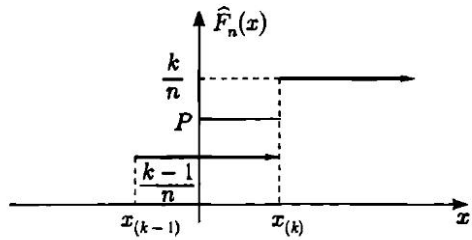


Рис. 2

Статистическим аналогом для ζ_p (*выборочной*, или *эмпирической*, *p-квантилью*) $\widehat{\zeta}_{n,p}$ называется p-квантиль эмпирической функции распределения $\widehat{F}_n(x)$, построенной по выборке $\boldsymbol{X} = (X_1, \quad , X_n)$ из $\mathcal{L}(\xi)$. Эти новые выборочные характеристики просто выражаются через порядковые статистики выборки (см. п. 1 § 2.1). Действительно (см. рис. 2), если $(k-1)/n < p \leqslant k/n$, или $k - 1 < np \leqslant k$, то $\widehat{\zeta}_{n,p} = \widehat{F}_n^{-1}(p) = X_{(k)}$. Поэтому, если число np целое, т. е. $np = k$, то $\widehat{\zeta}_{n,p} = X_{(np)}$; если же число np дробное (т. е. $k-1 < np < k$), то целая его часть $[np] = k-1$ (т. е. $k = [np]+1$) и потому $\widehat{\zeta}_{n,p} = X_{([np]+1)}$. Итак,

$$\widehat{\zeta}_{n,p} = \widehat{F}_n^{-1}(x) = \begin{cases} X_{([np]+1)} & \text{при } np \text{ дробном,} \\ X_{(np)} & \text{при } np \text{ целом.} \end{cases} \qquad (2)$$

Выборочная медиана



При $p = 1/2$ квантиль $\zeta_{1/2}$ называется *теоретической медианой*, а $\widehat{\zeta}_{n,\,1/2}$ — *выборочной медианой* (несколько отличается от определения медианы вариационного ряда в п. 1 § 2.1); при $p = 1/4$ и $3/4$ соответствующие квантили иногда называются *квартилями*, соответственно *нижними* и *верхними*.

Итак, поскольку, ввиду (2), выборочные квантили по существу есть соответствующие порядковые статистики выборки, то дальнейший разговор будет вестись в терминах именно порядковых статистик.

2. Распределение порядковых статистик

Предположим, что распределение случайной величины ξ абсолютно непрерывно и $f(x) = F'(x)$ — ее плотность. Найдем функцию распределения $F_k(x) = \mathrm{P}\{X_{(k)} \leqslant x\}$ и плотность $g_k(x) = F_k'(x)$ произвольной порядковой статистики $X_{(k)}$ $(k = 1, \quad , n)$ выборки $X = (X_1, \quad , X_n)$ из $\mathcal{L}(\xi)$.

Событие $\{X_{(k)} \leqslant x\}$ означает, что не менее k элементов выборки X имеют значения, не превосходящие x, но это есть событие $\{\mu_n(x) \geqslant k\}$ (см. (3) § 2.1), следовательно, эти два события эквивалентны. Как показано в п. 2 § 2.1, случайная величина $\mu_n(x)$ при любом $x \in R^1$ распределена по биномиальному закону $\mathrm{Bi}\big(n, F(x)\big)$ Отсюда имеем

$$F_k(x) = \mathrm{P}\{\mu_n(x) \geqslant k\} = \sum_{r=k}^{n} C_n^r\, F^r(x)(1 - F(x))^{n-} \tag{3}$$

В частности, из (3) имеем, что распределения экстремумов имеют вид

$$F_1(x) = \mathrm{P}\{X_{(1)} \leqslant x\} = 1 - (1 - F(x))^n,$$
$$F_n(x) = \mathrm{P}\{X_{(n)} \leqslant x\} = F^n(x).$$

Неполная бета-функция

Таким образом, в любой модели функции распределения порядковых статистик легко выписываются. Далее, правую часть (3) можно записать через так называемую *неполную бета-функцию*

$$B(z; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^z t^{a-1}(1-t)^{b-1}\, dt, \quad 0 \leqslant z \leqslant 1, \tag{4}$$

которая представляет собой функцию распределения закона $\mathrm{Be}(a, b)$ (см. п. 4 § 1.2). Именно, интегрированием по частям нетрудно установить формулу

$$B(z; k, n-k+1) = \frac{n!}{(k-1)!(n-k)!} \int_0^z t^{k-1}(1-t)^{n-k}\, dt =$$

$$\sum_{r=k}^{n} C_n^r z^r (1-z)^{n-r}, \tag{4'}$$

из которой следует, что формуле (3) можно придать также вид

Функция распределения $X_{(k)}$

$$F_k(x) = B\big(F(x); k, n-k+1\big). \qquad (5)$$

Неполная бета-функция подробно протабулирована, поэтому представление (5) полезно для проведения численных расчетов, связанных с порядковыми статистиками.

До сих пор нигде не предполагалось существование плотности $f(x) = F'(x)$, поэтому предыдущие формулы справедливы для любых распределений $\mathcal{L}(\xi)$. Если же ξ имеет плотность распределения $f(x)$, то соотношение (5) можно продифференцировать по x и получить в итоге, что плотность $g_k(x) = F_k'(x)$ существует и имеет вид

Плотность распределения $X_{(k)}$

$$g_k(x) = \frac{n!}{(k-1)!(n-k)!} F^{k-1}(x)(1-F(x))^{n-k} f(x). \qquad (6)$$

Можно также выписать совместную плотность для любого числа порядковых статистик (см. упр. 31). Укажем, например, вид двумерной плотности $g_{kr}(x_1, x_2)$ для пары порядковых статистик $(X_{(k)}, X_{(r)})$, $1 \leqslant k < r \leqslant n$:

$$g_{kr}(x_1, x_2) = \frac{n!}{(k-1)!(r-k-1)!(n-r)!} \times$$
$$\times F^{k-1}(x_1)\big(F(x_2)-F(x_1)\big)^{r-k-1}(1-F(x_2))^{n-r} f(x_1) f(x_2) \qquad (7)$$

при $x_1 < x_2$ и равна нулю в остальных точках.

Приведем, наконец, вид совместной плотности всех n порядковых статистик:

$$g_{1\ldots n}(x_1, \ldots, x_n) = n! f(x_1) \ldots f(x_n), \quad x_1 < x_2 < \quad < x_n. \qquad (8)$$

Рассмотрим один важный для последующего и поучительный пример работы с порядковыми статистиками.

Пример 1. Пусть $X = (X_1, \ldots, X_n)$ — выборка из стандартного показательного распределения $\Gamma(1, 1)$ (см. п. 3 § 1.2) т. е. когда плотность $f(x) = e^{-x}$ $x > 0$. В этом случае совместная плотность всех n порядковых статистик $X_{(1)}, \ldots, X_{(n)}$ по формуле (8) есть

$$g_{1\ldots n}(x_1, \ldots, x_n) = n! e^{-x_1 - \ldots - x_n}, \quad 0 < x_1 < x_2 < \quad < x_n. \qquad (9)$$

Рассмотрим новые статистики

Спейсинги показательной выборки

$$Y_r = (n-r+1)(X_{(r)} - X_{(r-1)}),$$
$$r = 1, \ldots, n, \quad X_{(0)} = 0,$$

называемые *спейсингами*, и найдем их совместное распределение. Для этого надо лишь воспользоваться соотношением (44) § 1.2. В данном случае якобиан

преобразования $y_1 = nx$, $y_2 = (n-1)(x_2 - x_1)$, $y_n = x_n - x_{n-1}$ равен, очевидно, $n!$, т. е. постоянен, а

$$\sum_{r=1}^{n} y_r = \sum_{r=1}^{n} (n - r + 1)(x_r - x_{r-1}) = \sum_{r=1}^{n} x_r.$$

Следовательно, для совместной плотности $\varphi(y_1, \ldots, y_n)$ преобразованных статистик имеем в виду (9) представление

$$\varphi(y_1, \quad , y_n) = \frac{g_{1\ldots n}(x_1, \quad , x_n)}{n!} = e^{-y_1 - \ldots - y_n}$$

Но это означает, что статистики $Y_1, \ldots, Y_n$ независимы и имеют одно и то же стандартное показательное распределение $\Gamma(1, 1)$. (!)

Отсюда можно сделать следующие интересные выводы. Выразим исходные порядковые статистики $X_{(k)}$ через спейсинги $Y_1, \quad , Y_n$:

$$X_{(k)} = \sum_{r=1}^{k} \frac{Y_r}{n - r + 1}, \quad k = 1, \quad , n. \tag{10}$$

Моменты порядковых статистик показательной выборки

Тем самым $X_{(k)}$ представляется в виде линейной комбинации независимых случайных величин, поэтому, в частности,

$$\mathbf{E}X_{(k)} = \sum_{r=1}^{k} \frac{1}{n - r + 1} \mathbf{E}Y_r,$$

$$\mathbf{D}X_{(k)} = \sum_{r=1}^{k} \frac{1}{(n - r + 1)^2} \mathbf{D}Y_r.$$

Но среднее и дисперсия закона $\Gamma(1, 1)$ равны 1 (см. (21) § 1.2), поэтому

$$\mathbf{E}X_{(k)} = \sum_{r=1}^{k} \frac{1}{n - r + 1} = \sum_{j=n-k+1}^{n} \frac{1}{j}, \quad \mathbf{D}X_{(k)} = \sum_{j=n-k+1}^{n} \frac{1}{j^2}. \tag{11}$$

В частности, для максимального значения выборки $X_{(n)}$ эти формулы принимают вид

$$\mathbf{E}X_{(n)} = \sum_{j=1}^{n} \frac{1}{j}, \quad \mathbf{D}X_{(n)} = \sum_{j=1}^{n} \frac{1}{j^2}. \tag{12}$$

Если объем выборки $n \to \infty$, то, используя известные асимптотические формулы из анализа:

$$\sum_{j=1}^{n} \frac{1}{j} = \ln n + c + o(1), \quad c = 0{,}5772\ldots \quad \text{— константа Эйлера}^{4)},$$

[4)] Эйлер Леонард (1707–1783) — гениальный швейцарский математик, механик, физик и астроном, академик Петербургской АН.

$$\sum_{j=1}^{n} \frac{1}{j^2} = \frac{\pi^2}{6} + o(1),$$

заключаем, что в среднем $\boldsymbol{X}_{(n)}$ неограниченно возрастает как $\ln n$, в то время как дисперсия $\mathbf{D}\boldsymbol{X}_{(n)} \longrightarrow \pi^2/6$, т. е. имеет конечный предел. •

Замечание. В рассмотренном примере нам удалось, «играя» на специфических особенностях показательной модели, вычислить в явном виде моменты (среднее и дисперсию) порядковых статистик. Но это исключительный случай. Вообще же проблема отыскания моментов порядковых статистик весьма сложна и ее удается решить лишь для некоторых моделей $\mathcal{L}(\xi)$ (см. упр. 32 данной главы и упр. 46, 56 гл. 1), в отличие от проблемы получения их распределений. Подчеркнем, что для выборочных моментов в этом аспекте — все наоборот.

Асимптотические свойства порядковых статистик

Перейдем теперь к исследованию асимптотических свойств порядковых статистик для больших выборок (при $n \to \infty$). Поскольку число порядковых статистик равно объему выборки и, следовательно, неограниченно растет вместе с n, то было бы наивным надеяться на то, что все они будут иметь однотипные асимптотические свойства. И, действительно, эта проблема распадается на две самостоятельные задачи: асимптотика «средних» членов вариационного ряда (2) § 2.1 и предельные распределения «крайних» порядковых статистик. Прежде всего поясним эту терминологию.

Под *средними членами* вариационного ряда выборки объема n при $n \to \infty$ понимаются порядковые статистики $X_{(k)}$, номера которых $k = k(n)$ удовлетворяют условию $k/n \to p$ при некотором p, $0 < p < 1$. Таким образом, это такие члены вариационного ряда, которые расположены «далеко» от его концов. *Крайние* же *порядковые статистики* — это члены вариационного ряда, расположенные в его концевых законах, т. е. статистики $X_{(r)}$ при r ограниченных и $X_{(n-s+1)}$ при s ограниченных.

3. Асимптотическая нормальность средних членов вариационного ряда

Зафиксируем некоторое p, $0 < p < 1$, и рассмотрим порядковую статистику $X_{(k)}$ с $k = [np]$. Справедлива следующая теорема о ее асимптотическом распределении.

Теорема 1. Если в некоторой окрестности точки ζ_p плотность $f(x)$ непрерывна вместе с производной и $f(\zeta_p) > 0$, то при $n \to \infty$

$$\mathcal{L}\left(X_{(\lfloor np \rfloor)}\right) \sim \mathcal{N}\left(\zeta_p, \frac{pq}{nf^2(\zeta_p)}\right), \quad q = 1 - p. \tag{13}$$

Доказательство. Рассмотрим случайную величину

$$\eta_n = (X_{(k)} - \zeta_p) f(\zeta_p) \sqrt{\frac{n}{pq}}$$

и покажем, что ее плотность $\varphi_n(x)$ сходится при $n \to \infty$ к стандартной нормальной плотности $\varphi(x) = (2\pi)^{-1/2} e^{-x^2/2}$ (см. (7) § 1.2). По формуле 43 § 1.2 из (6) имеем

$$\varphi_n(x) = \sqrt{\frac{pq}{n}} f^{-1}(\zeta_p) g_k(t_n), \quad \text{где} \quad t_n = \zeta_p + \frac{x}{f(\zeta_p)} \sqrt{\frac{pq}{n}}.$$

Запишем $\varphi_n(x)$ в виде произведения $A_1(n)A_2(n)A_3(n)$, где

$$A_1(n) = \frac{n!}{(k-1)!(n-k)!} \sqrt{\frac{pq}{n}} p^{k-1} q^{n-k} \qquad A_2(n) = \frac{f(t_n)}{f(\zeta_p)},$$

$$A_3(n) = \left(\frac{F(t_n)}{p}\right)^{k-1} \left(\frac{1 - F(t_n)}{q}\right)^{n-k}$$

и найдем предел каждого из этих трех множителей при $n \to \infty$ и $k = [np]$. Простым упражнением на применение формулы Стирлинга для гамма-функции (см. п. 3 § 1.2) является установление предела $A_1(n) \to (2\pi)^{-1/2}$ (оставляем это читателю). Далее, $A_2(n) \to 1$ в силу непрерывности функции $f(x)$. Наконец, поскольку $F(\zeta_p) = p$, из тейлоровского разложения в точке ζ_p

$$F(t_n) = p + x\sqrt{\frac{pq}{n}} + \frac{x^2 pq}{2n} \frac{f'(\zeta_p)}{f^2(\zeta_p)} + o\left(\frac{1}{n}\right)$$

нетрудно получить, что $\ln A_3(n) \to -x^2/2$. В итоге это дает нужный предел $\varphi_n(x) \to \varphi(x)$. ■

Состоятельность выборочных квантилей

Замечание. Ввиду связи (2) между порядковыми статистиками и выборочными квантилями соотношение (13) справедливо также и при замене $X_{(\lfloor np \rfloor)}$ на $\widehat{\zeta}_{n,p}$:

$$\mathcal{L}(\widehat{\zeta}_{n,p}) \sim \mathcal{N}\left(\zeta_p, \frac{pq}{nf^2(\zeta_p)}\right). \tag{14}$$

Отсюда можно сделать вывод о том, что выборочная p-квантиль $\widehat{\zeta}_{n,p}$ является асимптотически несмещенной и состоятельной оценкой теоретической p-квантили ζ_p.

Справедлив более общий результат (*многомерная предельная нормальная теорема*): если в некоторых окрестностях точек ζ_{p_i}, $i = 1, \ldots, r$, где $0 < p_1 < p_2 < \ldots < p_r < 1$, плотность $f(x)$ непрерывна вместе с производной и положительна, то при $n \to \infty$ совместное распределение выборочных квантилей $\widehat{\zeta}_{n,p_i}$, $i = 1, \ldots, r$, асимптотически нормально с вектором средних значений $(\zeta_{p_1}, \ldots, \zeta_{p_r})$ и матрицей вторых моментов

$$\frac{1}{n}\|\sigma_{ij}\|_1^r, \quad \text{где} \quad \sigma_{ij} = \frac{p_i(1-p_j)}{f(\zeta_{p_i}) f(\zeta_{p_j})}.$$

Распределение выборочной медианы в нормальной модели

Пример 2. Найдем асимптотическое распределение выборочной медианы $\widehat{\zeta}_{n,1/2}$ для выборки из нормального распределения $\mathcal{N}(\mu, \sigma^2)$. Соотношение (14) для выборочной медианы из любого распределения с плотностью $f(x)$ принимает вид

$$\mathcal{L}\left(\widehat{\zeta}_{n,1/2}\right) \sim \mathcal{N}\left(\zeta_{1/2}, \frac{1}{4nf^2(\zeta_{1/2})}\right). \tag{15}$$

Если же $f(x)$ имеет вид, указанный во Введении к гл. 1, то легко видеть, что функция $f(x)$ симметрична относительно точки μ, поэтому $\zeta_{1/2} = \mu$, а $f(\mu) = 1/(\sqrt{2\pi}\sigma)$. Подставляя эти значения в (15), получим

$$\mathcal{L}\left(\widehat{\zeta}_{n,1/2}\right) \sim \mathcal{N}\left(\mu, \frac{\pi\sigma^2}{2n}\right). \tag{16}$$

Для сравнения отметим, что соответствующее выборочное среднее $\overline{X}$ при любом n нормально $\mathcal{N}(\mu, \sigma^2/n)$. Таким образом, обе статистики $\overline{X}$ и $\widehat{\zeta}_{n,1/2}$ являются состоятельными оценками теоретического среднего μ, но $\overline{X}$ имеет меньшую дисперсию, поэтому выборочному среднему следует отдать предпочтение как более точной оценке. Систематически вопросы сравнения точности различных оценок одной и той же теоретической характеристики будут рассматриваться в гл. 3. •

Итак, средние члены вариационного ряда для больших выборок из достаточно гладких распределений являются асимптотически нормальными (как и выборочные моменты, см. п. 4 § 2.2), и ими можно оценивать теоретические квантили любых уровней p, $0 < p < 1$. Совершенно иначе устроены асимптотические распределения крайних порядковых статистик, к рассмотрению которых мы теперь переходим.

4. Асимптотическое поведение крайних порядковых статистик

Относительно асимптотического поведения порядковых статистик, находящихся на разных краях вариационного ряда, справедливо следующее общее утверждение.

Теорема 2. Для выборки из абсолютно непрерывного распределения статистики $X_{(r)}$ и $X_{(n-s+1)}$ при любых фиксированных r, $s \geqslant 1$ и $n \to \infty$ асимптотически независимы и при этом

$$\begin{aligned} &\mathcal{L}\left(nF(X_{(r)})\right) \longrightarrow \Gamma(1, r), \\ &\mathcal{L}\left(n\left(1 - F(X_{(n-s+1)})\right)\right) \longrightarrow \Gamma(1, s) \end{aligned} \tag{17}$$

(определение гамма-распределения $\Gamma(a, \lambda)$ см. в п. 3 § 1.2).

Доказательство. Пусть $\varkappa_n$ (каппа) и η_n обозначают рассматриваемые преобразованные статистики:

$$\varkappa_n = nF(X_{(r)}), \quad \eta_n = n\bigl(1 - F(X_{(n-s+1)})\bigr). \tag{18}$$

Вычислим сначала якобиан преобразования $y_1 = nF(x_1)$, $y_2 = n(1 - F(x_2))$: очевидно, он есть $J(x_1, x_2) = -n^2 f(x_1) f(x_2)$. С учетом этого, по формуле (44) § 1.2 и с использованием формулы (7) для совместной плотности $\varphi_n(y_1, y_2)$ преобразованных статистик $\varkappa_n$ и η_n имеем следующее представление:

$$\varphi_n(y_1, y_2) = \frac{g_{r,\,n-s+1}\left(F^{-1}\left(\dfrac{y_1}{n}\right), F^{-1}\left(1 - \dfrac{y_2}{n}\right)\right)}{n^2 f\left(F^{-1}\left(\dfrac{y_1}{n}\right)\right) f\left(F^{-1}\left(1 - \dfrac{y_2}{n}\right)\right)} =$$

$$= \frac{n!}{(r-1)!(n-r-s)!(s-1)!n^2}\left(\frac{y_1}{n}\right)^{k-1}\left(\frac{y_2}{n}\right)^{s-1}\left(1 - \frac{y_1 + y_2}{n}\right)^{n-r-}$$

В условиях теоремы при любых фиксированных $y_1, y_2 > 0$

$$\frac{n!}{(n-r-s)!n^{k+s}} \longrightarrow 1, \qquad \left(1 - \frac{y_1 + y_2}{n}\right)^{n-r-s} \longrightarrow e^{-y_1 - y_2},$$

что в итоге дает

$$\varphi_n(y_1, y_2) \longrightarrow \frac{y_1^{k-1} e^{-y_1}}{(k-1)!} \cdot \frac{y_2^{s-1} e^{-y_2}}{(s-1)!}.$$

Но это и означает, что в пределе случайные величины $\varkappa_n$, η_n (следовательно, и $X_{(r)}$, $X_{(n-s+1)}$) независимы и имеют указанные в (17) гамма-распределения, которые являются в данном случае распределениями Эрланга соответственно порядков r и s. ■

Обратим внимание на следующие моменты. Во-первых, предельные распределения крайних порядковых статистик не являются нормальными (в отличие от средних членов вариационного ряда). Во-вторых, теорема 2 определяет вид предельных распределений не самих порядковых статистик, а некоторых функций от них, определенных в (18). Конечно, если эти уравнения можно явно разрешить, т. е. получить явные представления

$$X_{(r)} = F^{-1}\left(\frac{\varkappa_n}{n}\right), \quad X_{(n-s+1)} = F^{-1}\left(1 - \frac{\eta_n}{n}\right), \tag{19}$$

то можно получить предельные распределения и самих статистик $X_{(r)}$, $X_{(n-s+1)}$ через предельные распределения $\varkappa_n$, η_n. В качестве иллюстрации такой возможности рассмотрим следующий пример.

Пример 3. Пусть $\mathcal{L}(\xi) = \Gamma(1, 1)$ — стандартное показательное распределение. Здесь

Распределение верхних экстремумов для показательной модели

$$F(x) = \int_0^x e^{-t}dt = 1 - e^{-x}, \quad x \geqslant 0,$$

и легко находим

$$F^{-1}(t) = -\ln(1-t), \quad 0 < t < 1.$$

Следовательно, в силу (19) $X_{(n-s+1)} = \ln n - \ln \eta_n$. Отсюда по теореме 2

$$\mathbf{P}\{X_{(n-s+1)} - \ln n \leqslant x\} = \mathbf{P}\{\eta_n \geqslant e^{-x}\} \longrightarrow \int_{e^{-x}}^{\infty} \pi_{s-1}(y)\,dy, \tag{20}$$

где обозначено $\pi_k(t) = e^{-t}t^k/k!$, $k = 0, 1, 2,$ — это обозначение будет использоваться и в дальнейшем. Интегрированием по частям интеграл в (20) может быть вычислен в явном виде, что приводит к окончательному результату:

$$\mathbf{P}\{X_{(n-s+1)} - \ln n \leqslant x\} \longrightarrow \sum_{k=0}^{s-1} \pi_k(e^{-x}), \quad -\infty < x < \infty. \tag{21}$$

В частности, для максимального значения выборки $X_{(n)}$ (21) принимает вид

$$\mathbf{P}\{X_{(n)} - \ln n \leqslant x\} \longrightarrow e^{-e^{-x}} \quad -\infty < x < \infty. \tag{22}$$

Напомним, что ранее, в примере 1, были получены асимптотические значения среднего и дисперсии статистики $X_{(n)}$.

Предельное распределение в (22) называется *дважды экспоненциальным*, его среднее и дисперсия равны соответственно c (константа Эйлера) и $\pi^2/6$, так что для

$$c = \lim_{n\to\infty}\left(\sum_{j=1}^{n}\frac{1}{j} - \ln n\right)$$

имеем одно из возможных интегральных представлений:

$$c = \int_{-\infty}^{\infty} x \exp\{-x - e^{-x}\}\,dx$$

(значение c вычислено с 260 десятичными знаками). •

Однако, в явном виде обратную функцию $F^{-1}(t)$ удается выписать лишь в редких случаях, поэтому теорема 2, вообще говоря, оставляет открытым вопрос о предельных распределениях для самих минимальных и максимальных значений выборки (или нижних и верхних экстремумов). Кроме того, эта теорема не охватывает случай выборок из дискретных распределений (но факт

асимптотической независимости для них также имеет место). Общее решение вопроса о возможных предельных распределениях для крайних порядковых статистик в больших выборках дается классической *теорией экстремальных значений*, основной вклад в которую внесен отечественными математиками Б. В. Гнеденко (1941–1943) и Н. В. Смирновым (1949) (о нем см. в § 2.1). Полное изложение этой теории выходит за рамки настоящего пособия, поэтому мы ограничимся лишь общим описанием основных ее результатов.

5. Асимптотическая теория для верхних экстремумов $X_{(n-m+1)}$, $m \geqslant 1$

Общая задача ставится так: пусть при некотором выборе нормализующих констант $a_n > 0$ и b_n имеет место сходимость

$$\mathbf{P}\left\{\frac{X_{(n-m+1)} - b_n}{a_n} \leqslant x\right\} \xrightarrow[n\to\infty]{} G_m(x), \tag{23}$$

где $G_m(x)$ — некоторая *невырожденная* (т. е. имеющая более чем одну точку роста) функция распределения. Нужно ответить на следующие два вопроса:

1) какой может быть предельная функция распределения $G_m(x)$? и
2) при каких условиях на общую функцию распределения $F(x)$ случайных величин $X_1, \ldots, X_n$ и каких a_n и b_n будет выполняться соотношение (23) для каждой возможной функции $G_m(x)$?

Распределения экстремальных значений

Вот полные ответы на эти вопросы.

1) $G_m(x)$ может быть лишь одного типа из следующих трех, называемых *распределениями экстремальных значений*: (обозначения см. в (20))

$$\Lambda_1^{(m)}(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 0, \\ \displaystyle\sum_{r=0}^{m-1} \pi_r(x^{-\alpha}) & \text{при} \quad x > 0, \end{cases} \qquad \alpha > 0;$$

$$\Lambda_2^{(m)}(x) = \begin{cases} \displaystyle\sum_{r=0}^{m-1} \pi_r\left(|x|^{\alpha}\right) & \text{при} \quad x \leqslant 0, \\ 1 & \text{при} \quad x > 0, \end{cases} \qquad \alpha > 0; \tag{24}$$

$$\Lambda_3^{(m)}(x) = \sum_{r=0}^{m-1} \pi_r(e^{-x}), \quad -\infty < x < \infty.$$

Если выполняется (23) с $G_m = \Lambda_j^{(m)}$ то говорят, что функция распределения $F(x)$ принадлежит *области притяжения* типа $\Lambda_j^{(m)}$

2) Необходимыми и достаточными условиями принадлежности функции распределения $F(x)$ области притяжения каждого из этих трех типов являются (ниже используются обозначения $x_F = \sup\{x \quad F(x) < 1\}$, $\gamma_n = \inf\{x \quad F(x) \geqslant 1 - 1/n\}$):

Области притяжения экстремальных типов

тип $\Lambda_1^{(m)}$:

$$x_F = \infty \quad \text{и} \quad \lim_{t \to \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha} \qquad \forall\, x > 0,$$

при этом $a_n = \gamma_n$, $b_n = 0$;

тип $\Lambda_2^{(m)}$:

$$x_F < \infty \quad \text{и} \quad \lim_{t \downarrow 0} \frac{1 - F(x_F - tx)}{1 - F(x_F - t)} = x^{\alpha}, \quad \forall\, x > 0,$$

при этом $a_n = x_F - \gamma_n$, $b_n = x_F$;

тип $\Lambda_3^{(m)}$: при всех $x \in R^1$ существует

$$\lim_{t \uparrow x_F} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x} \qquad \text{где} \quad g(t) = \int_t^{x_F} \frac{1 - F(u)}{1 - F(t)}\, du;$$

при этом $a_n = g(\gamma_n)$, $b_n = \gamma_n$.

Если $F(x)$ не относится ни к одной из указанных трех категорий, то соотношение (23) не может быть выполнено ни при каком выборе констант a_n и b_n.

3) Если существует $f(x) = F'(x)$, то достаточными условиями принадлежности $F(x)$ к соответствующим областям притяжения являются:

Условия принадлежности к экстремальным типам

тип $\Lambda_1^{(m)}$: $f(x) > 0$ для всех конечных $x \geqslant x_0$ и

$$\lim_{t \to \infty} \frac{t f(t)}{1 - F(t)} = \alpha > 0; \tag{25}$$

тип $\Lambda_2^{(m)}$ $f(x) > 0$ в некотором конечном интервале (x_0, x_F), $f(x) = 0$ при $x > x_F$ и

$$\lim_{t \uparrow x_F} \frac{(x_F - t) f(t)}{1 - F(t)} = \alpha > 0; \tag{26}$$

тип $\Lambda_3^{(m)}$ $f'(x) < 0$ в некотором интервале (x_0, x_F) $(x_F \leqslant \infty)$, $f(x) = 0$ при $x \geqslant x_F$ и

$$\lim_{t \uparrow x_F} \frac{f'(t)(1 - F(t))}{f^2(t)} = -1. \tag{27}$$

Замечание. Аналогичная теория существует и для нижних экстремумов $X_{(m)}$, $m \geqslant 1$ фиксировано; при этом, поскольку $X_{(m)} = -X'_{(n-m+1)}$, где $X'_{(1)} \leqslant \ldots \leqslant X'_{(n)}$ есть вариационный ряд выборки $(-X_1, \ldots, -X_n)$, то результаты для $X_{(m)}$ получаются простой переформулировкой соответствующих результатов для верхних экстремумов. Интересующихся деталями мы отсылаем к обзорной статье[5] где имеются также дополнительные сведения и комментарии по этой проблеме.

Приведенный выше пример 3 относится к области притяжения типа Λ_3. К такому же типу относится и следующий пример (далее мы ограничиваемся лишь максимумом $X_{(n)}$ и пишем Λ_j вместо $\Lambda_j^{(1)}$).

Примеры распределений экстремумов

Пример 4 (Нормальное распределение). Пусть $F(x) = \Phi(x)$ (см. (7) § 1.2). Здесь $x_F = \infty$ и $1 - \Phi(x) \sim \varphi(x)/x$ при $x \to \infty$, применим критерий (27) и при выборе

$$a_n = (2 \ln n)^{-1/2} \qquad b_n = (2 \ln n)^{1/2} - \frac{1}{2}(2 \ln n)^{-1/2}(\ln \ln n + \ln 4\pi)$$

имеем:

$$\mathcal{L}\left(\frac{X_{(n)} - b_n}{a_n}\right) \longrightarrow \Lambda_3. \qquad \bullet$$

Пример 5 (Распределение Парето, см. п. 9 § 1.2). Здесь $F(x) = 1 - (x_0/x)^\alpha$ $x > x_0$, и применим критерий (25), при этом $a_n = \gamma_n = x_0 n^{1/\alpha}$, $b_n = 0$. Следовательно, по (24)

$$\mathbf{P}\{\gamma_n^{-1} X_{(n)} \leqslant x\} \longrightarrow e^{-x^{-\alpha}} \qquad x > 0.$$

Таким образом, распределение Парето с параметрами x_0 и α принадлежит области притяжения типа Λ_1 с данным параметром α. $\bullet$

Пример 6 (Равномерное распределение). Пусть $F(x) = x$, $0 \leqslant x \leqslant 1$. Здесь $x_F = 1$ и применим, очевидно, критерий (26) с $\alpha = 1$, при этом $a_n = 1/n$, $b_n = 1$. Таким образом, в данном случае имеем предел типа Λ_2 (см. (24)):

$$\mathbf{P}\{n(X_{(n)} - 1) \leqslant x\} \longrightarrow e^x \qquad x < 0.$$

С примерами статистических применений экстремальных значений выборки мы встретимся в гл. 3. $\bullet$

§ 2.5. Линейные и квадратичные статистики от нормальных выборок

Выше (см. п. 1, 2 § 1.2) уже говорилось о той исключительной роли, которую играет в приложениях математической статистики нормальная модель. Различные свойства нормальных случайных величин обсуждались в § 1.2 и сформулированы в форме упражнений 26–32, 38, 40 к гл. 1. Здесь мы установим еще

[5] *Ивченко Г. И.* Предельные распределения экстремальных характеристик выборок // Обозрение прикладной и промышленной математики. 1997. Т. 4. Вып. 3. С. 443–469.

несколько общих результатов для нормальных выборок, относящихся к линейным и квадратичным формам от нормальных случайных величин, которые играют важную роль в статистике нормальной модели.

1. Линейные и квадратичные статистики, условия их независимости

Пусть $X = (X_1, \ldots, X_n)$ — выборка из стандартного нормального распределения $\mathcal{N}(0, 1)$. Рассмотрим квадратичную форму

$$Q = \sum_{i,j=1}^{n} a_{ij} X_i X_j = X'AX, \quad A = \|a_{ij}\|_1^n, \quad A' = A,$$

и m линейных форм $t_k = \sum_{i=1}^{n} b_{ki} X_i$, $k = 1, \ldots, m$, или в матричных обозначениях $\underline{t} = BX$. Здесь B — прямоугольная $(m \times n)$-матрица и $\underline{t} = (t_1, \ldots, t_m)$. Далее будем обозначать через O матрицу с нулевыми элементами и через $\mathbb{1}_n$ — единичную матрицу порядка n.

Следующая лемма дает критерий независимости статистик Q и $\underline{t}$.

Лемма 1. Если $BA = O$, то статистики Q и $\underline{t}$ независимы.

Доказательство. Матрица A действительна и симметрична, поэтому, как известно из алгебры, можно найти такую ортогональную матрицу U (т. е. $UU' = \mathbb{1}_n$), что $U'AU = \Lambda$. Здесь Λ — диагональная матрица с элементами $\lambda_1, \ldots, \lambda_n$, являющимися корнями характеристического уравнения

$$\det(A - \lambda \mathbb{1}_n) = 0.$$

Столбцами $\underline{u}_1, \ldots, \underline{u}_n$ матрицы U являются собственные векторы матрицы A:

$$A\underline{u}_i = \lambda_i \underline{u}_i, \quad i = 1, \ldots, n.$$

Пусть r — ранг матрицы A (т. е. число линейно независимых ее строк) и $\lambda_1, \ldots, \lambda_r$ — отличные от нуля характеристические числа. Эквивалентной формой записи

$$A = U\Lambda U' \tag{1}$$

является, так называемое, *спектральное представление* матрицы A:

Спектральное представление матрицы

$$A = \sum_{i=1}^{r} \lambda_i \underline{u}_i \underline{u}_i' \tag{2}$$

(напомним, что при матричных операциях векторы понимаются как вектор-столбцы). По условию леммы

$$O = BA = \sum_{i=1}^{r} \lambda_i (B\underline{u}_i)\underline{u}_i'.$$

Умножим это равенство на вектор $\underline{u}_j$ справа. В силу ортогональности собственных векторов ($\underline{u}_i'\underline{u}_j = 0$ при $i \neq j$) получим

$$B\underline{u}_i = 0, i = 1, \quad , r. \tag{3}$$

Рассмотрим теперь случайный вектор $(t_1, \ldots, t_m, \underline{u}_1'X, \quad , \underline{u}_r'X)$. Этот вектор распределен по нормальному закону, поскольку является линейным преобразованием нормального вектора X (ведь при линейных преобразованиях нормальность сохраняется, см. п. 2 § 1.2). В то же время из представления (2) имеем

$$Q = \sum_{i=1}^{r} \lambda_i (X'\underline{u}_i)(\underline{u}_i'X) = \sum_{i=1}^{r} \lambda_i (\underline{u}_i'X)^2 \tag{4}$$

Следовательно, утверждение будет доказано, если показать, что случайные величины $t_1, \quad , t_m$ некоррелированы с $\underline{u}_i'X$, $i = 1, \quad , r$ (тогда эти две группы величин будут независимыми, см. п. 2 § 1.2). Обозначая через $\underline{b}_i'$ строки матрицы B $(i = 1, \quad , m)$, в силу соотношений (3) имеем

$$\operatorname{cov}(t_i, \underline{u}_j'X) = \operatorname{cov}(\underline{b}_i'X, \underline{u}_j'X) = \mathrm{E}(\underline{b}_i'X\underline{u}_j'X) - \mathrm{E}(\underline{b}_i'X)\mathrm{E}(\underline{u}_j'X) =$$
$$= \mathrm{E}(\underline{b}_i'XX'\underline{u}_j) = \underline{b}_i'\mathrm{E}(XX')\underline{u}_j = \underline{b}_i'\mathbb{1}_n\underline{u}_j = \underline{b}_i'\underline{u}_j = 0. \quad \blacksquare$$

Установим теперь аналогичный критерий независимости двух квадратичных форм $Q_1 = X'AX$ и $Q_2 = X'BX$.

Лемма 2. Если $AB = BA = O$, то статистики Q_1 и Q_2 независимы.

Доказательство. Пусть для матрицы A справедливо спектральное представление (2), а аналогичное представление для B имеет вид

$$B = \sum_{j=1}^{s} \nu_j \underline{v}_j \underline{v}_j', \quad s = \operatorname{rank} B.$$

По условию,

$$O = AB = \sum_{i,j} \lambda_i \nu_j \underline{u}_i (\underline{u}_i'\underline{v}_j)\underline{v}_j'.$$

Умножив это равенство слева на $\underline{u}_k'$, а справа на $\underline{v}_l'$, получим $\underline{u}_k'\underline{v}_l = 0$, $k = 1, \quad , r$, $l = 1, \quad s$, т. е. векторы $\underline{u}_1, \quad , \underline{u}_r$ ортогональны всем векторам $\underline{v}_1, \quad , \underline{v}_s$. Отсюда, как и выше, имеем, что случайные величины $\underline{u}_i'X$ и $\underline{v}_j'X$ некоррелированы, а так как они совместно нормально распределены, то и независимы. Но

$$Q_1 = \sum_{i=1}^{r} \lambda_i (\underline{u}_i'X)^2, \quad Q_2 = \sum_{j=1}^{s} \nu_j (\underline{v}_j'X)^2,$$

следовательно, Q_1 и Q_2 независимы. $\blacksquare$

2. Распределения квадратичных статистик

Среди квадратичных статистик от выборки $X = (X_1, \ldots, X_n)$ из $\mathcal{L}(\xi) = \mathcal{N}(0, 1)$ простейшей является сумма квадратов $X_1^2 + \ldots + X_n^2$, о которой нам уже известно (см. п. 3 § 1.2 и упр. 38 к гл. 1), что ее распределение есть χ^2-распределение с n степенями свободы: $\mathcal{L}(X_1^2 + \ldots + X_n^2) = \chi^2(n)$. Что можно сказать о распределении произвольной квадратичной формы $Q = X'AX$ с действительной симметричной матрицей A? Этот вопрос мы и рассмотрим в данном пункте. Обозначим $\operatorname{tr} A$ *след* (track) квадратной матрицы A, т. е. сумму ее диагональных элементов. Имеет место следующее общее утверждение.

Теорема 1. Если матрица A идемпотентна ($A^2 = A$), то $\mathcal{L}(Q) = \chi^2(r)$, где $r = \operatorname{tr} A = \operatorname{rank} A$.

Доказательство. Пусть для A справедливо представление (2). Тогда, так как ненулевые характеристические числа матрицы A^2 равны λ_i^2, $i = 1, \ldots, r$, то, в силу идемпотентности A, $\lambda_1 = \quad = \lambda_r = 1$ и потому (см. (4))

$$Q = \sum_{i=1}^{r} (\underline{u}_i' X)^2$$

Из ортонормированности собственных векторов $\underline{u}_1, \ldots, \underline{u}_r$ следует, что случайные величины $\underline{u}_i' X$, $i = 1, \ldots, r$, независимы и нормальны $\mathcal{N}(0, 1)$, следовательно, $\mathcal{L}(Q) = \chi^2(r)$. Наконец, используя непосредственно проверяемое равенство $\operatorname{tr}(AB) = \operatorname{tr}(BA)$, из формулы (1) находим

$$\operatorname{tr} A = \operatorname{tr}(\mathcal{U}'\mathcal{U}\Lambda) = \operatorname{tr}\Lambda = \lambda_1 + \quad + \lambda_r = r. \quad \blacksquare$$

Замечание. В качестве следствия этого результата докажем упражнение 40 к гл. 1. Применив алгоритм п. 2 § 1.2, построим преобразование $\underline{Z} = \Lambda^{-1/2}\mathcal{U}'(\underline{\xi} - \underline{\mu})$, переводящее вектор $\underline{\xi}$ с $\mathcal{L}(\underline{\xi}) = \mathcal{N}(\underline{\mu}, \Sigma)$ в стандартный нормальный вектор $\underline{Z}$. Тогда $\underline{\xi} - \underline{\mu} = \mathcal{U}\Lambda^{1/2}\underline{Z}$ и, следовательно,

$$Q(\underline{\xi}) = (\underline{\xi} - \underline{\mu})'\Sigma^{-1}(\underline{\xi} - \underline{\mu}) = \underline{Z}'\Lambda^{1/2}\mathcal{U}'\Sigma^{-1}\mathcal{U}\Lambda^{1/2}\underline{Z} = \underline{Z}'\underline{Z},$$

так как $\mathcal{U}'\Sigma^{-1}\mathcal{U} = \Lambda^{-1}$. По теореме 1 $\mathcal{L}(Q(\underline{\xi})) = \chi^2(k)$, поскольку единичная матрица $\mathbb{1}_k$ является идемпотентной и $\operatorname{tr}\mathbb{1}_k = k$.

Таким образом, при некоторых условиях на матрицу квадратичной формы от стандартного нормального вектора она распределена типично по закону χ^2. Но что будет, если нормальный вектор $X = (X_1, \quad, X_n)$ не является стандартным? В таких случаях возникает новое распределение для квадратичных статистик от X, являющееся обобщением распределения хи-квадрат. Добавим это важное распределение к списку уже имеющихся.

Пусть $X_1, \quad , X_n$ — независимые нормальные случайные величины и при этом $\mathcal{L}(X_i) = \mathcal{N}(\mu_i, 1)$, $i = 1, \ldots, n$. Тогда, как мы знаем, сумма

$$\sum_{i=1}^{n} (X_i - \mu_i)^2$$

имеет распределение $\chi^2(n)$. Однако часто бывает необходимо знать распределение суммы квадратов самих случайных величин X_i. Распределение случайной величины $U^2 = X_1^2 + \quad + X_n^2$ называют *нецентральным распределением хи-квадрат с n степенями свободы и параметром нецентральности* $\lambda^2 = \mu_1^2 + \mu_n^2$ и обозначают символом $\chi^2(n; \lambda^2)$: $\mathcal{L}(U^2) = \chi^2(n; \lambda^2)$. Плотность этого распределения имеет вид

Нецентральное χ^2-распределение

$$f(x|n; \lambda^2) = \frac{x^{n/2-1}}{2^{n/2}} e^{-(x+\lambda^2)/2} \sum_{j=0}^{\infty} \frac{(x\lambda^2)^j}{j! 2^{2j} \Gamma(j + n/2)}, \quad x > 0, \tag{5}$$

а первые два момента равны

$$\mathbf{E}U^2 = n + \lambda^2 \qquad \mathbf{D}U^2 = 2(n + 2\lambda^2). \tag{6}$$

Если параметр $\lambda^2 = 0$, то $\chi^2(n; \lambda^2) = \chi^2(n)$, т.е. нецентральное распределение χ^2 переходит в обычное (центральное) распределение χ^2

В качестве иллюстрированного примера ситуаций, где возникает нецентральное распределение хи-квадрат, докажем следующее утверждение.

Лемма 3. Пусть n-мерный нормальный вектор Y имеет собственнораспределение $\mathcal{N}(\underline{\mu}, \Sigma)$. Тогда статистика $Q = Y'\Sigma^{-1}Y$ имеет распределение $\chi^2(n; \lambda^2)$, где $\lambda^2 = \underline{\mu}'\Sigma^{-1}\underline{\mu}$.

Доказательство. Рассмотрим вектор $X = \Lambda^{-1/2}\mathcal{U}'Y$, где $\mathcal{U}$ — ортогональная матрица, приводящая Σ к диагональному виду (см. п. 2 § 1.2): $\mathcal{U}'\Sigma\mathcal{U} = \Lambda$. Тогда $\mathcal{L}(X) = \mathcal{N}(\widetilde{\underline{\mu}}, \mathbb{1}_n)$, где $\widetilde{\underline{\mu}} = \Lambda^{-1/2}\mathcal{U}'\underline{\mu}$. Далее, $Y = \mathcal{U}\Lambda^{1/2}X$ и потому

$$Q = X'\Lambda^{1/2}\mathcal{U}'\Sigma^{-1}\mathcal{U}\Lambda^{1/2}X = X'X = X_1^2 + \ldots + X_n^2,$$

поскольку $\mathcal{U}'\Sigma^{-1}\mathcal{U} = \Lambda^-$ Следовательно,

$$\mathcal{L}(Q) = \mathcal{L}(X_1^2 + \quad + X_n^2) = \chi^2(n; \lambda^2),$$

где

$$\lambda^2 = \widetilde{\underline{\mu}}'\underline{\mu} = \underline{\mu}'\mathcal{U}\Lambda^{-1/2}\Lambda^{-1/2}\mathcal{U}'\underline{\mu} = \underline{\mu}'\Sigma^{-1}\underline{\mu},$$

поскольку $\mathcal{U}\Lambda^{-1}\mathcal{U}' = \Sigma^{-1}$. ■

3. Теорема Фишера

Наиболее часто используемыми статистиками от нормальных выборок являются выборочное среднее

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

и выборочная дисперсия

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Первая из них является линейной, а вторая — квадратичной функцией от $X_1, \ldots, X_n$. Полученные в предыдущих пунктах результаты позволяют весьма просто доказать знаменитую теорему Р. Фишера (1925).

Теорема 2 (Теорема Фишера). Пусть $X = (X_1, \ldots, X_n)$ — выборка из $\mathcal{L}(\xi) = \mathcal{N}(\mu, \sigma^2)$. Тогда статистики $\overline{X}$ и S^2 независимы и при этом

$$\mathcal{L}(\overline{X}) = \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{а} \quad \mathcal{L}\left(\frac{n}{\sigma^2}S^2\right) = \chi^2(n-1).$$

Доказательство. Перейдем к стандартным нормальным случайным величинам

$$Y_i = \frac{X_i - \mu}{\sigma}, \quad i = 1, \ldots, n.$$

Тогда легко проверить, что

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{\overline{X} - \mu}{\sigma} \quad \text{и} \quad S^2(Y) = \frac{S^2(X)}{\sigma^2}.$$

Поэтому достаточно показать, что $\overline{Y}$ и $S^2(Y)$ независимы и при этом

$$\mathcal{L}(\sqrt{n}\,\overline{Y}) = \mathcal{N}(0, 1), \quad \text{а} \quad \mathcal{L}\big(nS^2(Y)\big) = \chi^2(n-1).$$

Рассмотрим n-мерный вектор-столбец $\underline{b} = (1/n, \ldots, 1/n)'$ и $(n \times n)$-матрицу $B = \|\underline{b} \ldots \underline{b}\|$. Легко проверить, что $\overline{Y} = \underline{b}'Y$, а

$$nS^2(Y) = (Y - BY)'(Y - BY).$$

Отсюда $nS^2(Y) = Y'AY$, где матрица $A = \mathbb{1}_n - B$ идемпотентна и $\operatorname{tr} A = n - 1$. Теперь $\underline{b}'A = \underline{b}' - \underline{b}'B = \underline{b}' - \underline{b}' = \underline{O}$ и, следовательно, по лемме 1 статистики $\overline{Y}$ и $S^2(Y)$ независимы. Далее, закон распределения $\overline{Y}$ очевиден (см. (5) § 1.2), а закон распределения $nS^2(Y)$ следует из теоремы 1. ■

Замечание. Так как для распределения $\chi^2(n-1)$ среднее и дисперсия равны соответственно $n-1$ и $2(n-1)$ (см. п. 3 § 1.2), то из доказанной теоремы следует, что для нормальных выборок

$$\mathrm{E}S^2 = \sigma^2\frac{n-1}{n}, \quad \mathrm{D}S^2 = 2\sigma^4\frac{n-1}{n^2} \tag{7}$$

(сравни с соответствующими формулами (11) и (13) § 2.2 для выборки из произвольного распределения, при этом используйте формулу (20) § 2.2 для эксцесса γ_2 и тот факт, что для нормального распределения $\gamma_2 = 0$).

Приведем несколько важных примеров статистических применений теоремы Фишера.

Пример 1 (Оценивание параметров нормальной модели). Рассмотрим задачу построения доверительных интервалов (см. определение (26) § 2.2) для неизвестных параметров θ_1 и θ_2^2 общей нормальной статистической модели $\mathcal{N}(\theta_1, \theta_2^2)$.

Доверительный интервал для дисперсии

Доверительный интервал для дисперсии θ_2^2 может быть построен на основе того факта, что выборочная дисперсия $S^2 = S^2(X)$ для выборки $X = (X_1, \ldots, X_n)$ из распределения $\mathcal{N}(\theta_1, \theta_2^2)$ имеет по теореме 2 следующее распределение:

$$\mathcal{L}\left(\frac{nS^2}{\theta_2^2}\right) = \chi^2(n-1)$$

(независимо от значений неизвестных параметров!). Поэтому, обозначая через $\chi^2_{p,n-1}$ p-квантиль распределения $\chi^2(n-1)$ и выбирая при заданном доверительном уровне γ два числа α_1 и α_2 так, чтобы $\alpha_1 + \alpha_2 = 1-\gamma$, будем иметь

$$\gamma = \mathbf{P}\left\{\chi^2_{\alpha_1,n-1} < \frac{nS^2}{\theta_2^2} < \chi^2_{1-\alpha_2,n-1}\right\} =$$
$$= \mathbf{P}\left\{\frac{nS^2}{\chi^2_{1-\alpha_2,n-1}} < \theta_2^2 < \frac{nS^2}{\chi^2_{\alpha_1,n-1}}\right\}. \tag{8}$$

Но это и означает, что γ-доверительным интервалом для θ_2^2 является интервал

$$\left(\frac{nS^2}{\chi^2_{1-\alpha_2,n-1}}, \frac{nS^2}{\chi^2_{\alpha_1,n-1}}\right), \quad \alpha_1 + \alpha_2 = 1-\gamma. \tag{9}$$

На самом деле — это целое семейство интервалов, так как при заданном γ числа α_1 и α_2 можно выбрать разными способами. На практике обычно рекомендуют выбирать их так, чтобы были одинаковыми вероятности попадания величины nS^2/θ_2^2 левее и правее указанных в (8) границ:

$$\mathbf{P}\left\{\frac{nS^2}{\theta_2^2} \leqslant \chi^2_{\alpha_1,n-1}\right\} = \mathbf{P}\left\{\frac{nS^2}{\theta_2^2} \geqslant \chi^2_{1-\alpha_2,n-1}\right\} = \frac{1-\gamma}{2},$$

т. е. $\alpha_1 = \alpha_2 = (1-\gamma)/2$. В этом случае интервал (9) называют *центральным γ-доверительным интервалом*, и он имеет вид

$$\left(\frac{nS^2}{\chi^2_{\frac{1+\gamma}{2},n-1}}, \frac{nS^2}{\chi^2_{\frac{1-\gamma}{2},n-1}}\right). \tag{10}$$

В частности для выборки объема $n=10$ и доверительного уровня $\gamma=0{,}9$ из таблиц квантилей χ^2-распределения имеем $\chi^2_{0{,}05;9}=3{,}3251$, $\chi^2_{0{,}95;9}=16{,}9190$, поэтому центральный 0,9-доверительный интервал для θ_2^2 имеет вид $(0{,}5911S^2, 3{,}007S^2)$

Доверительный интервал для среднего

Рассмотрим теперь задачу доверительного оценивания среднего θ_1. Как уже отмечалось в п. 6 § 1.2 в связи с обсуждением распределения Стьюдента, в силу теоремы Фишера стьюдентово отношение (см. (29) § 1.2)

$$t_{n-1} = \sqrt{n-1}\,\frac{\overline{X}-\theta_1}{S} \tag{11}$$

имеет при любых значениях параметров θ_1 и θ_2^2 распределение Стьюдента $S(n-1)$, при этом оно симметрично относительно нуля (см. (28) § 1,2). Поэтому, обозначая через $t_{p,n-1}$ p-квантиль распределения $S(n-1)$, из (11) имеем

$$\gamma = \mathbf{P}\left\{|t_{n-1}| < t_{\frac{1+\gamma}{2},n-1}\right\} = \mathbf{P}\left\{\overline{X} - \frac{S}{\sqrt{n-1}}t_{\frac{1+\gamma}{2},n-1} < \theta_1 < \overline{X} + \frac{S}{\sqrt{n-1}}t_{\frac{1+\gamma}{2},n-1}\right\},$$

т. е. γ-доверительным интервалом для θ_1 является симметричный относительно случайной точки $\overline{X}$ интервал

$$\left(\overline{X} \mp t_{\frac{1+\gamma}{2},n-1}\frac{S}{\sqrt{n-1}}\right). \tag{12}$$

В частности, из таблиц квантилей распределения Стьюдента имеем $t_{0{,}975;9} = 2{,}262$, поэтому для выборки объема $n=10$ и доверительного уровня $\gamma = 0{,}95$ этот интервал имеет вид $(\overline{X} \mp 0{,}754S)$.

Доверительная область для среднего и дисперсии

Иногда необходимо знать оценку одновременно и для среднего θ_1 и для дисперсии θ_2^2, т. е. необходимо оценить пару $\theta = (\theta_1, \tau)$, $\tau = \theta_2^2$. В этом случае требуется построить такое случайное (т. е. зависящее от выборки $\boldsymbol{X}$) подмножество $\mathcal{G}_\gamma(\boldsymbol{X})$ в $R_+^2 = \{\theta = (\theta_1, \tau) \quad -\infty < \theta_1 < \infty, \tau > 0\}$, которое бы с заданной вероятностью γ накрывало неизвестное значение θ, т. е. чтобы выполнялось соотношение

$$\mathbf{P}\{\theta \in \mathcal{G}_\gamma(\boldsymbol{X})\} = \gamma. \tag{13}$$

Такое подмножество может быть построено на основании уже полученных результатов. Именно, сопоставим каждой паре $\theta = (\theta_1, \tau)$ подмножество выборочного пространства

$$\boldsymbol{H}(\theta) = \left\{\underline{\boldsymbol{x}} \quad \sqrt{\frac{n}{\tau}}|\overline{x} - \theta_1| < t,\ t' < \frac{n}{\tau}s^2 < t''\right\},$$

где

$$t = c_{\gamma_1} = \Phi^{-1}\left(\frac{1+\gamma_1}{2}\right), \quad t' = \chi^2_{\frac{1-\gamma_2}{2}, n-1}, \quad t'' = \chi^2_{\frac{1+\gamma_2}{2}, n-1},$$

$$s^2 = S^2(\underline{x}) \quad \text{и} \quad \gamma_1\gamma_2 = \gamma.$$

Тогда в силу независимости $\overline{X}$ и S^2

$$\mathbf{P}\{X \in H(\theta)\} = \mathbf{P}\left\{\sqrt{\frac{n}{\tau}}|\overline{X} - \theta_1| < c_{\gamma_1}\right\}\mathbf{P}\left\{t' < \frac{n}{\tau}S^2 < t''\right\} = \gamma_1\gamma_2 = \gamma.$$

Но условие $\{X \in H(\theta)\}$ эквивалентно условию $\{\theta \in \mathcal{G}_\gamma(X)\}$, где

$$\mathcal{G}_\gamma(\underline{x}) = \left\{\theta = (\theta_1, \tau) : \tau > (\overline{x} - \theta_1)^2\frac{n}{t^2}, \ \frac{ns^2}{t''} < \tau < \frac{ns^2}{t'}\right\}.$$

Область $\mathcal{G}_\gamma(\underline{x})$ представляет собой часть полуплоскости R^2_+, ограниченную параболой $\tau = (\overline{x} - \theta_1)^2 n/t^2$ и двумя прямыми $\tau = ns^2/t''$ и $\tau = ns^2/t'$ (см. рис. 1). Таким образом, построена γ-доверительная область $\mathcal{G}_\gamma(X)$ для пары $\theta = (\theta_1, \theta_1^2)$, удовлетворяющая (13). •

Пример 2 (Задача сравнения двух средних). Среди прикладных задач встречаются такие, при решении которых приходится сравнивать средние двух независимых выборок. Для такого сравнения полезно иметь доверительный интервал для разности соответствующих выборкам математических ожиданий. В случае нормальных выборок такой интервал можно легко построить на основании теоремы Фишера.

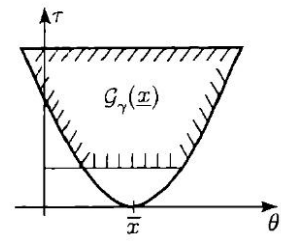


Рис. 1

Пусть $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ — две независимые выборки, причем первая — из распределения $\mathcal{N}(\theta_1^{(1)}, \theta_2^2)$, а вторая — из $\mathcal{N}(\theta_1^{(2)}, \theta_2^2)$. Таким образом, соответствующие модели различаются только своими средними и требуется оценить их разность $\Delta = \theta_1^{(1)} - \theta_1^{(2)}$. Эту задачу называют *задачей сравнения двух средних*. Построим выборочные средние $\overline{X}$, $\overline{Y}$ и выборочные дисперсии $S^2(X)$, $S^2(Y)$ этих выборок. Из независимости выборок и в силу теоремы 2 эти четыре статистики независимы. Кроме того,

$$\mathcal{L}\left(\frac{\overline{X} - \theta_1^{(1)}}{\theta_2}\right) = \mathcal{N}\left(0, \frac{1}{n}\right), \quad \mathcal{L}\left(\frac{\overline{Y} - \theta_1^{(2)}}{\theta_2}\right) = \mathcal{N}\left(0, \frac{1}{m}\right),$$

$$\mathcal{L}\left(\frac{n}{\theta_2^2}S^2(X)\right) = \chi^2(n-1), \quad \mathcal{L}\left(\frac{m}{\theta_2^2}S^2(Y)\right) = \chi^2(m-1).$$

Отсюда (см. (5) § 1.2)

$$\mathcal{L}\left(\frac{\overline{X} - \theta_1^{(1)}}{\theta_2} - \frac{\overline{Y} - \theta_1^{(2)}}{\theta_2}\right) = \mathcal{L}\left(\frac{\overline{X} - \overline{Y} - \Delta}{\theta_2}\right) = \mathcal{N}\left(0, \frac{1}{n} + \frac{1}{m}\right)$$

или

$$\mathcal{L}\left(\sqrt{\frac{mn}{m+n}}\ \frac{\overline{X}-\overline{Y}-\Delta}{\theta_2}\right)=\mathcal{N}(0,1),$$

а также (см. п 3 § 1.2, свойство воспроизводимости χ^2-распределения)

$$\mathcal{L}\left(\frac{n}{\theta_2^2}S^2(X)+\frac{m}{\theta_2^2}S^2(Y)\right)=\chi^2(m+n-2).$$

Следовательно, стьюдентово отношение (27) § 1.2 имеет в данном случае вид

$$t_{m+n-2}=\frac{\sqrt{\frac{mn}{m+n}}\ \frac{\overline{X}-\overline{Y}-\Delta}{\theta_2}}{\sqrt{\frac{nS^2(X)+mS^2(Y)}{\theta_2^2(m+n-2)}}}=$$

$$=\sqrt{\frac{mn(m+n-2)}{m+n}}\times\frac{\overline{X}-\overline{Y}-\Delta}{\sqrt{nS^2(X)+mS^2(Y)}},\qquad(14)$$

и оно имеет распределение Стьюдента $S(m+n-2)$ при любых $\theta_1^{(1)}$, $\theta_1^{(2)}$ и θ_2^2. Основываясь на (14), γ-доверительный интервал для Δ строится так же, как и интервал (13) в предыдущем примере: он имеет вид

Доверительный интервал для разности двух средних

$$\left(\overline{X}-\overline{Y}\mp t_{\frac{1+\gamma}{2},m+n-2}\left[\frac{m+n}{mn(m+n-2)}\left(nS^2(X)+mS^2(Y)\right)\right]^{1/2}\right).\qquad(15)$$

•

Пример 3 (Доверительный интервал для отношения дисперсий двух нормальных моделей). Рассмотрим ситуацию, когда $X=(X_1,\ldots,X_n)$ — выборка из распределения $\mathcal{N}(\theta_{11},\theta_{12}^2)$, а $Y=(Y_1,\quad,Y_m)$ — из распределения $\mathcal{N}(\theta_{21},\theta_{22}^2)$ и выборки независимы. Требуется оценить отношение $\tau=\theta_{12}^2/\theta_{22}^2$ неизвестных дисперсий. Построим выборочные дисперсии $S^2(X)$ и $S^2(Y)$. По условию они независимы, а по теореме Фишера

$$\mathcal{L}\left(\frac{n}{\theta_{12}^2}S^2(X)\right)=\chi^2(n-1),\quad\mathcal{L}\left(\frac{m}{\theta_{22}^2}S^2(Y)\right)=\chi^2(m-1).$$

Поэтому мы можем построить статистику Снедекора (32) § 1.2, которая в данном случае принимает вид

$$F_{n-1,m-1}=\frac{n(m-1)}{m(n-1)}\frac{S^2(X)}{S^2(Y)}\frac{1}{\tau}.\qquad(16)$$

Эта статистика имеет распределение Снедекора $S(n-1,m-1)$ при любых значениях параметров $\theta_{11},\theta_{12}^2,\theta_{21},\theta_{22}^2$, поэтому, обозначая через $F_{p,n-1,m-1}$

p-квантиль распределения $S(n-1, m-1)$ и задаваясь доверительным уровнем γ, из (16) будем иметь

$$\gamma = \mathbf{P}\left\{ F_{(1-\gamma)/2,\, n-1,\, m-1} < F_{n-1,\, m-1} < F_{(1+\gamma)/2,\, n-1,\, m-1} \right\} =$$

$$= \mathbf{P}\left\{ \frac{n(m-1)}{m(n-1)} \frac{S^2(X)}{S^2(Y)} \frac{1}{F_{(1+\gamma)/2,\, n-1,\, m-1}} < \tau < \right.$$

$$\left. < \frac{n(m-1)}{m(n-1)} \frac{S^2(X)}{S^2(Y)} \frac{1}{F_{(1-\gamma)/2,\, n-1,\, m-1}} \right\}.$$

Доверительный интервал для отношения дисперсий

Тем самым построен центральный γ-доверительный интервал для τ: он имеет вид

$$\left(\frac{n(m-1)}{m(n-1)} \frac{S^2(X)}{S^2(Y)} \frac{1}{F_{(1+\gamma)/2,\, n-1,\, m-1}}, \right.$$

$$\left. \frac{n(m-1)}{m(n-1)} \frac{S^2(X)}{S^2(Y)} \frac{1}{F_{(1-\gamma)/2,\, n-1,\, m-1}} \right) \qquad (17)$$

В частности, из таблиц квантилей распределения Снедекора имеем

$$F_{0,95;\,8,8} = F^{-1}_{0,05;\,8,8} = 3{,}44, \quad F_{0,95;\,12,12} = F^{-1}_{0,05;\,12,12} = 2{,}69,$$

поэтому 0,9-доверительный интервал для выборок объемов $n = m = 9$ есть

$$\left(\frac{0{,}29S^2(X)}{S^2(Y)}, \frac{3{,}44S^2(X)}{S^2(Y)} \right),$$

а для выборок объемов $n = m = 13$ —

$$\left(\frac{0{,}37S^2(X)}{S^2(Y)}, \frac{2{,}69S^2(X)}{S^2(Y)} \right) \quad \bullet$$

Упражнения

Там, где есть воля, всегда есть дорога.

Хадсон[6]

1 Пусть ν_n — число успехов в n испытаниях Бернулли с вероятностью успеха p, $0 < p < 1$. При больших n вычислить границу δ_γ такую, чтобы событие $|\nu_n/n - p| \leqslant \delta_\gamma$ имело вероятность $\approx \gamma$. Укладываются ли в эти границы при $\gamma = 0{,}98$ результаты следующего эксперимента (эксперимент Бюффона[7]): при $n = 4040$ бросаниях монеты наблюдалось 2048 выпадений «герба»?

[6] Хадсон — английский альпинист XIX в.

[7] Бюффон Жорж Луи Леклерк (1707–1788) — французский естествоиспытатель, почетный член Петербургской АН (1776).

◄ *Указание.* Применить теорему Муавра—Лапласа (см. замечание к упр. 33 гл. 1). ►

2 (*Продолжение*). Смоделировать выборку объема n 1000 из распределения Bi(1, 2/5) и проверить аналогично предыдущему, укладываются ли в указанные границы результаты полученных экспериментальных данных. Построить график, соединив прямыми последовательные точки

$$\left(k, \frac{\nu_k}{k}\right), \quad k = 100, 200, \ldots \quad 1000.$$

3 Смоделировать выборку из биномиального распределения Bi(4, 1/3). Построить график частот ν_r/n, где ν_r — число опытов, в которых наблюдалось значение $r = 0, 1, 2, 3, 4$, и вычислить выборочные среднее, дисперсию и коэффициенты асимметрии и эксцесса. Найти δ из условия $\mathbf{P}\{|\overline{X} - \alpha_1| \leqslant \delta\} = 0{,}998$ и сравнить с ним вычисленное по экспериментальным данным отклонение выборочного среднего от теоретического α_1

◄ *Указание.* Воспользоваться теоремой 3 § 2.2. ►

4 Смоделировать выборку из полиномиального распределения $M(500; p_i$ 1/5, $i = 1,$, 5), и, рассматривая эти данные как наблюдения над случайной величиной ξ, принимающей значения $-2, -1, 0, 1, 2$, построить полигон частот и вычислить выборочные среднее и дисперсию. Найти δ из условия

$$\mathbf{P}\{|\overline{X} - \alpha_1| \leqslant \delta\} = 0{,}98$$

и сравнить с ним наблюденное значение отклонения $|\overline{X} - \alpha_1|$.

5 Смоделировать выборку объема $n = 100$ из показательного распределения $\Gamma(1, 1)$. Вычислить ее основные характеристики, построить эмпирическую функцию распределения и гистограмму, вычислить выборочные моменты $\widehat{\alpha}_j$, $j = 1, 2$, и сравнить их с теоретическими значениями.

6 Пусть $X_{N1},$, X_{Nn} — приближенно нормальные $\mathcal{N}(\mu, \sigma^2)$ числа, каждое из которых получено суммированием N равномерно распределенных на (0, 1) слагаемых. Получить три реализации (при $N = 2, 4, 12$) выборок с $n = 100$, $\mu = 0$, $\sigma^2 = 1$. Для каждой выборки построить эмпирические функции распределения и гистограммы; получить оценки μ и σ^2; вычислить 3-й и 4-й выборочные центральные моменты и сравнить их с истинными значениями теоретических моментов.

7 Смоделировать выборку объема $n = 500$ из равномерного распределения $\mathcal{U}(0, 12)$. Группируя данные по интервалам $\mathcal{E}_r = (r, r + 1]$, $r = 0, 1,$ 11, построить полигон частот и сравнить его с графиком функции $f(x) = c$, $0 < x < 12$. Рассматривая группированные частоты как результаты наблюдений над дискретной случайной величиной, принимающей значения, совпадающие с серединами соответствующих интервалов, вычислить выборочные среднее и дисперсию и сравнить их с теоретическими значениями.

8 Для данных примера 2 в § 2.2 построить график частот ν_r/n и сравнить его с графиком функции $ce^{-x^2/2}$; вычислить выборочные коэффициенты асимметрии и эксцесса; найти δ из условия $\mathbf{P}\{|\overline{X} - \alpha_1| \leqslant \delta\} = 0{,}998$ и сравнить с ним вычисленное по приведенным данным отклонение выборочного среднего от теоретического α_1

9 (*Продолжение*). В том же эксперименте наблюдалась также случайная величина ξ, равная числу костей с 6 очками. Соответствующие данные приведены в таблице:

r	0	1	2	3	4	5	6	$\geqslant 7$	Σ
ν_r	447	1145	1181	796	380	115	24	8	4096

Ответить на те же вопросы, а также вычислить выборочные среднее и дисперсию.

◀ *Указание.* Считать, что $\mathcal{L}(\xi) = \mathrm{Bi}(12, 1/6)$. ▶

10 Наблюдались показания 500 наугад выбранных часов, выставленных в витринах. Пусть r — номер промежутка от r-го часа до $(r+1)$-го, $r = 0, 1, \ldots, 11$, а ν_r — число часов, показания которых принадлежат r-му промежутку. Результаты таким образом сгруппированных наблюдений оказались следующими [15, с. 459]:

r	0	1	2	3	4	5	6	7	8	9	10	11	Σ
ν_r	41	34	54	39	49	45	41	33	37	41	47	39	500

а) Построить полигон частот и сравнить его с графиком функции $f(x) = c$, $0 < x < 12$.

б) Рассматривая эти данные как независимые наблюдения над дискретной случайной величиной ξ, принимающей значения, совпадающие с серединами соответствующих интервалов (т. е. значения 0,5; 1,5; ..., 11,5), вычислить выборочные среднее и дисперсию.

в) Принимая, что в п. б) случайная величина ξ имеет равномерное распределение, найти δ из условия $\mathrm{P}\{|\overline{X} - \alpha_1| \leqslant \delta\} = 0{,}98$ и сравнить с ним наблюдавшееся значение отклонения $|\overline{X} - \alpha_1|$.

11 При проведении $n = 2608$ опытов по наблюдению числа α-частиц (ξ), излучаемых радиоактивным веществом за определенный период времени, получены следующие данные (ν_r — число опытов, для которых число $\xi = r$, $r = 0, 1, \ldots$):

r	0	1	2	3	4	5	6	7	8	9	10	11	$\geqslant 12$	Σ
ν_r	57	203	383	525	532	408	273	139	45	27	10	4	2	2608

Построить график частот ν_r/n и вычислить выборочные среднее и дисперсию.

12 Пусть $(-0{,}8; 2{,}9; 4{,}4; -5{,}6; 1{,}1; -3{,}2)$ — наблюдавшиеся значения случайной величины ξ. Вычислить основные характеристики выборки, построить эмпирическую функцию распределения и убедиться в том, что $\widehat{F}_6(-5) = 1/6$, $\widehat{F}_6(0) = 1/2$, $\widehat{F}_6(4) = 5/6$.

13 В эксперименте наблюдалась целочисленная случайная величина ξ. Соответствующие выборочные значения оказались равными: 3, 0, 4, 3, 6, 0, 3, 1. Вычислить основные характеристики выборки, построить эмпирическую функцию распределения и проверить, что $\widehat{F}_8(1) = 3/8$, $\widehat{F}_8(3) = 3/4$, $\widehat{F}_8(5) = 7/8$.

14 Когда функция $\widehat{F}_n(x)$ будет иметь скачки размера лишь $1/n$ (или, что эквивалентно, будет иметь ровно n скачков)?

15 Пусть $x_1 < x_2$. Чему равна вероятность $\mathrm{P}\{\widehat{F}_n(x_1) = \widehat{F}_n(x_2)\}$?

16 Пусть $\mathrm{P}\{\xi = j\} = p_j$, $i = 0, 1, \ldots, N$, и $X = (X_1, \ldots, X_n)$ — выборка из $\mathcal{L}(\xi)$. Найти вероятность $\mathrm{P}\{\widehat{F}_n(j) - \widehat{F}_n(j-0) = k/n\}$. Рассмотреть случай $\mathcal{L}(\xi) = \mathrm{Bi}(N, p)$.

17 Пусть $X = (X_1, \ldots, X_n)$ — выборка из непрерывного распределения F. Какому распределению будет соответствовать выборка $Y = (Y_1, \ldots, Y_n)$, где $Y_i = F(X_i)$?

Замечание. Если случайная величина X имеет функцию распределения $F(x)$ и $F(x)$ непрерывна, то переход к новой случайной величине $Y = F(X)$ называется *преобразованием Смирнова*.

18 Пусть X — выборка из равномерного распределения $\mathcal{U}(0, 1)$ и F — непрерывная и строго монотонная функция распределения. Положим $Y_i = F^{-1}(X_i)$, $i = 1, \ldots, n$. Из какого распределения будет выборка Y?

19 Пусть в точках x_i выполняются условия $0 < F(x_i) < 1$, $i = 1, \ldots, k$. Рассмотрим события $A_i = \{F(x_i) \in (\widehat{F}_n(x_i) \mp t_{\gamma_i}/\sqrt{n})\}$, где $(\widehat{F}_n(x_i) \mp t_{\gamma_i}/\sqrt{n})$ — асимптотический γ_i-доверительный интервал для $F(x_i)$, определенный в (8) § 2.1. Доказать, что вероятность одновременного осуществления событий $A_1, \ldots, A_k$ удовлетворяет условию

$$\mathbf{P}(A_1 \cap \ldots \cap A_k) \geqslant \sum_{i=1}^{k} \gamma_i.$$

Замечание. Система интервалов

$$\left\{\left(\widehat{F}_n(x_i) \mp \frac{t_{\gamma_i}}{\sqrt{n}}\right), \quad i = 1, \ldots, k\right\}$$

называется *системой совместных доверительных интервалов* для $F(x_i)$, $i = 1, \ldots, k$.

20 Пусть $x_1 < x_2$ — заданные точки на числовой прямой, такие, что $0 < F(x_1) \leqslant F(x_2) < 1$. Доказать формулу

$$\operatorname{cov}(\widehat{F}_n(x_1), \widehat{F}_n(x_2)) = \frac{F(x_1)(1 - F(x_2))}{n}.$$

◀ *Указание.* Представить случайную величиу

$$\Delta_n(x_1, x_2) = \mu_n(x_2) - \mu_n(x_1)$$

(см. (3) § 2.1) в виде суммы независимых индикаторов

$$\Delta_n(x_1, x_2) = \sum_{i=1}^{n} I(x_1 < X_i \leqslant x_2). \blacktriangleright$$

21 Пусть $X = (X_1, \ldots, X_n)$ — выборка из некоторого распределения $\mathcal{L}(\xi)$, для которого существует момент $\alpha_{2r} = \mathbf{E}\xi^{2r}$ Доказать, что дисперсионная матрица вектора $\underline{A}_n = (\widehat{\alpha}_{n1}, \ldots, \widehat{\alpha}_{nr})$ первых r выборочных моментов имест вид

$$\mathbf{D}\underline{A}_n = \left\|\operatorname{cov}(\widehat{\alpha}_{ni}, \widehat{\alpha}_{nj})\right\|_1^r = \frac{1}{n}\left\|\alpha_{i+j} - \alpha_i\alpha_j\right\|_1^r.$$

22 Доказать, что $\operatorname{cov}(\overline{X}, S^2) = \mu_3(n-1)/n^2$ Рассмотреть случай $\mathcal{L}(\xi) = \mathcal{N}(\mu, \sigma^2)$.

23 Доказать теорему 6 § 2.2 об асимптотической нормальности выборочной дисперсии. Убедиться в том, что при $n \to \infty$ $\mathbf{E}S^2 \sim \mu_2$, $\mathbf{D}S^2 \sim (\mu_4 - \mu_2^2)/n$.

◀ *Указание.* Учесть, что $S^2 = \varphi(\widehat{\alpha}_{n1}, \widehat{\alpha}_{n2})$, где $\varphi(x_1, x_2) = x_2 - x_1^2$ (см. (6) § 2.2), и применить теорему 5 § 2.2. ▶

24 Пусть $\alpha_1 \neq 0$, $\mu_2 < \infty$. Доказать, что при $n \to \infty$ $\mathcal{L}(\overline{X}^{-1}) \sim \mathcal{N}\left(\frac{1}{\alpha_1}, \frac{\mu_2}{\alpha_1^4 n}\right)$

◀ *Указание.* Применить теорему 5 § 2.2. ▶

25 Убедиться в справедливости следующих утверждений о распределении выборочного среднего для выборки $X = (X_1, \ldots, X_n)$ из соответствующих распределений:

а) $\mathcal{L}(\xi) = \Gamma(a, \lambda) \Rightarrow \mathcal{L}(n\overline{X}) = \Gamma(a, n\lambda)$,

б) $\mathcal{L}(\xi) = \mathrm{Bi}(k, p) \Rightarrow \mathcal{L}(n\overline{X}) = \mathrm{Bi}(nk, p)$,

в) $\mathcal{L}(\xi) = \Pi(\lambda) \Rightarrow \mathcal{L}(n\overline{X}) = \Pi(n\lambda)$,

г) $\mathcal{L}(\xi) = \overline{\mathrm{Bi}}(r, p) \Rightarrow \mathcal{L}(n\overline{X}) = \overline{\mathrm{Bi}}(nr, p)$.

◀ *Указание.* Воспользоваться свойством воспроизводимости рассматриваемых распределений (см. §§ 1.1 и 1.2). ▶

26 (*Продолжение*). Убедиться в том, что для выборки из распределения Коши $\mathcal{K}(a)$ (см. упр. 48 (б) гл. 1) $\mathcal{L}(\overline{X}) = \mathcal{K}(a)$.

27 Пусть в эксперименте наблюдается двумерная случайная величина $\xi = (\xi_1, \xi_2)$ и (X_{i1}, X_{i2}), $i = 1, \ldots, n$, — соответствующие выборочные данные. Для упрощения вычислений выборочного коэффициента корреляции $\widehat{\rho}_n$ (см. (5) § 2.3) часто бывает удобно предварительно преобразовать исходные данные с помощью некоторого линейного преобразования:

$$Y_{i1} = \frac{X_{i1} - a_1}{b_1}, \quad Y_{i2} = \frac{X_{i2} - a_2}{b_2}, \quad i = 1, \quad , n, \tag{1}$$

где a_1, a_2 — действительные, а b_1, b_2 — положительные числа. Доказать, что при любом выборе a_1, a_2, b_1, b_2

$$\widehat{\rho}_n(X) = \widehat{\rho}_n(Y).$$

28 Вычислить выборочный коэффициент корреляции для следующих данных $n = 26$ наблюдений над величиной $\xi = (\xi_1, \xi_2)$, записанных в виде статистического ряда:

ξ_1	23,0	24,0	24,5	24,5	25,0	25,5	26,0	26,0	26,0	26,5	26,5	27,0	27,0	28,0
ξ_2	0,48	0,50	0,49	0,50	0,51	0,52	0,49	0,51	0,53	0,50	0,52	0,54	0,52	0,53
ν_r	2	4	3	2	1	1	2	1	2	1	1	2	1	3

◀ *Указание.* Произвести линейное преобразование (1), выбрав $a_1 = 26{,}0$, $b_1 = 0{,}5$, $a_2 = 0{,}50$, $b_2 = 0{,}01$. ▶

Ответ: $\widehat{\rho}_{26} = 0{,}793$.

29 Используя независимые равномерно распределенные на $[0, 1]$ случайные числа $\{U_i\}$, построить последовательность пар случайных чисел

$$X_{i1} = \sqrt{-2 \ln U_{2i-1}} \sin(2\pi U_{2i}),$$

$$X_{i2} = \sqrt{-2 \ln U_{2i-1}} \cos(2\pi U_{2i}), \quad i = 1, 2, \quad , 50,$$

вычислить для этих данных выборочный коэффициент корреляции и проверить гипотезу независимости $H_0 : \rho = 0$, используя как точный, так и асимптотический алгоритмы, описанные в п. 4 § 2.3 (см. п. 8 § 1.3).

30 Доказать, что совместная функция распределения двух порядковых статистик $X_{(r)}$ и $X_{(s)}$ $(1 \leqslant r < s \leqslant n)$ при $x < y$ имеет вид

$$F_{rs}(x, y) = \sum_{i=r}^{n} \sum_{j=\max(0, s-i)}^{n-i} \frac{n!}{i!j!(n-i-j)!} F^i(x)[F(y) - F(x)]^j [1 - F(y)]^{n-i-j}$$

и

$$F_{rs}(x, y) = \mathrm{P}\{X_{(s)} \leqslant y\} = F_s(y) \quad \text{при} \quad x \geqslant y.$$

◀ *Указание.* Использовать эквивалентность событий

$$\{X_{(r)} \leqslant x,\ X_{(s)} \leqslant y\} \quad \text{и} \quad \{\mu_n(x) \geqslant r,\ \mu_n(y) \geqslant s\},$$

а также тот факт, что случайные величины

$$\nu_1 = \mu_n(x), \quad \nu_2 = \mu_n(y) - \mu_n(x), \quad \nu_3 = n - \mu_n(y)$$

(при $x < y$) имеют полиномиальное распределение $M(n; p_1, p_2, p_3)$, где $p_1 = F(x)$, $p_2 = F(y) - F(x)$, $p_3 = 1 - F(y)$. Если же $x \geqslant y$, то событие $\{X_{(r)} \leqslant x,\ X_{(s)} \leqslant y\} = \{X_{(s)} \leqslant y\}$. ▶

31 Пусть распределение $\mathcal{L}(\xi)$ абсолютно непрерывно и его плотность $F'(x) = f(x)$. Вывести следующую формулу для совместной плотности порядковых статистик $X_{(k_1)}, \ldots, X_{(k_r)}$ $(1 \leqslant k_1 < \quad < k_r \leqslant n)$:

$$g_{k_1 \ldots k_r}(x_1, \quad, x_r) = \frac{n!}{(k_1 - 1)!(k_2 - k_1 - 1)! \ldots (k_r - k_{r-1} - 1)!(n - k_r)!} \times$$

$$\times F^{k_1 - 1}(x_1)\big(F(x_2) - F(x_1)\big)^{k_2 - k_1 - 1}$$

$$\big(F(x_r) - F(x_{r-1})\big)^{k_r - k_{r-1} - 1}\big(1 - F(x_r)\big)^{n - k_r} f(x_1) \ldots f(x_r)$$

при $x_1 < x_2 < \quad < x_r$ и равна нулю в остальных точках.
Получить отсюда формулы (7) и (8) § 2.4.

◀ *Указание.* Записать вероятность события $\{X_{(k_i)} \in (x_i + dx_i),\ i = 1, \quad, r\}$ при бесконечно малых dx_i, $i = 1, \ldots, r$, и, разделив ее на $dx_1 \ldots dx_r$, получить в пределе при $dx_i \to 0$, $i = 1, \ldots, r$, вид искомой плотности. ▶

32 Убедиться в том, что для случая $\mathcal{L}(\xi) = \mathcal{U}(0, 1)$ распределения порядковых статистик имеют вид (при $1 \leqslant k < l \leqslant n$)

$$\mathcal{L}(X_{(k)}) = \mathrm{Be}(k, n - k + 1), \quad \mathcal{L}(X_{(l)} - X_{(k)}) = \mathrm{Be}(l - k, n - l + k + 1)$$

(см. п. 4 § 1.2). Вывести отсюда формулы для моментов

$$\mathbf{E}X_{(k)} = \frac{k}{n+1}, \quad \mathbf{D}X_{(k)} = \frac{k(n-k+1)}{(n+1)^2(n+2)}, \quad \operatorname{cov}(X_{(k)}, X_{(l)}) = \frac{k(n-l+1)}{(n+1)^2(n+2)}.$$

◀ *Указание.* Используя формулу (44) § 1.2, найти совместную плотность величин $Y_1 \quad X_{(k)}$ и $Y_2 = X_{(l)} - X_{(k)}$, интегрированием которой получить маргинальную плотность величины Y_2. При вычислении моментов использовать формулы (24) § 1.2, а также формулу $\mathbf{D}(X_{(l)} - X_{(k)}) = \mathbf{D}X_{(l)} + \mathbf{D}X_{(k)} - 2\operatorname{cov}(X_{(k)}, X_{(l)})$. ▶

33 (*Продолжение*). Используя эти результаты, получить формулы для моментов экстремумов выборки из произвольного равномерного распределения $\mathcal{U}(a, b)$, приведенные в упр. 46 гл. 1 (см. п. 5 § 1.2).

31 Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ — выборка из распределения с плотностью $f(x) = (1/2)\cos x$, $|x| \leqslant \pi/2$. Вывести формулу

$$\mathbf{E}X_{(n)} = \frac{\pi}{2}\big(1 - 2^{-2n+1}C_{2n}^{n}\big).$$

Доказать, что при $n \to \infty$

$$\mathbf{E}X_{(n)} = \frac{\pi}{2} - \sqrt{\frac{\pi}{n}}\left(1 + O\left(\frac{1}{n}\right)\right)$$

и

$$\mathbf{P}\left\{\frac{\sqrt{n}}{2}\left(X_{(n)} - \frac{\pi}{2}\right) \leqslant x\right\} \to e^{-x^2}, \quad x < 0.$$

◀ *Указание.* Воспользоваться тождеством

$$\sum_k C_n^{2k} C_{2k}^k 2^{-2k} = 2^{-n} C_{2n}^n$$

и также применить критерий (26) § 2.4. ▶

35 Пусть $X = (X_1, \quad, X_n)$ — выборка из распределения Коши (см. (31) § 1.2). Показать, что при $n \to \infty$ и $a_n = \operatorname{tg}(\pi/2 - \pi/n) \sim n/\pi$

$$\mathbf{P}\{a_n^{-1} X_{(n)} \leqslant x\} \longrightarrow e^{-1/x} \quad x > 0.$$

◀ *Указание.* Применить критерий (25) § 2.4 и учесть, что $\lim\limits_{t\to\infty} t(\pi/2 - \operatorname{arctg} t) = 1$ ▶

36 Пусть $X = (X_1, \ldots, X_n)$ — выборка из *усеченного показательного распределения*, задаваемого функцией распределения

$$F(x) = k(1 - e^{-x}), \quad 0 \leqslant x \leqslant x_0, \quad k = (1 - e^{-x_0})^{-1},$$

при некотором конечном $x_0 > 0$. Показать, что при $n \to \infty$

$$\mathbf{P}\{n(e^{x_0} - 1)^{-1}(X_{(n)} - x_0) \leqslant x\} \longrightarrow e^x, \quad x < 0.$$

◀ *Указание.* Применить критерий (26) § 2.4. ▶

Замечание. Обратим внимание на то, что здесь мы имеем предельное распределение типа Λ_2 (см. (24) § 2.4), в отличие от примера 3 в § 2.4, где для обычного показательного распределения (т. е. при $x_0 = \infty$) мы имеем предел типа Λ_3.

37 Пусть функция распределения наблюдаемой случайной величины ξ имеет вид

$$F(x) = \begin{cases} 1 - e^{1/x} & \text{при} \quad x < 0, \\ 1 & \text{при} \quad x \geqslant 0. \end{cases}$$

Доказать, что для распределения максимального значения выборки объема n из $\mathcal{L}(\xi)$ при $n \to \infty$ выполняется соотношение

$$\mathbf{P}\left\{\frac{X_{(n)} - b_n}{a_n} \leqslant x\right\} \longrightarrow e^{-e^{-x}} \quad -\infty < x < \infty,$$

где $a_n = (\ln n)^{-2}$, $b_n = -1/(\ln n)$.

◀ *Указание.* Применить критерий (27) § 2.4. ▶

38 Пусть $X = (X_1, \ldots, X_n)$ — стандартный нормальный вектор (т. е. выборка из стандартного нормального распределения $\mathcal{N}(0, 1)$). Рассмотрим три квадратичные формы от X:

$$Q_1 = \sum_{i=1}^n X_i^2, \quad Q_2 = \frac{1}{n}(X_1 + \ldots + X_n)^2, \quad Q_3 = \sum_{i=1}^n (X_i - \overline{X})^2$$

Применив теорему 1 § 2.5, убедиться в справедливости следующих соотношений:

$$\mathcal{L}(Q_1) = \chi^2(n), \quad \mathcal{L}(Q_2) = \chi^2(1), \quad \mathcal{L}(Q_3) = \chi^2(n-1).$$

◀ *Указание.* Получить представление $Q_1 = X' \mathbb{1}_n X$, $Q_2 = X' B X$, где $B = \|1/n\|_1^n$, $Q_3 = X' A X$, где $A = \mathbb{1}_n - B$. ▶

39 Пусть $\mathcal{L}(Y) = \mathcal{N}(\underline{\mu}, \Sigma)$, $|\Sigma| \neq 0$ и $Q = (Y - \underline{\mu})' A (Y - \underline{\mu})$, где матрица A удовлетворяет условию $A\Sigma A = A$. Доказать, что $\mathcal{L}(Q) = \chi^2(m)$, где $m = \operatorname{tr}(A\Sigma)$. Получить отсюда утверждение теоремы 1 § 2.5 и упражнения 40 к гл. 1.

◀ *Указание.* Запись Y в виде $Y = \underline{\mu} + BX$, где X — стандартный нормальный вектор (см. п. 2 § 1.2), тогда $Q = X' A_1 X$, где матрица $A_1 = B' AB$ идемпотентна. Далее применить теорему 1 § 2.5. ▶

40 (*Продолжение*). Доказать следующее обобщение леммы 3 § 2.5: квадратичная форма $\widetilde{Q} = Y' AY$ имеет распределение $\chi^2(r; \lambda^2)$, где $r = \operatorname{rank} A$ и $\lambda^2 = \underline{\mu}' A \underline{\mu}$. В частности, при $A = \Sigma^{-1}$ величина $r = n$ и $\lambda^2 = \underline{\mu}' \Sigma^{-1} \underline{\mu}$ (лемма 3).

◀ *Указание.* Использовать то же преобразование $Y = BX$, что и в доказательстве леммы 3. Тогда $Q = X' B' ABX$ и здесь матрица $A_1 = B' AB$ идемпотентна и $\operatorname{rank} A_1 = \operatorname{rank} A$. Далее, как и при доказательстве теоремы 1 § 2.5, привести Q к сумме квадратов независимых нормальных величин. ▶

41 Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения $\mathcal{N}(0, 1)$ и квадратичная форма $Q = X'X$ разложена на сумму двух квадратичных форм $Q = Q_1 + Q_2$, где $Q_i = X' A_i X$ и $\operatorname{rank} A_i = n_i$, $i = 1, 2$. Доказать, что если $n_1 + n_2 = n$, то Q_1 и Q_2 независимы и $\mathcal{L}(Q_i) = \chi^2(n_i)$, $i = 1, 2$.

◀ *Указание.* Проверить, что матрицы A_1 и A_2 идемпотентны и $A_1 A_2 = O$ ▶

Замечание. Справедливо более сильное утверждение: если $Q = Q_1 + \ldots + Q_k$, где $Q_i = X' A_i X$, $\operatorname{rank} A_i = n_i$, $i = 1, \ldots, k$, то $n_1 + \ldots + n_k = n \iff Q_1, \ldots, Q_k$ независимы и $\mathcal{L}(Q_i) = \chi^2(n_i)$, $i = 1, \ldots, k$.

42 Пусть $X = (X_1, \ldots, X_n)$ — выборка из нормального распределения $\mathcal{N}(\mu, \sigma^2)$. Показать, что функция распределения статистики $\eta = (X_1 - \overline{X})/(\sqrt{n-1}S)$ имеет вид $F_\eta(u) = 1 - 1/2B\left(1 - u^2; (n-2)/2, 1/2\right)$ при $0 < u < 1$ и $F_\eta(u) = 1 - F_\eta(-u)$ для $-1 < u < 0$ (здесь $B(x; a, b)$ — неполная бета-функция, см. п. 2 § 2.4).

◀ *Указание.* Получить сначала представление

$$\eta = \frac{Y_1}{\sqrt{Y_1^2 + \ldots + Y_{n-1}^2}},$$

где случайные величины $Y_1, \ldots, Y_{n-1}$ независимы и нормальны $\mathcal{N}(0, 1)$. Далее воспользоваться упр. 52 к гл. 1. ▶

43 Пусть имеется выборка объема n из двумерного собственного нормального распределения (см. п. 4 § 2.3) и $(\overline{X}_1, \overline{X}_2)$, (S_1^2, S_{12}, S_2^2) и $\widehat{\rho}_n$ — соответствующие выборочные характеристики.

а) Доказать, что $(\overline{X}_1, \overline{X}_2)$ и (S_1^2, S_{12}, S_2^2) независимы;

б) пусть $Q = n(\overline{X}_1 - \mu_1, \overline{X}_2 - \mu_2)' \Sigma^{-1} (\overline{X}_1 - \mu_1, \overline{X}_2 - \mu_2)$, доказать, что $\mathcal{L}(Q) = \chi^2(2)$;

в) пусть $n > 2$ и $T_n = \sqrt{n-2}\,\widehat{\rho}_n / \left(\sqrt{1 - \widehat{\rho}_n^2}\right)$, доказать, что $\mathcal{L}(T_n) = S(n-2)$.

◀ *Указание.* а) см. замечание к упр. 28 гл. 1; б) см. замечание в п. 2 § 2.5; в) см. замечание в п. 4 § 2.3. ▶

44 Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ — выборка из распределения $\mathcal{L}(\xi)$, $\mathbf{E}\xi = \mu$, $\mathbf{D}\xi = \sigma^2 > 0$, $\mathbf{E}\xi^4 < \infty$ и $\overline{X}, S^2$ — соответствующие выборочные среднее и дисперсия. Доказать, что при $n \to \infty$

$$\mathcal{L}\left(T_n \equiv \sqrt{n}(\overline{X} - \mu)/S\right) \longrightarrow \mathcal{N}(0, 1).$$

◀ *Указание.* Воспользоваться центральной предельной теоремой, сходимостью $S/\sigma \xrightarrow{\mathrm{P}} 1$ и следующим фактом: $\mathcal{L}(\eta_n) \to \mathcal{L}(\eta)$, $\zeta_n \xrightarrow{\mathrm{P}} 1 \Rightarrow \mathcal{L}(\zeta_n \eta_n) \to \mathcal{L}(\eta)$. ▶

Глава 3

Общая теория оценивания неизвестных параметров распределений

Оценивать — значит создавать. Именно оценка придает ценность и драгоценность всем оцененным вещам. Лишь через оценку появляется ценность: и без оценивания был бы пуст орех бытия.

Ф. Ницше [1] Так говорил Заратустра

Настоящая глава посвящена общей теории статистических оценок — одной из основных проблем, решаемых математической статистикой, как это отмечено в общем Введении. Здесь в рамках произвольной параметрической статистической модели излагаются основные подходы к построению эффективных оценок для неизвестных параметров модели и функций от них. Рассматриваются два традиционных подхода к решению этих задач: точечное и интервальное оценивание (более общо — оценивание с помощью доверительных множеств), а также некоторые другие современные концепции оценивания. При точечном оценивании в основном рассматриваются несмещенные оценки, и за меру их точности принимается величина дисперсии оценки. Основное внимание уделяется методам построения оптимальных оценок, при этом рассматриваются случаи как скалярного, так и векторного параметров. Излагаются различные подходы для построения доверительных интервалов и доверительных множеств, в том числе принцип отношения правдоподобия — один из наиболее универсальных современных принципов решения многих статистических задач. Рассматриваются также вопросы оценивания при выборе из конечных совокупностей. При решении рассматриваемых задач используются как точный, так и асимптотический (для больших выборок) подходы, что диктуется, в основном, спецификой конкретной анализируемой ситуации.

§ 3.1. Статистические оценки и общие требования к ним

1. Понятие статистической оценки и ее среднеквадратичная ошибка

Это понятие уже встречалось нам в гл. 2. Там говорилось, что значение эмпирической функции распределения $\widehat{F}_n(x)$ в каждой точке x можно рассматривать

[1] Ницше Фридрих (1844–1900) — немецкий философ, поэт-мыслитель.

в качестве приближенного значения (оценки) для теоретической функции распределения $F(x)$ наблюдаемой в эксперименте случайной величины ξ, а различные выборочные характеристики (моменты $\widehat{\alpha}_k$ и $\widehat{\mu}_k$ (см. (2)–(3) § 2.2), квантили $\zeta_{n,p}$ (см. (2) § 2.4) и т. д.) — как оценки соответствующих теоретических характеристик распределения $\mathcal{L}(\xi)$. При этом использование термина «оценка» обосновывалось тем, что для выборок большого объема значительная разница между значениями выборочной и теоретической характеристик маловероятна, и поэтому разумно (по крайней мере для больших выборок) принять выборочную характеристику за приближенное значение соответствующей теоретической характеристики, когда последняя неизвестна. Таким образом, в этот термин вкладывался определенный асимптотический смысл при $n \to \infty$ (напомним, что n — это объем выборки $X = (X_1, \quad , X_n)$ из распределения $\mathcal{L}(\xi)$). Выражением этого «разумного» свойства оценки является понятие ее *состоятельности* (см. определение 3 в § 2.2) — одного из основных понятий теории оценок.

В то же время при применении теории оценивания на практике обычно приходится строить приближенные значения для различных неизвестных теоретических характеристик изучаемой модели при конечных (и часто весьма умеренных) объемах выборки, и при этом обосновывать соответствующие рекомендации с точки зрения каких-либо критериев оптимальности. Общие методы решения подобных задач развиты в теории оценивания неизвестных параметров распределений, которой и посвящена настоящая глава.

Итак, пусть задана статистическая модель $\mathcal{F} = \{F\}$ для схемы повторных независимых наблюдений над некоторой случайной величиной ξ и $X = (X_1, \quad , X_n)$ есть выборка из распределения $\mathcal{L}(\xi) \in \mathcal{F}$. В дальнейшем рассматриваются различные функции $T = T(X)$ от выборки, сами являющиеся случайными величинами (т. е. для которых при всех t определены вероятности $F_T(t) = \mathbf{P}\{T(X) \leqslant t\}$). Если при этом функция T не зависит от неизвестного распределения наблюдаемой величины ξ, то ее принято называть *статистикой* (см. определение 1 в § 2.2). Часто требуется по выборке X сделать те или иные заключения об истинном значении g_0 некоторой (неизвестной) характеристики $g \quad g(\xi)$ наблюдаемой случайной величины. Прежде всего под этим понимается задача *оценить* это значение g_0, т. е. построить такую статистику $T(X)$, значение которой $t = T(\underline{x})$ при наблюдавшейся реализации $\underline{x}$ выборки X можно было бы считать разумным (в каком-то смысле) приближением для g_0: $t \approx g_0$. В этом случае говорят, что *статистика* $T(X)$ *оценивает* g или, что $T(X)$ *есть оценка* g (для g) и часто подчеркивают это обозначением $\widehat{g} = T(X)$. Так в общем виде формулируется задача *точечного оценивания* неизвестных параметров распределений.

Статистика и оценка

Ясно, что для оценивания заданной характеристики g можно использовать различные оценки и, чтобы выбрать лучшую из них, надо иметь критерий сравнения качества (точности) оценок. В свою очередь, и подобные критерии могут быть разными

Точность оценки

(в зависимости от целей, для которых строится оценка, удобства использования критерия и т. п.), но любой критерий определяется выбором *меры точности* оценок (*меры близости* оценки к истинному значению оцениваемой характеристики). Кроме того, обычно класс рассматриваемых в конкретной задаче оценок ограничивают некоторыми дополнительными требованиями. Таким образом, если определен некоторый класс оценок $\mathcal{T}_g$ (в задаче оценивания g) и выбрана мера точности, то оценка $T \in \mathcal{T}_g$, оптимизирующая эту меру, называется *оптимальной* (в классе $\mathcal{T}_g$).

Итак, для построения теории оптимального оценивания прежде всего надо договориться о мере точности оценок, т. е. уточнить смысл приближенного равенства $T(X) \approx g$. Если статистика $T(X)$ используется для оценивания величины g, то одной из разумных мер расхождения между ними является $(T(X)-g)^2$, или *квадратичная ошибка*. Но так как эта величина случайная, то для измерения точности оценки T используется *среднеквадратичная ошибка* (с. к. о.) $\Delta(T) = \mathrm{E}(T(X)-g)^2$, где E, как и ранее, — знак математического ожидания. С. к. о. — наиболее распространенная и традиционно используемая в прикладной математической статистике мера точности оценок, хотя, в принципе, можно было бы использовать и другие меры точности, например, *среднюю абсолютную ошибку* $\mathrm{E}|T(X)-g|$, однако здесь мы сталкиваемся с серьезными аналитическими трудностями вычисления этой меры в конкретных задачах. Мера $\Delta(T)$ порождает и соответствующий критерий оптимальности оценок: оценка, минимизирующая с. к. о. в заданном классе оценок $\mathcal{T}_g$, называется *оптимальной в среднеквадратичном смысле* (в классе $\mathcal{T}_g$) и обозначается символом T^* (или g^*, чтобы подчеркнуть, что эта оценка для g):

Среднеквадратичная ошибка оценки

Оптимальная оценка

$$T^* = \arg\min_{T \in \mathcal{T}_g} \Delta(T); \tag{1}$$

ее с. к. о. есть

$$\Delta^* = \Delta(T^*) = \mathrm{E}(T^*(X)-g)^2 = \inf_{T \in \mathcal{T}_g} \mathrm{E}(T(X)-g)^2 \tag{2}$$

Таким образом, в общем случае задача построения наилучшей (в среднеквадратичном смысле и в заданном классе оценок $\mathcal{T}_g$) оценки для характеристики g по выборке X сводится к решению экстремальной проблемы (1)–(2).

Ниже эти общие положения конкретизируются в рамках произвольной параметрической модели $\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$ (см. пример 4 Введения). Если модель $\mathcal{F}$ параметрическая, то любая теоретическая характеристика является функцией от параметра модели θ, и в таком случае речь идет об оценивании параметрических функций, для которых будем использовать обозначение $\tau(\theta)$ (тау от тэта).

Подчеркнем, наконец, следующее довольно очевидное, но весьма важное обстоятельство. При оценивании заданной параметрической функции $\tau(\theta)$

Необходимое условие оценивания

необходимо, чтобы область значений соответствующей оценки $T(X)$ совпадала с областью значений $\tau(\theta)$:

$$\{t : t = T(X),\ \underline{x} \in \mathfrak{X}\} = \{\tau : \tau = \tau(\theta),\ \theta \in \Theta\}, \quad (3)$$

так как в противном случае оценка может давать значения вообще не принимаемые оцениваемой величиной (оценка будет лишена практического смысла!). Напомним, что в (3) $\mathfrak{X}$ обозначает выборочное пространство (совокупность всех возможных реализаций выборки X), а Θ — параметрическое множество (совокупность всех допустимых значений параметра модели θ).

2. Несмещенные оценки

Одно из центральных понятий математической статистики — понятие несмещенной оценки уже было введено в § 2.2 (соотношение (22)), где также даны краткие комментарии и соответствующие примеры. Здесь мы обсудим это понятие с общих позиций и более детально.

Любая оценка $T = T(X)$ является случайной величиной, поэтому общим требованием к ней является достаточно высокая степень концентрации ее распределения около истинного значения оцениваемой характеристики. Чем выше степень этой концентрации, тем лучше (точнее) соответствующая оценка. Выражением этого в определенной степени и является условие несмещенности. Дадим его более детализированное (в контексте обсуждаемой проблематики) и общее определение. Но сначала условимся о некоторых дополнительных обозначениях.

В случае параметрической модели $\mathcal{F} = \{F(x; \theta),\ \theta \in \Theta\}$ распределение вероятностей на выборочном пространстве $\mathfrak{X}$, отвечающее параметру θ, обозначается символом $\mathbf{P}_\theta$. Аналогично, индекс θ при символах математического ожидания $\mathbf{E}$, дисперсии $\mathbf{D}$ и т. д. будет означать, что соответствующие величины вычисляются для распределения $\mathbf{P}_\theta$. Например, $\mathbf{E}_\theta T(X)$, $\mathbf{D}_\theta T(X)$ — обозначения соответствующих моментов статистики $T(X)$ в случае, когда функция распределения выборки $\boldsymbol{X} = (X_1, \quad , X_n)$ есть

$$F_X(\underline{x}; \theta) = F(x_1, \theta) \quad F(x_n, \theta)$$

(или плотность $f_X(\underline{x}; \theta) = f(x_1, \theta) \quad f(x_n; \theta)$). При этом для единообразия иногда будет использоваться сокращенная запись через интеграл Стилтьеса [2]:

$$\mathbf{E}_\theta T(X) = \int T(\underline{x})\, dF_X(\underline{x}; \theta).$$

Здесь интеграл понимается как n-мерный интеграл по всему выборочному пространству $\mathfrak{X}$, который в случае абсолютно непрерывной модели вычисляется как обычный интеграл Римана

$$\int T(x_1, \quad , x_n) f(x_1; \theta) \quad f(x_n; \theta)\, dx_1 \quad dx_n,$$

[2] Стилтьес Томас Ян (мл.) (1856–1894) — голландский математик и астроном.

а в случае дискретной модели — как соответствующая сумма

$$\Sigma T(x_1, \ldots, x_n) f(x_1; \theta) \qquad f(x_n; \theta).$$

Этого уже достаточно, чтобы перейти к точным формулировкам.

Определение Статистика $T = T(X)$ называется *несмещенной в среднем* (или просто *несмещенной*) *оценкой* для заданной параметрической функции $\tau(\theta)$, если она удовлетворяет условию

Условие несмещенности

$$\mathbf{E}_\theta T(X) = \tau(\theta), \quad \forall\, \theta \in \Theta. \tag{4}$$

Об условии (4) говорят как об *условии несмещенности*. Если оно не выполняется, то величина

$$b(\theta) = \mathbf{E}_\theta T(X) - \tau(\theta)$$

называется *смещением* оценки $T(X)$. Таким образом, можно также считать, что несмещенные оценки — это такие оценки, для которых смещение $b(\theta) = 0$, $\forall\, \theta \in \Theta$. В общем случае среднеквадратичная ошибка

$$\Delta(T; \theta) = \mathbf{E}_\theta (T(X) - \tau(\theta))^2$$

оценки $T = T(X)$ выражается через дисперсию оценки и ее смещение следующим образом

$$\Delta(T; \theta) = \mathbf{D}_\theta T + b^2(\theta). \tag{5}$$

Следовательно, для несмещенных оценок с. к. о. совпадает с дисперсией оценки.

В конкретных задачах часто ограничиваются рассмотрением лишь класса $\mathcal{T}_\tau$ несмещенных оценок для $\tau = \tau(\theta)$:

$$\mathcal{T}_\tau = \{T(X) \quad \mathbf{E}_\theta T(X) = \tau(\theta),\ \forall\, \theta \in \Theta\}.$$

Функция $\tau(\theta)$ называется *оцениваемой*, если класс $\mathcal{T}_\tau$ не пуст, т. е. если существует хотя бы одна несмещенная оценка для $\tau(\theta)$. Про любую несмещенную оценку можно сказать, что «в среднем» она приводит к желаемому результату. К тому же, как это будет показано далее, для несмещенных оценок можно построить достаточно простую и практически полезную теорию, в которой критерием точности оценки является ее дисперсия.

Оцениваемые функции



Несколько слов о философии «среднего». В пользу «среднего» можно привести следующие рассуждения из книги венгерского математика Габора Секея «Парадоксы в теории вероятностей и математической статистике» [23]. «Уравнивание противоположных значений и отклонений в „среднем", т. е. суммирование наблюдений в одно значение, имеет давние традиции. Эсхил писал в трагедии „Эвмениды": „Богу всегда середина любезна, и меру чтит божество", а последователи китайского философа Конфуция говорят, что „в неподвижности среднего (≡ Чжун Йюн) есть величайшее совершенство" Математически понятие „среднего" можно интерпретировать различными способами (арифметическое среднее, геометрическое среднее, медиана и т. д.). Однако в статистических применениях в течение долгого времени крайне важную роль играло арифметическое

среднее. Уже в первых выдающихся результатах в теории вероятностей и математической статистике изучалось арифметическое среднее статистической выборки и росло понимание важности его использования».

Однако условие несмещенности в среднем не следует абсолютизировать: в некоторых случаях требование несмещенности (4) может оказаться слишком «жестким» и привести к нежелательным результатам. Так, может оказаться, что несмещенные оценки (в данной модели и для данной параметрической функции) вообще не существуют (т. е. класс $\mathcal{T}_\tau = \varnothing$ — пустое множество), а в других случаях, хотя и имеются, но практически бесполезны. Ниже эти общие положения иллюстрируются примерами. Но предварительно отметим, что существует также и критическая философия «среднего». Так имеет место (и некоторыми активно используется) такой прием искажения истины, как использование «средних цифр» (как заметил Ленин, «нет ничего глупее так называемых „средних цифр“»), когда понятие среднего используется как синоним типичного. Примерами таких подмен могут служить средний доход населения, средняя температура больных в некоторой больнице, среднее количество осадков за несколько сезонов в некотором регионе (в области N в прошлом году была летом засуха, и все поля выгорели, а этим летом прошли ливни, которые все смыли с полей, но в среднем количество осадков за два сезона — в пределах нормы, или: в некоторой стране есть всего лишь несколько очень богатых семей и большое количество бедных, чьи доходы соответственно огромны и малы, поэтому здесь средний доход вовсе нетипичен). Таким образом, средние характеристики часто вводят в заблуждение. Как отмечается в [23], «средний человек — одна из них». Неудивительно, что исследования бельгийского ученого Л. А. Ж. Кетле по этому вопросу стали источником горячих споров. Худшее в «среднем человеке» не его серость, а возникающие противоречия. Например, средний рост не соответствует среднему весу и т. д. Только по одной этой причине можно усомниться в справедливости слов Дж. Рейнольдса (первого президента Королевской академии художеств в Лондоне), когда он сказал, что «в среднем источник прекрасного».

Перейдем к обещанным примерам, оттеняющим различные нюансы несмещенности в среднем.

Что примера лучше действует?
Что людьми сильней ворочает?

А. С. Пушкин. Бова (1814)

Пример 1 (Несмещенное оценивание в схеме Бернулли). Пусть

$$\underline{X} = (X_1, \quad , X_n)$$

— выборка из распределения Бернулли $\mathrm{Bi}(1, \theta)$, т. е. X_i принимает значения 0 и 1 с вероятностями $f(x;\theta) = \theta^x(1-\theta)^{1-x} \quad x \quad 0, 1$ (см. п. 1 § 1.1). Опишем класс оцениваемых функций в рассматриваемой модели и найдем соответствующие несмещенные оценки. Для этого запишем для произвольной статистики $T(\underline{X})$ ее математическое ожидание:

$$\mathbf{E}_\theta T(\underline{X}) = \sum_{\underline{x}=(x_1,\ldots,x_n)} T(\underline{x}) f(x_1;\theta) \quad f(x_n;\theta) =$$

$$\sum_{\underline{x}} T(\underline{x})\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i} = \sum_{r=0}^{n} \theta^r (1-\theta)^{n-} \sum_{\underline{x}:\,\Sigma x_i = r} T(\underline{x}).$$

Здесь правая часть представляет собой полином от θ степени не выше n, следовательно, в данной модели несмещенные оценки можно строить лишь для параметрических функций $\tau(\theta)$, являющихся полиномами степени не выше n — числа наблюдений. Отсюда, в частности, следует, что для функций вида $\tau(\theta) = \theta^{-a}$ $a > 0$ и $\tau(\theta) = \theta^b$ $b > n$, несмещенных оценок по выборке объема n не существует.

Пусть теперь

$$\tau(\theta) = \sum_{j=0}^{n} a_j \theta^j$$

— произвольный полином степени n (или ниже, если старшие коэффициенты a_j равны нулю). Построим сначала несмещенную оценку для отдельной степени θ^j при $j \leqslant n$. Из упр. 19 к гл. 1 имеем, что если случайная величина ν имеет биномиальное распределение $\mathrm{Bi}(n, p)$, то для ее j-го факториального момента при $j \leqslant n$ справедлива формула $\mathbf{E}(\nu)_j = (n)_j p^j$ В нашем случае статистика $X = X_1 + \quad + X_n$ имеет распределение $\mathrm{Bi}(n, \theta)$ (см. п. 1 § 1.1), следовательно, $\mathbf{E}_\theta (X)_j = (n)_j \theta^j$. Это означает, что статистика $(X)_j/(n)_j$ является несмещенной оценкой для θ^j при любом $j \leqslant n$. Отсюда, в свою очередь, имеем, что несмещенной оценкой полинома $\tau(\theta)$ является статистика

$$\widehat{\tau} = T(\underline{X}) = \sum_{j=0}^{n} \frac{a_j (X)_j}{(n)_j}. \tag{6}$$

Итак, в бернуллиевской модели $\mathrm{Bi}(1, \theta)$ класс оцениваемых параметрических функций по n наблюдениям совпадает с классом полиномов степени не выше n, и для любой такой функции $\tau(\theta)$ ее несмещенная оценка $\widehat{\tau}$ дается формулой (6). •

Отсутствие несмещенной оценки в пуассоновской модели

Пример 2. Пусть $\mathcal{L}(\xi) \in \Pi(\theta)$ — семейство пуассоновских распределений (см. п. 3 § 1.1) и $n = 1$, т. е. произведено одно наблюдение X над пуассоновской случайной величиной. Требуется оценить параметрическую функцию $\tau(\theta) = 1/\theta$.

Если $T(X)$ — несмещенная оценка $\tau(\theta)$, то условие (4) принимает вид

$$\sum_{x=0}^{\infty} T(x) e^{-\theta} \frac{\theta^x}{x!} = \frac{1}{\theta}, \quad \forall\, \theta > 0,$$

или

$$\sum_{x=0}^{\infty} T(x) \frac{\theta^{x+1}}{x!} = e^{\theta} = \sum_{r=0}^{\infty} \frac{\theta^r}{r!}, \quad \forall\, \theta > 0.$$

Очевидно, что функция $T(x)$, удовлетворяющая последнему условию и независящая от θ, как это требуется в соответствии с определением понятия оценки, не существует, т. е. в данном случае несмещенных оценок для $\tau(\theta)$ вообще нет. •

Несмещенные оценки в отрицательной биномиальной модели

Пример 3. Пусть $\mathcal{L}(\xi) \in \overline{\mathrm{Bi}}(r, \theta)$ — семейство отрицательных биномиальных распределений (см. п. 2 § 1.1) и $n = 1$. Построим несмещенную оценку для параметрической функции $\tau_s(\theta) = \theta^s$ где $s \geqslant 1$ — целое число. Здесь условие несмещенности (4) принимает вид

$$\sum_{x=0}^{\infty} T(x) f(x|r, \theta) = \tau_s(\theta), \quad \forall\, \theta \in (0, 1),$$

или

$$\sum_{x=0}^{\infty} T(x) C_{r+x-1}^{x} \theta^x = \frac{\theta^s}{(1-\theta)^r} = \sum_{j=0}^{\infty} C_{r+j-1}^{j} \theta^{s+j}, \quad \forall\, \theta \in (0, 1).$$

Но из тождественного равенства двух степенных рядов следует равенство их коэффициентов при соответствующих степенях θ. Следовательно, функция $T(x)$ здесь однозначно определена, и она есть

$$T_{r,s}(x) = \frac{C_{r+x-s-1}^{x-s}}{C_{r+x-1}^{x}} = \begin{cases} 0 & \text{при} \quad x < s, \\ \displaystyle\prod_{j=0}^{s-1} \frac{x-j}{x+r-j-1} & \text{при} \quad x \geqslant s \end{cases}$$

(здесь учтено, что при $b < 0$ $C_a^b = 0$).

Таким образом, несмещенная оценка для $\tau_s(\theta)$ существует, единственна и имеет вид

$$\widehat{\tau}_s = T_{r,s}(X) = \prod_{j=0}^{s-1} \frac{X-j}{X+r-j-1} \quad I(X \geqslant s). \tag{7}$$

Обратим внимание на то, что эта оценка может принимать нулевое значение (при $X < s$), которое не принадлежит области значений функции $\tau_s(\theta)$ (ведь $0 < \tau_s(\theta) < 1$ при $0 < \theta < 1$), тем самым для нее не выполняется условие (3). Если же $r = 1$, то оценка $T_{1,s}(X)$ вообще принимает лишь два значения:

$$T_{1,s}(X) = \begin{cases} 0, & X < s, \\ 1, & X \geqslant s, \end{cases}$$

не принадлежащие области значений функции $\tau_s(\theta)$, т. е. такая оценка лишена какого-либо практического смысла. Этот пример показывает, что иногда принцип несмещенности может привести к бесполезному результату. •

Пример 4 (Общая нормальная модель, оценивание дисперсии). Пусть

$$X = (X_1, \ldots, X_n), \quad n \geqslant 2,$$

— выборка из нормального распределения $\mathcal{N}(\theta_1, \theta_2^2)$ (оба параметра неизвестны). Требуется оценить параметрическую функцию $\tau(\theta) = \tau(\theta_1, \theta_2) = \theta_2^2$. Мы уже знаем (см. теорему 2 § 2.2), что несмещенной оценкой теоретической дисперсии является (в любой модели!) статистика

$$S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

при этом (см. замечание к теореме Фишера в § 2.5, соотношение (7))

$$\mathbf{D}_\theta S_0^2 = \mathbf{D}\left(\frac{n}{n-1}S^2\right) = \frac{2\theta_2^4}{n-1}. \tag{8}$$

Ниже (см. § 3.2, пример 9) будет показано, что по критерию минимума дисперсии оценка S_0^2 является оптимальной среди всех несмещенных оценок функции $\tau(\theta) = \theta_2^2$ в модели $\mathcal{N}(\theta_1, \theta_2^2)$, т. е. ее дисперсия (8) меньше дисперсии любой другой несмещенной оценки величины θ_2^2.

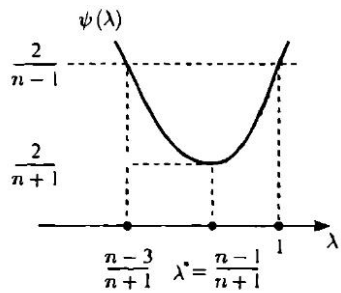


Рис. 1

Рассмотрим теперь класс оценок вида $T_\lambda = \lambda S_0^2$. Так как

$$\mathbf{E}_\theta T_\lambda = \lambda \mathbf{E}_\theta S_0^2 = \lambda\theta_2^2,$$

то в этом классе имеется лишь единственная несмещенная оценка $\tau(\theta)$, соответствующая значению $\lambda = 1$, т. е. оценка S_0^2.

Вычислим среднеквадратичную ошибку произвольной оценки T_λ:

$$\Delta(T_\lambda; \theta) = \mathbf{E}_\theta(T_\lambda - \theta_2^2)^2 = \mathbf{E}_\theta\left[\lambda(S_0^2 - \theta_2^2) + (\lambda - 1)\theta_2^2\right]^2 = \lambda^2\mathbf{D}_\theta S_0^2 + (\lambda-1)^2\theta_2^4.$$

Отсюда из формулы (8) получим

$$\Delta(T_\lambda; \theta) = \psi(\lambda)\theta_2^4, \quad \psi(\lambda) = \frac{2\lambda^2}{n-1} + (\lambda - 1)^2$$

Здесь функция $\psi(\lambda)$ есть парабола, график которой изображен на рис. 1. Функция $\psi(\lambda)$ достигает минимума при $\lambda^* = (n-1)/(n+1)$ и $\psi(\lambda^*) = 2/(n+1)$. Отсюда, учитывая (8), имеем

$$\Delta(T_{\lambda^*}; \theta) = \frac{2\theta_2^4}{n+1} < \frac{2\theta_2^4}{n-1} = \Delta(S_0^2; \theta).$$

Таким образом, по критерию минимума с. к. о. оптимальной в классе $\mathcal{T} = \{T_\lambda\}$ является оценка

$$T_{\lambda^*} = \lambda^* S_0^2 = \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \overline{X})^2. \tag{9}$$

Эта оценка смещенная, но она точнее оценивает теоретическую дисперсию θ_2^2 модели $\mathcal{N}(\theta_1, \theta_2^2)$, чем несмещенная оценка S_0^2. Найдем еще смещение $b_n(\theta)$ оценки T_{λ^*}:

$$b_n(\theta) = \mathbf{E}T_{\lambda^*} - \theta_2^2 = (\lambda^* - 1)\theta_2^2 = -\frac{2\theta_2^2}{n+1}.$$

Это смещение мало при большом объеме выборки n и $b_n(\theta) \to 0$ при $n \to \infty$, т. е. T_{λ^*} — асимптотически несмещенная оценка.

Отметим дополнительно, что при любом $\lambda \in \big((n-3)/(n+1), 1\big)$ оценка T_λ имеет меньшую с. к. о., нежели $T_1 = S_0^2$, но в классе несмещенных оценок S_0^2 — оптимальная оценка.

Этот пример показывает, что не может быть единственного критерия, по которому следует сравнивать все оценки, как не существует единственной оценки данной теоретической характеристики, подходящей для всех случаев.•

Статистические задачи не имеют однозначных ответов.

Э. Леман

3. Оптимальные оценки

Требование несмещенности в силу сделанных выше замечаний нельзя рассматривать как универсальное. Тем не менее во многих встречающихся на практике случаях оно уместно и обоснованно и далее мы будем рассматривать в основном именно несмещенные оценки.

Итак, пусть требуется оценить заданную параметрическую функцию $\tau = \tau(\theta)$ в модели $\mathcal{F} = \{F(x;\theta), \theta \in \Theta\}$ по соответствующей выборке $X = (X_1, \ldots, X_n)$. Обозначим $\mathcal{T}_\tau$ класс всех несмещенных оценок $T = T(X)$ для $\tau(\theta)$ и предположим, что он не пуст (что это не всегда имеет место, говорит нам пример 2 выше). Дополнительно предположим, что дисперсии всех оценок из класса $\mathcal{T}_\tau$ конечны: $\mathbf{D}_\theta T = \mathbf{E}_\theta(T - \tau(\theta))^2 < \infty$, $\forall T \in \mathcal{T}_\tau$ и $\forall \theta \in \Theta$. В этом случае, как отмечено выше (см. соотношение (5)), мерой точности оценок является их дисперсия.

Замечание. Напомним, что в теории вероятностей для любой случайной величины ξ с конечной дисперсией мерой «разброса» ее значений вокруг среднего значения $\mu = \mathbf{E}\xi$ служит *стандартное отклонение* $\sigma = \sqrt{\mathrm{D}\xi}$, а знаменитое *правило «трех сигм»* гласит, что значение ξ попадает в отрезок $[\mu - 3\sigma, \mu + 3\sigma]$ с вероятностью, мало отличающейся от 1. Так, для нормальной $\mathcal{N}(\mu, \sigma^2)$ величины

$$\mathrm{P}\{\mu - 3\sigma \leqslant \xi \leqslant \mu + 3\sigma\} = \mathrm{P}\left\{\left|\frac{\xi - \mu}{\sigma}\right| \leqslant 3\right\} = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 \approx 0{,}997,$$

т. е. ξ попадает в отрезок $[\mu \mp 3\sigma]$ с вероятностью, лишь немного отличающейся от 1.

Итак, пусть T^* и T — оценки из класса $\mathcal{T}_\tau$. Если

$$\mathbf{D}_\theta T^* \leqslant \mathbf{D}_\theta T, \quad \forall \theta \in \Theta, \tag{10}$$

со строгим неравенством хотя бы при одном θ, то по критерию минимума дисперсии следует отдать предпочтение T^*, как более точной оценке. Если условие (10) выполняется для любой оценки $T \in \mathcal{T}_\tau$, то T^* называют *несмещенной оценкой с равномерно минимальной дисперсией* (сокращенно, н. о. р. м. д.). Такую оценку T^* (когда она существует) будем для краткости называть также *оптимальной* и иногда обозначать τ^* чтобы подчеркнуть, что она относится к функции $\tau(\theta)$.

Условие оптимальности несмещенной оценки

Итак, оптимальной (среди несмещенных) является оценка $\tau^* \in \mathcal{T}_\tau$, для которой выполняется условие

$$\mathbf{D}_\theta \tau^* = \inf_{T \in \mathcal{T}_\tau} \mathbf{D}_\theta T, \quad \forall\, \theta \in \Theta. \tag{11}$$

Требование равномерной (по параметру θ) минимальности дисперсии весьма сильное и не всегда имеет место. Может оказаться, что из двух несмещенных оценок T_1 и T_2 дисперсия $\mathbf{D}_\theta T_1$ минимальна (в классе $\mathcal{T}_\tau$) для одних значений параметра θ, а дисперсия $\mathbf{D}_\theta T_2$ — для других значений θ. В таких случаях с помощью одного критерия минимума дисперсии эти оценки сравнить нельзя. Однако это требование выделяет оптимальную оценку в классе $\mathcal{T}_\tau$ однозначно, если такая оценка существует, о чем говорит следующая теорема.

Теорема 1. Пусть $T_i = T_i(X)$, $i = 1, 2$, — две оптимальные оценки для $\tau = \tau(\theta)$. Тогда $T_1 = T_2$ [3)].

Доказательство. Рассмотрим новую оценку $T_3 = (T_1 + T_2)/2$. Ясно, что $T_3 \in \mathcal{T}_\tau$ и

$$\mathbf{D}_\theta T_3 = \frac{1}{4}\big(\mathbf{D}_\theta T_1 + \mathbf{D}_\theta T_2 + 2\,\mathrm{cov}_\theta\,(T_1, T_2)\big). \tag{12}$$

Воспользуемся известным в теории вероятностей неравенством Коши—Буняковского: для любых случайных величин η_1 и η_2 с конечными дисперсиями

$$\big|\mathrm{cov}(\eta_1, \eta_2)\big| \leqslant \sqrt{\mathbf{D}\eta_1 \mathbf{D}\eta_2},$$

причем знак равенства имеет место только если η_1 и η_2 линейно связаны. Отсюда и из (12), положив $\mathbf{D}_\theta T_1 = \mathbf{D}_\theta T_2 = v = v(\theta)$, получим

$$\mathbf{D}_\theta T_3 \leqslant \frac{1}{2}\big(v + \big|\mathrm{cov}_\theta\,(T_1, T_2)\big|\big) \leqslant v. \tag{13}$$

Поскольку T_i, $i = 1, 2$, — оптимальные оценки, то $v = \mathbf{D}_\theta T_i \leqslant \mathbf{D}_\theta T_3$, откуда $\mathbf{D}_\theta T_3 = v$, т. е. T_3 — также оптимальная оценка. Но так как в неравенствах (13) имеют место знаки равенства, то $\mathrm{cov}_\theta\,(T_1, T_2) \geqslant 0$ и, более того, $\mathrm{cov}_\theta\,(T_1, T_2) = \mathbf{D}_\theta T_1 = v$. Следовательно, T_1 и T_2 линейно связаны,

[3)] Здесь и далее равенство статистик понимается в том смысле, что

$$\mathbf{P}_\theta\{X \in \{\underline{x} : T_1(\underline{x}) \neq T_2(\underline{x})\}\} = 0, \quad \forall\, \theta \in \Theta.$$

т. е. $T_1 = kT_2 + a$. Из условия несмещенности оценок имеем $\tau = k\tau + a$, т. е. $a = (1-k)\tau$, и следовательно, $T_1 - \tau = k(T_2 - \tau)$. Здесь коэффициент $k = k(\theta)$ определяется цепочкой равенств

$$v = \operatorname{cov}_\theta(T_1, T_2) = \mathbf{E}_\theta(T_1 - \tau)(T_2 - \tau) = k\mathbf{E}_\theta(T_2 - \tau)^2 = k\mathbf{D}_\theta T_2 = kv.$$

Отсюда имеем $k \equiv 1$, следовательно, $T_1 = T_2$. ■

Приведем пример существования оптимальной оценки в конкретной модели.

Пример 5 (Бернуллиевская модель, оценивание параметра). Пусть

$$X = (X_1, \quad , X_n)$$

— выборка из $\mathcal{L}(\xi) \in \mathrm{Bi}(1, \theta)$ и требуется оценить параметр θ.

Здесь $\mathbf{E}_\theta X_i = \theta$, поэтому выборочное среднее $\overline{X}$ является несмещенной и состоятельной оценкой θ (см. теорему 2 § 2.2). Однако $\overline{X}$ не единственная несмещенная оценка θ. Например, всякая статистика

$$T = \frac{1}{n}\sum_{i=1}^{n} b_i X_i \quad \text{при} \quad b_1 + \quad + b_n = n$$

также является несмещенной оценкой θ. При этом, так как (см. (2) § 1.1)

$$\mathbf{D}_\theta T = \frac{1}{n^2}\sum_{i=1}^{n} b_i^2 \theta(1-\theta) \leqslant \frac{b^2}{n}\theta(1-\theta) \leqslant \frac{b^2}{4n} \longrightarrow 0 \quad \text{при} \quad n \to \infty$$

(здесь $b = \max\limits_i |b_i|$), то оценка T также и состоятельна. Таким образом, такие оценки так же «хороши», как и $\overline{X}$.

Оптимальность выборочного среднего в бернуллиевской модели

Итак, в данной задаче класс $\mathcal{T}_\tau$ содержит много оценок и потому возникает вопрос о выборе среди них наилучшей. Покажем, что в данном случае оптимальная оценка T^* существует и при этом $T^* = \overline{X}$. Поскольку $\mathbf{D}_\theta \overline{X} = \theta(1-\theta)/n$, то в соответствии с определением (11) достаточно показать, что для любой оценки $T \in \mathcal{T}_\tau$ выполняется неравенство

$$\mathbf{D}_\theta T \geqslant \frac{\theta(1-\theta)}{n}, \quad \forall\, \theta \in (0, 1). \tag{14}$$

В данном случае распределение выборки задается плотностью (см. пример 1)

$$L(\underline{x}; \theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}, \quad \underline{x} = (x_1, \ldots, x_n). \tag{15}$$

Так как

$$1 \equiv \sum_{\underline{x}} L(\underline{x}; \theta) \quad \text{и} \quad \theta \equiv \mathbf{E}_\theta T(X) = \sum_{\underline{x}} T(\underline{x}) L(\underline{x}; \theta),$$

то, дифференцируя эти тождества по θ, получим

$$0 = \sum_{\underline{x}} \frac{\partial L(\underline{x};\theta)}{\partial\theta} = \sum_{\underline{x}} \frac{\partial \ln L(\underline{x};\theta)}{\partial\theta} L(\underline{x};\theta) = \mathbf{E}_\theta\left(\frac{\partial \ln L(X;\theta)}{\partial\theta}\right),$$

$$1 = \sum_{\underline{x}} T(\underline{x}) \frac{\partial \ln L(\underline{x};\theta)}{\partial\theta} L(\underline{x};\theta) = \mathbf{E}_\theta\left(T(X)\frac{\partial \ln L(X;\theta)}{\partial\theta}\right).$$

Отсюда можно записать

$$1 = \mathbf{E}_\theta\left[(T(X)-\theta)\frac{\partial \ln L(X;\theta)}{\partial\theta}\right]$$

и, согласно неравенству Коши—Буняковского,

$$1 \leqslant \mathbf{E}_\theta(T(X)-\theta)^2 \mathbf{E}_\theta\left(\frac{\partial \ln L(X;\theta)}{\partial\theta}\right)^2$$

Но $\mathbf{E}_\theta(T(X)-\theta)^2 = \mathbf{D}_\theta T$, поэтому из последнего неравенства следует, что

$$\mathbf{D}_\theta T \geqslant \frac{1}{\mathbf{E}_\theta\left(\frac{\partial \ln L(X;\theta)}{\partial\theta}\right)^2}, \quad \forall\,\theta \in (0,1). \tag{16}$$

В рассматриваемом случае (см. (15))

$$\frac{\partial \ln L(X;\theta)}{\partial\theta} = \frac{1}{\theta}\sum_{i=1}^n x_i - \frac{1}{1-\theta}\left(n - \sum_{i=1}^n x_i\right) = \frac{1}{\theta(1-\theta)}\sum_{i=1}^n (x_i-\theta),$$

поэтому

$$\mathbf{E}_\theta\left(\frac{\partial \ln L(X,\theta)}{\partial\theta}\right)^2 = \frac{1}{\theta^2(1-\theta)^2}\mathbf{E}_\theta\left[\sum_{i=1}^n (X_i-\theta)\right]^2$$

$$= \frac{n}{\theta^2(1-\theta)^2}\mathbf{D}_\theta X_1 = \frac{n}{\theta(1-\theta)}. \tag{17}$$

Отсюда и из неравенства (16) получаем соотношение (14).

Учитывая важность для приложений бернуллиевской модели $\mathrm{Bi}(1,\theta)$, сформулируем полученный результат в виде теоремы. •

Теорема 2. Относительная частота произвольного события в n независимых испытаниях является оптимальной несмещенной оценкой для вероятности этого события.

Как следствие этой теоремы отметим, что значение эмпирической функции распределения $\widehat{F}_n(x)$ в каждой точке x является оптимальной несмещенной оценкой для значения в этой точке теоретической функции распределения $F(x)$.



Приведем следующую юмористическую историю, принадлежащую Д. Пойа [20] и показывающую, как не следует интерпретировать частотную концепцию вероятности. Д-р Тел (доктор телепатии), закончив осмотр пациента, покачал головой. «Вы очень серьезно больны, — сказал он, — из десяти человек с такой болезнью выживает только один». Пациента эта информация изрядно испугала, и д-р Тел начал его успокаивать: «Но Вам очень повезло, что Вы пришли ко мне, сэр. У меня уже умерли от этой болезни девять пациентов, так что Вы выживете».

Установим следующее важное свойство оптимальных оценок.

Теорема 3. Пусть T_1^* и T_2^* — оптимальные оценки функций $\tau_1 = \tau_1(\theta)$ и $\tau_2 = \tau_2(\theta)$ соответственно. Тогда статистика $T^* = a_1T_1^* + a_2T_2^*$ является оптимальной оценкой функции $\tau = a_1\tau_1 + a_2\tau_2$ для любых постоянных a_1 и a_2.

Доказательство. Установим сначала следующее свойство оптимальных оценок, представляющее самостоятельный интерес: для любой статистики $\psi = \psi(X)$ с $\mathbf{E}_\theta\psi = 0$, $\forall\,\theta \in \Theta$ (такая статистика $\psi(X)$ называется *несмещенной оценкой нуля*), выполняется равенство $\operatorname{cov}_\theta(T_1^*, \psi) = 0$, $\forall\,\theta \in \Theta$, т. е. *оптимальная оценка некоррелирована с любой несмещенной оценкой нуля.*

Для доказательства рассмотрим статистику $\widetilde{T}_1 = T_1^* + \lambda\psi$. При любом λ это несмещенная оценка τ_1, поэтому в силу оптимальности оценки T_1^*

$$\mathbf{D}_\theta\widetilde{T}_1 = \mathbf{D}_\theta T_1^* + \lambda^2\mathbf{D}_\theta\psi + 2\lambda\operatorname{cov}_\theta(T_1^*, \psi) \geqslant \mathbf{D}_\theta T_1^*, \quad \forall\,\theta \in \Theta. \tag{18}$$

Отсюда следует, что $\operatorname{cov}_\theta(T_1^*, \psi) = 0$, так как в противном случае при $\lambda = \lambda_0 = -\operatorname{cov}_\theta(T_1^*, \psi)/\mathbf{D}_\theta\psi$ левая часть (18) была бы равна $\mathbf{D}_\theta T_1^* - \operatorname{cov}_\theta^2(T_1^*, \psi)/\mathbf{D}_\theta\psi$, что меньше, чем $\mathbf{D}_\theta T_1^*$, при $\operatorname{cov}_\theta(T_1^*, \psi) \neq 0$ (см. рис. 2).

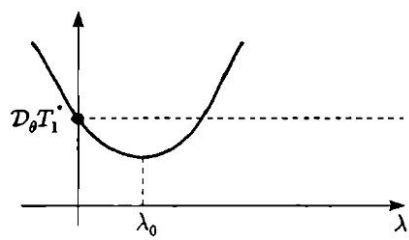


Рис. 2

Перейдем непосредственно к доказательству теоремы. Пусть T — произвольная несмещенная оценка τ. Тогда $\psi = T^* - T$ будет несмещенной оценкой нуля и, по предыдущему, $\operatorname{cov}_\theta(T_i^*, \psi) = 0$, $i = 1, 2$. Отсюда

$$0 = a_1\operatorname{cov}_\theta(T_1^*, \psi) + a_2\operatorname{cov}_\theta(T_2, \psi) = \operatorname{cov}_\theta(T^*, \psi) = \mathbf{D}_\theta T^* - \operatorname{cov}_\theta(T^*, T),$$

т. е.

$$\mathbf{D}_\theta T^* = \operatorname{cov}_\theta(T^*, T) \leqslant \sqrt{\mathbf{D}_\theta T^* \cdot \mathbf{D}_\theta T} \quad \text{или} \quad \mathbf{D}_\theta T^* \leqslant \mathbf{D}_\theta T, \quad \forall\,\theta. \ \blacksquare$$

Таким образом, в конкретных задачах, когда оцениваемая параметрическая функция $\tau(\theta)$ раскладывается в сумму более простых функций $\tau_i(\theta)$, задачу построения оптимальной несмещенной оценки для $\tau(\theta)$ можно сводить к аналогичной задаче для ее слагаемых $\tau_i(\theta)$.

В заключение этого параграфа рассмотрим один конкретный пример применения полученных результатов.

Оптимальная оценка вероятности «успеха» в схеме Бернулли

Пример 6. Рассмотрим ситуацию, описанную в примере 1 Введения, и предположим, что в эксперименте реализуется схема выбора шаров с возвращением. Тогда, как показано в примере 1 § 1.1, мы имеем дело со схемой (моделью) Бернулли $\mathrm{Bi}(1, \theta)$, где неизвестная вероятность «успеха» θ a_1/a — доля белых шаров в урне. По теореме 2 доля X/n наблюдавшихся белых шаров при n извлечениях (относительная частота «успеха» $\equiv$ извлечение белого шара) является оптимальной несмещенной оценкой θ: $\theta^* = X/n$. Таким образом, предложенная в примере 1 Введения интуитивная оценка на самом деле оказывается наилучшей с рассматриваемых позиций несмещенности и оптимальности. •

Перейдем теперь к общим методам построения оптимальных несмещенных оценок.

§ 3.2. Критерии оптимальности оценок, основанные на неравенстве Рао—Крамера. Эффективные оценки

Чтобы продвинуться вперед, нам необходимо ввести несколько новых важных понятий, которыми оперирует теория оптимальных оценок. С этого мы и начнем.

1. Функция правдоподобия, вклад выборки, функция информации

Пусть, как обычно, $f(x;\theta)$ — плотность наблюдаемой случайной величины ξ, $X = (X_1, \ldots, X_n)$ — выборка из $\mathcal{L}(\xi) \in \mathcal{F} = \{F(x;\theta), \theta \in \Theta\}$ и $\underline{x} = (x_1, \ldots, x_n)$ — реализация X. Функция $L(\underline{x};\theta) = f(x_1;\theta) \quad f(x_n;\theta)$ является, очевидно, плотностью случайного вектора X. Рассматриваемая при фиксированном $\underline{x}$ как функция параметра θ $L(\underline{x};\theta)$ называется *функцией правдоподобия* (иногда об $L(\underline{x};\theta)$ говорят как о *правдоподобии данных*).

В дальнейшем предполагается, что $L(\underline{x};\theta) > 0$ при всех $\underline{x} \in \mathfrak{X}$ и $\theta \in \Theta$, и дифференцируема по параметру θ. Пусть сначала параметр θ — скалярный. Случайная величина

Вклад выборки

$$V(X;\theta) = \frac{\partial \ln L(X;\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial \theta} \tag{1}$$

называется *вкладом выборки*, а i-е слагаемое в правой части (1) называется *вкладом i-го наблюдения*, $i = 1, \quad , n$. В дальнейшем предполагается, что

$$0 < \mathbf{E}_\theta V^2(X;\theta) < \infty, \quad \forall \theta \in \Theta. \tag{2}$$

В последующем изложении нам неоднократно придется дифференцировать по θ интегралы типа

$$\int T(\underline{x})L(\underline{x};\theta)\,d\underline{x}$$

(интегрирование ведется по всему выборочному пространству $\mathfrak{X}$) и предполагать при этом, что можно менять порядок интегрирования и диффернцирования. Модели, для которых все перечисленные условия выполняются, называются кратко *регулярными*. Отметим, в частности, общее необходимое условие регулярности, состоящее в том, что выборочное пространство $\mathfrak{X}$ не должно зависеть от неизвестного параметра θ.

Регулярные модели

Рассмотрим некоторые свойства вклада $V(\boldsymbol{X};\theta)$ для регулярных моделей. Всегда имеет место тождество (по θ)

$$\int L(\underline{x};\theta)\,d\underline{x} \equiv 1 \quad (d\underline{x} = dx_1 \quad dx_n)$$

(здесь и далее везде для дискретных моделей интегрирование заменяется суммированием). Так как модель регулярна, то, дифференцируя это тождество по θ, получим

$$0 = \int \frac{\partial \ln L(\underline{x};\theta)}{\partial\theta}\,d\underline{x} = \int \frac{\partial \ln L(\underline{x};\theta)}{\partial\theta} L(\underline{x};\theta)\,d\underline{x} = \mathbf{E}_\theta V(X;\theta).$$

Таким образом, для регулярной модели

$$\mathbf{E}_\theta V(X;\theta) = 0, \quad \forall\,\theta \in \Theta. \tag{3}$$

Далее, в силу предположения (2) и (3) определена функция

$$i_n(\theta) = \mathbf{D}_\theta V(X;\theta) = \mathbf{E}_\theta V^2(X\ \ \theta), \tag{4}$$

которая носит название *информации Фишера* о параметре θ, содержащейся в выборке X. Величину

Информация Фишера

$$i(\theta) = i_1(\theta) = \mathbf{E}_\theta\left(\frac{\partial \ln f(\xi;\theta)}{\partial\theta}\right)^2 \tag{5}$$

называют также *количеством информации*, содержащейся в одном наблюдении. Из (1) имеем

$$i_n(\theta) = \mathbf{D}_\theta V(X;\theta) = \sum_{i=1}^{n} \mathbf{D}\left(\frac{\partial \ln f(X_i;\theta)}{\partial\theta}\right) = ni(\theta), \tag{6}$$

т. е. *количество информации, содержащейся в выборке, возрастает пропорционально объему выборки.*

Если плотность $f(x;\theta)$ дважды дифференцируема по θ, то продифференцировав при $n = 1$ тождество (3) еще раз, получим

$$\int \frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2} f(x;\theta)\,dx + \int \left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2 f(x;\theta)\,dx = 0,$$

т. е. (см. (5))

$$i(\theta) = -\mathbf{E}_\theta\left(\frac{\partial^2 \ln f(\xi;\theta)}{\partial\theta^2}\right). \qquad (5')$$

Пример вычисления информации Фишера для бернуллиевской модели $\mathrm{Bi}(1,\theta)$ имеется в § 3.1 (формула (17)): для этой модели $i(\theta) = 1/(\theta(1-\theta))$.

Рассмотрим еще один пример вычисления функции $i(\theta)$ для нормальной модели $\mathcal{N}(\theta,\sigma^2)$. Здесь вклад одного наблюдения

$$V(X_1;\theta) = \frac{\partial \ln f(X_1;\theta)}{\partial\theta} = \frac{X_1-\theta}{\sigma^2}, \qquad \frac{\partial^2 \ln f(X_1;\theta)}{\partial\theta^2} = -\frac{1}{\sigma^2}, \qquad (7)$$

откуда по формуле $(5')$ получаем $i(\theta) = \sigma^{-2}$. Вид функции $i(\theta)$ для ряда стандартных моделей приведен в табл. 1.

Таблица 1

Модель	$\mathcal{N}(\theta,\sigma^2)$	$\mathcal{N}(\mu,\theta^2)$	$\Gamma(\theta,\lambda)$	$\mathrm{Bi}(k,\theta)$	$\overline{\mathrm{Bi}}(r,\theta)$	$\Pi(\theta)$	$\mathcal{K}(\theta)$
$i(\theta)$	σ^{-2}	$2\theta^{-2}$	$\lambda\theta^{-2}$	$k/(\theta(1-\theta))$	$r/(\theta(1-\theta))^2$	θ^{-1}	$1/2$

(определение указанных моделей см. в § 1.1 § 1.2, общая модель Коши $\mathcal{K}(\theta)$ определена в упр. 48 (б) к гл. 1).

Приведем еще пример нерегулярной модели. Пусть $\mathcal{L}(\xi) \in \mathcal{U}(0,\theta)$. Здесь из тождества

Нерегулярность равномерной модели

$$\int_0^\theta \frac{d\boldsymbol{x}}{\theta} \equiv 1$$

не следует, что

$$\int_0^\theta \frac{\partial}{\partial\theta}\frac{1}{\theta}\,d\boldsymbol{x} = 0,$$

так как при дифференцировании интеграла по верхнему пределу появляется еще одно слагаемое. В данном случае причина нерегулярности заключается в том, что выборочное пространство (интервал $(0,\theta)$) зависит от неизвестного параметра θ.

2. Неравенство Рао—Крамера

В случае регулярных моделей для дисперсий несмещенных оценок можно установить абсолютную нижнюю границу, что и будет показано в настоящем пункте. Рассмотрим задачу оценивания некоторой параметрической функции $\tau(\theta)$ в регулярной модели $\mathcal{F} = \{F(x;\theta),\ \theta\in\Theta\}$ со скалярным параметром θ. Предположим, что класс $\mathcal{T}_\tau$ несмещенных оценок $\tau(\theta)$ не пуст, и функция $\tau(\theta)$ дифференцируема. Тогда имеет место следующее утверждение.

Теорема 1 (Неравенство Рао—Крамера). Для любой оценки $T = T(X) \in \mathcal{T}_\tau$ справедливо неравенство

$$\mathbf{D}_\theta T \geqslant \frac{[\tau'(\theta)]^2}{ni(\theta)}. \tag{8}$$

Равенство здесь имеет место тогда и только тогда, когда T — линейная функция вклада выборки, т. е.

$$T(X) - \tau(\theta) = a(\theta)V(X;\theta), \tag{9}$$

где $a(\theta)$ — некоторая функция от θ; в этом случае

$$\mathbf{D}_\theta T = a(\theta)\tau'(\theta). \tag{10}$$

Доказательство. По условию

$$\mathbf{E}_\theta T(X) = \int T(\underline{x})L(\underline{x};\theta)\,d\underline{x} = \tau(\theta), \quad \forall\,\theta \in \Theta. \tag{11}$$

Модель $\mathcal{F}$ регулярна, поэтому, дифференцируя это тождество по θ и учитывая (1), (3), получим

$$\begin{aligned}\tau'(\theta) &= \int T(\underline{x})\frac{\partial \ln L(\underline{x};\theta)}{\partial\theta}L(\underline{x};\theta)\,d\underline{x} = \\ &= \mathbf{E}_\theta\big(T(X)V(X;\theta)\big) = \operatorname{cov}_\theta\big(T(X)V(X;\theta)\big).\end{aligned} \tag{12}$$

Используя неравенство Коши—Буняковского и определение (4), из этого состояния получаем оценку $[\tau'(\theta)]^2 \leqslant i_n(\theta)\mathbf{D}_\theta T$ причем равенство здесь имеет место лишь в случае, когда (при каждом θ) статистика $T(X)$ и вклад выборки $V(X;\theta)$ линейно связаны, т. е. имеет место (9). Отсюда, во-первых, имеем неравенство (8), так как $i_n(\theta) = ni(\theta)$, а, во-вторых, подставляя представление (9) в (12) имеем $\tau'(\theta) = a^{-1}(\theta)\mathbf{D}_\theta T$, т. е. приходим к (10). ■

Замечание. Если $T = T(X)$ — оценка со смещением $b(\theta)$ и функция $b(\theta)$ дифференцируема, то, используя вместо соотношения (11) тождество

$$\int T(\underline{x})L(\underline{x};\theta)\,d\underline{x} = \tau(\theta) + b(\theta),$$

с помощью аналогичных рассуждений можно установить неравенство (также называемое неравенством Рао—Крамера)

$$\mathbf{D}_\theta T \geqslant \frac{[\tau'(\theta) + b'(\theta)]^2}{ni(\theta)},$$

обобщающее (8).

Эффективные оценки

Если существует оценка $T^* \in \mathcal{T}_\tau$, для которой граница Рао—Крамера (8) достигается, то ее назы-

вают *эффективной*. Эффективная оценка является оптимальной, и, согласно теореме 1 § 3.1, она единственна. Из теоремы 1 следует, что *критерием эффективности оценки является представление* (9). Будем называть этот критерий оптимальности оценки *критерием Рао—Крамера*. Для эффективной оценки, т. е. удовлетворяющей представлению (9), ее дисперсия может быть вычислена также по формуле (10).

Отметим следующее обстоятельство. Вклад выборки $V(X;\theta)$ однозначно определяется моделью. Поэтому представление (9) (когда оно имеет место) единственно и, следовательно, эффективная оценка может существовать только для одной определенной параметрической функции $\tau(\theta)$ и не существует ни для какой другой функции параметра θ, отличной от $a\tau(\theta)+b$, где a и b — константы.

С примером эффективной оценки мы уже встречались в примере 1 § 3.1. Там мы имели представление $\theta(1-\theta)/nV(X;\theta)=\overline{X}-\theta$, т. е. $a(\theta)=\theta(1-\theta)/n$, $\tau(\theta)=\theta$, и потому статистика $\overline{X}$ в модели $\mathrm{Bi}(1,\theta)$ является эффективной оценкой параметра θ. Рассмотрим еще несколько примеров на применение критерия Рао—Крамера.

Пример 1. Пусть X — выборка из нормального распределения $\mathcal{N}(\theta,\sigma^2)$. Из (7) следует, что в данном случае $\sigma^2/nV(X;\theta)=\overline{X}-\theta$, т. е. по критерию Рао—Крамера выборочное среднее $\overline{X}$ в модели $\mathcal{N}(\theta,\sigma^2)$ является эффективной оценкой θ и $\mathbf{D}_\theta\overline{X}=\sigma^2/n$, что совпадает с нижней границей в (8). •

Пример 2. Пусть $X=(X_1,\ldots,X_n)$ — выборка из $\mathcal{L}(\xi)\in\mathcal{N}(\mu,\theta^2)$. Здесь

$$L(X;\theta)=\frac{1}{\left(\sqrt{2\pi}\theta\right)^n}\exp\left\{-\frac{n}{2\theta^2}T^*(X)\right\},\quad T^*(X)=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2,$$

и потому

$$V(X;\theta)=\frac{\partial\ln L(X;\theta)}{\partial\theta}=\frac{n}{\theta^3}T^*(X)-\frac{n}{\theta},$$

или

$$\frac{\theta^3}{n}V(X;\theta)=T^*(X)-\theta^2$$

Но это по критерию Рао—Крамера означает, что статистика $T^*(X)$ является эффективной оценкой $\tau(\theta)=\theta^2$ и $\mathbf{D}_\theta T^*=2\theta^4/n$. •

Определение Назовем *эффективностью* оценки $T=T(X)\in\mathcal{T}_\tau$ отношение

Эффективность оценки

$$e(T)=e(T;\theta)=\frac{[\tau'(\theta)]^2}{\mathbf{D}_\theta T\cdot ni(\theta)}. \tag{13}$$

Из неравенства (8) следует, что $0\leqslant e(T)\leqslant 1$, а для эффективной оценки T^* ее эффективность $e(T^*)=1$. Таким образом, эффективность оценки — это отношение минимально возможной дисперсии в классе всех несмещенных оценок $\mathcal{T}_\tau$ к дисперсии данной оценки $T\in\mathcal{T}_\tau$.

Важным понятием в теории оценок является также асимптотическая эффективность, которая является аналогом понятия эффективности для больших выборок (при $n \to \infty$). Оценка $T_n(X)$, $X = (X_1, \quad , X_n)$, для $\tau(\theta)$ при $n \to \infty$ может оказаться асимптотически нормальной:

$$\mathcal{L}_\theta(T_n(X)) \sim \mathcal{N}\left(\tau(\theta), \frac{\sigma_T^2(\theta)}{n}\right)$$

Асимптотическая эффективность оценки

даже в тех случаях, когда $\mathbf{D}_\theta T_n$ не существует. *Асимптотической эффективностью* такой оценки назовем величину

$$e_0(T_n; \theta) = \frac{[\tau'(\theta)]^2}{i(\theta)}\sigma_T^2(\theta). \tag{14}$$

Оценку T_n^* называют *асимптотически эффективной*, если $e_0(T_n^*; \theta) \equiv 1$. Таким образом, в этом случае за среднее и дисперсию оценки T_n мы принимаем среднее и дисперсию аппроксимирующего нормального закона распределения.

Пример 3 (продолжение примера 2). Рассмотрим еще одну несмещенную оценку функции $\tau(\theta) = \theta^2$ именно, оценку S_0^2 (см. (17), (18) § 2.2). Как следует из (7) § 2.5, для нее $\mathbf{D}_\theta S_0^2 = 2\theta^4/(n-1)$, а функция информации в данном случае есть $i(\theta) = 2\theta^{-2}$ (см. табл. 1). Следовательно, по формуле (13) эффективность $e(S_0^2) = (n-1)/n < 1$. Таким образом, статистика S_0^2 не является эффективной оценкой теоретической дисперсии (таковой является указанная в примере 2 статистика T^*), но она асимптотически эффективна. •

Пример 4 (продолжение примера 1). Рассмотрим в качестве оценки θ выборочную медиану $\widehat{\zeta}_{n,\,1/2}$. Как показано в примере 2 § 2.4

$$\mathcal{L}_\theta\left(\widehat{\zeta}_{n,\,1/2}\right) \sim \mathcal{N}\left(\theta, \frac{\pi\sigma^2}{2n}\right),$$

поэтому асимптотическая эффективность (14) в данном случае есть

$$e_0\left(\widehat{\zeta}_n, \quad ; \theta\right) = \frac{2}{\pi} \approx 0{,}637.$$

Эффективной же оценкой является выборочное среднее $\overline{X}$. Это означает, что при больших n статистика $\overline{X}$ для выборки объема $n' = 0{,}637n$ оценивает истинное значение θ с такой же точностью, что и статистика $\widehat{\zeta}_{n,\,1/2}$, независимо от значений θ и σ^2

Сверхэффективные оценки

Интересно отметить, что при нарушении условий регулярности дисперсия оценки при $n \to \infty$ может убывать значительно быстрее, чем граница Рао—Крамера в (8), т. е. могут существовать «сверхэффективные» оценки. •

Пример 5. Рассмотрим нерегулярную модель $\mathcal{U}(0, \theta)$. Из упр. 46 к гл. I получаем, что максимальное значение выборки $X_{(n)}$ имеет моменты

$$\mathbf{E}_\theta X_{(n)} = \frac{n}{n+1}\,\theta, \quad \mathbf{D}_\theta X_{(n)} = \frac{n\theta^2}{(n+1)^2(n+2)},$$

т.е. $T_n = (n+1)/nX_{(n)}$ — несмещенная оценка θ, а ее дисперсия

$$\mathbf{D}_\theta T_n = \frac{\theta^2}{n(n+2)}$$

убывает с ростом n как n^{-2} в то время как граница Рао—Крамера убывает как n^{-1} •

Пример 6. Пусть требуется оценить неизвестный параметр сдвига θ распределения Вейбулла $W(\theta, \alpha, b)$ (см. п. 8 § 1.2). Из упр. 56 к гл. I получаем, что несмещенная оценка θ по выборке объема n есть

$$T_n = X_{(1)} - b\Gamma\left(1 + \frac{1}{\alpha}\right) n^{-1/\alpha},$$

где $X_{(1)}$ — минимальное значение выборки. Эта оценка имеет дисперсию

$$\mathbf{D}_\theta T_n = \mathbf{D}_\theta X_{(1)} = b^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right] n^{-2/\alpha},$$

порядок которой $n^{-2/\alpha}$ может быть сколь угодно высоким при достаточно малых $\alpha < 2$. В данном случае модель также нерегулярна, поскольку ее выборочное пространство зависит от θ. •

3. Экспоненциальная модель

Введем важный класс параметрических моделей, называемых *экспоненциальными*, для которых всегда существует эффективная оценка некоторой параметрической функции. По определению, модель $\mathcal{F} = \{F(x;\theta), \theta \in \Theta\}$ — *экспоненциальная*, если плотность $f(x;\theta)$ имеет вид

$$f(x;\theta) = \exp\{A(\theta)B(x) + C(\theta) + D(x)\}. \tag{15}$$

Многие статистические модели удовлетворяют этому условию. В частности, это известные нам модели $\mathcal{N}(\theta, \sigma^2)$, $\mathcal{N}(\mu, \theta^2)$, $\Gamma(\theta, \lambda)$, $\mathrm{Bi}(k, \theta)$, $\overline{\mathrm{Bi}}(r, \theta)$, $\Pi(\theta)$, приведенные в табл. 1. Модель Коши $\mathcal{K}(\theta)$, очевидно, не экспоненциальная, а равномерная модель $\mathcal{U}(0, \theta)$, хотя и удовлетворяет условию (15), но не является регулярной.

Вклад выборки для экспоненциальной модели равен

$$V(X;\theta) = A'(\theta)\sum_{i=1}^{n} B(X_i) + nC'(\theta) = nA'(\theta)\left[\frac{1}{n}\sum_{i=1}^{n} B(X_i) + \frac{C'(\theta)}{A'(\theta)}\right].$$

Это равенство можно записать в виде (9), если положить

$$T^* = T^*(X) = \frac{1}{n}\sum_{i=1}^{n} B(X_i), \quad \tau(\theta) = -\frac{C'(\theta)}{A'(\theta)}, \quad a(\theta) = \frac{1}{nA'(\theta)}. \tag{16}$$

Эффективные оценки в экспоненциальных моделях

По критерию Рао—Крамера заключаем, что статистика T^* является эффективной оценкой для параметрической функции $\tau(\theta)$, при этом согласно формулам (10) и (16)

$$\mathbf{D}_\theta T^* = \frac{\tau'(\theta)}{nA'(\theta)}. \tag{17}$$

Верно и обратное: если эффективная оценка для некоторой функции $\tau(\theta)$ существует, то модель является экспоненциальной. Действительно, интегрируя по θ равенство (9), получаем

$$\ln L(X;\theta) = A_1(\theta)T(X) + C_1(\theta) + D_1(X).$$

Отсюда следует, что плотность $f(x;\theta)$ должна иметь вид (15).

Таким образом, эффективные оценки существуют для некоторых функций $\tau(\theta)$ только в случае экспоненциальных моделей.

Для перечисленных выше шести регулярных экспоненциальных моделей вид функции $\tau(\theta)$, эффективной оценки τ^* для нее и соответствующей дисперсии (17) приведены в табл. 2.

Таблица 2

Модель	$\tau(\theta)$	τ^*	$\mathbf{D}_\theta\tau^*$
$\mathcal{N}(\theta, \sigma^2)$	θ	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\frac{\sigma^2}{n}$
$\mathcal{N}(\mu, \theta^2)$	θ^2	$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$	$\frac{2\theta^4}{n}$
$\Gamma(\theta, \lambda)$	θ	$\frac{\overline{X}}{\lambda}$	$\frac{\theta^2}{\lambda n}$
$\mathrm{Bi}(k, \theta)$	θ	$\frac{\overline{X}}{k}$	$\frac{\theta(1-\theta)}{kn}$
$\overline{\mathrm{Bi}}(r, \theta)$	$\frac{\theta}{1-\theta}$	$\frac{\overline{X}}{r}$	$\frac{\theta}{rn(1-\theta)^2}$
$\Pi(\theta)$	θ	$\overline{X}$	$\frac{\theta}{n}$

Отметим также, что для регулярных экспоненциальных моделей функцию информации $i(\theta)$ можно вычислять по формуле

$$i(\theta) = \tau'(\theta)A'(\theta) = \frac{C'(\theta)A''(\theta)}{A'(\theta)} - C''(\theta), \tag{18}$$

что следует из сравнения границы Рао—Крамера в (8) с (17) и вида $\tau(\theta)$ в (16).

4. Критерий Бхаттачария оптимальности оценки

Неравенство Рао—Крамера (8) дает неулучшаемую границу дисперсии несмещенных оценок (для регулярных моделей) в тех случаях, когда существует эффективная оценка. Однако, если такой оценки не существует (представление (9) не имеет места ни для какой статистики $T(X)$), то можно найти лучшую (т. е. бóльшую) нижнюю границу для дисперсии несмещенных оценок. Основное условие достижения нижней границы дисперсии в неравенстве Рао—Крамера состоит в существовании оценки T, для которой разность $T - \tau(\theta)$ является линейной функцией от

$$\frac{\partial \ln L}{\partial \theta} = \frac{1}{L}\frac{\partial L}{\partial \theta}$$

(для краткости мы пишем просто L вместо $L(X;\theta)$). Предположим, что такой оценки не существует, но зато имеется оценка T^*, для которой $T^* - \tau(\theta)$ — линейная функция от $\frac{1}{L}\frac{\partial L}{\partial \theta}, \frac{1}{L}\frac{\partial^2 L}{\partial \theta^2}, \ldots \frac{1}{L}\frac{\partial^r L}{\partial \theta^r}$:

$$T^* - \tau(\theta) = \sum_{i=1}^{r} a_i(\theta)\frac{L^{(i)}}{L}, \quad L^{(i)} = \frac{\partial^i L}{\partial \theta^i}, \tag{19}$$

при некотором $r \geqslant 2$. Тогда T^* — оптимальная оценка для $\tau(\theta)$. Этот критерий оптимальности, обобщающий критерий Рао—Крамера, называется *критерием Бхаттачария*.

Практически этот критерий используют так: учитывая старшие производные функции правдоподобия, подбирают такую их линейную комбинацию, чтобы получить представление вида (19); при этом последовательно полагают r 2, 3, Если при некотором значении r это удается сделать, то получающаяся при этом пара $\tau(\theta)$ и $T^* = T^*(X)$ и есть соответственно параметрическая функция и оптимальтная оценка для нее (случай $r = 1$ в этой методике соответствует критерию Рао—Крамера).

Мы не будем приводить точных формулировок (ввиду их громоздкости), а ограничимся лишь примером, иллюстрирующим изложенную идеологию.

Пример 7. Рассмотрим модель $\mathcal{N}(\theta, \sigma^2)$ и оценим функцию $\tau(\theta) = \theta^2$ Мы уже знаем (см. пример 1), что оптимальной оценкой для среднего θ является выборочное среднее

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Конечно, было бы ошибкой оценивать θ^2 посредством $\overline{X}^2$. Найдем здесь оптимальную оценку с помощью критерия Бхаттачария.

Здесь

$$L(X;\theta)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\theta)^2\right\}$$

и непосредственно проверяется, что

$$\frac{1}{L}\left[\frac{2\theta\sigma^4}{n}L^{(1)}+\frac{\sigma^4}{n^2}L^{(2)}\right]=\overline{X}^2-\frac{\sigma^2}{n}-\theta^2$$

По критерию Бхаттачария отсюда следует, что статистика $T^*=\overline{X}^2-\sigma^2/n$ — оптимальная (т. е. несмещенная с равномерно минимальной дисперсией) оценка для $\tau(\theta)=\theta^2$. Несмещенность этой оценки проверяется непосредственно:

$$\mathbf{E}_\theta T^*=\mathbf{E}_\theta\overline{X}^2-\frac{\sigma^2}{n}=\mathbf{D}_\theta\overline{X}+(\mathbf{E}_\theta\overline{X})^2-\frac{\sigma^2}{n}=\frac{\sigma^2}{n}+\theta^2-\frac{\sigma^2}{n}=\theta^2 \qquad \bullet$$

5. Критерии оптимальности в случае векторного параметра

Выше рассматривался случай скалярного параметра, однако все введенные понятия и критерии оптимальности непосредственно переносятся на случай векторного параметра $\theta=(\theta_1,\ \ldots,\ \theta_r)$. В этом случае под вкладом выборки понимается вектор $\underline{V}=(V_1,\ \ldots,\ V_r)$, где

$$V_i=V_i(X;\theta)=\frac{\partial\ln L(X;\theta)}{\partial\theta_i},\quad i=1,\ \ldots,\ r,$$

Информационная матрица

а аналогом функции информации (4) является *информационная матрица Фишера*

$$I_n(\theta)=\mathbf{D}_\theta\underline{V}=\left\|\mathbf{E}_\theta(V_iV_j)\right\|_1^r. \tag{20}$$

Информационную матрицу одного наблюдения

$$I(\theta)=I_1(\theta)=\left\|g_{ij}\right\|_1^r$$

можно вычислить по формулам

$$g_{ij}=g_{ij}(\theta)=\mathbf{E}_\theta\left(\frac{\partial\ln f(X_1;\theta)}{\partial\theta_i}\,\frac{\partial\ln f(X_1;\theta)}{\partial\theta_j}\right)=-\mathbf{E}_\theta\left(\frac{\partial^2\ln f(X_1,\theta)}{\partial\theta_i\partial\theta_j}\right) \tag{21}$$

(последнее равенство справедливо для дважды дифференцируемых по параметрам $\theta_1,\ \ldots,\ \theta_r$ функций (плотностей) $f(x;\theta)$). Для схемы повторных независимых наблюдений $I_n(\theta)=nI(\theta)$, т. е. здесь имеется полная аналогия со случаем скалярного параметра (см. (6)). В определении регулярной модели здесь дополнительно предполагается, что матрица $I(\theta)$ невырождена при всех $\theta\in\Theta$.

Предположим, что ищется оценка для дифференцируемой *числовой* функции $\tau(\theta)$. Тогда имеют место следующий *многомерный аналог неравенства Рао—Крамера* и соответствующий критерий оптимальности.

Теорема 2. Если модель $\mathcal{F}$ регулярна, то для дисперсии любой несмещенной оценки $T = T(X)$ функции $\tau(\theta)$ справедливо неравенство

$$\mathbf{D}_\theta T \geqslant \frac{1}{n}\underline{b}'(\theta)I^{-1}(\theta)\underline{b}(\theta) \equiv \frac{\sigma_\tau^2(\theta)}{n}, \quad \forall\, \theta \in \Theta, \tag{22}$$

где

$$\underline{b}(\theta) = \operatorname{grad} \tau(\theta) = \left(\frac{\partial \tau(\theta)}{\partial \theta_1}, \quad \frac{\partial \tau(\theta)}{\partial \theta_r}\right).$$

Равенство здесь имеет место тогда и только тогда, когда

$$T(X) - \tau(\theta) = \sum_{i=1}^{r} a_i(\theta) V_i(X; \theta) \tag{23}$$

при некоторых функциях $a_i(\theta)$, $i = 1, \quad , r$.

По аналогии со случаем скалярного параметра назовем несмещенную оценку $T^* = T^*(X)$ для $\tau(\theta)$ *эффективной*, если ее дисперсия совпадает при всех θ с правой частью неравенства Рао—Крамера (22). Аналогично вводится и *асимптотическая эффективность* $e_0(T_n)$ для асимптотически нормальных $\mathcal{N}\left(\tau(\theta), \sigma_T^2(\theta)/n\right)$ оценок $T_n = T_n(X)$, $X = (X_1, \quad , X_n)$, функции $\tau(\theta)$:

Асимптотическая эффективность

$$e_0(T_n; \theta) = \frac{\sigma_\tau^2(\theta)}{\sigma_T^2(\theta)}. \tag{24}$$

Оценка T_n называется *асимптотически эффективной*, если $e_0(T_n) \equiv 1$, т. е. если ее асимптотическая дисперсия $\sigma_T^2(\theta)/n$ совпадает с границей Рао—Крамера (22). С применением этого понятия мы встретимся немного позднее, в § 3.5.

Из теоремы 2 следует, что критерием эффективности (а значит, и оптимальности) оценки T является представление (23). Приведем пример использования этого критерия.

Пример 8 (Общая нормальная модель, оценивание среднего). Пусть

$$X = (X_1, \quad , X_n)$$

— выборка из нормального распределения $\mathcal{N}(\theta_1, \theta_2^2)$ (оба параметра неизвестны). Требуется оценить теоретическое среднее $\tau(\theta) = \theta_1$. Как мы уже знаем, выборочное среднее $\overline{X}$ всегда является несмещенной оценкой теоретического среднего. Покажем, что в данном случае $\overline{X}$ — оптимальная оценка. Для этого достаточно убедиться, что справедливо представление

$$\frac{\theta_2^2}{n} V_1(X; \theta) = \overline{X} - \theta_1,$$

и применить критерий (23).

Оптимальность выборочного среднего в нормальной модели

Ранее (см. пример 1) было показано, что $\overline{X}$ — оптимальная оценка теоретического среднего и для модели $\mathcal{N}(\theta, \sigma^2)$ (с известной дисперсией). Таким образом, *статистика $\overline{X}$ наилучшим образом оценивает неизвестное среднее нормальной модели независимо от того, известна дисперсия или нет.* •

Пример 9 (Общая нормальная модель, оценивание дисперсии). В ситуации предыдущего примера 8 оценим $\tau(\theta) = \theta_2^2$. Мы уже знаем (см. пример 4 § 3.1), что несмещенная оценка S_0^2 имеет дисперсию $2\theta_2^4/(n-1)$, $n \geqslant 2$. Сравним эту дисперсию с границей Рао—Крамера в (22). Для этого вычислим сначала информационную матрицу для модели $\mathcal{N}(\theta_1, \theta_2^2)$. Здесь

$$f(x;\theta) = \frac{1}{\sqrt{2\pi}\theta_2} \exp\left\{-\frac{(x-\theta_1)^2}{2\theta_2^2}\right\}$$

и по формулам (21) имеем

$$\left.\begin{aligned} g_{11} &= \mathbf{E}_\theta \theta_2^{-2} = \theta_2^{-2} \\ g_{12} = g_{21} &= 2\theta_2^{-3}\mathbf{E}_\theta(X_1 - \theta_1) = 0, \\ g_{22} &= 3\theta_2^{-4}\mathbf{E}_\theta(X_1-\theta_1)^2 - \theta_2^{-2} = 2\theta_2^{-2} \end{aligned}\right\} \Rightarrow I(\theta) = \left\| \begin{matrix} \theta_2^{-2} & 0 \\ 0 & 2\theta_2^{-2} \end{matrix} \right\| \qquad (25)$$

Отсюда находим, что правая часть в неравенстве (22) равна $2\theta_2^4/n$, что меньше, чем $\mathbf{D}_\theta S_0^2$. Таким образом, S_0^2 не является эффективной оценкой. Это, конечно, еще не означает, что существует более точная (с меньшей дисперсией) оценка, нежели S_0^2. Мы покажем, что в данной задаче граница Рао—Крамера в (22) не достигается ни для какой несмещенной оценки θ_2^2, но оптимальная оценка (н. о. р. м. д.) существует, и это S_0^2.

Оптимальное оценивание дисперсии в нормальной модели

Для этого мы воспользуемся аналогом критерия Бхаттачария для векторного параметра, который имеет следующий вид: если при некотором целом $s \geqslant 2$ и коэффициентах c_i $c_i(\theta)$ имеет место представление

$$T - \tau = \frac{1}{L}\left[\sum_i c_i \frac{\partial L}{\partial \theta_i} + \sum_{i,j} c_{ij} \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} + \ldots + \sum_{i_1,\ldots,i_s} c_{i_1\ldots i_s} \frac{\partial^s L}{\partial \theta_{i_1} \ldots \partial \theta_{i_s}}\right], \qquad (26)$$

то статистика $T = T(X)$ есть оптимальная оценка функции $\tau = \tau(\theta)$.

В нашем случае

$$\frac{1}{L}\frac{\partial L}{\partial \theta_2} = \frac{\partial \ln L}{\partial \theta_2} = \frac{1}{\theta_2^3}\sum_{i=1}^n (X_i - \theta_1)^2 - \frac{n}{\theta_2} =$$

$$= \frac{1}{\theta_2^3}\sum_{i=1}^n (X_i - \overline{X})^2 + \frac{n}{\theta_2^3}(\overline{X} - \theta_1)^2 - \frac{n}{\theta_2},$$

$$\frac{1}{L}\frac{\partial^2 L}{\partial\theta_1^2} = \frac{\partial^2 \ln L}{\partial\theta_1^2} + \left(\frac{\partial \ln L}{\partial\theta_1}\right)^2 = \frac{n^2}{\theta_2^4}(\overline{X} - \theta_1)^2 \sim \frac{n}{\theta_2^2}.$$

Отсюда

$$\frac{1}{L}\left(\frac{\theta_2^3}{n-1}\frac{\partial L}{\partial\theta_2} - \frac{\theta_2^4}{n(n-1)}\frac{\partial^2 L}{\partial\theta_1^2}\right) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 - \theta_2^2 = S_0^2 - \theta_2^2,$$

и по критерию (26) S_0^2 — оптимальная оценка θ_2^2.

Итак, в задаче оценивания дисперсии $\tau(\theta) = \theta_2^2$ в модели $\mathcal{N}(\theta_1, \theta_2^2)$ имеет место более точная, чем (22), граница для дисперсии несмещенных оценок функции $\tau(\theta)$:

$$\mathbf{D}_\theta T \geqslant \frac{2\theta_2^4}{n-1};$$

при этом равенство достигается для статистики $T^* = S_0^2$. •

Замечание. При оценивании самого параметра (вектора) $\theta = (\theta_1, \ldots, \theta_r)$ в качестве оценки выступает векторная статистика $\underline{T} = \underline{T}(X) = (T_1(X), \ldots, T_r(X))$, и тогда свойство ее несмещенности определяется равенствами $\mathbf{E}_\theta T_i(X) = \theta_i$, $i = 1, \ldots, r$, или в векторной форме: $\mathbf{E}_\theta \underline{T}(X) = \theta$, $\forall\,\theta$, а аналогом неравенства (22) является следующее неравенство для дисперсионной матрицы $\mathbf{D}_\theta \underline{T} = \left\|\mathrm{cov}_\theta\,(T_i, T_j)\right\|_1^r$:

$$\mathbf{D}_\theta \underline{T} \geqslant I_n^{-1}(\theta). \tag{27}$$

Здесь неравенство $A \geqslant B$ между матрицами (одинаковой размерности) означает, что матрица $A - B$ является неотрицательно определенной, т. е. для любого вектора $\underline{c}$ соответствующей размерности выполняется условие $\underline{c}'(A-B)\underline{c} \geqslant 0$ или $\underline{c}' A\underline{c} \geqslant \underline{c}' B\underline{c}$.

Пример 10 (Полиномиальная модель, оценивание параметрического вектора). В приложениях часто возникает модель с конечным числом $N \geqslant 2$ возможных исходов (которые можно считать числами $1, 2, \ldots, N$), вероятности которых $p_1, \ldots, p_N$ неизвестны ($p_1 + \ldots + p_N = 1$) — полиномиальная модель $M(n; \underline{p})$, $\underline{p} = (p_1, \ldots, p_N)$ (см. п. 6 § 1.1). Здесь в качестве неизвестного параметра следует взять вектор $\theta = (p_1, \ldots, p_{N-1})$, поскольку $p_N = 1 - p_1 - \ldots - p_{N-1}$. Другими словами, в данном случае речь идет об n испытаниях $X_1, \ldots, X_n$ над дискретной случайной величиной ξ с плотностью $f(x;\theta) = \mathbf{P}_\theta\{\xi = x\} = p_x$, $x = 1, \ldots, N$. Записав функцию $f(x;\theta)$ с помощью индикаторов в виде

$$f(x;\theta) = \prod_{j=1}^{N} p_j^{I(x=j)} = p_N \prod_{j=1}^{N-1} (p_j/p_N)^{I(x=j)} \tag{28}$$

(так как $\sum_{j=1}^{N} I(x=j) = 1$), получим

$$-\frac{\partial^2 \ln f(x;\theta)}{\partial p_i \partial p_j} = \begin{cases} I(x=i)p_i^{-2} + I(x=N)p_N^{-2}, & i=j, \\ I(x=N)p_N^{-2}, & i \neq j, \end{cases} \quad i, j = 1, \ldots, N-1.$$

Информационная матрица для полиномиальной модели

Отсюда по формуле (21) найдем информационную матрицу $I(\theta) = \|g_{ij}\|_1^{N-1}$:

$$g_{ij} = -\mathbf{E}_\theta\left(\frac{\partial^2 \ln f(X_1;\theta)}{\partial p_i \partial p_j}\right) = \begin{cases} p_i^{-1} + p_N^{-1}, & i = j, \\ p_N^{-1}, & i \neq j. \end{cases} \tag{29}$$

Непосредственными вычислениями можно убедиться (это мы оставляем читателю в качестве упражнения), что

$$|I(\theta)| = (p_1 \quad p_N)^{-1} > 0 \quad \text{и} \quad I^{-1}(\theta) = \|\sigma_{ij}\|_1^{N-1} \equiv \Sigma_{N-1},$$

где

$$\sigma_{ij} = \begin{cases} p_i(1 - p_i), & i = j, \\ -p_i p_j, & i \neq j. \end{cases}$$

Если теперь в качестве оценки для θ взять вектор относительных частот исходов , 2, , N $\underline{T} = (\nu_1/n, \quad , \nu_{N-1}/n)$, где $\nu_j \quad \sum_{i=1}^{n} I(X_i = j)$, $j = 1, \ldots, N$ (как это предполагается в примере 4 в § 1.1), то, во-первых, это будет несмещенная оценка, а, во-вторых, ее дисперсионная матрица, в силу соотношений (15) § 1.1, есть $\mathbf{D}_\theta \underline{T} = 1/n\Sigma_{N-1} = 1/nI^{-1}(\theta)$, т. е. для этой оценки в неравенстве (27) достигается равенство. Следовательно, $\underline{T}$ есть эффективная (оптимальная) оценка θ. •

§ 3.3. Достаточные статистики и оптимальные оценки

Рассмотренные в предыдущем параграфе критерии оптимальности оценок имеют сравнительно ограниченную применимость. Во-первых, они требуют жестких условий регулярности исходной модели и, во-вторых, в лучшем случае позволяют находить оптимальные оценки для отдельных параметрических функций $\tau(\theta)$. Более эффективным способом построения оптимальных оценок является использование достаточных статистик, рассмотрению которых и посвящен настоящий параграф. Достаточность является одним из важнейших понятий в математической статистике. Ввел ее Р. Фишер в 20-е гг. XX в. выдвинув идею о том, что для статистического анализа, касающегося неизвестных параметров, не всегда нужно знать все элементы выборки в отдельности, а достаточно знать лишь некоторые функции от выборки, называемые достаточными статистиками.

1. Достаточные статистики и критерий факторизации

По определению, статистика $T = T(X)$, скалярная или векторная, называется *достаточной* для модели $\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$ (или для параметра θ), если для любого события $A \subset \mathcal{X}$ условная вероятность $\mathbf{P}_\theta\{X \in A \,|\, T(X) = t\}$

не зависит от θ. Это свойство статистики T означает, что она содержит всю информацию о параметре θ, имеющуюся в выборке. Действительно, вероятность любого события, которое может произойти при фиксированном T, не зависит от θ, следовательно, оно не может нести дополнительной информации о неизвестном параметре θ. Эквивалентное определение достаточности в терминах плотности выборки $X = (X_1, \quad , X_n)$ формулируется так: статистика T достаточна, если условная плотность $L(\underline{x}|t;\theta)$ случайного вектора X при условии $T(X) = t$ не зависит от параметра θ. Сама выборка X, очевидно, является достаточной статистикой ($\mathbf{P}_\theta(X \in A|X = \underline{x})$ при любом θ есть либо 1, если $\underline{x} \in A$, либо 0 в противном случае), но обычно стремятся найти достаточную статистику наименьшей размерности, представляющую исходные данные в наиболее сжатом виде, в этом смысле говорят о *минимальной достаточной статистике*. При наличии достаточной статистики все статистические выводы об исследуемой модели в конечном итоге формулируются в терминах этой статистики; следовательно, минимальная достаточная статистика дает оптимальный способ представления исходных данных, что особенно важно при обработке больших массивов статистической информации.

Критерий факторизации

Практически достаточные статистики находят на основании следующего утверждения, которое называется *критерием факторизации*.

Теорема 1 (Критерий факторизации). Для того чтобы статистика $T(X)$ была достаточной для параметра θ, необходимо и достаточно, чтобы функция правдоподобия $L(\underline{x};\theta)$ выборки имела вид

$$L(\underline{x};\theta) = g(T(\underline{x});\theta)h(\underline{x}), \tag{1}$$

где g и h — неотрицательные функции, и h не зависит от θ.

Доказательство. Доказательство приведем только для дискретной модели. Если статистика T достаточна, то при любом t из области значений $T(\underline{x})$ функция $L(\underline{x}|t;\theta)$ не зависит от θ и ее можно записать в виде $h(\underline{x};t)$. Пусть $\mathbf{P}_\theta\{T(X) = t\} = g(t;\theta)$ и $T(\underline{x}) = t$; тогда событие $\{X = \underline{x}\} \subseteq \{T(X) = t\}$, поэтому

$$\begin{aligned} L(\underline{x};\theta) = \mathbf{P}_\theta\{X = \underline{x}\} &= \mathbf{P}_\theta\{X = \underline{x},\ T(X) = t\} = \\ &= \mathbf{P}_\theta\{T(X) = t\}\,\mathbf{P}_\theta\{X = \underline{x} \mid T(X) = t\} = \\ &= g(t;\theta)L(\underline{x}|t;\theta) = g(T(\underline{x});\theta)h(\underline{x};t), \end{aligned}$$

т. е. имеем представление (1).

Обратно: пусть имеет место разложение (факторизация) (1). Тогда при любом $\underline{x}$ таком, что $T(\underline{x}) = t$,

$$L(\underline{x}|t;\theta) = \mathbf{P}_\theta\{X = \underline{x} \mid T(X) = t\} = \frac{\mathbf{P}_\theta\{X = \underline{x}, T(X) = t\}}{\mathbf{P}_\theta\{T(X) = t\}} =$$

$$= \frac{L(\underline{x};\theta)}{\sum\limits_{\underline{x}':T(\underline{x}')=t} L(\underline{x}';\theta)} = \frac{g(t;\theta)h(\underline{x})}{\sum\limits_{\underline{x}':T(\underline{x}')=t} g(t;\theta)h(\underline{x}')} = \frac{h(\underline{x})}{\sum\limits_{\underline{x}':T(\underline{x}')=t} h(\underline{x}')},$$

т. е. не зависит от θ. Если же $\underline{x}$ таково, что $T(\underline{x}) \neq t$, то, очевидно, $L(\underline{x}|t;\theta) = 0$. Таким образом, при любом $\underline{x}$ условная вероятность $L(\underline{x}|t;\theta)$ не зависит от θ. ■

Отметим, что всякая эффективная оценка является одновременно достаточной статистикой, так как (см. (9) § 3.2) из представления

$$T(\boldsymbol{X}) - \tau(\theta) = a(\theta)\frac{\partial \ln L(\boldsymbol{X};\theta)}{\partial\theta}$$

следует форма (1) для $L(\underline{x};\theta)$. Итак, эффективная оценка существует только тогда, когда имеется достаточная статистика. Но последняя может существовать и при отсутствии эффективной оценки, т. е. условие достаточности является менее ограничительным, чем условие существования эффективной оценки.

Заметим также, что если T достаточна, то таковой же является и любая взаимно однозначная функция от T. Действительно, если $H = \varphi(T)$, и φ — взаимно однозначная функция, то существует обратная функция

$$\varphi^{-1} \quad T = \varphi^{-1}(H)$$

и из представления (1) имеем

$$L(\underline{x};\theta) = g\big(\varphi^{-1}(H);\theta\big)h(\underline{x}) = g_1(H;\theta)h(\underline{x}),$$

т. е. H — достаточная статистика.

Примерами достаточных статистик для ряда моделей со скалярным параметром являются (с учетом сделанного замечания) эффективные оценки τ^* приведенные в табл. 2 § 3.2. Рассмотрим еще несколько примеров нахождения достаточных статистик для моделей с векторным параметром и для нерегулярных моделей.

Пример 1 (Общая нормальная модель, достаточная статистика для нее). Функцию правдоподобия для модели $\mathcal{N}(\theta_1, \theta_2^2)$ можно записать в виде

$$L(\underline{x};\theta) = \frac{1}{(2\pi\theta_2^2)^{n/2}} \exp\left\{-\frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i - \overline{x})^2 - \frac{n(\overline{x} - \theta_1)^2}{2\theta_2^2}\right\}, \qquad (2)$$

откуда по критерию факторизации (1) заключаем, что в данном случае достаточной статистикой является пара

$$T = (T_1, T_2) = (\overline{X}, S^2)$$

или пара

$$\left(\overline{X}, \sum_{i=1}^{n} X_i^2\right),$$

поскольку они однозначно определяют друг друга. •

Пример 2 (Двухпараметрическая показательная модель, достаточная статистика для нее). Рассматривается частный случай $W(\theta_1, 1, \theta_2)$ модели Вейбулла (см. п. 8 § 1.2) с плотностью

$$f(x;\theta) = \frac{1}{\theta_2} e^{-(x-\theta_1)/\theta_2}, \quad x \geqslant \theta_1 \quad \big(\theta = (\theta_1, \theta_2) \quad \theta_1 \in R^1 \quad \theta_2 > 0\big). \tag{3}$$

Запишем плотность с помощью индикатора:

$$f(x;\theta) = \theta_2^{-1} e^{-(x-\theta_1)/\theta_2} I(x \geqslant \theta_1),$$

тогда функция правдоподобия выборки $X = (X_1, \ldots, X_n)$ примет вид

$$L(\underline{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \theta_2^{-n} \exp\left\{-\sum_{i=1}^{n} \frac{x_i - \theta_1}{\theta_2}\right\} \prod_{i=1}^{n} I(x_i \geqslant \theta_1).$$

Здесь

$$\prod_{i=1}^{n} I(x_i \geqslant \theta_1) = I(x_i \geqslant \theta_1,\ i = 1, \quad , n) = I\big(\min_i x_i \geqslant \theta_1\big) = I(x_{(1)} \geqslant \theta_1),$$

поэтому

$$L(\underline{x};\theta) = \theta_2^{-n} \exp\left\{-\frac{n}{\theta_2}(\bar{x} - \theta_1)\right\} I(x_{(1)} \geqslant \theta_1).$$

Отсюда по критерию факторизации следует, что достаточной статистикой для параметра $\theta = (\theta_1, \theta_2)$ является пара $(X_{(1)}, \overline{X})$. Как будет ясно из дальнейшего, удобнее для работы оказывается пара

$$T = (T_1, T_2) = (X_{(1)}, \overline{X} - X_{(1)}),$$

которая также является достаточной статистикой. •

Пример 3 (Равномерная модель, достаточная статистика для нее). Пусть $X = (X_1, \quad , X_n)$ — выборка из равномерного распределения $U(0, \theta)$, $\theta > 0$ неизвестно. Записав плотность равномерного распределения (см. п. 5 § 1.2) с помощью индикатора: $f(x;\theta) = \theta^{-1} I(0 \leqslant x \leqslant \theta)$, получим для функции правдоподобия выборки представление

$$L(\underline{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \theta^{-n} \prod_{i=1}^{n} I(0 \leqslant x_i \leqslant \theta) = \theta^{-n} I(x_{(n)} \leqslant \theta) I(x_{(1)} \geqslant 0).$$

Отсюда следует, что в данном случае достаточной статистикой является максимальное значение выборки $X_{(n)}$.

Аналогично, для общей равномерной модели $U(\theta_1, \theta_2)$, $\theta_1 < \theta_2$, можно записать

$$L(\underline{x};\theta) = (\theta_2 - \theta_1)^{-n} I(\theta_1 \leqslant x_{(1)} \leqslant x_{(n)} \leqslant \theta_2),$$

следовательно, статистика

$$T = (T_1, T_2) = (X_{(1)}, X_{(n)})$$

достаточная для $\theta = (\theta_1, \theta_2)$. •

Пример 4 (Равномерная модель (продолжение)). Рассмотрим модель

$$\mathcal{U}\big(a(\theta), b(\theta)\big),$$

где $a(\theta) < b(\theta)$, $\forall\,\theta$, — заданные непрерывные функции скалярного параметра θ. Как и в предыдущем примере 3, можно записать

$$L(\underline{x};\theta) = \big(b(\theta) - a(\theta)\big)^{-n} I\big(a(\theta) \leqslant x_{(1)} \leqslant x_{(n)} \leqslant b(\theta)\big) \tag{4}$$

Следовательно, двумерная статистика $T = (X_{(1)}, X_{(n)})$ является достаточной для θ. Выясним теперь, нельзя ли здесь найти одномерную, т. е. более экономную, достаточную статистику. Индикатор в (4) равен 1 тогда и только тогда, когда

$$\{x_{(1)} \geqslant a(\theta),\ b(\theta) \geqslant x_{(n)}\}. \tag{5}$$

Пусть $a(\theta) \uparrow$, $b(\theta) \downarrow$ с возрастанием θ Тогда условие (5) эквивалентно условию

$$\big\{\theta \leqslant a^{-1}(x_{(1)}),\ \theta \leqslant b^{-1}(x_{(n)})\big\} \Longleftrightarrow \big\{\theta \leqslant T_1(\underline{x}) = \min\big(a^{-1}(x_{(1)}),\ b^{-1}(x_{(n)})\big)\big\}.$$

Таким образом, соотношение (4) можно записать в виде

$$L(\underline{x};\theta) = \big(b(\theta) - a(\theta)\big)^{-n} I(T_1(\underline{x}) \geqslant \theta),$$

откуда следует, что

$$T_1 = T_1(X) = \min\big(a^{-1}(X_{(1)}),\ b^{-1}(X_{(n)})\big)$$

— одномерная достаточная статистика для θ.

Аналогично, если $a(\theta) \downarrow$, $b(\theta) \uparrow$ с возрастанием θ, то одномерная достаточная статистика также имеется и имеет вид

$$T_2 = T_2(X) = \max\big(a^{-1}(X_{(1)}),\ b^{-1}(X_{(n)})\big).$$

Этими двумя случаями исчерпываются ситуации, когда в модели $\mathcal{U}\big(a(\theta), b(\theta)\big)$, существует одномерная достаточная статистика. Обратим внимание на то, что как T_1, так и T_2 являются функциями исходной двумерной достаточной статистики $T = (X_{(1)}, X_{(n)})$. Это обстоятельство имеет общий характер, т. е. *минимальная достаточная статистика есть функция любых других достаточных статистик.*

В частности, для равномерной модели с нулевым средним $\mathcal{U}(-\theta, \theta)$ одномерная достаточная статистика T_2 принимает вид

$$T_2 = \max(-X_{(1)}, X_{(n)}) = \max\big(|X_{(1)}|, |X_{(n)}|\big),$$

а для модели $\mathcal{U}(\theta, \theta+1)$ одномерной достаточной статистики не существует. •

Пример 5 (Модель Коши). Для модели $\mathcal{K}(\theta)$ (см. упр. 48 б) к гл. 1) функция правдоподобия выборки $X = (X_1, \quad, X_n)$ имеет вид

$$L(\underline{x};\theta) = \frac{1}{\pi^n} \prod_{i=1}^{n} \frac{1}{1 + (x_i - \theta)^2},$$

и здесь нельзя найти статистику T, дающую факторизацию (1), размерность которой была бы меньше n. •

2. Теорема Рао—Блекуэлла—Колмогорова

Прежде всего напомним определение условного математического ожидания. Пусть X и Y — две случайные величины и $p(x)$, $q(y)$ — их плотности. Обозначим $f(x, y)$ совместную плотность величин X и Y, тогда функция

Условная плотность

$$p(x|y) = \frac{f(x, y)}{q(y)},$$

определенная там, где $q(y) > 0$, называется *условной плотностью величины X при условии, что $Y = y$*. Далее, обозначим

$$h(y) = \mathbf{E}(X|Y = y) =$$

$$= \begin{cases} \Sigma\, x\, p(x|y), & \text{если } X \text{ – дискретная случайная величина,} \\ \displaystyle\int x\, p(x|y)\, dx, & \text{если } X \text{ – непрерывная случайная величина.} \end{cases}$$

По определению, случайную величину $h(Y)$ называют условным математическим ожиданием случайной величины X при фиксированном значении Y Имеет место следующая формула полного математического ожидания

$$\mathbf{E}\big(\mathbf{E}(X|Y)\big) = \mathbf{E}X \tag{6}$$

(предполагается, что среднее значение $\mathbf{E}X$ существует).

Роль достаточных статистик в теории оценивания раскрывает следующее утверждение.

Теорема 2 (Теорема Рао—Блекуэлла—Колмогорова). Оптимальная оценка, если она существует, является функцией от достаточной статистики.

Доказательство. Пусть $T = T(X)$ — достаточная статистика и $T_1 = T_1(X)$ — произвольная несмещенная оценка заданной параметрической функции $\tau(\theta)$. Рассмотрим функцию

$$H(t) = \mathbf{E}_\theta(T_1|T = t) = \int T_1(\underline{x}) L(\underline{x}|t; \theta)\, d\underline{x}.$$

Эта функция не зависит от θ, так как условная плотность $L(\underline{x}|t; \theta)$ от параметра не зависит. Далее, статистика $H(T(X))$ является также несмещенной оценкой $\tau(\theta)$. Действительно, по формуле полного математического ожидания (6)

$$\mathbf{E}_\theta H(T) = \mathbf{E}_\theta\big(\mathbf{E}_\theta(T_1|T)\big) = \mathbf{E}_\theta T_1 = \tau(\theta).$$

Наконец, справедливо неравенство

$$\mathbf{D}_\theta H(T) \leqslant \mathbf{D}_\theta T_1, \quad \forall\, \theta,$$

причем равенство имеет место в том и только в том случае, когда $T_1 = H(T)$. Для доказательства этого заметим, что

$$\mathbf{D}_\theta T_1 = \mathbf{E}_\theta\big[T_1 - H(T) + H(T) - \tau(\theta)\big]^2 = \mathbf{E}_\theta\big[T_1 - H(T)\big]^2 + \mathbf{D}_\theta H(T) \geqslant \mathbf{D}_\theta H(T),$$

поскольку, в силу (6),

$$\mathbf{E}\big[T_1 - H(T)\big]\big[H(T) - \tau(\theta)\big] = \mathbf{E}\big[(T_1 - H(T))H(T)\big] =$$
$$= \mathbf{E}\big\{H(T)\mathbf{E}\big[(T_1 - H(T))|T\big]\big\} = \mathbf{E}\big\{H(T)\big[\mathbf{E}(T_1|T) - H(T)\big]\big\} = 0.$$

Таким образом, для любой несмещенной оценки функции $\tau(\theta)$, не являющейся функцией от достаточной статистики, всегда можно указать несмещенную оценку, которая зависит от достаточной статистики и имеет дисперсию меньшую, чем исходная оценка. Следовательно, *оптимальную оценку надо искать среди функций от достаточной статистики.* ■

При отыскании явного вида оптимальных оценок важную роль играет свойство полноты достаточной статистики. По определению, достаточная статистика $T = T(X)$ называется *полной*, если для всякой функции $\varphi(T)$ из того, что

Полная достаточная статистика

$$\mathbf{E}_\theta\varphi(T) = 0, \quad \forall\,\theta, \tag{7}$$

следует $\varphi(t) \equiv 0$ на всем множестве значений статистики T.

Теорема 3. Если существует полная достаточная статистика, то всякая функция от нее является оптимальной оценкой своего математического ожидания.

Доказательство. Пусть $T = T(X)$ — полная достаточная статистика и $H(T)$ — произвольная функция от T. Обозначим

$$\mathbf{E}_\theta H(T) = \tau(\theta). \tag{8}$$

Из условия полноты следует, что $H(T)$ — единственная функция от T, удовлетворяющая соотношению (8), так как если бы была еще какая-нибудь функция $H_1(T)$, удовлетворяющая (8), то мы бы имели, что

$$\mathbf{E}_\theta\big[H(T) - H_1(T)\big] = 0, \quad \forall\,\theta,$$

откуда $H_1(T) = H(T)$.

По теореме 2 оптимальную оценку для $\tau(\theta)$ надо искать в классе функций, зависящих от T. Но $H(T)$ — единственная функция от T, несмещенно оценивающая $\tau(\theta)$; следовательно, она и является искомой оптимальной оценкой. ■

Подведем итог. Итак, пусть в рассматриваемой модели

$$\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$$

существует полная достаточная статистика T (т. е. уравнение (7) имеет единственное (нулевое) решение) и требуется оценить заданную параметрическую функцию $\tau(\theta)$. Тогда:

Построение оптимальных оценок

1) *если существует какая-то несмещенная оценка* $\tau(\theta)$, *то существует и несмещенная оценка, являющаяся функцией от* T; можно так же сказать, что *если нет несмещенных оценок вида* $H(T)$ (т. е. уравнение (8) не имеет решения), *то класс* $\mathcal{T}_\tau$ *несмещенных оценок* $\tau(\theta)$ *пуст*;

2) *оптимальная* (т. е. *н. о. р. м. д.*) *оценка, когда она существует, всегда является функцией от* T, *и она однозначно определяется уравнением* (8);

3) *оптимальную оценку* τ^* *можно искать по формуле*

$$\tau^* = H(T) = \mathbf{E}_\theta(T_1|T), \tag{9}$$

исходя из любой несмещенной оценки T_1 *функции* $\tau(\theta)$.

В конкретных задачах при отыскании явного вида оптимальных оценок критерий (9) используется редко, так как вычисление условного математического ожидания обычно сопряжено со значительными аналитическими трудностями. Чаще решают непосредственно уравнение (8), которое в развернутой форме для абсолютно непрерывной модели имеет вид

$$\int H(T)g(t;\theta)\,dt = \tau(\theta), \quad \forall\,\theta \in \Theta, \tag{10}$$

где $g(t;\theta)$ — плотность статистики T (для дискретной модели интегрирование в (10) заменяется суммированием $\sum_t H(T)g(t;\theta)$). Это уравнение в дальнейшем будем называть *уравнением несмещенности*. Его можно решать различными способами (см. дальнейшие примеры в § 3.4), например, разлагая правую и левую части в ряды по степеням θ (в случае аналитических функций $\tau(\theta)$ и $g(t;\theta)$) и приравнивая соответствующие коэффициенты.

Дальнейшее обсуждение роли достаточных статистик в теории оптимального оценивания будет посвящено различным примерам их применения в конкретных задачах. Но прежде мы выделим один важный класс параметрических моделей, для которых вопросы о виде достаточной статистики и ее полноте решаются весьма просто. Это модели экспоненциального типа, которые нам уже встречались ранее (в п. 3 § 3.2) для случая скалярного параметра. Дадим их определение для общего случая r-мерного параметра $\theta = (\theta_1, \quad, \theta_r)$, $r \geqslant 1$.

3. Экспоненциальные семейства и достаточные статистики

Модель, задаваемая плотностью вида (сравни с (15) § 3.2)

$$f(x;\theta) = \exp\left\{\sum_{j=1}^{r} \theta_j B_j(x) + C(\theta) + D(x)\right\} \tag{11}$$

(или приводимой к такому виду заменой параметров) называется *r-параметрическим экспоненциальным семейством*. В качестве примера можно указать

нормальную модель $\mathcal{N}(\theta_1, \theta_2^2)$, для которой

$$f(x;\theta) = \exp\left\{-\frac{x^2}{2\theta_2^2} + \frac{\theta_1 x}{\theta_2^2} - \frac{\theta_1^2}{2\theta_2^2} - \ln\left(\theta_2\sqrt{2\pi}\right)\right\},$$

что приводится к виду (11) заменой $\theta_1' = -\theta_2^{-2}$ $\theta_2' = \theta_1\theta_2^{-2}$

Если имеет место (11), то по критерию факторизации статистика

$$T = (T_1, \quad , T_r), \quad \text{где} \quad T_j = T_j(X) = \sum_{i=1}^{n} B_j(X_i), \quad j = 1, \quad , r, \quad (12)$$

является достаточной для параметра $\theta = (\theta_1, \quad , \theta_r)$, и она будет полной, если размерность параметрического множества Θ равна r (т. е. Θ содержит некоторый r-мерный параллелепипед).

Так, указанные в табл. 2 § 3.2 оценки τ^* являются полными достаточными статистиками для соответствующих моделей; для модели же $\mathcal{N}(\theta_1, \theta_2^2)$ полной достаточной статистикой является в соответствии с (12), пара $(\Sigma X_i, \Sigma X_i^2)$ или эквивалентная ей пара $(\overline{X}, S^2)$ (см. пример 1).

Пример 6 (Оценивание нормальной функции распределения).
Пусть $\mathcal{L}(\xi) = \mathcal{N}(\theta, \sigma^2)$, т. е. наблюдается нормальная случайная величина с известной дисперсией σ^2 и неизвестным средним θ Требуется по выборке $X = (X_1, \quad , X_n)$ оценить вероятность $\tau(\theta) = \mathbf{P}_\theta\{\xi \leqslant x_0\}$ в заданной точке x_0. Как мы знаем (см. замечание к теореме 2 § 3.1), когда у нас нет никакой априорной информации о теоретической функции распределения $F_\xi(x)$, то оптимальной несмещенной оценкой для $F_\xi(x_0)$ является значение в точке x_0 эмпирической функции распределения

$$\widehat{F}_n(x_0) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leqslant x_0).$$

Но в нашем случае вид оцениваемой функции нам известен:

$$\tau(\theta) = F_\xi(x_0;\theta) = \mathbf{P}_\theta\left\{\frac{\xi - \theta}{\sigma} \leqslant \frac{x_0 - \theta}{\sigma}\right\} = \Phi\left(\frac{x_0 - \theta}{\sigma}\right) \quad (13)$$

(см. п. 1, (7) § 1.2), поэтому, применяя изложенную теорию достаточных статистик, мы построим более точную оценку, учитывающую имеющуюся априорную информацию о модели, в которой мы работаем.

Именно, как отмечено выше, выборочное среднее $\overline{X}$ является здесь полной достаточной статистикой, и если мы возьмем в качестве исходной несмещенной оценки функции (13) простейшую статистику

$$T_1 = I(X_1 \leqslant x_0) \quad \mathbf{E}_\theta T_1 = \mathbf{P}_\theta(X_1 \leqslant x_0) = \tau(\theta),$$

то оптимальная оценка может быть найдена по формуле (см. (9))

$$\tau^* = \mathbf{E}_\theta(T_1|\overline{X}) = \mathbf{P}_\theta\{X_1 \leqslant x_0|\overline{X}\} = \mathbf{P}_\theta\{X_1 - \overline{X} \leqslant x_0 - \overline{X}|\overline{X}\}.$$

Здесь (см. упр. 28 к гл. 1) случайная величина $X_1 - \overline{X}$ не зависит от условия $\overline{X}$, поэтому указанная условная вероятность совпадает с безусловной вероятностью $\mathbf{P}_\theta\{X_1 - \overline{X} \leqslant x_0 - \overline{X}\}$. Но линейная функция $X_1 - \overline{X}$ от нормальной выборки распределена по нормальному закону со средним

$$\mathbf{E}_\theta(X_1 - \overline{X}) = \mathbf{E}_\theta X_1 - \mathbf{E}_\theta \overline{X} = \theta - \theta = 0$$

и дисперсией

$$\mathbf{D}_\theta(X_1 - \overline{X}) = \mathbf{D}_\theta\left(\frac{n-1}{n}X_1 - \frac{1}{n}\sum_{i=2}^{n} X_i\right) = \left[\left(\frac{n-1}{n}\right)^2 + \frac{n-1}{n^2}\right]\sigma^2 = \frac{n-1}{n}\sigma^2$$

поэтому

$$\begin{aligned}\tau^* &= \mathbf{P}_\theta\{X_1 - \overline{X} \leqslant x_0 - \overline{X}\} = \\ &= \mathbf{P}_\theta\left\{\sqrt{\frac{n}{n-1}}\frac{X_1 - \overline{X}}{\sigma^2} \leqslant \sqrt{\frac{n}{n-1}}\frac{x_0 - \overline{X}}{\sigma}\right\} = \\ &= \Phi\left(\sqrt{\frac{n}{n-1}}\frac{x_0 - \overline{X}}{\sigma}\right). \qquad (14)\end{aligned}$$

Обратим внимание на то, что оптимальная (н. о. р. м. д) оценка для функции (13) получается не просто заменой неизвестного параметра θ на достаточную статистику $\overline{X}$, но введением еще и дополнительного множителя $\sqrt{n/(n-1)}$ в аргумент функции Φ, — догадаться до такой «экзотики» из каких-то других соображений представляется весьма проблематичным. •

Еще более «экзотичен» результат следующего примера, полученный с использованием таких же идей А. Н. Колмогоровым в 1950 г. в более общей задаче.

Оценивание функции распределения в общей нормальной модели

Пример 7 (Задача Колмогорова). Пусть по выборке $\boldsymbol{X} = (X_1, \ldots, X_n)$, $n > 2$, из нормального распределения $\mathcal{N}(\theta_1, \theta_2^2)$ (оба параметра неизвестны) требуется оценить значение теоретической функции распределения $F_\xi(x; \theta)$ в заданной точке x_0, т. е. параметрическую функцию (сравни с (13))

$$\tau(\theta) = F_\xi(x_0; \theta) = \mathbf{P}_\theta\left\{\frac{\xi - \theta_1}{\theta_2} \leqslant \frac{x_0 - \theta_1}{\theta_2}\right\} = \Phi\left(\frac{x_0 - \theta_1}{\theta_2}\right). \qquad (15)$$

Следуя изложенной выше идеологии, будем искать оптимальную несмещенную оценку в виде (9), где в качестве исходной несмещенной оценки снова возьмем индикатор $T_1 = I(X_1 \leqslant x_0)$. В данном случае полной достаточной статистикой является, как мы знаем, $T = (\overline{X}, S^2)$, поэтому можем записать

$$\tau^* = \mathbf{E}_\theta(T_1|T) = \mathbf{P}_\theta\{X_1 \leqslant x_0|T\} = \mathbf{P}\left\{\frac{X_1 - \overline{X}}{\sqrt{n-1}S} \leqslant u_0 \middle| T\right\}, \qquad (16)$$

где

$$u_0 = u_0(T) = \frac{x_0 - \overline{X}}{\sqrt{n-1}S}.$$

Если бы случайная величина

$$\eta = \frac{X_1 - \overline{X}}{\sqrt{n-1}S}$$

не зависела от условия T, то правая часть (16) была бы равна $F_\eta(u_0)$, где вид функции распределения $F_\eta(u)$ указан в упр. 42 к гл. 2. Таким образом, нам остается установить независимость статистик η и T. Для этого сделаем небольшое отступление и добавим еще кое-что из теории (изложенного выше нам не хватает).

Свободные статистики

Определение Статистика $\mathcal{U} = \mathcal{U}(X)$ называется *свободной*, если ее распределение не зависит от параметра $\theta \in \Theta$ (в произвольной параметрической модели $\mathcal{F}$).

Теорема 4 (Теорема Басу). Пусть $T(X)$ — полная достаточная статистика, $\mathcal{U}(X)$ — свободная статистика. Тогда $T(X)$ и $\mathcal{U}(X)$ независимы.

Доказательство. Пусть A — произвольное измеримое (т. е. для которого определены выписываемые далее вероятности) событие. Тогда вероятности: условная — $\varphi_A(T) = \mathbf{P}_\theta\{\mathcal{U} \in A | T\}$ и безусловная — $\gamma_A = \mathbf{P}_\theta\{\mathcal{U} \in A\}$ не зависят от θ по условию теоремы, при этом $\mathbf{E}_\theta \varphi_A(T) = \gamma_A$. Следовательно, обозначая $g(T) = \varphi_A(T) - \gamma_A$, имеем $\mathbf{E}_\theta g(T) = 0$, $\forall\, \theta$. Из полноты T следует, что $g(T) \equiv 0$, т. е. безусловная γ_A и условная $\varphi_A(T)$ вероятности совпадают. Но это и означает независимость статистик $T(X)$ и $\mathcal{U}(X)$. ■

Вернемся к нашей задаче и заметим, что если ввести нормированные случайные величины $Y_i = (X_i - \theta_1)/\theta_2$, $i = 1, \quad , n$, которые имеют уже стандартное нормальное распределение $\mathcal{N}(0, 1)$ (т. е. не зависящее от параметра θ), то легко проверить (это мы оставляем в качестве простого упражнения для читателя), что статистика η в терминах величин Y_i сохранит свой вид:

$$\eta = \frac{Y_1 - \overline{Y}}{\sqrt{n-1}S(Y)}, \quad \text{где} \quad S^2(Y) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

Отсюда следует, что распределение статистики η не зависит от параметра модели $\theta = (\theta_1, \theta_2)$, значит, она является свободной (!) статистикой, и по теореме Басу η и $T = (\overline{X}, S^2)$ независимы (!).

Отсюда и из (16) имеем $\tau^* = F_\eta(u_0)$. Используя результат упр. 42 к гл. 2, можем записать явный вид этой оценки через неполную бета-функцию:

$$\tau^* = \tau^*(T) = \begin{cases} 1 - \frac{1}{2} B\left(1 - u_0^2(T); \frac{n-2}{2}, \frac{1}{2}\right), & \text{если} \quad 0 < u_0(T) < 1, \\ \frac{1}{2} B\left(1 - u_0^2(T); \frac{n-2}{2}, \frac{1}{2}\right), & \text{если} \quad -1 < u_0(T) < 0. \end{cases}$$

С помощью индикаторов этой формуле можно придать следующий окончательный вид: н. о. р. м. д для функции (15) является статистика

$$\tau^* = \left[1 - \frac{1}{2} B\left(1 - u_0^2(T); \frac{n-2}{2}, \frac{1}{2}\right)\right] I(\overline{X} < x_0) + \\ + \frac{1}{2} B\left(1 - u_0^2(T); \frac{n-2}{2}, \frac{1}{2}\right) I(\overline{X} \geqslant x_0). \tag{17}$$

Это и есть знаменитый результат А. Н. Колмогорова. •

Наш следующий пример относится к оцениванию в полиномиальной модели $M(n; \underline{p})$ (см. пример 10 в § 3.2).

Пример 8 (Полиномиальная модель: оценивание параметрических функций). Записав плотность $f(x; \theta)$ (см. (28) § 3.2) в виде

$$f(x; \theta) = p_N \prod_{j=1}^{N-1} \left(\frac{p_j}{p_N}\right)^{I(x=j)} = \exp\left\{\sum_{j=1}^{N-1} I(x = j) \ln \frac{p_j}{p_N} + \ln p_N\right\}, \tag{18}$$

видим, что она приводится к виду (11) заменой

$$\theta'_j = \ln\left(\frac{p_j}{p_N}\right), \quad j = 1, \ldots, N-1.$$

Таким образом, эта модель укладывается в схему r-параметрического (при $r = N - 1$) экспоненциального семейства, и, следовательно, для нее полной достаточной статистикой является вектор $\underline{T} = (T_1, \quad , T_{N-1})$, где

$$T_j = \sum_{i=1}^{n} I(X_i = j) = \nu_j$$

— число реализаций исхода «j» в выборке $X = (X_1, \ldots, X_n)$, $j = 1, \ldots, N-1$.

Заметим, что число реализаций исхода «N» есть $\nu_N = n - \nu_1 - \quad - \nu_{N-1}$, а распределение полного случайного вектора $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ есть полиномиальное распределение $M(n; p_1, \quad , p_N)$.

Из изложенной в п. 2 теории следует, что в данной модели несмещенные оценки существуют лишь для таких параметрических функций, которые имеют вид $\mathbf{E}_\theta H(\underline{T})$, или, что то же самое, вид $\mathbf{E}_\theta H(\underline{\nu})$. Но (см. (14) § 1.1)

$$\mathbf{E}_\theta H(\underline{\nu}) = \sum_{x_1 + \ldots + x_N = n} H(x_1, \ldots, x_N) \frac{n!}{x_1! \ldots x_N!} p_1^{x_1} \ldots p_N^{x_N},$$

и при любой функции H правая часть здесь представляет собой полином от $p_1,\ \ ,p_N$ степени не выше n. Следовательно, оцениваемыми (т. е. для которых существуют несмещенные оценки) параметрическими функциями в данной модели являются лишь полиномы степени не выше n от $p_1, \ldots, p_N$ Так, из упр. 19 к гл. I следует, что несмещенной оптимальной оценкой функции $\tau(\theta) = p_1^{r_1} \ldots p_N^{r_N}$ при $r_1 +\ \ +r_N \leqslant n$ является статистика

$$\tau^* = \frac{(\nu_1)_{r_1}\ \ (\nu_N)_{r_N}}{(n)_{r_1+\ldots+r_N}}$$

(напомним, что $(a)_k = a(a-1)\ \ (a-k+1)$, $(a)_0 = 1$). Оценки для произвольных полиномов (степени не выше n) можно строить с помощью линейных комбинаций этих статистик на основе теоремы 3 § 3.1.

Как частный случай этих результатов при $N = 2$ имеем соответствующие выводы для бернуллиевской модели. В примере 1 в § 3.1 были построены несмещенные оценки для произвольных полиномов степени не выше n (n — объем выборки) от параметра θ модели $\mathrm{Bi}(1, \theta)$ (соотношение (6) в § 3.1). Эти оценки представляют собой функции от полной достаточной статистики $X = X_1 +\ \ +X_n$ — числа «успехов» в n испытаниях, следовательно они являются оптимальными (н. о. р. м. д). •

В заключение этого параграфа приведем пример достаточной статистики, не являющейся полной.

Пример 9. Пусть произведено одно наблюдение над дискретной случайной величиной X с распределением

$$f(x;\theta) = \begin{cases} \theta & \text{при } x = -1, \\ \theta^x(1-\theta)^2 & \text{при } x = 0, 1, 2, \ldots, \end{cases} \qquad \theta \in (0, 1).$$

Тогда X — достаточная, но не полная статистика. Действительно, пусть $\varphi(x)$ — заданная на множестве $\{-1, 0, 1, 2, \ldots\}$ функция, удовлетворяющая условию $\mathbf{E}_\theta\varphi(x) = 0$, $\forall\,\theta \in (0, 1)$. Это условие можно записать в виде

$$\varphi(-1)\theta(1-\theta)^{-2} + \sum_{x=0}^{\infty} \varphi(x)\theta^x \equiv 0$$

или, так как

$$(1-\theta)^{-2} = \sum_{x=0}^{\infty}(x+1)\theta^x \qquad \varphi(0) + \sum_{x=1}^{\infty}\left[\varphi(x) + x\varphi(-1)\right]\theta^x \equiv 0.$$

Но этому условию удовлетворяет любая функция, для которой $\varphi(0)\ \ 0$, $\varphi(x) = -x\varphi(-1)$, $x = 1, 2,\ \ $ т. е. значение $\varphi(-1)$ можно задать произвольно. Следовательно, в данном случае критерий полноты не выполняется, поэтому уравнение несмещенности не будет иметь однозначного решения.

Рассмотрим, например, задачу оценивания функции $\tau(\theta) = (1-\theta)^2$. Непосредственно можно убедиться в том, что любая статистика $T = T_a(X)$, где

$$T_a(x) = \begin{cases} 1 & \text{при } x = 0, \\ ax & \text{при } x = -1, 1, 2, \end{cases}$$

a произвольно, является несмещенной оценкой $\tau(\theta)$ (проверку этого мы оставляем читателю в качестве простого упражнения). •

§ 3.4. Способы решения уравнения несмещенности

А математику уже затем учить следует, что она ум в порядок приводит.

М. В. Ломоносов[4)]

Настоящий параграф посвящен применению теории достаточных статистик при отыскании оптимальных несмещенных оценок в конкретных задачах. Ключевым моментом при этом является отыскание решения уравнения несмещенности (10) § 3.3. Имеются классы статистических моделей, для которых разработаны специфические методы решения уравнения несмещенности, позволяющие получать явные выражения для оптимальных оценок произвольных оцениваемых параметрических функций. Изложение этих методов и составляет содержание данного параграфа. Подчеркнем, что прежде чем решать уравнение несмещенности, необходимо в каждой конкретной задаче:

а) определить достаточную статистику $T = T(X)$;

б) найти ее распределение: $F_T(t;\theta) = \mathbf{P}_\theta\{T(X) \leqslant t\}$ или плотность $g(t;\theta)$;

в) проверить, обладает ли она свойством полноты, т. е. решить однородное уравнение (7) § 3.3 и убедиться, что оно имеет единственное (нулевое) решение.

Только убедившись, что в рассматриваемой модели несмещенная оценка нуля единственна, можно приступать к решению уравнения несмещенности для заданной параметрической функции $\tau(\theta)$. По этой программе мы и будем в дальнейшем действовать.

Полезное замечание

Но прежде сделаем одно полезное замечание. В конкретной задаче, найдя полную достаточную статистику $T = T(X)$, полезно вычислить ее среднее $\mu(\theta) = \mathbf{E}_\theta T(X)$, т. е. определить тем самым ту параметрическую функцию, для которой сама статистика T является оптимальной оценкой. Может оказаться, что $\mu(\theta) = k\theta$, т. е. линейно зависит от параметра θ рассматриваемой модели. Тогда оптимальной оценкой θ^* для параметра θ будет, очевидно, статистика $\theta^* = T(X)/k$. Например,

[4)] Ломоносов Михаил Васильевич (1711–1765) — русский ученый-энциклопедист и мыслитель, поэт, академик Петербургской АН.

пусть модель $\mathcal{F} = \mathcal{U}(0; \theta)$ — равномерное распределение с неизвестным правым концом носителя (*носителем* распределения называют то множество, на котором соответствующая плотность $f(x; \theta) > 0$). Как показано в примере 3 § 3.3 достаточной статистикой в этом случае является максимальное значение $X_{(n)}$ выборки $X = (X_1, \ldots, X_n)$. То, что эта статистика полна, будет установлено чуть ниже (см. пример 3), а так как $\mathbf{E}_\theta X_{(n)} = \dfrac{n}{n+1}\theta$ (см. упр. 46 к гл. 1), то отсюда следует, что $\theta^* = \dfrac{n+1}{n} X_{(n)}$ — оптимальная оценка θ.

1. Модель степенного ряда, оценивание параметрических функций

Эта модель, охватывающая большое число конкретных дискретных распределений, введена и описана в п. 8 § 1.1. Из вида плотности (20) § 1.1 следует, что она относится к экспоненциальному семейству (см. п. 3 § 3.3) и потому обладает полной достаточной статистикой $T = T(X) = X_1 + \quad + X_n$. Распределение этой статистики указано в (24) § 1.1. Следовательно, здесь у нас все готово, чтобы приступить к основной проблеме — решению уравнения несмещенности. Предположим, что требуется оценить параметрическую функцию $\tau(\theta)$, представимую в виде сходящегося на Θ степенного ряда:

$$\tau(\theta) = \sum_{j=r}^{\infty} b_j \theta^j$$

Тогда из формулы (24) § 1.1 следует, что в данном случае уравнение несмещенности $\mathbf{E}_\theta H(T) = \tau(\theta)$ принимает вид

$$\sum_{t=nl}^{\infty} H(t) a_n(t) \theta^t = f^n(\theta)\tau(\theta) = \sum_{x=nl}^{\infty} a_n(x)\theta^x \sum_{j=r}^{\infty} b_j \theta^j = \\ = \sum_{k=nl+r}^{\infty} \theta^k \sum_{j=r}^{k-nl} b_j a_n(k-j). \qquad (1)$$

Поскольку это тождество по θ, приравнивая соответствующие коэффициенты, находим

$$H(t)a_n(t) = \begin{cases} \displaystyle\sum_{j=r}^{t-nl} b_j a_n(t-j) & \text{при} \quad t \geqslant nl + r, \\ 0 & \text{при} \quad t < nl + r. \end{cases}$$

Общее решение

Отсюда получаем, что оптимальная оценка τ^* для функции $\tau(\theta)$ имеет вид

$$\tau^* = H(T) = a_n^{-1}(T) \sum_{j=r}^{T-nl} b_j a_n(T-j)\, I(T \geqslant nl + r). \qquad (2)$$

Это чрезвычайно общий результат, поскольку здесь функция $f(\theta)$, задающая модель, и оцениваемая функция $\tau(\theta)$ — произвольные аналитические функции на Θ. Выделим особо важные специальные случаи решения (2).

Специальные случаи параметрических функций

Пусть $\tau(\theta) = f(\theta)$. Тогда в исходном соотношении (1) правая часть есть

$$f^{n+1}(\theta) = \sum_{x=(n+1)l}^{\infty} a_{n+1}(x)\theta^x$$

и потому оптимальная оценка f^* этой функции есть

$$f^* = \frac{a_{n+1}(T)}{a_n(T)} I(T \geqslant (n+1)l). \tag{3}$$

Если $\tau_r(\theta) = \theta^r$ то из (2) следует, что

$$\tau_r^* = \frac{a_n(T-r)}{a_n(T)} I(T \geqslant nl + r). \tag{4}$$

Пусть теперь $\tau(\theta; x) = \theta^x / f(\theta)$, $x \geqslant l$, т. е. речь фактически идет об оценивании теоретической плотности $f(x|\theta)$ (см. (20) § 1.1). В этом случае исходное соотношение (1) запишется в виде

$$\sum_{t \geqslant nl} H(t) a_n(t) \theta^t = \theta^x f^{n-1}(\theta) = \theta^x \sum_{j=(n-1)l}^{\infty} a_{n-1}(j)\theta^j,$$

откуда аналогично предыдущему имеем

$$\tau^*(x) = \frac{a_{n-1}(T-x)}{a_n(T)} I(T \geqslant (n-1)l + x). \tag{5}$$

В частности, оптимальная оценка $f^*(x)$ теоретической плотности $f(x|\theta)$ получается умножением правой части (5) на коэффициент $a(x)$:

$$f^*(x) = a(x)\frac{a_{n-1}(T-x)}{a_n(T)} I(T \geqslant (n-1)l + x), \quad x \geqslant l. \tag{6}$$

Наконец, если требуется оценить параметрическую функцию вида

$$\tau_\varphi(\theta) = \mathbf{E}_\theta \varphi(\xi) = \sum_{x \geqslant l} \varphi(x) f(x|\theta),$$

то оптимальная оценка получается заменой в правой части этого равенства теоретической плотности $f(x|\theta)$ ее оценкой (6):

$$\tau_\varphi^* = \sum_{x \geqslant l} \varphi(x) f^*(x) = \sum_{x=l}^{T-(n-1)l} \varphi(x) a(x) \frac{a_{n-1}(T-x)}{a_n(T)}. \tag{7}$$

Отметим также, что по критерию для экспоненциальной модели (см. п. 3 § 3.2) выборочное среднее $\overline{X} = T(X)/n$ является эффективной (значит, и оптимальной) оценкой теоретического среднего (см. (23) § 1.1) $\theta f'(\theta)/f(\theta)$, и это есть единственная эффективная оценка в модели степенного ряда.

Итак, для важного класса дискретных распределений типа степенного ряда можно строить оптимальные оценки для произвольных параметрических функций, представимых в виде степенного ряда от θ.

Рассмотрим несколько конкретных примеров на применение полученного общего результата.

Пример 1 (Пуассоновская модель, оценивание параметрических функций). Пусть $X = (X_1, \quad , X_n)$ — выборка из распределения $\mathcal{L}(\xi) \in \Pi(\theta)$. Оценим параметрическую функцию $\tau_{r,t}(\theta) = \theta^r e^{-t\theta}$ при заданных $t < n$ и $r \geqslant 0$ целом. Здесь (см. п. 8 § 1.1) $f(\theta) = e^{\theta}$, значит,

$$f^n(\theta) = e^{n\theta} = \sum_{x=0}^{\infty} \frac{(n\theta)^x}{x!}, \quad \text{т. е.} \quad a_n(x) = \frac{n^x}{x!}, \quad x \geqslant 0.$$

Далее,

$$\tau_{r,t}(\theta) = \theta^r e^{-t\theta} = \theta^r \sum_{i=0}^{\infty} \frac{(-t\theta)^i}{i!} = \sum_{j=r}^{\infty} \frac{(-t)^{j-}}{(j-r)!}\theta^j,$$

т. е.

$$b_j = \frac{(-t)^{j-r}}{(j-r)!}, \quad j \geqslant r.$$

Поэтому по формуле (2)

$$\tau^*_{r,t} = \frac{T!}{n^T} \sum_{j=r}^{T} \frac{(-t)^{j-r}}{(j-r)!} \frac{n^{T-j}}{(T-j)!} I(T \geqslant r) =$$

$$= \frac{T!}{(T-r)!n^T} \sum_{i=0}^{T-r} C^i_{T-r}(-t)^i n^{T-r-i} I(T \geqslant r) =$$

$$= r! C^r_T \frac{1}{n^T}(n-t)^{T-r}$$

Общий результат

Этому результату можно придать следующую окончательную форму:

$$\tau^*_{r,t} = r! C^r_T \frac{1}{n^r}\left(1 - \frac{t}{n}\right)^{T-} \tag{8}$$

Подчеркнем, что в данном случае оцениваемая функция $\tau(\theta)$ зависит от двух свободных параметров: $t < n$ и целого $r \geqslant 0$. Придавая этим параметрам конкретные значения, можно из (8) получать самостоятельные интересные результаты.

Специальные случаи

Так, при $r = 0$, $t = -1$ функция $\tau_{r,t}(\theta) = f(\theta)$ и из (8) имеем оценку

$$f^* = \left(1 + \frac{1}{n}\right)^T$$

согласующуюся с общим результатом (3).

При $t = 1$ функция $\tau_{r,t}(\theta)/r! = e^{-\theta}\theta^r/r!$ есть теоретическая вероятность $\mathbf{P}_\theta\{\xi = r\} = f(r|\theta)$ и для нее из (8) имеем оптимальную оценку

$$f^*(r) = C_T^r \frac{1}{n^r}\left(1 - \frac{1}{n}\right)^{T-r} \qquad r = 0, 1, \quad , T$$

Если, например, требуется оценить вероятность

$$\tau(\theta) = \mathbf{P}_\theta\{\xi \leqslant k\} = \sum_{r=0}^{k} f(r|\theta),$$

то оптимальной оценкой будет статистика

$$\tau^* = \sum_{r=0}^{k} f^*(r) = \begin{cases} \displaystyle\sum_{r=0}^{k} C_T^r \frac{1}{n^r}\left(1 - \frac{1}{n}\right)^{T-r} & \text{при} \quad T > k, \\ 1 & \text{при} \quad T \leqslant k. \end{cases}$$

Наконец, при $t = 0$ из (8) имеем вид оптимальной оценки для $\tau_r(\theta) = \theta^r$

$$\tau_r^* = r!\frac{C_T^r}{n^r} = \frac{T(T-1) \quad (T-r+1)}{n^r} = \frac{(T)_r}{n^r}. \tag{9}$$

В частности, оптимальная оценка самого параметра θ есть выборочное среднее: $\theta^* = T/n = \overline{X}$, что согласуется с результатом, указанным в табл. 2 § 3.2. •

Пример 2 (Отрицательная биномиальная модель, оценивание параметрических функций). По выборке $X = (X_1, \ldots, X_n)$ требуется оценить параметрическую функцию $\tau_{s,t}(\theta) = \theta^s(1-\theta)^t \quad t < nr$, модели $\overline{\mathrm{Bi}}(r, \theta)$. Здесь (см. п. 8 § 1.1)

$$f^n(\theta) = (1-\theta)^{-nr} = \sum_{x=0}^{\infty} C_{nr+x-1}^x \theta^x$$

т. е. $a_n(x) = C_{nr+x-1}^x$, $x = 0, 1, 2, \qquad$ а

$$f^n(\theta)\tau_{s,t}(\theta) = \theta^s(1-\theta)^{-(nr-t)} = \sum_{i=0}^{\infty} C_{nr-t+i-1}^i \theta^{s+i}$$

Следовательно, уравнение несмещенности (1) запишется в виде

$$\sum_{x=0}^{\infty} H(x)a_n(x)\theta^x = \sum_{i=0}^{\infty} C_{nr-t+i-1}^i \theta^{s+i},$$

откуда находим, что

$$H(x) = \frac{C_{nr-t+x-s-1}^{x-s}}{C_{nr+x-1}^{x}}.$$

Общий результат

Таким образом, оптимальная оценка $\tau_{s,t}^*$ имеет вид

$$\tau_{s,t}^* = H(T) = \frac{C_{nr+T-s-t-1}^{T-s}}{C_{nr+T-1}^{T}}. \tag{10}$$

Специальные случаи

Это значительное усиление результата примера 3 § 3.1. В частности, полагая $s = 1$, $t = 0$, получим вид оптимальной оценки по n наблюдениям для параметра θ:

$$\theta^* = \tau_{1,0}^* = \frac{C_{nr+T-2}^{T-1}}{C_{nr+T-1}^{T}} = \frac{T}{nr+T-1} \tag{11}$$

(сравни с (7) § 3.1 при $s = 1$).

Кроме того, поскольку (см. п. 2 § 1.1) $\mathcal{L}_\theta(T)$ $\overline{\text{Bi}}(nr, \theta)$, то $\mathbf{E}_\theta T = nr\theta/(1-\theta)$, т.е. статистика $T/nr = \overline{X}/r$ является оптимальной оценкой для параметрической функции $\tau(\theta) = \theta/(1-\theta)$, что согласуется с результатом, указанным в табл. 2 § 3.2. •

2. Модели с выборочным пространством, зависящим от параметра θ

Пусть носителем распределения наблюдаемой случайной величины ξ является отрезок $[a, \theta]$ и теоретическая плотность имеет вид

$$f(x;\theta) = Q_1(\theta)M_1(x), \quad a \leqslant x \leqslant \theta \tag{12}$$

(вне отрезка $[a, \theta]$ плотность равна нулю). Здесь функция $Q_1(\theta) > 0$ для всех $\theta \geqslant a$ и дифференцируема, кроме того, выполняется условие нормировки

$$\int_a^\theta f(x;\theta)\,dx = 1, \quad \text{т.е.} \quad \int_a^\theta M_1(x)\,dx = \frac{1}{Q_1(\theta)}.$$

Дифференцируя это тождество по θ, имеем

$$M_1(\theta) = -\frac{Q_1'(\theta)}{Q_1^2(\theta)}$$

— этим соотношением связаны функции M_1 и Q_1 в (12).

Пусть теперь дана выборка $X = (X_1, \quad, X_n)$ из указанного распределения. Записав функцию правдоподобия с помощью индикаторов в виде

$$L(\underline{x};\theta) = \prod_{i=1}^n f(x_i;\theta) = Q_1^n(\theta)\prod_{i=1}^n M_1(x_i)\prod_{i=1}^n I(a \leqslant x_i \leqslant \theta) =$$

$$= Q_1^n(\theta)\, I(x_{(n)} \leqslant \theta)\prod_{i=1}^n M_1(x_i)\, I(x_{(1)} \geqslant a),$$

по критерию факторизации получаем, что достаточной статистикой в данной модели является $T = X_{(n)}$ — максимальное значение выборки. Найдем ее распределение: при $t \in [a, \theta]$

$$F_T(t;\theta) = \mathrm{P}_\theta\{T \leqslant t\} = \mathrm{P}_\theta^n\{\xi \leqslant t\} = Q_1^n(\theta)\Big(\int_a^t M_1(x)\,dx\Big)^n = \frac{Q_1^n(\theta)}{Q_1^n(t)},$$

откуда для плотности $g_1(t;\theta) = F_T'(t;\theta)$ получаем представление

$$g_1(t;\theta) = -\frac{nQ_1^n(\theta)Q_1'(t)}{Q_1^{n+1}(t)} = \frac{nM_1(t)Q_1^n(\theta)}{Q_1^{n-1}(t)}, \quad t \in [a, \theta]. \tag{13}$$

Для проверки полноты статистики T надо убедиться в том, что несмещенная оценка нуля в этой модели единственна, т. е. решить однородное уравнение $\mathrm{E}_\theta\varphi(T) = 0$, $\forall\,\theta \geqslant a$. В нашем случае оно эквивалентно уравнению

$$\int_a^\theta \varphi(t)M_1(t)Q_1^{1-n}(t)\,dt = 0, \quad \forall\,\theta \geqslant a.$$

Дифференцируя это тождество по θ, получаем $\varphi(\theta)M_1(\theta)Q_1^{1-n}(\theta) = 0, \forall\,\theta \geqslant a$, откуда $\varphi(\theta) = 0$, $\forall\,\theta \geqslant a$. Таким образом, свойство полноты здесь имеет место, и можно приступать к решению уравнения несмещенности для заданной параметрической функции $\tau(\theta)$:

$$\int_a^\theta H(t)g_1(t;\theta)\,dt = \tau(\theta), \quad \forall\,\theta \geqslant a. \tag{14}$$

Далее мы считаем, что функция $\tau(\theta)$ дифференцируема. Переписав уравнение (14) с учетом (13) в виде

$$n\int_a^\theta H(t)M_1(t)Q_1^{1-n}(t)\,dt = \frac{\tau(\theta)}{Q_1^n(\theta)}, \quad \forall\,\theta \geqslant a,$$

и продифференцировав это тождество по θ, получим

$$nH(\theta)M_1(\theta)Q_1^{1-n}(\theta) = \frac{\tau'(\theta)Q_1(\theta) - n\tau(\theta)Q_1'(\theta)}{Q_1^{n+1}(\theta)},$$

или

$$H(\theta) = \frac{\tau'(\theta)Q_1(\theta) - n\tau(\theta)Q_1'(\theta)}{nM_1(\theta)Q_1^2(\theta)} = \tau(\theta) + \frac{\tau'(\theta)}{nM_1(\theta)Q_1(\theta)} = \tau(\theta) + \frac{\tau'(\theta)}{nf(\theta;\theta)}.$$

Общий результат

В итоге, мы имеем следующий общий результат: *в модели* (12) *оптимальной оценкой для произвольной*

дифференцируемой функции $\tau(\theta)$ является статистика

$$\tau^* = H\big(X_{(n)}\big) = \tau(X_{(n)}) + \frac{\tau'(X_{(n)})}{nf(X_{(n)};\, X_{(n)})}. \tag{15}$$

Замечание. Если рассмотреть «симметричную» модель (неизвестен левый конец носителя распределения):

$$f(x;\theta) = Q_2(\theta)M_2(x), \quad \theta \leqslant x \leqslant b, \tag{16}$$

то аналогичными рассуждениями можно установить, что полной достаточной статистикой теперь является минимальное значение выборки $X_{(1)}$, ее плотность имеет вид

$$g_2(t;\theta) = nM_2(t)\frac{Q_2^n(\theta)}{Q_2^{n-1}(t)}, \quad t \in [\theta, b], \tag{17}$$

а оптимальной оценкой для произвольной дифференцируемой функции $\tau(\theta)$ является статистика

$$\tau^* = \tau(X_{(1)}) - \frac{\tau'(X_{(1)})}{nf(X_{(1)}, X_{(1)})} \tag{18}$$

Проверку этих утверждений мы оставляем читателю в качестве упражнения, а далее мы проиллюстрируем изложенную теорию несколькими примерами.

Пример 3 (Равномерная модель, оценивание параметрических функций). Пусть $\mathcal{L}(\xi) \in \mathcal{U}(0,\theta)$, $\theta > 0$. Это модель вида (12) с $a = 0$, $Q_1(\theta) = 1/\theta$, $M_1(x) \equiv 1$, поэтому для нее «работает» результат (15). В частности, если $\tau_r(\theta) = \theta^r$ то оптимальная оценка для этой функции по n наблюдениям есть

$$\tau_r^* = X_{(n)}^r + \frac{rX_{(n)}^{r-1}}{nf(X_{(n)};\, X_{(n)})} = \frac{n+r}{n}X_{(n)}^r \tag{19}$$

(раньше, см. введение к данному параграфу, мы установили этот результат для $r = 1$). Обратим внимание на то, что здесь $\alpha_1 = \mathbf{E}_\theta\xi = \theta/2$, и, следовательно, оптимальная оценка для теоретического среднего α_1 есть

$$\alpha_1^* = \frac{1}{2}\tau_1^* = \frac{n+1}{2n}X_{(n)},$$

а не выборочное среднее $\overline{X}$(!). При этом

$$\mathbf{D}_\theta\overline{X} = \frac{1}{n}\mathbf{D}_\theta\xi = \frac{\theta^2}{12n},$$

а в силу упр. 46 к гл. I

$$\mathbf{D}_\theta\alpha_1^* = \frac{(n+1)^2}{4n^2}\mathbf{D}_\theta X_{(n)} = \frac{(n+1)^2}{4n^2}\ \frac{n\theta^2}{(n+1)^2(n+2)} = \frac{\theta^2}{4n(n+2)}.$$

Следовательно, эффективность выборочного среднего по отношению к оптимальной оценке равна

$$\frac{\mathbf{D}_\theta\alpha_1^*}{\mathbf{D}_\theta\overline{X}} = \frac{3}{n+2} < 1 \quad \text{при} \quad n \geqslant 2.$$

Этот пример показывает, что не всегда выборочное среднее наилучшим образом оценивает теоретическое среднее модели. •

Пример 4 (Показательная модель с параметрами сдвига, оценивание параметрических функций). Пусть плотность наблюдаемой случайной величины ξ есть $f(x;\theta) = e^{-(x-\theta)}$, $x \geqslant 0$. Требуется оценить $\tau_r(\theta) = \theta^r$ по выборке объема n. Это модель вида (16) с $b = \infty$, $Q_2(\theta) = e^{\theta}$ $M_2(x) = e^{-x}$ поэтому по формуле (18)

$$\tau_r^* = X_{(1)}^r - \frac{r}{n} X_{(1)}^{r-1} \tag{20}$$

•

Пример 5 (Приведенное показательное распределение, оценивание параметрических функций). Приведенное к отрезку $[\theta, b]$ показательное распределение задается плотностью $f(x;\theta) = e^{-x}/(e^{-\theta} - e^{-b})$, $x \in [\theta, b]$. Это распределение имеет вид (16), поэтому, применяя (18), получим, что оптимальной оценкой произвольной степени параметра $\tau_r(\theta) = \theta^r$ является следующая статистика от выборки объема n:

$$\tau_r^* = X_{(1)}^r - \frac{r}{n} X_{(1)}^{r-1}(1 - e^{X_{(1)}-b}). \tag{21}$$

•

3. Метод подгонки

Здесь речь будет идти о весьма простом, но очень эффективном в ряде случаев приеме решения уравнения несмещенности $\mathbf{E}_\theta H(T) = \tau(\theta)$. Предположим, что $\tau(\theta) > 0$ для всех $\theta \in \Theta$, тогда можем записать тождество

$$\mathbf{E}_\theta \left[\frac{H(T)}{\tau(\theta)} \right] \equiv 1, \quad \text{или} \quad \int \frac{H(t) g(t;\theta)}{\tau(\theta)}\, dt \equiv 1.$$

Пусть здесь (ключевой момент!) имеет место факторизация

$$\frac{g(t;\theta)}{\tau(\theta)} = u(t) g_1(t;\theta), \tag{22}$$

где функция $g_1(t;\theta)$ принадлежит тому же семейству плотностей, что и плотность $g(t;\theta)$ полной достаточной статистики T. Тогда мы имеем тождество

$$\int H(t) u(t) g_1(t;\theta)\, dt \equiv 1,$$

которому удовлетворяет (в силу полноты статистики T) единственная функция $H(t) = u^{-1}(t)$. Это дает решение исходного уравнения несмещенности: $\tau^* = u^{-1}(T)$ (для дискретной модели, в этих рассуждениях интегрирование надо заменить суммированием). Таким образом, если в конкретной задаче удается установить факторизацию (22), то мы тем самым получаем одновременно и вид искомой оптимальной оценки для $\tau(\theta)$. Продемонстрируем эффективность такого приема соответствующими примерами.

Пример 6 (Нормальная модель с неизвестным средним, оценивание параметрических функций). Пусть $X=(X_1,\dots,X_n)$ — выборка из $\mathcal{L}(\xi)\in\mathcal{N}(\theta,\sigma^2)$. Здесь, как мы уже знаем, полной достаточной статистикой T является выборочное среднее $\overline{X}$ и $\mathcal{L}_\theta(\overline{X})=\mathcal{N}(\theta,\sigma^2/n)$, т. е.

$$g(t;\theta)\equiv f_{\sigma^2/n}(t;\theta)=\left(\frac{n}{2\pi\sigma^2}\right)^{1/2}\exp\left\{-\frac{n}{2\sigma^2}(t-\theta)^2\right\}.$$

Рассмотрим сначала задачу оценивания теоретической плотности $\tau(\theta)=f_{\sigma^2}(x;\theta)$ в произвольной точке x. Здесь уравнение факторизации (22) принимает вид

$$\frac{f_{\sigma^2/n}(t;\theta)}{f_{\sigma^2}(x;\theta)}=\sqrt{n}\exp\left\{-\frac{n}{2\sigma^2}(t-\theta)^2+\frac{1}{2\sigma^2}(x-\theta)^2\right\}=$$

$$=\sqrt{n}\exp\left\{\frac{n}{2(n-1)\sigma^2}(x-t)^2-\frac{n^2}{2(n-1)\sigma^2}\left(t-\frac{(n-1)\theta+x}{n}\right)^2\right\}=$$

$$=f^{-1}_{\sigma_1^2}(x;t)f_{\sigma_1^2/n}\left(t;\frac{(n-1)\theta+x}{n}\right),\quad\text{где}\quad\sigma_1^2=\frac{n-1}{n}\sigma^2,$$

т. е.

$$u(t)=f^{-1}_{\sigma_1^2}(x;t),\quad\text{а}\quad g_1(t;\theta)=f_{\sigma_1^2/n}\left(t;\frac{(n-1)\theta+x}{n}\right)$$

Таким образом, $g_1(t;\theta)$ является нормальной плотностью такого же вида, как и $g(t;\theta)$, и тем самым решением уравнения несмещенности

$$\int H(t)g(t;\theta)\,dt=\tau(\theta)$$

Оптимальная оценка нормальной плотности

является функция $H(t)=u^{-1}(t)=f_{\sigma_1^2}(x;t)$ — нормальная плотность со средним t и дисперсией σ_1^2. В итоге имеем, что оптимальной оценкой теоретической плотности $f_{\sigma^2}(x;\theta)$ является статистика

$$f^*(x)=f_{\sigma_1^2}(x;\overline{X}).\tag{23}$$

Специальный случай

Отсюда можно вывести следующее важное заключение: если оценивается параметрическая функция вида

$$\tau(\theta;\sigma^2)=\mathbf{E}_\theta\psi(\xi)=\int\psi(x)f_{\sigma^2}(x;\theta)\,dx,\tag{24}$$

то оптимальная оценка для нее получается по формуле

$$\tau^*=\int\psi(x)f^*(x)\,dx=\int\psi(x)f_{\sigma_1^2}(x;\overline{X})\,dx=\tau(\overline{X};\sigma_1^2),\tag{25}$$

т. е. в выражении для $\tau(\theta;\sigma^2)$ надо просто заменить параметр θ на статистику $\overline{X}$, а величину σ^2 — на $\sigma_1^2=((n-1)/n)\sigma^2$.

Так, если $\psi(x) = e^{itx}$ то получаем характеристическую функцию

$$\tau(\theta; \sigma^2) = \mathbf{E}_\theta e^{it\xi} = \exp\left\{it\theta - \frac{\sigma^2}{2}t^2\right\}$$

(см. упр. 26 к гл. 1), следовательно оптимальной оценкой для нее является статистика

Оптимальная оценка нормальной характеристической функции

$$\tau^* = \exp\left\{it\overline{X} - \frac{\sigma_1^2}{2}t^2\right\} \qquad (26)$$

В качестве еще одного конкретного примера возьмем функцию $\psi(x) = x^2$ Тогда из (24) имеем

$$\tau(\theta; \sigma^2) = \mathbf{E}_\theta \xi^2 = \mathbf{D}_\theta \xi + \left(\mathbf{E}_\theta \xi\right)^2 = \sigma^2 + \theta^2$$

и, следовательно, в силу (25), $\tau^* = \sigma_1^2 + \overline{X}^2$ или $(\theta^2)^* = \tau^* - \sigma^2 = \overline{X}^2 - \sigma^2/n$ — оптимальная оценка для θ^2

Наконец, если взять $\psi(x) = I(x \leqslant x_0)$, то

$$\tau(\theta; \sigma^2) = \mathbf{P}_\theta\{\xi \leqslant x_0\} = \Phi\left(\frac{x_0 - \theta}{\sigma}\right)$$

и по формуле (25)

Оптимальная оценка нормальной функции распределения

$$\tau^* = \Phi\left(\sqrt{\frac{n}{n-1}}\,\frac{x_0 - \overline{X}}{\sigma}\right)$$

— с этим результатом мы уже встречались в примере 6 § 3.3 (см. там соотношение (14)). •

Пример 7 (Биномиальная модель, оценивание параметрических функций). Пусть требуется по выборке $\boldsymbol{X} = (X_1, \ldots, X_n)$ из биномиального распределения $\mathrm{Bi}(k, \theta)$ оценить параметрическую функцию $\tau_{r,s}(\theta)$ $\theta^r(1-\theta)^s$ где r, s — заданные неотрицательные целые числа такие, что $r + s \leqslant kn$. Поскольку модель $\mathrm{Bi}(k, \theta)$ входит в экспоненциальное семейство, то статистика $T = X_1 + \ldots + X_n$ является достаточной и полной, и ее распределение $\mathcal{L}_\theta(T) = \mathrm{Bi}(kn, \theta)$ (см. п. 1 § 1.1). В данном случае уравнение факторизации (22) запишется в виде

$$C_{kn}^t \theta^{t-r}(1-\theta)^{kn-t-s} = C_{kn}^t \left(C_{kn-r-s}^{t-r}\right)^{-1} C_{kn-r-s}^{t-r} \theta^{t-r}(1-\theta)^{kn-r-}$$

т. е. здесь $g_1(t; \theta) = C_{kn-r-s}^{t-r}\theta^{t-r}(1-\theta)^{kn-r-\ -(t-r)}$ — биномиальные вероятности в точках $t = r,\ r+1,\quad , kn - s$, а функция $u(t) = C_{kn}^t \left(C_{kn-r-s}^{t-r}\right)^{-1}$ Значит, оптимальная оценка функции $\tau_{r,s}(\theta)$ есть

$$\tau_{r,s}^* = u^{-1}(T) = \frac{C_{kn-r-s}^{T-r}}{C_{kn}^T}. \qquad (27)$$

В частности, при $r = 1$, $s = 0$ получаем оптимальную оценку параметра θ:

$$\theta^* = \tau^*_{1,0} = \frac{C^{T-1}_{kn-1}}{C^T_{kn}} = \frac{T}{kn} = \frac{1}{k}\overline{X}$$

— этот результат мы уже имели в § 3.2 (см. табл. 2). •

Пример 8 (Оценивание размера конечной совокупности: выбор с возвращением). Пусть в урне находятся занумерованные числами $1, 2, \quad , \theta$ шары, но их число θ нам неизвестно. По схеме случайного выбора с возвращением из урны извлекается n шаров и при этом каждый раз фиксируется номер извлеченного шара. Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ — номера извлекавшихся шаров. Требуется по этой информации оценить общее число шаров в урне θ и, более общо, прозвольную степень $\tau_r(\theta) = \theta^r$ где $r > -n$ (целое). Если ξ обозначает номер случайно извлеченного шара, то распределение этой случайной величины задается вероятностями

$$f(x;\theta) = \mathrm{P}_\theta\{\xi = x\} = \frac{1}{\theta}, \quad x = 1, \quad , \theta, \quad \text{или} \quad f(x;\theta) = \frac{1}{\theta} I(x \leqslant \theta).$$

Отсюда для функции правдоподобия имеем представление

$$L(\underline{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \theta^{-n} I\big(\max_i x_i \leqslant \theta\big) = \theta^{-n} I(x_{(n)} \leqslant \theta).$$

По критерию факторизации отсюда делаем вывод, что максимальный наблюдавшийся номер $X_{(n)}$ является достаточной статистикой для параметра θ. Найдем ее распределение. Имеем

$$\mathbf{P}_\theta\big\{X_{(n)} \leqslant t\big\} = \mathbf{P}_\theta\big\{X_i \leqslant t,\ i = 1, \quad , n\big\} = \left(\frac{t}{\theta}\right)^n \quad t \leqslant \theta.$$

Следовательно,

$$g(t;\theta) = \mathbf{P}_\theta\{X_{(n)} = t\} = \left(\frac{t}{\theta}\right)^n \quad \left(\frac{t-1}{\theta}\right)^n \quad t = 1, 2, \quad , \theta. \qquad (28)$$

Убедимся теперь в полноте статистики $X_{(n)}$. Для этого запишем уравнение несмещенного оценивания нуля:

$$\sum_{t=1}^{\theta} \varphi(t) g(t;\theta) = 0, \quad \forall\, \theta \geqslant 1$$

Полагая здесь последовательно $\theta = 1$, $\theta = 2$, очевидно, получим, что $\varphi(1) = \varphi(2) = \quad = 0$, т. е. этому условию удовлетворяет лишь нулевая (на множестве значений статистики $X_{(n)}$) функция, следовательно, статистика $X_{(n)}$ — полная.

Для оценивания функции $\tau_r(\theta) = \theta^r$ применим метод подгонки. Здесь уравнение факторизации (22) записывается в виде

$$\frac{t^n - (t-1)^n}{\theta^{n+r}} = \frac{t^n - (t-1)^n}{t^{n+r} - (t-1)^{n+r}} \left[\left(\frac{t}{\theta}\right)^{n+r} - \left(\frac{t-1}{\theta}\right)^{n+r}\right],$$

т. е. новая плотность

$$g_1(t;\theta) = \left(\frac{t}{\theta}\right)^{n+r} - \left(\frac{t-1}{\theta}\right)^{n+r}$$

имеет такой же вид, как и $g(t;\theta)$, а функция

$$u(t) = \frac{t^n - (t-1)^n}{t^{n+r} - (t-1)^{n+r}}.$$

Следовательно, оптимальная оценка для $\tau_r(\theta)$ есть

$$\tau_r^* = u^{-1}(X_{(n)}) = \frac{X_{(n)}^{n+r} - \left(X_{(n)} - 1\right)^{n+r}}{X_{(n)}^n - \left(X_{(n)} - 1\right)^n} = X_{(n)}^r \frac{1 - \left(1 - X_{(n)}^{-1}\right)^{n+r}}{1 - \left(1 - X_{(n)}^{-1}\right)^n}. \quad (29)$$

Отметим, что если $X_{(n)}$ велико, то, воспользовавшись приближением

$$\left(1 - X_{(n)}^{-1}\right)^k \approx 1 - kX_{(n)}^{-1},$$

формуле (29) можно придать приближенную форму

$$\tau_r^* \approx \frac{n+r}{n} X_{(n)}^r.$$

Оптимальная оценка числа шаров в урне

В частности, оптимальная оценка числа θ шаров в урне имеет вид

$$\theta^* = \tau_1^* = X_{(n)} \frac{1 - \left(1 - X_{(n)}^{-1}\right)^{n+1}}{1 - \left(1 - X_{(n)}^{-1}\right)^n} \approx \frac{n+1}{n} X_{(n)}. \quad (30)$$

Игровую интерпретацию этой задачи дает В. Феллер [25, с. 240], предлагая для оценки числа зарегистрированных в городе автомобилей встать на перекрестке и записывать номера $X_1, \ldots, X_n$ проезжающих автомобилей. Если в городе имеется 1000 машин и производится случайная выборка объема $n = 10$, то математическое ожидание максимального наблюденного регистрационного номера $\mathbf{E}_\theta X_{(n)} \approx (n/(n+1))\theta$ приближенно равно 910.

Иллюстрацией статистической процедуры (30) являются следующие результаты, полученные моделированием с помощью таблиц случайных чисел: при $\theta = 10\,000$, $n = 10$ значения оценки (30) для десяти независимых выборок оказались равными (с точностью до единицы) 10 549, 9841, 10 486, 10 891, 10 462, 10 873, 8789, 10 753, 9960, 10 287. Эти данные свидетельствуют о хорошем качестве оценки (30). •

Замечание. Мы знаем, что оценка (30) имеет наименьшую дисперсию среди всех несмещенных оценок θ. Но какова величина этой дисперсии $\mathbf{D}_\theta\theta^*$? На практике, чтобы судить о точности оценки, важно уметь оценивать ее дисперсию по имеющимся данным (по выборке). В таких случаях оптимальную несмещенную оценку $\mathbf{D}^*$ для $\mathbf{D}_\theta\theta^*$ можно построить по формуле

$$\mathbf{D}^* = (\theta^*)^2 - (\theta^2)^*. \quad (31)$$

Действительно, так как $\mathbf{E}_\theta\theta^* = \theta$, $\mathbf{E}_\theta(\theta^2)^* = \theta^2$ то

$$\mathbf{E}_\theta \mathbf{D}^* = \mathbf{E}_\theta(\theta^*)^2 - \mathbf{E}_\theta(\theta^2)^* = \mathbf{E}_\theta(\theta^*)^2 - \theta^2 = \mathbf{E}_\theta(\theta^*)^2 - (\mathbf{E}_\theta\theta^*)^2 = \mathbf{D}_\theta\theta^*,$$

а поскольку $\mathbf{D}^*$ есть функция от полной достаточной статистики, то $\mathbf{D}^*$ есть н. о. р. м. д. Так, в рассматриваемом случае на основании (29) имеем

$$\mathbf{D}^* = (\tau_1^*)^2 - \tau_2^* = X_{(n)}^2\left[\left(\frac{1-\left(1-X_{(n)}^{-1}\right)^{n+1}}{1-\left(1-X_{(n)}^{-1}\right)^{n}}\right)^2 - \frac{1-\left(1-X_{(n)}^{-1}\right)^{n+2}}{1-\left(1-X_{(n)}^{-1}\right)^{n}}\right] \approx \frac{1}{n^2}X_{(n)}^2$$

(приближенное равенство справедливо при больших значениях $X_{(n)}$).

Замечание. В данной модели теоретическое среднее есть

$$\alpha_1 = \mathbf{E}_\theta\xi = \frac{1}{\theta}\sum_{x=1}^{\theta} x = \frac{\theta+1}{2}$$

и оптимальной оценкой для него является статистика $\alpha_1^* = (\tau_1^* + 1)/2$, но не выборочное среднее $\overline{X}$ (!).

Пример 9 (Оценивание размера конечной совокупности: выбор без возвращения). Предположим, что в ситуации предыдущего примера производится случайный выбор без возвращения n шаров или, что то же самое, извлекается сразу комплект из n шаров, причем все C_θ^n возможных вариантов выбора равновероятны. Снова пусть $X_{(n)}$ обозначает максимальный наблюденный номер. Тогда $X_{(n)}$ будет также полной достаточной статистикой и

$$g(t;\theta) = \mathbf{P}_\theta\{X_{(n)} = t\} = \frac{C_{t-1}^{n-1}}{C_\theta^n}, \quad t = n, n+1, \quad , \theta. \tag{32}$$

Здесь метод подгонки удобен для оценивания параметрических функций вида

$$\tau_r(\theta) = (\theta+1) \quad (\theta+r) = \frac{(\theta+r)!}{\theta!}, \quad r > -n \quad \text{(целое)}. \tag{33}$$

Записав уравнение факторизации (22):

$$\frac{C_{t-1}^{n-1}(C_\theta^n)^{-1}\theta!}{(\theta+r)!} = \frac{n}{n+r}\,\frac{(t-1)!}{(t+r-1)!}\,\frac{C_{t+r-1}^{n+r-1}}{C_{\theta+r}^{n+r}},$$

имеем, что

$$g_1(t;\theta) = \frac{C_{t+r-1}^{n+r-1}}{C_{\theta+r}^{n+r}}$$

— плотность такого же вида, как и $g(t;\theta)$, а функция

$$u(t) = \frac{n}{n+r}\quad\frac{(t-1)!}{(t+r-1)!}$$

Следовательно, оптимальная оценка для функции $\tau_r(\theta)$ имеет вид

$$\tau_r^* = u^{-1}(X_{(n)}) = \frac{n+r}{n}\quad\frac{(X_{(n)}+r-1)!}{(X_{(n)}-1)!} =$$

$$= \frac{n+r}{n} X_{(n)}(X_{(n)}+1) \quad (X_{(n)}+r-1). \tag{34}$$

Если $X_{(n)}$ велико, то формуле (34) можно придать приближенную форму

$$\tau_r^* \approx \frac{n+r}{n} X_{(n)}^r,$$

т. е. получаем такой же результат, как и в предыдущем примере для схемы выбора с возвращением.

В частности, оптимальная оценка числа θ шаров в урне есть

Оптимальная оценка числа шаров в урне

$$\theta^* = \tau_1^* - 1 = \frac{n+1}{n} X_{(n)} - 1 \tag{35}$$

(сравни с (30)), а оптимальная оценка для дисперсии $\mathbf{D}_\theta \theta^*$ имеет вид (см. (31) и (33))

$$\mathbf{D}^* = (\theta^*)^2 - (\theta^2)^* = (\tau_1^* - 1)^2 - \tau_2^* + 3\tau_1^* - 1 = (\tau_1^*)^2 + \tau_1^* - \tau_2^*.$$

Учитывая (34), окончательно находим

$$\mathbf{D}^* = \frac{X_{(n)}(X_{(n)} - n)}{n^2}. \quad \bullet$$

Замечание. Поскольку (34) есть несмещенная оценка для (33), то мы имеем общую формулу для рассматриваемой модели:

$$\mathbf{E}_\theta X_{(n)}(X_{(n)}+1)\dots(X_{(n)}+r-1) = \frac{n}{n+r}(\theta+1) \quad (\theta+r). \tag{36}$$

В частности,

$$\mathbf{E}_\theta X_{(n)} = \frac{n}{n+1}(\theta+1),$$

$$\mathbf{D}_\theta X_{(n)} = \mathbf{E}_\theta X_{(n)}(X_{(n)}+1) - \mathbf{E}_\theta X_{(n)} - (\mathbf{E}_\theta X_{(n)})^2 = \frac{n}{(n+1)^2(n+2)}(\theta+1)(\theta-n).$$

Учитывая (35), можем также записать

$$\mathbf{D}_\theta \theta^* = \frac{(n+1)^2}{n^2} \mathbf{D}_\theta X_{(n)} = \frac{(\theta+1)(\theta-n)}{n(n+2)}.$$

Следовательно, оптимальная оценка для $\tau(\theta) = \mathbf{D}_\theta X_{(n)}$ есть

$$\tau^* = \frac{n^2}{(n+1)^2} \mathbf{D}^* = \frac{(X_{(n)} - n)}{X_{(n)}(n+1)^2}.$$

4. Гамма-модель с неизвестным параметром масштаба, оценивание параметрических функций

Определение и основные свойства модели $\Gamma(\theta, \lambda)$, где $\theta > 0$ — неизвестный параметр масштаба (параметр формы λ далее считается известным) приведены в п. 3 § 1.2 (см. также пример 1 в § 2.4 о свойствах порядковых статистик

для выборки из $\Gamma(1, 1)$). Отдельные результаты об оценивании для этой модели нам уже встречались ранее. Так в табл. 2 § 3.2 указана эффективная (следовательно, оптимальная) оценка для самого параметра θ — это статистика $\overline{X}/\lambda$. Там же отмечено, что $\Gamma(\theta, \lambda)$ — это модель экспоненциального типа, и для нее полной достаточной статистикой является сумма наблюдений $T = T(X) = X_1 + . \quad + X_n$. По воспроизводящему свойству гамма-распределения $\mathcal{L}_\theta(T(X)) = \Gamma(\theta, n\lambda)$, т. е. (см. (18) § 1.2) плотность статистики T есть

$$g(t; \theta) = f(t|\theta, n\lambda) = \frac{t^{n\lambda-1}e^{-t/\theta}}{\Gamma(n\lambda)\theta^{n\lambda}}, \quad t \geqslant 0. \tag{37}$$

Основываясь на этих фактах, в данном пункте мы решим задачу оптимального оценивания произвольных параметрических функций для этой модели.

Мы начнем с построения оценки для параметрической функции

$$\tau_{r,x}(\theta) = \theta^r e^{-x/\theta} \qquad r > -n\lambda, \quad x \geqslant 0. \tag{38}$$

Здесь два свободных параметра: r и x, поэтому (38) на самом деле задает двухпараметрическое семейство функций от θ. Для отыскания оптимальной оценки $\tau^*_{r,x}$ мы воспользуемся методом подгонки. Уравнение факторизации (22) этого метода в данном случае принимает вид

$$\frac{t^{n\lambda-1}e^{-(t-x)/\theta}}{\Gamma(n\lambda)\theta^{n\lambda+r}} = \frac{\Gamma(n\lambda+r)}{\Gamma(n\lambda)} \frac{t^{n\lambda-1}}{(t-x)^{n\lambda+r-1}} f(t-x|\theta, n\lambda+r), \quad t \geqslant x,$$

т. е. $g_1(t; \theta) = f(t - x|\theta, n\lambda + r)$ — гамма-плотность вида (37), а функция

$$u(t) = \frac{\Gamma(n\lambda+r)}{\Gamma(n\lambda)} \frac{t^{n\lambda-1}}{(t-x)^{n\lambda+r-1}} I(t \geqslant x).$$

Общий результат

Следовательно, искомая оптимальная оценка есть

$$\tau^*_{r,x} = u^{-1}(T) = \frac{\Gamma(n\lambda)}{\Gamma(n\lambda+r)} \frac{(T-x)^{n\lambda+r-1}}{T^{n\lambda-1}} I(T \geqslant x). \tag{39}$$

Выделим важный частный случай этого результата. Если в (38) положить $r = -\lambda$ (при $n \geqslant 2$) и домножить функцию $\tau_{-\lambda,x}(\theta)$ на множитель $x^{\lambda-1}/\Gamma(\lambda)$, то получим теоретическую плотность $f(x|\theta, \lambda)$. Следовательно, из (39) полу-

Оптимальная оценка гамма-плотности

чаем оптимальную несмещенную оценку $f^*(x)$ для теоретической плотности гамма-модели $\Gamma(\theta, \lambda)$:

$$f^*(x) = \frac{\Gamma(n\lambda)}{\Gamma(\lambda)\Gamma\big((n-1)\lambda\big)} \frac{x^{\lambda-1}(T-x)^{(n-1)\lambda-1}}{T^{n\lambda-1}} I(T \geqslant x). \tag{40}$$

Специальный случай

В свою очередь, этот результат можно применить для оценивания параметрических функций вида

$$\tau(\theta) = \mathbf{E}_\theta \varphi(\xi) = \int_0^\infty \varphi(x) f(x|\theta, \lambda)\, dx$$

— оптимальная оценка для $\tau(\theta)$ получается просто заменой в этом интеграле теоретической плотности $f(x|\theta, \lambda)$ ее оценкой (40). В итоге имеем

$$\tau^* = \int_0^\infty \varphi(x) f^*(x)\, dx = \frac{\Gamma(n\lambda) T^{1-n\lambda}}{\Gamma(\lambda)\Gamma\big((n-1)\lambda\big)} \int_0^T \varphi(x) x^{\lambda-1} (T-x)^{(n-1)\lambda-1}\, dx,$$

что после замены переменных $x \to xT$ приводит к формуле

$$\tau^* = \frac{\Gamma(n\lambda)}{\Gamma(\lambda)\Gamma\big((n-1)\lambda\big)} \int_0^1 \varphi(xT) x^{\lambda-1} (1-x)^{(n-1)\lambda-1}\, dx. \tag{41}$$

Пример 10. Пусть по выборке $\boldsymbol{X} = (X_1, \ldots, X_n)$, $n \geqslant 2$, из $\mathcal{L}(\xi) \in \Gamma(\theta, \lambda)$ требуется оценить преобразование Лапласа

$$\tau(\theta) = \mathbf{E}_\theta e^{-t\xi} = (1+\theta t)^{-\lambda}$$

при $t > 0$. По формуле (41) имеем

$$\tau^* = \frac{\Gamma(n\lambda)}{\Gamma(\lambda)\Gamma\big((n-1)\lambda\big)} \int_0^1 e^{-xtT} x^{\lambda-1} (1-x)^{(n-1)\lambda-1}\, dx.$$

Этот интеграл можно вычислить, разложив экспоненту e^{-xtT} в ряд и воспользовавшись формулой для бета-функции $B(a, b)$ (см. п. 4 § 1.2). В итоге получим

Оценивание преобразования Лапласа

$$\tau^* = \frac{\Gamma(n\lambda)}{\Gamma(\lambda)} \sum_{j=0}^\infty \frac{\Gamma(\lambda+j)}{j!\Gamma(n\lambda+j)} (-tT)^j \tag{42}$$

•

Пример 11 (Оценивание функции надежности). Функцией надежности называется вероятность $\tau_c(\theta) = \mathbf{P}_\theta\{\xi \geqslant c\} = \mathbf{E}_\theta I(\xi \geqslant c)$, где $c > 0$ — заданное число. Чтобы построить оптимальную оценку для $\tau(\theta)$ по $n \geqslant 2$ наблюдениям над ξ, нужно вычислить интеграл в (41) при $\varphi(x) = I(x \geqslant c)$, т. е.

$$\tau_c^* = \begin{cases} \dfrac{\Gamma(n\lambda)}{\Gamma(\lambda)\Gamma\big((n-1)\lambda\big)} \displaystyle\int_{c/T}^1 x^{\lambda-1}(1-x)^{(n-1)\lambda-1}\, dx & \text{при} \quad T > c, \\ 0 & \text{при} \quad T \leqslant c. \end{cases}$$

Используя неполную бета-функцию $B(z; a, b)$ (см. (4) § 2.4), этой формуле можно придать вид

$$\tau_c^* = \left[1 - B\left(\frac{c}{T}; \lambda, (n-1)\lambda\right)\right] I(T > c). \tag{43}$$

Выделим частный случай этого результата при $\lambda = 1$: в этом случае $\tau_c(\theta) = e^{-c/\theta}$ и

$$\tau_c^* = \left(1 - \frac{c}{T}\right)^{n-1} I(T > c). \tag{44}$$

•

Пример 12 (Нормальная модель с неизвестной дисперсией: оценивание параметрических функций). Полученные для гамма-модели результаты можно применять и к другим моделям, именно, к таким, в которых соответствующая достаточная статистика имеет гамма-распределение: ведь ключевым моментом выше при выводе оценки (39) было то, что достаточная статистика $T = T(X)$ имела гамма-распределение $\Gamma(\theta, n\lambda)$. Продемонстрируем это на примере исследования нормальной модели $\mathcal{N}(\mu, \theta^2)$. Эта модель экспоненциального типа, и как отмечено в п. 3 § 3.3 (см. табл. 2 § 3.2) полной достаточной статистикой для нее является

$$T = \sum_{i=1}^{n} (X_i - \mu)^2$$

Для нее, очевидно, $\mathbf{E}_\theta T = n\mathbf{D}_\theta X_1 = n\theta^2$ и потому оптимальной оценкой для θ^2 — теоретической дисперсии модели — является T/n, что соответствует результату, указанному в табл. 2 § 3.2. Но теперь мы можем продвинуться значительно дальше и построить оптимальные оценки для произвольных функций от параметра θ. Для этого найдем сначала распределение статистики T. Заметим, что $\mathcal{L}_\theta((X_i - \mu)/\theta) = \mathcal{N}(0, 1)$, $\forall\, i$, следовательно (см. упр. 38 к гл. 1), $\mathcal{L}_\theta(T/\theta^2) = \chi^2(n) = \Gamma(2, n/2)$. Отсюда имеем (см. п. 3 § 1.2), что $\mathcal{L}_\theta(T) = \Gamma(2\theta^2, n/2)$. Поэтому все полученные выше результаты для гамма-модели заменой $\theta \to 2\theta^2$, $\lambda \to 1/2$ переносятся на рассматриваемую модель.

Например, рассмотрим параметрическую функцию $\tau(\theta) = (2\theta^2)^r$ Ее оптимальная оценка получается из (39) при $x = 0$, $\lambda = 1/2$ и имеет, следовательно, вид

$$\tau^* = T^r \frac{\Gamma(n/2)}{\Gamma(n/2 + r)},$$

Общий результат для модели $\mathcal{N}(\mu, \theta^2)$

причем r здесь — любое действительное число, большее $-(n/2)$. В частности, любая степень θ^k при $k > -n$ оценивается статистикой

$$(\theta^k)^* = \left(\frac{T}{2}\right)^{k/2} \frac{\Gamma(n/2)}{\Gamma((n+k)/2)}. \tag{45}$$

При $k = 2$ имеем отмеченный выше результат для дисперсии θ^2 поскольку $\Gamma(n/2 + 1) = (n/2)\Gamma(n/2)$. •

Пример 13 (Общая нормальная модель: оценивание функций от дисперсии). Использованную в предыдущем примере идеологию можно применить и в более общей ситуации, когда оба параметра нормальной модели неизвестны.

Итак, пусть $\boldsymbol{X} = (X_1, \quad, X_n)$, $n \geqslant 2$, — выборка из $\mathcal{L}(\xi) = \mathcal{N}(\theta_1, \theta_2^2)$. Для этой модели полной достаточной статистикой является пара

$$T = (T_1, T_2) = \left(\overline{X}, \sum_{i=1}^{n}(X_i - \overline{X})^2\right)$$

(см. пример 1 в § 3.3, а также п. 3), при этом по теореме Фишера (см. п. 3 § 2.5) $\mathcal{L}_\theta(T_2/\theta_2^2) = \chi^2(n-1) = \Gamma(2, (n-1)/2)$, т. е. $\mathcal{L}_\theta(T_2) = \Gamma(2\theta_2^2, (n-1)/2)$. Поэтому, если $\tau(\theta_1, \theta_2) = \tau(\theta_2)$, т. е. параметрическая функция зависит только от теоретической дисперсии, то для ее оценивания можно применить известные результаты для гамма-модели с учетом замены $\theta \to 2\theta_2^2$, $\lambda \to (n-1)/(2n)$.

В качестве конкретного примера, как и выше, рассмотрим задачу оценивания степени θ_2^k при $k > -n+1$. Полагая в (39) $r \quad k/2$, $x = 0$, $n\lambda = (n-1)/2$ получим вид искомой оптимальной оценки

Общий результат для модели $\mathcal{N}(\theta_1, \theta_2^2)$

$$(\theta_2^k)^* = \left(\frac{T_2}{2}\right)^{k/2} \frac{\Gamma((n-1)/2)}{\Gamma((n+k-1)/2)}. \tag{46}$$

В частности, при $k = 2$ эта оценка принимает вид

$$\frac{T_2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S_0^2,$$

и мы приходим к уже известному нам из примера 9 § 3.2 результату, что статистика S_0^2 является оптимальной несмещенной оценкой теоретической дисперсии θ_2^2 нормальной модели $\mathcal{N}(\theta_1, \theta_2^2)$. Однако результат (46) гораздо более общий. •

5. Другие применения гамма-модели для оценивания неизвестных параметров

В ряде случаев для наблюдаемой случайной величины ξ с плотностью $f(x; \theta)$ можно подобрать такое функциональное преобразование $\eta = h(\xi)$, что преобразованная величина η будет иметь гамма-распределение с неизвестным параметром масштаба, и тогда задача оценивания для исходной модели $f(x; \theta)$ может быть решена с использованием изложенных в п. 4 результатов для гамма-модели. Несколько таких примеров будут рассмотрены в данном пункте.

Пример 14 (Бета-распределение). Пусть ξ имеет плотность

$$f(x; \theta) = \theta x^{\theta-1}, \quad x \in [0, 1], \quad \theta > 0,$$

т. е. (см. п. 4 § 1.2) $\mathcal{L}_\theta(\xi) = \mathrm{Be}(\theta, 1)$. Рассмотрим преобразование $\eta = -\ln \xi$, тогда по формуле (45) § 1.2 найдем, что плотность $g(y; \theta)$ случайной величины η есть $g(y; \theta) = \theta e^{-\theta y}$, $y \geqslant 0$, т. е. $\mathcal{L}_\theta(\eta) = \Gamma(\theta^{-1}, 1)$. Следовательно, если

имеется выборка $\boldsymbol{X} = (X_1, \quad , X_n)$ из $\mathcal{L}_\theta(\xi)$, то статистика

$$T = -\sum_{i=1}^{n} \ln X_i$$

имеет распределение $\Gamma(\theta^{-1}, n)$, и в терминах этой статистики можно строить оптимальные оценки для параметрических функций от θ^{-1}, используя результаты п. 4. •

Пример 15 (Распределение Вейбулла). Пусть ξ имеет плотность

$$f(x;\theta) = \frac{\alpha}{\theta} x^{\alpha-1} e^{-x^\alpha/\theta} \qquad x \geqslant 0,$$

т. е. (см. п. 8 § 1.2) $\mathcal{L}_\theta(\xi) = W(0, \alpha, \theta^{1/\alpha})$. Тогда аналогично предыдущему примеру имеем $\mathcal{L}_\theta(\xi^\alpha) = \Gamma(\theta, 1)$. Следовательно, статистика

$$T = \sum_{i=1}^{n} X_i^\alpha$$

имеет распределение $\Gamma(\theta, n)$ и «работают» все результаты п. 4. •

Пример 16 (Распределение Парето). Пусть плотность наблюдаемой случайной величины ξ имеет вид

$$f(x;\theta) = \frac{\theta}{x^{\theta+1}}, \quad x \geqslant 1, \quad \theta > 0$$

(см. п. 9 § 1.2). Тогда $\mathcal{L}_\theta(\ln \xi) = \Gamma(\theta^{-1}, 1)$, и мы имеем ситуацию примера 14 с достаточной статистикой

$$T = \sum_{i=1}^{n} \ln X_i. \qquad \bullet$$

Пример 17 (Распределение Лапласа). Это распределение задается плотностью

$$f(x;\theta) = \frac{e^{-|x|/\theta}}{2\theta}, \quad x \in R^1 \qquad \theta > 0.$$

Нетрудно убедиться в том, что здесь $\mathcal{L}_\theta(|\xi|) \quad \Gamma(\theta, 1)$ (это мы оставляем в качестве упражнения читателю), поэтому мы имеем ситуацию примера 15 с достаточной статистикой

$$T = \sum_{i=1}^{n} |X_i|. \qquad \bullet$$

Я слышу, и я забываю.
Я вижу, и я запоминаю.
Я делаю, и я понимаю.

Китайская пословица

§ 3.5. Оценки максимального правдоподобия

Оптимальность — это еще не все; оптимальность — лишь одна сторона очень сложной картины.

Э. Леман

Метод оценивания, основанный на использовании достаточных статистик, наряду с сильными сторонами обладает и своими слабостями. Прежде всего это связано с тем, что этот метод ориентирован на построение несмещенных оценок, а, как уже отмечалось выше, несмещенные оценки не всегда существуют или практически мало пригодны; в таких случаях приходится использовать смещенные оценки. Далее, даже при наличии несмещенных оценок этот метод не всегда приводит к однозначному решению (достаточная статистика не всегда обладает свойством полноты), и, следовательно, в таких случаях для выбора окончательной оценки необходимо привлекать те или иные дополнительные соображения. Наконец, при реализации метода достаточных статистик мы часто сталкиваемся с трудными аналитическими задачами, связанными с решением уравнения несмещенности (10) § 3.3, и этот метод часто приводит к сложным для практического использования формулам для оптимальных несмещенных оценок (см. например, формулу (17) § 3.3 для оценки нормальной функции распределения (15) § 3.3). В таких случаях предпочтение может быть отдано другим методам, которые приводят пусть не к оптимальным, но зато просто конструируемым и легче вычислимым оценкам (особенно, если они по эффективности мало уступают оптимальным оценкам). Такими преимуществами в ряде случаев обладают оценки максимального правдоподобия, которым посвящен настоящий параграф.

1. Определение и примеры оценок максимального правдоподобия (о. м. п.)

Одним из наиболее универсальных и эффективных методов оценивания неизвестных параметров распределений является *метод максимального правдоподобия*, который и приводит к *оценкам максимального правдоподобия* (далее мы будем использовать для них сокращенное обозначением о. м. п.). Этот метод получил распространение в 20-е гг. XX в. благодаря работам английского статистика Р. Фишера (хотя у него были и предшественники). Суть этого метода состоит в следующем.

Пусть, как обычно, имеется выборка $X = (X_1, \quad , X_n)$ из некоторого распределения $\mathcal{L}(\xi) \in \mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$ и $L(\underline{x};\theta)$ — функция правдоподобия (см. п. 1 § 3.2). Если для двух значений θ_1, θ_2 параметра θ выполняется условие $L(\underline{x};\theta_1) > L(\underline{x};\theta_2)$, то говорят, что значение θ_1 «более правдоподобно», чем θ_2 (при $\theta = \theta_1$ «вероятность получить реализацию $\underline{x}$ больше», чем при $\theta = \theta_2$). Наиболее правдоподобное значение параметра при данной реализации $\underline{x}$ выборки X обозначается $\widehat{\theta}(\underline{x})$, а статистика $\widehat{\theta} = \widehat{\theta}(X)$ и есть о. м. п. для параметра θ. Таким образом, по определению, $\widehat{\theta}(\underline{x})$ — это та-

кая точка параметрического множества Θ, в которой функция правдоподобия $L(\underline{x};\theta)$ достигает максимума:

$$L\big(\underline{x};\widehat{\theta}(\underline{x})\big) \geqslant L(\underline{x};\theta), \quad \forall\,\theta, \quad \text{или} \quad L\big(\underline{x};\widehat{\theta}(\underline{x})\big) = \sup_{\theta} L(\underline{x};\theta). \qquad (1)$$

Если для каждого $\underline{x}$ из выборочного пространства $\mathfrak{X}$ максимум $L(\underline{x};\theta)$ достигается во внутренней точке Θ и функция $L(\underline{x};\theta)$ дифференцируема по θ, то $\widehat{\theta}(\underline{x})$ удовлетворяет уравнению

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = 0 \quad \text{или} \quad \frac{\partial \ln L(\underline{x};\theta)}{\partial\theta} = 0.$$

Уравнения правдоподобия

Для векторного параметра $\theta = (\theta_1, \ldots, \theta_r)$ это уравнение заменяется системой уравнений

$$\frac{\partial \ln L(\underline{x};\theta)}{\partial\theta_i} = 0, \quad i = 1, \quad , r.$$

Эти уравнения называются *уравнениями правдоподобия*.

Свойства о. м. п.

Отметим следующие свойства о. м. п.

1) *Если существует эффективная оценка $T(X)$ для скалярного параметра θ, то $\widehat{\theta} = T(X)$.* Это очевидное следствие критерия эффективности Рао—Крамера (9) § 3.2:

$$\frac{\partial \ln L(X;\theta)}{\partial\theta} = \frac{1}{a(\theta)}(T(X) - \theta).$$

2) *Если имеется достаточная статистика $T(X)$, а о. м. п. существует и единственна, то она является функцией от $T(X)$.* Действительно, из представления (1) § 3.3 следует, что в данном случае максимизация $L(\underline{x};\theta)$ сводится к максимизации $g(T(\underline{x});\theta)$ по θ. Следовательно, $\widehat{\theta}(\underline{x}) = \widehat{\theta}(T(\underline{x}))$.

Примерами о. м. п. $\widehat{\theta}$ для ряда моделей со скалярным параметром являются приведенные в табл. 2 § 3.2 соответствующие эффективные оценки. Так для моделей $\mathcal{N}(\theta,\sigma^2)$, $\Gamma(\theta,1)$, $\mathrm{Bi}(1,\theta)$, $\Pi(\theta)$ о. м. п. $\widehat{\theta} = \overline{X}$. Рассмотрим еще несколько примеров нахождения о. м. п.

Пример 1 (Общая нормальная модель, о. м. п. ее параметров).
Пусть $\mathcal{L}(\xi) \in \mathcal{N}(\theta_1,\theta_2^2)$. Функция правдоподобия для этой модели указана в примере 1 в § 3.3. Обозначив

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

запишем ее в виде

$$L(\underline{x};\theta) = (2\pi e s^2)^{-n/2} \exp\left\{-n\left[\frac{(\overline{x}-\theta_1)^2}{2\theta_2^2} + \frac{1}{2}\left(\frac{s^2}{\theta_2^2} - 1\right) - \ln\frac{s}{\theta_2}\right]\right\}. \qquad (2)$$

Воспользуемся теперь элементарным неравенством $\ln a \leqslant a - 1$, $\forall a > 0$, из которого следует, что $\ln a^2 \leqslant a^2 - 1$ или $\ln a \leqslant (a^2 - 1)/2$; знак равенства достигается только при $a = 1$. Учитывая это, имеем, что выражение в квадратных скобках в (2) всегда неотрицательно и обращается в нуль лишь при $\theta_1 = \bar{x}$, $\theta_2 = s$. Но это означает, что $\widehat{\theta}(\underline{x}) = \big(\widehat{\theta}_1(\underline{x}), \widehat{\theta}_2(\underline{x})\big) = (\bar{x}, s)$. Таким образом, в данном случае о. м. п. существует, единственна и при этом $\widehat{\theta} = (\overline{X}, S)$, где S^2 — выборочная дисперсия. Отметим также полезный факт, что

$$\sup_{\theta} L(\underline{x}; \theta) = L\big(\underline{x}; \widehat{\theta}(\underline{x})\big) = (2\pi e s^2)^{-n/2} \tag{3}$$

Подчеркнем, что полученная о. м. п. $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2) = (\overline{X}, S)$ — функция достаточной статистики $T = (T_1, T_2) = (\overline{X}, S^2)$, определенной в примере 1 § 3.3.•

Пример 2 (Многомерная нормальная модель, о. м. п. ее параметров). Предположим, что наблюдается многомерная случайная величина $\underline{\xi} = (\xi_1, \quad, \xi_k)$, распределенная по невырожденному нормальному закону $\mathcal{N}(\underline{\mu}, \Sigma)$, где вектор средних значений $\underline{\mu} = (\mu_1, \quad, \mu_k)$ и дисперсионная матрица $\Sigma = \|\sigma_{ij}\|_1^k$ — неизвестные параметры (см. п. 2 § 1.2). Здесь общее число неизвестных параметров (с учетом симметричности матрицы Σ) равно $k + k(k+1)/2$. Пусть $X = (\underline{X}_1, \quad, \underline{X}_n)$ — выборка из $\mathcal{L}(\underline{\xi})$, т. е. $\underline{X}_l$ — независимые k-мерные случайные величины с плотностью, указанной в (9) § 1.2. Найдем оценки параметров $\theta = (\underline{\mu}, \Sigma)$ такой модели по методу максимального правдоподобия. В данном случае функция правдоподобия

$$L(\underline{x}; \theta) = \prod_{l=1}^{n} f(\underline{x}_l | \underline{\mu}, \Sigma) =$$

$$= (2\pi)^{-kn/2} |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{l=1}^{n} (\underline{x}_l - \underline{\mu})' \Sigma^{-1} (\underline{x}_l - \underline{\mu})\right\}. \tag{4}$$

Пусть $\underline{\bar{x}} = (\underline{x}_1 + \quad + \underline{x}_n)/n$ (сложение векторов производится покоординатно), тогда

$$\sum_{l=1}^{n} (\underline{x}_l - \underline{\mu})' \Sigma^{-1} (\underline{x}_l - \mu) = \sum_{l=1}^{n} (\underline{x}_l - \underline{\bar{x}})' \Sigma^{-1} (\underline{x}_l - \underline{\bar{x}}) + n(\underline{\bar{x}} - \underline{\mu})' \Sigma^{-1} (\underline{\bar{x}} - \underline{\mu}) \tag{5}$$

(здесь $\underline{\bar{x}} = (\bar{x}_1, \ \ldots, \bar{x}_k)$, где $\bar{x}_j$ — среднее арифметическое j-х координат векторов $\underline{x}_1, \quad, \underline{x}_n$). Введем также выборочную дисперсионную матрицу

$$\widehat{\Sigma}(\underline{x}) = \frac{1}{n}\sum_{l=1}^{n} (\underline{x}_l - \underline{\bar{x}})(\underline{x}_l - \underline{\bar{x}})' = \|s_{ij}\|_1^k$$

(см. п. 2 § 2.3). Воспользовавшись равенством

$$\underline{y}' B \underline{y} = \operatorname{tr}(BY), \quad Y = \underline{y}\underline{y}'$$

и линейностью оператора tr (взятия следа от матрицы), получим

$$\sum_{l=1}^{n}(\underline{x}_l - \underline{\bar{x}})'\Sigma^{-1}(\underline{x}_l - \underline{\bar{x}}) = n\,\mathrm{tr}\left(\Sigma^{-1}\widehat{\Sigma}(\underline{x})\right). \tag{6}$$

Учитывая равенства (5) и (6), перепишем формулу (4) в виде

$$L(\underline{x};\theta) = (2\pi)^{-kn/2}\exp\left\{-\frac{n}{2}\left[(\underline{\bar{x}} - \underline{\mu})'\Sigma^{-1}(\underline{\bar{x}} - \underline{\mu}) + \mathrm{tr}\left(\Sigma^{-1}\widehat{\Sigma}(\underline{x})\right) + \ln|\Sigma|\right]\right\}$$

Отсюда следует, что максимизация по θ функции правдоподобия эквивалентна минимизации по $\underline{\mu}$ и Σ функции

$$\psi(\underline{x};\underline{\mu},\Sigma) = (\underline{\bar{x}} - \underline{\mu})'\Sigma^{-1}(\underline{\bar{x}} - \underline{\mu}) + \left[\mathrm{tr}\left(\Sigma^{-1}\widehat{\Sigma}(\underline{x})\right) - k - \ln\left|\Sigma^{-1}\widehat{\Sigma}(\underline{x})\right|\right]$$

(постоянные k и $\ln\left|\widehat{\Sigma}(\underline{x})\right|$ введены здесь для упрощения дальнейших преобразований). Обозначим $\lambda_1, \quad, \lambda_k$ собственные числа матрицы $\Sigma^{-1}\widehat{\Sigma}(\underline{x})$ (все они — положительны), тогда след (tr) этой матрицы равен сумме $\lambda_1 + \quad + \lambda_k$, а ее определитель — произведению $\lambda_1 \quad \lambda_k$, следовательно, можем записать, что

$$\psi(\underline{x};\underline{\mu},\Sigma) = (\underline{\bar{x}} - \underline{\mu})'\Sigma^{-1}(\underline{\bar{x}} - \underline{\mu}) + \sum_{i=1}^{k}(\lambda_i - 1 - \ln\lambda_i).$$

Поскольку Σ^{-1} — положительно определенная матрица и $a - 1 - \ln a \geqslant 0$, $\forall\, a > 0$, из последнего соотношения получаем $\psi(\underline{x};\underline{\mu},\Sigma) \geqslant 0$, причем равенство имеет место только при $\underline{\mu} = \underline{\bar{x}}$, $\lambda_1 \quad = \lambda_k = 1$, т. е. при $\Sigma = \widehat{\Sigma}(\underline{x})$. Но это означает, что о. м. п. параметров $\underline{\mu}$ и Σ являются соответственно статистики $\overline{X} = (\overline{X}_1, \quad, \overline{X}_k)$ и $\widehat{\Sigma}(X) = \left\|S_{ij}\right\|_1^k$. Как показано в п. 2 § 2.3, вектор выборочных средних $\overline{X} = (\overline{X}_1, \quad, \overline{X}_k)$ является несмещенной оценкой для вектора $\underline{\mu} = (\mu_1, \quad, \mu_k)$, несмещенной же оценкой для Σ является подправленная выборочная дисперсионная матрица $(n/(n-1))\widehat{\Sigma}(X)$. •

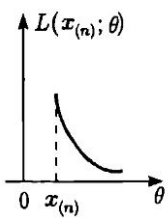


Рис. 1

Пример 3 (Равномерная модель, о. м. п. ее параметра). Если $X = (X_1, \quad, X_n)$ — выборка из равномерного распределения $U(0,\theta)$, то из формулы для $L(\underline{x};\theta)$, полученной в примере (3) § 3.3, следует, что $L(\underline{x};\theta)$ монотонно убывает по θ для $\theta \geqslant x_{(n)}$, а при $\theta = x_{(n)}$ функция правдоподобия достигает максимума (см. рис. 1). Таким образом, о. м. п. $\widehat{\theta} = X_{(n)}$, т. е. совпадает с достаточной статистикой. В этой точке функция правдоподобия разрывна, поэтому производная $L(\underline{x};\theta)$ по θ в этой точке не существует. Таким образом, в данном случае о. м. п. не является решением уравнения правдоподобия. Это характерно для ситуаций, когда выборочное пространство $\mathfrak{X}$ зависит от неизвестного параметра.

Если же X — выборка распределения $\mathcal{U}(\theta, \theta+1)$, то из формулы (4) § 3.3 имеем $L(\underline{x};\theta) = I(\theta \leqslant x_{(1)} \leqslant x_{(n)} \leqslant \theta+1)$. Отсюда следует, что любое $\theta \in [x_{(n)}-1, x_{(1)}]$ максимизирует функцию правдоподобия; таким образом, здесь о. м. п. не единственна. В качестве решения можно выбрать, например, $\widehat{\theta} = (X_{(1)} + X_{(n)} - 1)/2$; в этом случае $\widehat{\theta}$ — функция достаточной статистики $T = (X_{(1)}, X_{(n)})$ (см. пример 4 в § 3.3). •

Пример 4 (Двухпараметрическая показательная модель, о. м. п. ее параметров). Эта модель определена в примере 2 в § 3.3. Из полученной там формулы

$$L(\underline{x};\theta) = \theta_2^{-n} \exp\left\{-\frac{n}{\theta^2}\left(T_1(\underline{x}) + T_2(\underline{x}) - \theta_1\right)\right\} I(T_1(\underline{x}) \leqslant \theta_1) \qquad (7)$$

следует, что при любом θ_2 максимум по θ_1 достигается при $\theta_1 = T_1(\underline{x}) = x_{(1)}$, и он равен $\exp\left\{-n\left(\ln\theta_2 + T_2(\underline{x})/\theta_2\right)\right\}$; максимум же последнего выражения по θ_2, как легко видеть, достигается при $\theta_2 = T_2(\underline{x}) = \overline{x} - x_{(1)}$. Таким образом, о. м. п. $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2) = (X_{(1)}, \overline{X} - X_{(1)})$, т. е. совпадает с достаточной статистикой $T = (T_1, T_2) = (X_{(1)}, \overline{X} - X_{(1)})$. •

2. Принцип инвариантности для о. м. п.

Важным свойством оценок максимального правдоподобия является их инвариантность относительно преобразований параметра. Это означает, что если $q \quad q(\theta)$ произвольная функция $\theta \quad (\theta_1, \quad, \theta_r)$, взаимно однозначно отображающая параметрическое множество Θ рассматриваемой модели в множество $Q = \{q = (q_1, \quad, q_r)\} \subset R^r$ то о. м. п. $\widehat{q} = q(\widehat{\theta})$. Действительно, поскольку q — взаимно однозначное отображение, существует обратная функция $\theta = \theta(q)$. Тогда

$$\sup_{\theta\in\Theta} L(\underline{x};\theta) = \sup_{q\in Q} L\left(\underline{x};\theta(q)\right),$$

и если максимум по θ функции $L(\underline{x};\theta)$ достигается в точке $\widehat{\theta}$, то максимум по q функции $L\left(\underline{x};\theta(q)\right)$ имеет место в точке $\widehat{q}$, удовлетворяющей уравнению $\theta(\widehat{q}) = \widehat{\theta}$, т. е. в точке $\widehat{q} = q(\widehat{\theta})$.

Согласно этому, в каждой конкретной задаче можно выбрать в качестве исходной наиболее удачную параметризацию, а о. м. п. получать затем с помощью соответствующих преобразований. Таким образом, в методе максимального правдоподобия по существу достаточно построить лишь оценку $\widehat{\theta}$ параметра модели, а оценки различных параметрических функций получаются как результат подстановки $\theta = \widehat{\theta}$ в оцениваемую функцию. Отметим, что в методе достаточных статистик наилучшая несмещенная оценка ищется для каждой функции $\tau(\theta)$ независимо.

Продемонстрируем эффективность принципа инвариантности примерами.

Пример 5 (Общая нормальная модель, о. м. п. функции распределения). Пусть в условиях примера 1 требуется оценить параметрическую функцию

$$\tau(\theta) = \mathrm{P}_\theta\{\xi \leqslant x_0\} = \Phi\left(\frac{x_0 - \theta_1}{\theta_2}\right) \tag{7}$$

(см. пример 7 в § 3.3). Поскольку о. м. п. $\widehat{\theta} = (\overline{X}, S)$, то по **принципу** инвариантности о. м. п.

$$\widehat{\tau} = \tau(\widehat{\theta}) = \Phi\left(\frac{x_0 - \overline{X}}{S}\right)$$

(сравни с оценкой τ^* в примере 7 в § 3.3, соотношение (17)). •

Пример 6 (Двумерная нормальная модель, о. м. п. ее параметров). Пусть $X_l = (X_{l1}, X_{l2})$, $l = 1, \quad , n$, — независимые наблюдения над двумерной случайной величиной $\underline{\xi} = (\xi_1, \xi_2)$ с распределением

$$\mathcal{N}\left((0,0), \left\|\begin{matrix}\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2\end{matrix}\right\|\right),$$

где $\sigma^2 > 0$ и $\rho \in (-1, 1)$ неизвестны. Здесь теоретическая плотность есть (далее $\theta = (\sigma^2, \rho)$)

$$f(x_1, x_2; \theta) = \frac{1}{2\pi\sqrt{1-\rho^2\sigma^2}} \exp\left\{-\frac{x_1^2 + x_2^2}{2\sigma^2(1-\rho^2)} + \frac{\rho x_1 x_2}{\sigma^2(1-\rho^2)}\right\} \tag{8}$$

(это частный случай общей формулы (9) § 1.2). Найдем о. м. п. $\widehat{\sigma}^2$ и $\widehat{\rho}$. Здесь плотность сложным образом зависит от параметров, поэтому удобно перейти к новым параметрам $q = (q_1, q_2)$, положив $q_1 = q_1(\theta) \quad -\left[2\sigma^2(1-\rho^2)\right]^{-1}$ $q_2 = q_2(\theta) = \rho\left[\sigma^2(1-\rho^2)\right]^{-1}$ Тогда формула (8) перепишется в виде

$$f(x_1, x_2; q) = \frac{1}{2\pi} \exp\left\{q_1(x^2 + y^2) + q_2 xy + \tau(q)\right\},$$

где $\tau(q) = (1/2)\ln(4q_1^2 - q_2^2)$. Отсюда сразу получаем, что уравнения правдоподобия для нахождения о. м. п. $\widehat{q}_1$ и $\widehat{q}_2$ имеют вид

$$\begin{aligned} \frac{1}{n}\sum_{l=1}^{n}(x_{l1}^2 + x_{l2}^2) &= -\frac{\partial\tau(q)}{\partial q_1} = -\frac{4q_1}{4q_1^2 - q_2^2}, \\ \frac{1}{n}\sum_{l=1}^{n} x_{l1}x_{l2} &= -\frac{\partial\tau(q)}{\partial q_2} = \frac{q_2}{4q_1^2 - q_2^2}. \end{aligned} \tag{9}$$

Но

$$\sigma^2 = -\frac{2q_1}{4q_1^2 - q_2^2}, \quad \rho = -\frac{q_2}{2q_1},$$

поэтому из соотношений (9) находим, что

$$\widehat{\sigma}^2 = \frac{1}{2n}\sum_{l=1}^{n}(X_{l1}^2 + X_{l2}^2), \quad \widehat{\rho} = \frac{2\sum_{l=1}^{n} X_{l1}X_{l2}}{\sum_{l=1}^{n}(X_{l1}^2 + X_{l2}^2)}. \tag{10}$$

•

Пример 7 (Логнормальное распределение, о. м. п. параметрических функций). Пусть $X = (X_1, \quad , X_n)$ — выборка из логнормального распределения, т. е. (см. п. 1 § 1.2) $X_i = e^{Y_i}$, где $\mathcal{L}_\theta(Y_i) \in \mathcal{N}(\theta_1, \theta_2^2)$. Построим о. м. п. для теоретических среднего и дисперсии этого распределения, т. е. (см. (8) § 1.2) для функций

$$\tau_1(\theta) = \exp\left\{\theta_1 + \frac{\theta_2^2}{2}\right\} \quad \text{и} \quad \tau_2(\theta) = (e^{\theta_2^2} - 1)\exp\{2\theta_1 + \theta_2^2\}. \tag{11}$$

Прежде всего отметим, что моменты $\mathbf{E}_\theta X_1^k$ можно вычислить следующим образом. Поскольку характеристическая функция случайной величины Y_1 есть (см. упр. 26 к гл. 1)

$$\psi(t) = \mathbf{E}_\theta e^{itY_1} = \exp\left\{it\theta_1 - \frac{\theta_2^2 t^2}{2}\right\},$$

то

$$\mathbf{E}_\theta X_1^k = \mathbf{E}_\theta e^{kY_1} = \psi\left(\frac{k}{i}\right) = \exp\left\{k\theta_1 + \frac{k^2 t^2}{2}\right\}.$$

Отсюда для $\tau_1(\theta) = \mathbf{E}_\theta X_1$ и $\tau_2(\theta) = \mathbf{D}_\theta X_1 = \mathbf{E}_\theta X_1^2 - (\mathbf{E}_\theta X_1)^2$ имеем представления (11). Перейдем теперь к построению оценок. Мы уже знаем о. м. п. $(\widehat{\theta}_1, \widehat{\theta}_2) = (\overline{Y}, S(Y))$, но тогда по принципу инвариантности искомые о. м. п. $\widehat{\tau}_1$ и $\widehat{\tau}_2$ строятся сразу в соответствии с формулами (11):

$$\widehat{\tau}_1 = \exp\left\{\overline{Y} + \frac{S^2(Y)}{2}\right\}, \quad \widehat{\tau}_2 = \widehat{\tau}_1^2(e^{S^2(Y)} - 1).$$

•

3. Метод накопления для приближенного вычисления о. м. п.

Оценки максимального правдоподобия в явном виде не всегда удается получить; в таких случаях прибегают к приближенным методам решения уравнений правдоподобия (когда такие уравнения можно составить). Мы опишем один из таких методов, который называется *методом накопления Фишера*. Он заключается в следующем. Пусть θ — скалярный параметр и функция правдоподобия $L(\theta) = L(\underline{x}; \theta)$ дважды дифференцируема по θ. Разложим функцию вклада $V(\theta) = V(\underline{x}; \theta)$ (см. (1) § 3.2) в ряд Тейлора в окрестности точки θ_0, выбранной в качестве начального приближения для $\widehat{\theta}(\underline{x}) = \widehat{\theta}$, и вычислим это

Метод накопления Фишера

разложение при $\theta = \widehat{\theta}$:

$$V(\widehat{\theta}) = V(\theta_0) + (\widehat{\theta} - \theta_0)V'(\theta^*), \tag{12}$$

где θ^* — некоторая промежуточная точка между $\widehat{\theta}$ и θ_0. Но $V(\widehat{\theta}) = 0$ по определению о. м. п. $\widehat{\theta}$, поэтому из (12) имеем

$$\widehat{\theta} = \theta_0 - \frac{V(\theta_0)}{V'(\theta^*)}$$

Если заменить теперь в этом равенстве θ^* на θ_0 и $V'(\theta_0)$ на $\mathbf{E}_\theta V'(X;\theta_0) = -ni(\theta_0)$ (см. соотношение (5′) § 3.2), то получим первое приближение

$$\theta_1 = \theta_0 + \frac{V(\theta_0)}{ni(\theta_0)}$$

Теперь этот процесс можно повторить, взяв в качестве нового начального приближения θ_1 и т. д. Таким образом, $(k+1)$-е приближение по методу накопления вычисляют рекуррентно по формуле

Рекуррента метода накопления

$$\theta_{k+1} = \theta_k + \frac{V(\theta_k)}{ni(\theta_k)}, \quad k = 0, \ , 2, \ldots \tag{13}$$

Процесс вычислений продолжают до достижения желаемой точности: $|\theta_{k+1} - \theta_k| < \varepsilon$. Важным в этом методе является выбор начальной точки θ_0. Обычно в качестве θ_0 берут значение какой-нибудь легко вычисляемой состоятельной оценки (см. определение 3 в п. 3 § 2.2); тогда все три точки θ_0, θ^* и $\widehat{\theta}$ при больших объемах выборки n будут находиться вблизи истинного значения параметра. Метод накопления, как правило, быстро приводит к цели, хотя в некоторых случаях этот процесс может и не сходиться.

Пример 8 (Модель Коши, оценка параметра по методу накопления). Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения Коши $K(\theta)$ (см. упр. 48 (б) к гл. 1).

Здесь функция вклада

$$V(\theta) = 2\sum_{i=1}^{n} \frac{x_i - \theta}{1 + (x_i - \theta)^2}$$

и получить точное решение уравнения правдоподобия $V(\theta) = 0$ невозможно. Функция информации $i(\theta)$ для модели $K(\theta)$ равна $1/2$ (см. табл. 1 § 3.2), поэтому последовательность (13) принимает в данном случае вид

$$\theta_{k+1} = \theta_k + \frac{2}{n}V(\theta_k), \quad k = 0, 1, 2, \ldots$$

В качестве начального приближения θ_0 здесь можно взять выборочную медиану $\widehat{\zeta}_{n,1/2}$, которая в силу соотношения (15) § 2.4 является состоятельной оценкой теоретической медианы $\zeta_{1/2}$, совпадающей в данном случае с θ. •

Пример 9 (Модель степенного ряда, метод накопления для нее). Эта модель определена в п. 8 § 1.1. Из формулы (20) § 1.1 имеем для функции вклада представление

$$V(\theta) = \frac{\partial}{\partial\theta} \ln\left(\theta^{n\bar{x}} f^{-n}(\theta) \prod_{i=1}^{n} a(x_i)\right) = \frac{n\bar{x}}{\theta} - n\frac{f'(\theta)}{f(\theta)} = n\frac{\bar{x} - \mu(\theta)}{\theta},$$

где $\mu(\theta)$ — теоретическое среднее (см. (23) § 1.1). Таким образом, уравнение правдоподобия $V(\theta) = 0$ здесь имеет вид

$$\mu(\theta) = \bar{x}. \tag{14}$$

Вычислим функцию информации $i(\theta)$. Имеем (см. (4), (5) § 3.2)

$$i(\theta) = \mathbf{E}_\theta V^2(X_1; \theta) = \frac{1}{\theta^2}\mathbf{E}_\theta\big(X_1 - \mu(\theta)\big)^2 = \frac{1}{\theta^2}\mathbf{D}_\theta X_1 = \frac{\sigma^2(\theta)}{\theta^2} = \frac{\mu'(\theta)}{\theta},$$

где $\sigma^2(\theta)$ — теоретическая дисперсия (см. (23) § 1.1). Если уравнение (14) нельзя решить точно, то для приближенного определения значения о. м. п. $\widehat{\theta}$ можно применить метод накопления.

В частности, для усеченного в нуле распределения Пуассона (см. (22) § 1.1) функция $f(\theta) = e^\theta - 1$ и уравнение (14) принимает вид $\theta = \bar{x}(1 - e^{-\theta})$, а функция информации равна $i(\theta) = \dfrac{1 - (1+\theta)e^{-\theta}}{\theta(1 - e^{-\theta})^2}$ •

4. Асимптотические свойства о. м. п.

Метод максимального правдоподобия и возникающие в нем оценки максимального правдоподобия никак не связаны с обсуждавшимися ранее свойствами несмещенности и оптимальности оценок. Более того, в рассмотренных выше примерах вычисления о. м. п. мы обнаружили, что о. м. п. не всегда являются несмещенными оценками. Так из примера 1 следует, что о. м. п. дисперсии θ_2^2 нормальной модели $\mathcal{N}(\theta_1, \theta_2^2)$ является выборочная дисперсия S^2, которая, как мы знаем, является смещенной оценкой (см. п. 3 § 2.2), однако ее смещение $\mathbf{E}_\theta S^2 - \theta_2^2 = -\theta_2^2/n$ убывает с ростом объема выборки, т. е. она является асимптотически несмещенной. Тем самым, можно сказать, что свойства этой оценки «улучшаются» с увеличением числа наблюдений n. Это обстоятельство имеет весьма общий характер, именно,

Большие выборки

о. м. п. обладают хорошими асимптотическими свойствами для больших выборок. Можно сказать, что широкое использование этих оценок связано именно с их асимптотическими свойствами, проявляющимися при неограниченном накоплении данных.

О роли асимптотического подхода (предельных теорем) в теории вероятностей и математической статистике уже вкратце говорилось ранее (см. общее замечание в п. 1 § 1.2). Добавим еще к этому, что асимптотический подход позволяет исследовать единым методом различные задачи и получать при этом достаточно общие результаты. Эти результаты в дальнейшем могут использо-

ваться в задачах с выборками конечного объема, и тогда они рассматриваются как приближенные. Иногда асимптотический подход является единственным математическим подходом, позволяющим в конкретной статистической задаче предлагать некоторые процедуры и исследовать их свойства. При этом (что очень существенно!) окончательные результаты удается получить в простой и наглядной форме, и (что не менее важно) именно асимптотический подход позволяет вскрыть глубокую математическую сущность проблемы. Изложение асимптотической теории оценок максимального правдоподобия и составляет дальнейшее содержание этого параграфа. Чтобы подчеркнуть зависимость рассматриваемых ниже статистик от объема выборки, будем отмечать их индексом n.

Когда говорят об асимптотических свойствах оценок (или свойствах для больших выборок), то прежде всего имеют в виду их *состоятельность*. Это свойство нам уже встречалось ранее (см. определение 3 в п. 3 § 2.2 и комментарии к нему). Дадим его общее определение в рассматриваемом контексте и сформулируем простой, но важный, критерий состоятельности, обобщающий критерий (23) § 2.2.

Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения $\mathcal{L}(\xi) \in \mathcal{F} = \{F(x;\theta), \theta \in \Theta\}$, где параметрическое множество Θ — в общем случае некоторый невырожденный открытый интервал в R^r. По определению, оценка $T_n = T_n(X)$ для заданной параметрической функции $\tau(\theta)$ называется *состоятельной*, если при $n \to \infty$

$$T_n \xrightarrow{\mathbf{P}_\theta} \tau(\theta), \quad \forall\, \theta \in \Theta, \tag{15}$$

т. е., каково бы ни было истинное значение параметра θ, оценка T_n сходится по вероятности (относительно распределения $\mathbf{P}_\theta$) к истинному значению оцениваемой функции. Подчеркнем еще раз, что свойство состоятельности в статистических задачах является обязательным для любого правила оценивания.

Критерий состоятельности оценки

Если статистика имеет конечную дисперсию, то имеет место следующий критерий состоятельности.

Теорема 1. Пусть $\mathbf{E}_\theta T_n = \tau(\theta) + \varepsilon_n$, $\mathbf{D}_\theta T_n = \delta_n$ и $\varepsilon_n = \varepsilon_n(\theta) \to 0$, $\delta_n = \delta_n(\theta) \to 0$ при $n \to \infty$ для всех $\theta \in \Theta$. Тогда T_n — состоятельная оценка функции $\tau = \tau(\theta)$.

Доказательство. Поскольку

$$|T_n - \mathbf{E}_\theta T_n| = \big|(T_n - \tau) + (\tau - \mathbf{E}_\theta T_n)\big| \geqslant |T_n - \tau| - |\mathbf{E}_\theta T_n - \tau|,$$

событие $\{|T_n - \tau| \geqslant \varepsilon\}$ влечет событие $\{|T_n - \mathbf{E}_\theta T_n| \geqslant \varepsilon - |\varepsilon_n|\}$. Отсюда по неравенству Чебышева имеем

$$\mathbf{P}_\theta\big\{|T_n - \tau| \geqslant \varepsilon\big\} \leqslant \mathbf{P}_\theta\big\{|T_n - \mathbf{E}_\theta T_n| \geqslant \varepsilon - |\varepsilon_n|\big\} \leqslant \frac{\delta_n}{(\varepsilon - |\varepsilon_n|)^2} \to 0$$

при $n \to \infty$ и всех $\theta \in \Theta$. ■

С помощью этой теоремы во многих случаях легко доказывается состоятельность конкретных оценок.

Основные предельные теоремы теории вероятностей

Сейчас будет своевременным привести список-справочник основных предельных теорем теории вероятностей о сходимости функций от случайных величин, которые интенсивно используются в различных статистических задачах, связанных со схемой «больших выборок» (пример одной из таких теорем нам встречался в п. 3 § 2.2 (теорема 1)).

Справочник предельных теорем (обозначения см. в п. 1 § 1.2)

1° Если $\eta_n \xrightarrow{\mathrm{P}} \eta$ и функция φ непрерывна, то $\varphi(\eta_n) \xrightarrow{\mathrm{P}} \varphi(\eta)$, а также $\varphi(\eta_n) \xrightarrow{\mathcal{L}} \varphi(\eta)$.

2° Пусть $\{\eta_n, \zeta_n\}$, $n = 1, 2, \ldots$, — последовательность пар случайных величин. Тогда

а) $\eta_n - \zeta_n \xrightarrow{\mathrm{P}} 0$, $\zeta_n \xrightarrow{\mathrm{P}} \zeta \Longrightarrow \eta_n \xrightarrow{\mathrm{P}} \zeta$;

б) $\eta_n - \zeta_n \xrightarrow{\mathrm{P}} 0$, $\zeta_n \xrightarrow{\mathcal{L}} \zeta \Longrightarrow \eta_n \xrightarrow{\mathcal{L}} \zeta$;

в) $\eta_n \xrightarrow{\mathcal{L}} \eta$, $\zeta_n \xrightarrow{\mathrm{P}} 0 \Longrightarrow \eta_n\zeta_n \xrightarrow{\mathrm{P}} 0$;

г) $\eta_n \xrightarrow{\mathcal{L}} \eta$, $\zeta_n \xrightarrow{\mathrm{P}} c = \mathrm{const} \Longrightarrow \eta_n + \zeta_n \xrightarrow{\mathcal{L}} \eta + c$,
$\eta_n\zeta_n \xrightarrow{\mathcal{L}} c\eta$, $\eta_n/\zeta_n \xrightarrow{\mathcal{L}} \eta/c$ при $c \neq 0$;

д) $\eta_n - \zeta_n \xrightarrow{\mathrm{P}} 0$, $\zeta_n \xrightarrow{\mathcal{L}} \zeta$,

функция φ непрерывна $\Longrightarrow \varphi(\eta_n) - \varphi(\zeta_n) \xrightarrow{\mathrm{P}} 0$.

3° Пусть $T_n = T_n(X)$ — оценка скалярного параметра θ такая, что

$$\mathcal{L}_\theta\big(\sqrt{n}(T_n - \theta)\big) \to \mathcal{N}(0, \sigma^2(\theta))$$

(или, что то же, $\mathcal{L}_\theta(T_n) \sim \mathcal{N}(\theta, \sigma^2(\theta)/n)$) при $n \to \infty$ и всех $\theta \in \Theta$. Пусть, далее, функция φ дифференцируема и $\varphi' \neq 0$. Тогда

$$\mathcal{L}_\theta\big(\sqrt{n}\big(\varphi(T_n) - \varphi(\theta)\big)\big) \longrightarrow \mathcal{N}\big(0, [\varphi'(\theta)]^2\sigma^2(\theta)\big); \tag{16}$$

кроме того, если $\varphi'(\theta)$ и $\sigma(\theta)$ непрерывны, то

$$\mathcal{L}_\theta\left(\sqrt{n}\,\frac{\varphi(T_n) - \varphi(\theta)}{\varphi'(T_n)\sigma(T_n)}\right) \longrightarrow \mathcal{N}(0, 1). \tag{17}$$

4° (аналог утверждения 3° для случая векторного параметра $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)$).

Пусть $\underline{T}_n = (T_{1n}, \ldots, T_{rn})$ — оценка параметра $\boldsymbol{\theta}$, удовлетворяющая условию $\mathcal{L}_{\boldsymbol{\theta}}\big(\sqrt{n}(\underline{T}_n - \boldsymbol{\theta})\big) \to \mathcal{N}(\underline{0}, \Sigma(\boldsymbol{\theta}))$ при $n \to \infty$ и всех $\boldsymbol{\theta} \in \Theta$. Тогда для любой дифференцируемой функции φ от r переменных

$$\mathcal{L}_{\boldsymbol{\theta}}\big(\sqrt{n}\big(\varphi(\underline{T}_n) - \varphi(\boldsymbol{\theta})\big)\big) \longrightarrow \mathcal{N}(0, v^2(\boldsymbol{\theta})) \tag{18}$$

при условии, что $v(\theta) \neq 0$, где

$$v^2(\theta) = \underline{b}'(\theta)\Sigma(\theta)\underline{b}(\theta) \quad \text{и} \quad \underline{b}(\theta) = \left(\frac{\partial\varphi(\theta)}{\partial\theta_1}, \qquad \frac{\partial\varphi(\theta)}{\partial\theta_r}\right);$$

если, кроме того, функция φ непрерывно дифференцируема, и все элементы дисперсионной матрицы $\Sigma(\theta)$ непрерывны по θ, то

$$\mathcal{L}_\theta\left(\frac{\sqrt{n}\,(\varphi(\underline{T}_n) - \varphi(\theta))}{v(\underline{T}_n)}\right) \longrightarrow \mathcal{N}(0, 1). \tag{19}$$

Вернемся к оценкам максимального правдоподобия и сформулируем их основные асимптотические свойства. Будем предполагать, что модель $\mathcal{F}$ является регулярной в смысле § 3.2, а функция правдоподобия $L_n(\underline{x}; \theta)$ при всех $n \geqslant 1$ и $\underline{x} \in \mathcal{X}$ имеет лишь один локальный максимум по θ, лежащий внутри Θ, и притом достижимый (т. е. о. м. п. $\widehat{\theta}_n$ существует и совпадает с локальным максимумом). Кроме того, плотность $f(x; \theta)$ трижды дифференцируема по θ и при этом существует не зависящая от θ функция $M(x)$ такая, что для всех $\theta \in \Theta$

Асимптотические свойства о. м. п.

$$\left|\frac{\partial^3 \ln f(x; \theta)}{\partial\theta_i\,\partial\theta_j\,\partial\theta_k}\right| \leqslant M(x), \quad \mathbf{E}_\theta M(\xi) < \infty \quad (i, j, k = 1, \ldots, r).$$

Теорема 2. Пусть выполнены перечисленные условия. Тогда:

1) о. м. п. $\widehat{\theta}_n$ является состоятельной оценкой параметра θ;
2) эта оценка асимптотически нормальна, именно, при $n \to \infty$

$$\mathcal{L}_\theta\left(\sqrt{n}(\widehat{\theta}_n - \theta)\right) \longrightarrow \mathcal{N}\left(\underline{0}, I^{-1}(\theta)\right), \tag{20}$$

где $I(\theta)$ — информационная матрица модели, определенная в (21) § 3.2; кроме того, если $\tau(\theta)$ — непрерывно дифференцируемая функция от θ и $\widehat{\tau}_n = \tau(\widehat{\theta}_n)$ — ее о. м. п., то при $n \to \infty$

$$\mathcal{L}_\theta\left(\sqrt{n}\left(\widehat{\tau}_n - \tau(\theta)\right)\right) \longrightarrow \mathcal{N}\left(0, \sigma_\tau^2(\theta)\right), \tag{21}$$

если

$$\sigma_\tau^2(\theta) \neq 0, \text{ где } \sigma_\tau^2(\theta) = \underline{b}'(\theta)I^{-1}(\theta)\underline{b}(\theta), \ \underline{b}(\theta) = \left(\frac{\partial\tau(\theta)}{\partial\theta_1}, \qquad \frac{\partial\tau(\theta)}{\partial\theta_r}\right),$$

более того, если функция $\sigma_\tau^2(\theta)$ непрерывна по θ, то

$$\mathcal{L}_\theta\left(\frac{\sqrt{n}\left(\widehat{\tau}_n - \tau(\theta)\right)}{\sigma_\tau(\widehat{\theta}_n)}\right) \to \mathcal{N}(0, 1); \tag{22}$$

3) $\widehat{\tau}_n$ является асимптотически эффективной оценкой $\tau(\theta)$ (в смысле п. 5 § 3.2)).

Выделим специально случай скалярного параметра, ввиду его особой распространенности в приложениях. Для этого случая соотношения (20)–(22) принимают соответственно вид

$$\mathcal{L}_\theta\left(\sqrt{n}(\widehat{\theta}_n - \theta)\right) \longrightarrow \mathcal{N}\left(0, \frac{1}{i(\theta)}\right), \tag{23}$$

$$\mathcal{L}_\theta\left(\sqrt{n}\left(\widehat{\tau}_n - \tau(\theta)\right)\right) \longrightarrow \mathcal{N}\left(0, \frac{[\tau'(\theta)]^2}{i(\theta)}\right) \tag{24}$$

и

$$\mathcal{L}_\theta\left(\frac{\sqrt{n}\left(\widehat{\tau}_n - \tau(\theta)\right)}{\tau'(\widehat{\theta}_n)\sqrt{i(\widehat{\theta}_n)}}\right) \longrightarrow \mathcal{N}(0, 1) \tag{25}$$

(здесь $i(\theta)$ — функция информации, определенная в (5) и (5′) § 3.2, и $\tau'(\theta) \neq 0$).

Доказательство. Приведем доказательство асимптотической нормальности (23). Для этого разложим функцию вклада $V_n(\theta) = V_n(X; \theta)$ в ряд Тейлора в окрестности истинной точки θ и вычислим это разложение в точке $\widehat{\theta}_n$

$$0 = V_n(\widehat{\theta}_n) = V_n(\theta) + (\widehat{\theta}_n - \theta)V_n'(\theta) + \frac{1}{2}(\widehat{\theta}_n - \theta)^2 V_n''(\theta^*),$$

где θ^* — некоторая промежуточная точка между θ и $\widehat{\theta}_n$. Это равенство можно записать в виде

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{V_n(\theta)}{\sqrt{n}\, i(\theta)}\left(-\frac{V_n'(\theta)}{n\, i(\theta)} + \varepsilon_n\right) \tag{26}$$

где

$$|\varepsilon_n| = \frac{|\widehat{\theta}_n - \theta|\ \ |V_n''(\theta^*)|}{2ni(\theta)} \leqslant \frac{|\widehat{\theta}_n - \theta|}{2i(\theta)}\ \ \frac{1}{n}\sum_{j=1}^{n} M(X_j).$$

Поскольку $\widehat{\theta}_n \xrightarrow{\mathbf{P}} \theta$, из условия на функцию M на основании свойства 2° в) следует, что $\varepsilon_n \xrightarrow{\mathbf{P}} 0$. К величине

$$\frac{1}{n}V_n'(\theta) = \frac{1}{n}\sum_{j=1}^{n}\frac{\partial^2 \ln f(X_j; \theta)}{\partial\theta^2}$$

применим закон больших чисел, согласно которому (см. (5′) § 3.2) эта величина сходится по вероятности к $-i(\theta)$, следовательно, все выражение в знаменателе в (25) сходится по вероятности к 1. К случайной же величине

$$\frac{V_n(\theta)}{\sqrt{n}\, i(\theta)} = \left(\sum_{j=1}^{n}\frac{\partial \ln f(X_j; \theta)}{\partial\theta}\right)\frac{1}{\sqrt{n}\, i(\theta)}$$

применима центральная предельная теорема, по которой в силу соотношений (3) и (5) § 3.2 при $n \to \infty$ эта величина асимптотически нормальна $\mathcal{N}(0, i^{-1}(\theta))$.

Отсюда и из (26) по свойству 2° г) следует, что такое же предельное распределение имеет и $\sqrt{n}\,(\widehat{\theta}_n - \theta)$.

Утверждения (24) и (25) следуют из (23) на основании свойства 3°

Наконец, утверждение об асимптотической эффективности оценки $\widehat{\tau}_n$ следует из того, что величина $e_0(\widehat{\tau}_n; \theta)$ в (14) § 3.2 тождественно равна 1. ■

Комментарии к свойствам о. м. п.

Добавим некоторые комментарии к изложенным результатам. Факт асимптотической нормальности о. м. п. — это очень сильный результат. Например, утверждение (24) говорит о том, что для больших выборок распределение о. м. п. $\widehat{\tau}_n$ концентрируется вокруг истинного значения оцениваемой параметрической функции $\tau(\theta)$ (предельное распределение имеет среднее значение $\tau(\theta)$), и степень этой концентрации, характеризуемая асимптотической дисперсией $\dfrac{[\tau'(\theta)]^2}{n\,i(\theta)}$, возрастает с ростом объема выборки n. Это можно интерпретировать так, что оценка $\widehat{\tau}_n$ является асимптотически несмещенной и среди всех таких оценок обладает наименьшей асимптотической дисперсией, т. е. она является асимптотически эффективной. При этом, однако, надо иметь в виду, что сама оценка $\widehat{\tau}_n$ может вообще не иметь моментов, т. е. $\tau(\theta)$ может не быть асимптотическим значением среднего $\mathbf{E}_\theta \widehat{\tau}_n$, а $\dfrac{[\tau'(\theta)]^2}{n\,i(\theta)}$ — асимптотическим значением дисперсии $\mathbf{D}_\theta \widehat{\tau}_n$ (в теории вероятностей известно, что из сходимости $\eta_n \xrightarrow{\mathcal{L}} \eta$, вообще говоря, не следует сходимость соответствующих моментов).

Рассмотрим, например, задачу оценивания параметрической функции $\tau(\theta) \quad 1/\theta$ в пуассоновской модели $\Pi(\theta)$. Здесь, как уже отмечалось в п. 1, о. м. п. $\widehat{\theta} = \overline{X}$, поэтому о. м. п. $\widehat{\tau}_n = 1/\overline{X}$, и для нее соотношение (24) принимает вид

$$\mathcal{L}_\theta\left(\sqrt{n}\,(\widehat{\tau}_n - \theta^{-1})\right) \longrightarrow \mathcal{N}(0, \theta^{-3}), \quad \forall\, \theta > 0.$$

В то же время $\mathcal{L}_\theta(n\overline{X}) = \Pi(n\theta)$ (см. п. 3 § 1.1), откуда $\mathbf{P}_\theta\{\overline{X} = 0\} = e^{-n\theta} > 0$. Таким образом, с положительной вероятностью $\overline{X}$ принимает нулевое значение и поэтому статистика $\widehat{\tau}_n$ ни при каком n не имеет конечных моментов.

Тем не менее для больших выборок (при $n \to \infty$) среднее и дисперсию аппроксимирующего нормального закона распределения можно трактовать как среднее и дисперсию самой оценки и говорить о ее свойствах несмещенности и оптимальности в асимптотическом смысле.

Далее, определение асимптотически эффективной оценки предполагает, что невозможно построить оценку с меньшей асимптотической дисперсией. Как правило, это имеет место для регулярных моделей, удовлетворяющих условиям, при которых справедлива теорема 2, хотя существуют и исключения, как это видно из приводимого ниже примера 10. Для нерегулярных моделей (когда неприменимо неравенство Рао—Крамера) асимптотическая дисперсия о. м. п. может иметь порядок, больший n^{-1}. Так, из результатов примеров 5

§ 3.2 и 3 следует, что для равномерной модели $\mathcal{U}(0, \theta)$ о. м. п. $\widehat{\theta}_n = X_{(n)}$ имеет дисперсию, порядок которой n^{-2}

Пример 10 (Нормальная модель с неизвестным средним, сверхэффективная оценка ее параметра). Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ — выборка из $\mathcal{N}(\theta, 1)$. Тогда о. м. п. $\widehat{\theta}_n = \overline{X}$ и $\mathcal{L}_\theta(\widehat{\theta}_n) = \mathcal{N}(\theta, 1/n)$. Рассмотрим оценку

$$T_n = \begin{cases} \overline{X} & \text{при} \quad |\overline{X}| \geqslant a_n, \\ b\overline{X} & \text{при} \quad |\overline{X}| < a_n, \end{cases}$$

где константа $a_n \to 0$, но $\sqrt{n}\, a_n \to \infty$ при $n \to \infty$. Найдем предельный закон распределения этой оценки. Имеем

$$\mathbf{P}_\theta\{\sqrt{n}(T_n - \theta) \leqslant x\} = p_n\, \mathbf{P}_\theta\{\sqrt{n}(\overline{X} - \theta) \leqslant x\} + (1 - p_n)\, \mathbf{P}_\theta\{\sqrt{n}(b\overline{X} - \theta) \leqslant x\},$$

где

$$p_n = \mathbf{P}_\theta\{|\overline{X}| \geqslant a_n\} = \Phi\big(\sqrt{n}(\theta - a_n)\big) + \Phi\big(-\sqrt{n}(\theta + a_n)\big) \xrightarrow[n\to\infty]{} \begin{cases} 1 & \text{при} \quad |\theta| > 0, \\ 0 & \text{при} \quad \theta = 0. \end{cases}$$

Таким образом, при $n \to \infty$ для $|\theta| > 0$ и $\theta = 0$ соответственно имеем

$$\mathcal{L}_\theta\big(\sqrt{n}(T_n - \theta)\big) \longrightarrow \mathcal{N}(0, 1), \quad \mathcal{L}_0\big(\sqrt{n}\, T_n\big) \longrightarrow \mathcal{N}(0, b^2)$$

или $\mathcal{L}_\theta(T_n) \sim \mathcal{N}\big(\theta, \sigma_T^2(\theta)/n\big)$, где

$$\sigma_T^2(\theta) = \begin{cases} 1 & \text{при} \quad |\theta| > 0, \\ b^2 & \text{при} \quad \theta = 0. \end{cases}$$

Следовательно, ее асимптотическая эффективность (14) § 3.2 (см. табл. 1 § 3.2) есть

$$e_0(T_n; \theta) = \frac{1}{\sigma_T^2(\theta)} = \begin{cases} 1 & \text{при} \quad |\theta| > 0, \\ b^{-2} & \text{при} \quad \theta = 0. \end{cases}$$

При $|b| < 1$ $e_0(T_n; \theta) \geqslant 1$ со строгим неравенством в точке $\theta = 0$, т. е. T_n — сверхэффективная оценка. Причина этого заключается в данном случае в нарушении непрерывности функции $\sigma_T^2(\theta)$ в нуле.

Сверхэффективная оценка

Точки θ, в которых $e_0(T_n; \theta) > 1$, называют *точками сверхэффективности*. В рассматриваемом примере это точка $\theta = 0$. Как показал Ле Кам (1953) множество точек сверхэффективности не более, чем счетно, поэтому явление сверхэффективности не играет существенной роли в теории оценивания. Известны также дополнительные условия регулярности на статистическую модель, которые исключают это явление. Так, С. Р. Рао доказал, что для этого достаточно потребовать равномерности по параметру θ в предельном соотношении

$$\mathcal{L}_\theta(T_n) \sim \mathcal{N}\left(\tau(\theta), \frac{\sigma_T^2(\theta)}{n}\right), \quad \forall\, \theta \in \Theta.$$

Приведем пример, иллюстрирующий роль свойства регулярности модели: при его нарушении о. м. п. не являются асимптотически нормальными. •

Пример 11 (Равномерная модель, распределение о. м. п. ее параметра). Пусть $X = (X_1, \quad , X_n)$ — выборка из распределения $U(0, \theta)$. Тогда о. м. п. $\widehat{\theta}_n = X_{(n)}$, а как было показано в § 2.4, экстремальные значения выборки не являются асимптотически нормальными. В частности, в рассматриваемом случае, из примера 6 § 2.4 имеем (поскольку $\mathcal{L}_\theta(\xi/\theta) = U(0, 1)$)

$$\mathbf{P}_\theta\left\{n\left(\frac{\widehat{\theta}_n}{\theta} - 1\right) \leqslant x\right\} \to e^x \quad x < 0. \qquad \bullet$$

В заключение этого пункта применим изложенную теорию о. м. п. к полиномиальной модели $M(n; p_1, \ldots, p_N)$ (см. примеры 10 в § 3.2 и 8 в § 3.3).

Пример 12 (Полиномиальная модель, о. м. п. для нее). Здесь функция правдоподобия

$$L(\theta) = \prod_{i=1}^{N} p_i^{\nu_i}, \quad \theta = (p_1, \ldots, p_{N-1}),$$

поэтому уравнения правдоподобия имеют вид $\nu_i/p_i - \nu_N/p_N = 0$, $i = 1, \ldots, N-1$. Они определяют единственное решение $\widehat{p}_i = \nu_i/n$, $i = 1, \ldots, N-1$, максимизирующее функцию правдоподобия, следовательно, о. м. п. $\widehat{\theta}_n = (\nu_1/n, \quad , \nu_{N-1}/n)$, т. е. выражается через полную достаточную статистику $(\nu_1, \quad , \nu_{N-1})$ (см. пример 8 в § 3.3). Подчеркнем, что о. м. п. вероятностей p_i являются относительные частоты ν_i/n реализаций соответствующих исходов в n испытаниях. Информационная матрица модели вычислена в примере 10 в § 3.2, где также указана обратная к ней матрица Σ_{N-1}. В данном случае все условия регулярности выполняются, следовательно, по теореме 2

$$\mathcal{L}_\theta\left(\sqrt{n}(\widehat{\theta}_n - \theta)\right) \longrightarrow \mathcal{N}(\underline{0}, \Sigma_{N-1}) \quad \text{при} \quad n \to \infty, \tag{27}$$

или

$$\mathcal{L}_\theta\left(\underline{\nu}^* \equiv \frac{\nu_i - np_i}{\sqrt{n}}, \quad i = 1, \ldots, N-1\right) \longrightarrow \mathcal{N}(\underline{0}, \Sigma_{N-1}).$$

Этот результат есть центральная предельная теорема для полиномиального распределения $M(n; p_1, \quad , p_N)$, сформулированная в упр. 33 к гл. 1. •

5. Асимптотические доверительные интервалы, основанные на о. м. п.

Основное значение утверждения теоремы 2 об асимптотической нормальности оценок максимального правдоподобия состоит в том, что оно позволяет сконструировать универсальный алгоритм построения доверительных интервалов и доверительных множеств (в случае векторного параметра) как для самого параметра модели, так и для произвольных достаточно хороших параметрических функций. С подобной задачей мы уже встречались в п. 5 § 2.2,

где на основании теоремы об асимптотической нормальности выборочных моментов и функций от них были построены асимптотические доверительные интервалы для теоретических моментов и функций от них. Общее определение γ-доверительного интервала для заданной теоретической характеристики наблюдаемой случайной величины дано в (26) § 2.2. Когда мы работаем в рамках параметрической модели $\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$, то оно конкретизируется следующим образом: γ-доверительным интервалом для параметрической функции $\tau(\theta)$ называется такой случайный интервал $(T_1(X), T_2(X))$, $T_1 < T_2$, который удовлетворяет условию

$$\mathbf{P}_\theta\{T_1(X) < \tau(\theta) < T_2(X)\} \geqslant \gamma, \quad \forall\, \theta \in \Theta. \tag{28}$$

Расчет таких интервалов на основе теоремы 2 (в этом случае мы будем говорить об *асимптотических доверительных интервалах*) осуществляется по той же схеме, что и в п. 5 § 2.2. Именно, на основании (22) имеем, что при $n \to \infty$ и всех θ

$$\mathbf{P}_\theta\left\{\frac{\sqrt{n}\,|\widehat{\tau}_n - \tau(\theta)|}{\sigma_\tau(\widehat{\theta}_n)} < c_\gamma\right\} \longrightarrow \Phi(c_\gamma) - \Phi(-c_\gamma) = 2\Phi(c_\gamma) - 1 = \gamma,$$

если $c_\gamma = \Phi^{-1}((1+\gamma)/2)$. Переписав это соотношение в виде

Асимптотический доверительный интервал: общее решение

$$\mathbf{P}_\theta\left\{\widehat{\tau}_n - \frac{c_\gamma \sigma_\tau(\widehat{\theta}_n)}{\sqrt{n}} < \tau(\theta) < \widehat{\tau}_n + \frac{c_\gamma \sigma_\tau(\widehat{\theta}_n)}{\sqrt{n}}\right\} \longrightarrow \gamma, \tag{29}$$

получим, что $(\widehat{\tau}_n \mp c_\gamma \sigma_\tau(\widehat{\theta}_n)/\sqrt{n})$ и есть искомый асимптотический γ-доверительный интервал для $\tau(\theta)$. Центр этого интервала находится в случайной точке $\widehat{\tau}_n = \tau(\widehat{\theta}_n)$, являющейся о. м. п. для $\tau(\theta)$, и он имеет случайную длину $2c_\gamma \sigma_\tau(\widehat{\theta}_n)/\sqrt{n}$. При этом, поскольку $\widehat{\tau}_n$ — асимптотически эффективная оценка $\tau(\theta)$, то построенный интервал будет асимптотически кратчайшим среди всех подобных интервалов, порождаемых аналогичным образом другими состоятельными и асимптотически нормальными оценками функции $\tau(\theta)$. Выделим также представляющие самостоятельный интерес частные случаи общего результата (29).

Асимптотический доверительный интервал: скалярный параметр

Если θ — скалярный параметр, то из (23) следует, что асимптотически кратчайшим γ-доверительным интервалом для него является интервал

$$\left(\widehat{\theta}_n - \frac{c_\gamma}{\sqrt{n\, i(\widehat{\theta}_n)}}, \widehat{\theta}_n + \frac{c_\gamma}{\sqrt{n\, i(\widehat{\theta}_n)}}\right), \quad c_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right) \tag{30}$$

Более того, если $\tau(\theta)$ — непрерывно дифференцируемая функция и $\tau'(\theta) \neq 0$, то $\left(\widehat{\tau}_n \mp c_\gamma \tau'(\widehat{\theta}_n)/\sqrt{n\, i(\widehat{\theta}_n)}\right)$ — аналогичный интервал для $\tau(\theta)$.

Преобразование, стабилизирующее дисперсию

Если функция τ выбрана так, что

$$\frac{\tau'(\theta)}{\sqrt{i(\theta)}} \equiv a = \text{const}, \quad \text{или} \quad \tau(\theta) = a \int \sqrt{i(\theta)}\, d\theta, \tag{31}$$

то в этом случае асимптотическая дисперсия о. м. п. $\widehat{\tau}_n$ не зависит от неизвестного параметра (такую функцию называют *преобразованием, стабилизирующим дисперсию*) и вид соответствующего интервала значительно упрощается: $(\widehat{\tau}_n \mp ac_\gamma/\sqrt{n})$, т. е. он имеет фиксированную длину. Так как функция τ монотонна (ведь $\tau'(\theta) \neq 0$), то неравенства

$$\widehat{\tau}_n - \frac{ac_\gamma}{\sqrt{n}} < \tau(\theta) < \widehat{\tau}_n + \frac{ac_\gamma}{\sqrt{n}}$$

можно разрешить относительно θ и получить γ-доверительный интервал для самого параметра θ. Например, если $\tau'(\theta) > 0$, то

$$\tau^{-1}\left(\widehat{\tau}_n - \frac{ac_\gamma}{\sqrt{n}}\right) < \theta < \tau^{-1}\left(\widehat{\tau}_n + \frac{ac_\gamma}{\sqrt{n}}\right), \tag{32}$$

где τ^{-1} — обратная функция к τ Использование таких интервалов в ряде случаев предпочтительнее, нежели (30).

Продемонстрируем эти общие принципы примерами.

Пример 13 (Общая нормальная модель, доверительный интервал для теоретической функции распределения). Построим приближенный (асимптотический) доверительный интервал для параметрической функции $\tau(\theta) = \Phi\big((x_0-\theta_1)/\theta_2\big)$ в модели $\mathcal{N}(\theta_1, \theta_2^2)$ (см. (7)). О. м. п. $\widehat{\tau}_n$ указана в примере 5, поэтому нам достаточно вычислить величину $\sigma_\tau^2(\theta)$ в (21). Информационная матрица нашей модели вычислена в примере 9 в § 3.2 (соотношение (25)). Используя этот результат находим

$$\sigma_\tau^2(\theta) = \left[\left(\frac{\partial\tau(\theta)}{\partial\theta_1}\right)^2 + \frac{1}{2}\left(\frac{\partial\tau(\theta)}{\partial\theta_2}\right)\right]\theta_2^2 = \frac{1}{2\pi}\left[1 + \frac{(x_0-\theta_1)^2}{2\theta_2^2}\right]\exp\left\{-\frac{(x_0-\theta_1)^2}{\theta_2^2}\right\}$$

Итак, искомый γ-доверительный интервал есть

$$\left(\tau(\widehat{\theta}_n) \mp \frac{c_\gamma \sigma_\tau(\widehat{\theta}_n)}{\sqrt{n}}\right),$$

где $\widehat{\theta}_n = (\widehat{\theta}_{1n}, \widehat{\theta}_{2n}) = (\overline{X}, S)$. •

Пример 14 (Доверительный интервал для параметра модели степенного ряда). Функция информации $i(\theta)$ для этой модели вычислена в примере 9. Используя этот результат и общее соотношение (30), имеем для искомого асимптотического γ-доверительного интервала для θ представления

$$\left(\widehat{\theta}_n \mp c_\gamma\sqrt{\frac{\widehat{\theta}_n}{n\mu'(\widehat{\theta}_n)}}\right) = \left(\widehat{\theta}_n \mp c_\gamma\frac{\widehat{\theta}_n}{\sigma(\widehat{\theta}_n)\sqrt{n}}\right), \tag{33}$$

где $\widehat{\theta}_n$ — решение уравнения $\mu(\theta) = \overline{X}$, а $\mu(\theta)$ и $\sigma^2(\theta)$ — теоретические среднее и дисперсия соответственно. •

Пример 15 (Пуассоновская модель, доверительный интервал для параметра). Поскольку пуассонвская модель является моделью типа степенного ряда и для нее $\mu(\theta) = \sigma^2(\theta) = \theta$, то из (33) имеем, что в данном случае асимптотический γ-доверительный интервал для θ есть $\left(\overline{X} \mp c_\gamma \sqrt{\overline{X}/n}\right)$ Далее, уравнению (31) удовлетворяет здесь функция $\tau(\theta) = \sqrt{\theta}$ при a $1/2$ (ведь $i(\theta) = 1/\theta$, см. табл. 1 § 3.2), а соответствующий интервал для нее есть $\left(\sqrt{\overline{X}} \mp c_\gamma/(2\sqrt{n})\right)$. Отсюда по формуле (32) получаем другой доверительный интервал для θ:

$$\left(\left(\sqrt{\overline{X}} - \frac{c_\gamma}{2\sqrt{n}}\right)^2 \quad \left(\sqrt{\overline{X}} + \frac{c_\gamma}{2\sqrt{n}}\right)^2\right),$$

который совпадает с предыдущим, если пренебречь членами порядка n^{-1} •

Пример 16 (Геометрическая модель, доверительный интервал для параметра). Оценим параметр θ геометрической модели $\overline{\mathrm{Bi}}(1, \theta)$. Это также модель типа степенного ряда, поэтому применимо общее решение (33). Здесь (см. (7) § 1.1) $\mu(\theta) = \theta/(1-\theta)$, $\sigma^2(\theta) = \theta/(1-\theta)^2$ решением уравнения правдоподобия $\mu(\theta) = \overline{X}$ является о. м. п. $\widehat{\theta}_n = \overline{X}/(1+\overline{X})$, поэтому искомый γ-доверительный интервал имеет вид

$$\left(\widehat{\theta}_n \mp c_\gamma(1-\widehat{\theta}_n)\sqrt{\frac{\widehat{\theta}_n}{n}}\right) = \left(\frac{\overline{X}}{1+\overline{X}} \mp c_\gamma \sqrt{\frac{\overline{X}}{n(1+\overline{X})^3}}\right) \quad \bullet$$

Пример 17 (Бернуллиевская модель, асимптотический доверительный интервал для параметра). В рассматриваемом случае о. м. п. $\widehat{\theta}_n = \overline{X}$ и функция информации $i(\theta) = [\theta(1-\theta)]^{-1}$ (см. табл. 1 § 3.2), поэтому из (30) имеем, что $\left(\overline{X} \mp c_\gamma \sqrt{\overline{X}(1-\overline{X})/n}\right)$ — асимптотический γ-доверительный интервал для θ Решением уравнения (31) является здесь функция $\tau(\theta) = \arcsin\sqrt{\theta}$ при $a = 1/2$, поэтому соответствующий интервал для нее есть $\left(\tau(\overline{X}) \mp \frac{c_\gamma}{2\sqrt{n}}\right)$ Если воспользоваться соотношениями (32) и при этом пренебречь членами порядка n^{-1} то получим предыдущее решение для θ (это мы предлагаем проверить читателю самостоятельно). •

Доверительные области: многомерный параметр

В случае векторного параметра $\theta = (\theta_1, \ldots, \theta_r)$ может представлять практический интерес доверительное оценивание самого этого вектора. В этом случае нужно указать такое случайное (т. е. зависящее от выборки X) подмножество $\mathcal{G}_\gamma(X)$ в параметрическом множестве Θ рассматриваемой модели,

которое накрывает истинное значение неизвестного параметра θ с вероятностью, не меньшей заданного доверительного уровня γ, т. е. чтобы выполнялось соотношение, аналогичное (28):

$$\mathbf{P}_\theta\{\theta \in \mathcal{G}_\gamma(X)\} \geqslant \gamma, \quad \forall\, \theta \in \Theta. \tag{34}$$

Такое подмножество $\mathcal{G}_\gamma(X) \subset \Theta$ называется *γ-доверительной областью* для θ. В случае регулярных моделей, удовлетворяющих условиям теоремы 2, для больших выборок легко построить приближенные доверительные области, основанные на оценках максимального правдоподобия. Для этого надо воспользоваться следующим утверждением, представляющим собой асимптотический вариант результата, сформулированного в упр. 40 к гл. 1 (см. также Замечание в п. 2 § 2.5): если для k-мерного случайного вектора $\underline{\xi}_n$ выполняется асимптотическое (при $n \to \infty$) соотношение $\mathcal{L}(\underline{\xi}_n) \sim \mathcal{N}(\underline{\mu}_n, \Sigma_n)$ и $|\Sigma_n| \neq 0$ при всех n, то

$$\mathcal{L}\Big((\underline{\xi}_n - \underline{\mu}_n)'\Sigma_n^{-1}(\underline{\xi}_n - \underline{\mu}_n)\Big) \longrightarrow \chi^2(k). \tag{35}$$

В силу этого утверждения из (20) имеем, что квадратичная форма

$$Q_n(\theta) = n(\widehat{\theta}_n - \theta)I(\theta)(\widehat{\theta}_n - \theta) \tag{36}$$

при всех θ имеет предельное при $n \to \infty$ распределение $\chi^2(r)$. Более того, если информационная матрица $I(\theta)$ непрерывна по θ, то на основании утверждения 4° можно заключить, что в (36) $I(\theta)$ можно заменить на $I(\widehat{\theta}_n)$ с сохранением при этом формы предельного распределения. Итак, при выполнении указанного условия имеет место предельное соотношение

$$\mathcal{L}_\theta\Big(\widehat{Q}_n(\theta) = n(\widehat{\theta}_n - \theta)'I(\widehat{\theta}_n)(\widehat{\theta}_n - \theta)\Big) \longrightarrow \chi^2(r). \tag{37}$$

Пусть, как обычно, $\chi^2_{\gamma,r}$ обозначает γ-квантиль распределения $\chi^2(r)$. Тогда из (37) имеем, что при $n \to \infty$

$$\mathbf{P}_\theta\{\widehat{Q}_n(\theta) \leqslant \chi^2_{\gamma,r}\} \longrightarrow \gamma. \tag{38}$$

Асимптотический доверительный эллипсоид для параметра регулярной модели

Определим теперь область $\mathcal{G}_\gamma(X) \subset \Theta$ условием

$$\mathcal{G}_\gamma(X) = \{\theta \;\; \widehat{Q}_n(\theta) \leqslant \chi^2_{\gamma,r}\} \tag{39}$$

Это есть некоторый случайный эллипсоид с центром в точке $\widehat{\theta}_n$ и, по построению, при $n \to \infty$

$$\mathbf{P}_\theta\{\theta \in \mathcal{G}_\gamma(X)\} = \mathbf{P}_\theta\{\widehat{Q}_n(\theta) \leqslant \chi^2_{\gamma,r}\} \longrightarrow \gamma,$$

т. е. $\mathcal{G}_\gamma(X)$ есть асимптотическая γ-доверительная область для θ

Пример 18 (Доверительная область для параметров полиномиальной модели). Если в полиномиальной модели $M(n; p_1, \quad , p_N)$, $\theta = (p_1, \ldots, p_{N-1})$, число испытаний n велико, то для одновременного оценивания неизвестных

вероятностей p_i можно воспользоваться изложенной асимптотической теорией. Все предварительные вычисления нами уже проведены в примере 12 и примере 10 в § 3.2, используя которые вычисляем квадратичную форму в (37):

$$\widehat{Q}_n(\theta) = n\sum_{i,j=1}^{N-1}(\widehat{p}_i - p_i)(\widehat{p}_j - p_j)\widehat{p}_N^{-1} + n\sum_{i=1}^{N-1}(\widehat{p}_i - p_i)^2\widehat{p}_i^{-1} =$$

$$= \frac{\left[\sum_{i=1}^{N-1}(\nu_i - np_i)\right]^2}{\nu_N} + \sum_{i=1}^{N-1}\frac{(\nu_i - np_i)^2}{\nu_i} = \sum_{i=1}^{N}\frac{(\nu_i - np_i)^2}{\nu_i}. \qquad (40)$$

При этом параметр r в (37) равен здесь $N-1$. В результате искомая область (39) может быть записана в виде

$$\mathcal{G}_\gamma(X) = \mathcal{G}_\gamma(\underline{\nu}) = \left\{(p_1, \quad , p_N) \ \sum_{i=1}^{N}\frac{(\nu_i - np_i)^2}{\nu_i} \leqslant \chi^2_{\gamma,\,N-1}, \sum_{i=1}^{N} p_i = 1\right\}. \qquad (41)$$

Эта область представляет собой пересечение N-мерного эллипсоида с центром в точке $\underline{\nu} = (\nu_1, \quad , \nu_N)$ с гиперплоскостью $p_1 + \quad + p_N = 1$.

Если $N = 2$, то речь идет об оценивании параметра бернуллиевской модели $\mathrm{Bi}(1, \theta)$ по n испытаниям, и легко видеть, что область (41) сводится к интервалу, построенному в примере 17. •

§ 3.6. Другие методы и принципы построения оценок

Правильно в философии рассматривать сходство даже в вещах, далеко отстоящих друг от друга.

Аристотель [5)]

Кроме рассмотренных в предыдущих параграфах общих подходов для оценивания неизвестных параметров распределений на практике используют и другие методы построения оценок. В настоящем параграфе мы лишь кратко затронем соответствующие проблемы, чтобы показать их многообразие и одновременно расширить горизонты для читателя.

1. Метод моментов

Исторически первым общим методом точечного оценивания неизвестных параметров является *метод моментов*, предложенный К. Пирсоном (1894). Суть этого метода состоит в следующем.

Пусть, как обычно, $X = (X_1, \ldots, X_n)$ — выборка из распределения $\mathcal{L}(\xi) \in \mathcal{F} = \{F(x, \theta),\ \theta \in \Theta\}$, где $\theta = (\theta_1, \quad , \theta_r)$ и $\Theta \subseteq R^r$ Предположим,

[5)] Аристотель (384–322 гг. до н. э.) — древнегреческий философ, ученый-энциклопедист.

что у наблюдаемой случайной величины ξ существуют первые r моментов $\alpha_k = \mathbf{E}\xi^k \quad k = 1, \quad , r$ Они являются функциями от неизвестных параметров θ: $\alpha_k = \alpha_k(\theta)$. Рассмотрим соответствующие выборочные моменты $\widehat{\alpha}_k$ (см. п. 1 § 2.2) и составим систему уравнений относительно θ:

$$\alpha_k(\theta) = \widehat{\alpha}_k, \quad k = 1, . \quad , r \tag{1}$$

Предположим, что эта система однозначно разрешима и обозначим ее решение $\widetilde{\theta}_i = \widetilde{\theta}_i(X) = \varphi_i(\widehat{\alpha}_1, \quad , \widehat{\alpha}_r)$, $i = 1, \quad , r$ Тогда $\widetilde{\theta}_1, \quad , \widetilde{\theta}_r$ и есть оценки параметров $\theta_1, \quad , \theta_r$ по методу моментов. Если функции φ_i, $i = 1, \quad , r$, непрерывны, то из состоятельности выборочных моментов (см. п. 3 § 2.2) по теореме 1 § 2.2 следует, что статистики $\widetilde{\theta}_i(X)$ являются состоятельными оценками θ_i: при $n \to \infty$

Оценки по методу моментов

$$\widetilde{\theta}_i(X) \xrightarrow{\mathrm{P}} \varphi_i(\alpha_1, . \quad , \alpha_r) = \theta_i, \quad i = 1, . . \quad , r$$

Таким образом, метод моментов при определенных условиях приводит к состоятельным оценкам; при этом уравнения (1) во многих случаях просты и их решение (в отличие, например, от метода максимального правдоподобия) не связано с большими вычислительными трудностями.

Пример 1. Пусть $\mathcal{L}(\xi) \in \mathcal{N}(\theta_1, \theta_2^2)$. Здесь $\alpha_1 = \mathbf{E}\xi = \theta_1$, $\theta_2^2 = \mathbf{D}\xi = \alpha_2 - \alpha_1^2$, поэтому $\widetilde{\theta}_1 = \widehat{\alpha}_1 = \overline{X}$, $\widetilde{\theta}_2 = \sqrt{\widehat{\alpha}_2 - \widehat{\alpha}_1^2} = S$, т. е. $(\widetilde{\theta}_1, \widetilde{\theta}_2)$ совпадает с оценками максимального правдоподобия $(\widehat{\theta}_1, \widehat{\theta}_2) = (\overline{X}, S)$ (см. пример 1 § 3.5). •

Пример 2. Пусть $\mathcal{L}(\xi) \in \mathcal{U}(\theta_1, \theta_2)$. Здесь

$$\alpha_1 = \frac{\theta_1 + \theta_2}{2}, \quad \alpha_2 = \frac{\theta_1^2 + \theta_1\theta_2 + \theta_2^2}{3}$$

и система (1) имеет решение $\widetilde{\theta}_1 = \overline{X} - \sqrt{3}\,S$, $\widetilde{\theta}_2 = \overline{X} + \sqrt{3}\,S$. Эти оценки не являются функциями достаточной статистики $(X_{(1)}, X_{(n)})$ (см. пример 3 § 3.3), поэтому они мало эффективны. •

Пример 3 (Гамма-модель, оценивание параметров методом моментов). Пусть $\mathcal{L}(\xi) \in \Gamma(\theta_1, \theta_2)$ (см. п. 3 § 1.2). Здесь $\Theta = \{\theta = (\theta_1, \theta_2) \quad \theta_i > 0, \ i = 1, 2\}$ и $\alpha_k = \theta_1^k \Gamma(\theta_2 + k)/\Gamma(\theta_2) = \theta_1^k \theta_2(\theta_2 + 1) \quad (\theta_2 + k - 1)$, в частности, $\alpha_1 = \theta_1\theta_2$, $\alpha_2 = \theta_1^2\theta_2(\theta_2 + 1)$. Система (1) имеет здесь решение

$$\widetilde{\theta}_1(X) = \frac{\widehat{\alpha}_2 - \widehat{\alpha}_1^2}{\widehat{\alpha}_1} = \frac{S^2}{\overline{X}}, \quad \widetilde{\theta}_2(X) = \frac{\widehat{\alpha}_1^2}{\widehat{\alpha}_2 - \widehat{\alpha}_1^2} = \frac{\overline{X}^2}{S^2}.$$

Гамма-модель с двумя незвестными параметрами весьма сложна для анализа какими-нибудь другими известными нам методами, метод же моментов быстро приводит к простым формулам для оценок ее параметров. •

Отметим, что метод моментов неприменим, когда теоретические моменты нужного порядка не существуют (например, для распределения Коши). Кроме того, оценки метода моментов, вообще говоря, мало эффективны, поэтому

их обычно используют лишь в качестве первых приближений, основываясь на которых можно находить последующие приближения с бóльшей эффективностью.

2. Метод минимума хи-квадрат

Систему различных методов построения оценок можно получить, конструируя тем или иным способом меру D отклонения эмпирической функции распределения $\widehat{F}_n(x)$ от $F(x;\theta)$ — истинной (но неизвестной) теоретической функции распределения выборки. В любом случае такая мера является функцией от выборки (данных) X и неизвестного параметра θ: $D = D(X;\theta)$. Если такая мера сконструирована, то в качестве оценки параметра θ естественно взять значение, минимизирующее меру D: $\widehat{\theta}(X) = \arg\min_{\theta} D(X;\theta)$.

Одной из наиболее известных таких мер является предложенная К. Пирсоном *мера хи-квадрат*. Эта мера конструируется следующим образом. Разобьем множество $\mathcal{E}$ возможных значений наблюдаемой случайной величины ξ на N непересакающихся подмножеств («интервалов»)

$$\mathcal{E}_1, \ldots, \mathcal{E}_N \quad (\mathcal{E} = \mathcal{E}_1 \cup . \quad \cup \mathcal{E}_N, \quad \mathcal{E}_i \cap \mathcal{E}_j = \varnothing, \quad i \neq j),$$

и пусть ν_j — число элементов выборки $X = (X_1, \ldots, X_n)$, попавших в интервал $\mathcal{E}_j$:

$$\nu_j = \sum_{i=1}^{n} I(X_i \in \mathcal{E}_j), \quad j = 1, \ldots, N \quad (\nu_1 + \quad + \nu_N = n).$$

Обозначим, далее, через $p_j(\theta)$ теоретическую вероятность попадания в $\mathcal{E}_j$:

$$p_j(\theta) = \mathbf{P}_\theta\{\xi \in \mathcal{E}_j\} = \int_{\mathcal{E}_j} dF(x;\theta), \quad j = 1, \ldots, N \quad (p_1(\theta) + \ldots + p_N(\theta) = 1).$$

Поскольку относительная частота ν_j/n попадания в интервал $\mathcal{E}_j$ является состоятельной оценкой вероятности $p_j(\theta)$, то одной из возможных мер отклонения эмпирических (выборочных) данных от теоретических значений может быть мера

$$D = \sum_{j=1}^{N} c_j \left(\frac{\nu_j}{n} - p_j(\theta)\right)^2$$

где $c_1, \quad , c_N$ — некоторые «веса».

Если здесь положить «веса» $c_j = n/p_j(\theta)$, то мы и получим меру хи-квадрат:

$$\chi^2 = \sum_{j=1}^{N} \frac{n}{p_j(\theta)} \left(\frac{\nu_j}{n} - p_j(\theta)\right)^2 = \sum_{j=1}^{N} \frac{\left(\nu_j - np_j(\theta)\right)^2}{np_j(\theta)} = \sum_{j=1}^{N} \frac{\nu_j^2}{np_j(\theta)} - n. \qquad (2)$$

Выбор именно таких «весовых» коэффициентов в D оправдан тем, что, как это будет показано позже (см. п. 2 и 3 § 4.2), полученная при этом мера хи-квадрат (2) обладает следующим замечательным свойством: ее асимптотическое

распределение для больших выборок (при $n \to \infty$) зависит лишь от числа N интервалов (подмножеств) разбиения и является стандартным распределением хи-квадрат.

Оценка по методу минимума χ^2

Оценку θ, полученную из условия обращения меры χ^2 в минимум, называют *оценкой по методу минимума* χ^2. При достаточно общих условиях эта оценка оказывается состоятельной, асимптотически нормальной и асимптотически эффективной (как и оценка максимального правдоподобия).

Для нахождения указанной оценки надо решить систему уравнений

$$\sum_{j=1}^{N} \frac{\nu_j^2}{p_j^2(\theta)} \frac{\partial p_j(\theta)}{\partial \theta_k} = 0, \quad k = 1, \quad , r \quad (\theta = (\theta_1, \quad , \theta_r)).$$

Однако это трудно сделать даже в простейших случаях, поэтому, учитывая, что

$$\frac{\nu_j}{n p_j(\theta)} \xrightarrow[n\to\infty]{\mathrm{P}} 1, \quad j = 1, \ldots, N,$$

эту систему обычно заменяют более простой системой

$$\sum_{j=1}^{N} \frac{\nu_j}{p_j(\theta)} \frac{\partial p_j(\theta)}{\partial \theta_k} = 0, \quad k = 1, \ldots, r, \tag{3}$$

решение которой находить, как правило, гораздо проще. Оценки параметров θ, получающиеся как решение системы (3), называют *оценками по видоизмененному методу минимума* χ^2; для больших выборок их асимптотические свойства остаются такими же, как сказано выше.

Замечание. В описанном методе оценивания параметров используются не сами наблюдения $X_1, \ldots, X_n$, а лишь частоты $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ попадания элементов выборки в соответствующие подмножества $\mathcal{E}_1, \ldots, \mathcal{E}_N$. Такой «частотный» способ представления статистических данных называют *методом группировки наблюдений*, а подмножества $\mathcal{E}_1, \ldots, \mathcal{E}_N$ — *подмножествами* (или *интервалами*) группировки. При этом распределение вектора частот $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ является, очевидно, полиномиальным распределением $M(n; p_1(\theta), \quad , p_N(\theta))$ (см. п. 6 § 1.1). В этом контексте вероятность

$$\mathbf{P}_\theta\{\underline{\nu} = \underline{h}\} \equiv L^{(1)}(\underline{h}; \theta) = \frac{n!}{h_1! \ldots h_N!} \prod_{j=1}^{N} p_j^{h_j}(\theta), \quad h_1 + \quad + h_N = n, \tag{4}$$

называют *частотной функцией правдоподобия в методе группировки наблюдений* (в отличие от обычной функции правдоподобия $L(\underline{x}; \theta) \quad \prod_{i=1}^{n} f(x_i; \theta)$). Если искать оценку параметра θ по методу максимального правдоподобия, основываясь на функции $L^{(1)}(\underline{h}; \theta)$, т. е. искать ее точку максимума, решая систему уравнений

$$\frac{\partial \ln L^{(1)}(\underline{h}; \theta)}{\partial \theta_k} = 0, \quad k = 1, \ldots, r,$$

то, как легко видеть, эта система сводится к системе (3).

3. Модально несмещенные оценки

До сих пор мы в основном рассматривали несмещенные оценки, т. е. такие статистики $\delta(X)$, которые удовлетворяют условию (если оценивается сам параметр θ)

$$\mathbf{E}_\theta \delta(X) = \theta, \quad \forall\, \theta \in \Theta.$$

В силу того, что математическое ожидание случайной величины часто называют средним значением, несмещенную оценку $\delta(X)$ называют также *несмещенной в среднем*. Однако в статистике используются и другие понятия несмещенности, в частности, модальной несмещенности и медианной несмещенности, в основе определений которых лежат такие характеристики статистической оценки, как соответственно мода и медиана. Их мы кратко и обсудим в этом и следующем пунктах. Остановимся сначала на понятии модальной несмещенности.

Пусть X — случайная величина с плотностью $f(x)$, $x \in R^1$. Любая точка максимума функции $f(x)$ называется *модой* случайной величины X (ее закона распределения). Если плотность зависит от параметра: $f(x) = f(x; \theta)$, $\theta \in \Theta$, то и мода, вообще говоря, зависит от θ. Если мода единственна, то распределение величины X называют *унимодальным*. Унимодальные распределения играют важную роль в теории вероятностей и математической статистике. Примерами таких распределений являются нормальное, показательное, Пуассона, Коши и др. распределения.

Определение Пусть оценка $\delta(X)$ параметра θ имеет унимодальное распределение и $\varkappa(\theta)$ (каппа от тэта) — его мода. Если $\varkappa(\theta) = \theta$ при всех $\theta \in \Theta$, то $\delta(X)$ называется *модально несмещенной оценкой* параметра θ.

Такие оценки могут оказаться предпочтительнее других при некоторых функциях потерь. Здесь уместно сделать общее замечание о терминологии, используемой в общей теории оценивания. В этой теории при выборе критерия качества (или точности) оценки исходят из некоторой неотрицательной функции $L(\delta, \theta)$, измеряющей расхождение между параметром θ и его оценкой $\delta = \delta(X)$ — такая функция называется *функцией потерь*. Средние потери $\mathbf{E}_\theta L(\delta, \theta) = R(\delta, \theta)$ называются *функцией риска* (или просто *риском*) оценки δ, и риск является мерой точности оценки. Если $L(\delta, \theta) = (\delta - \theta)^2$, то говорят о *квадратичной функции потерь*, тогда *квадратичный риск* $R(\delta, \theta) = \mathbf{E}_\theta(\delta - \theta)^2$ есть среднеквадратичная ошибка (с. к. о.). Иногда используются и другие функции потерь, например, $L(\delta, \theta) = |\delta - \theta|$ или $L(\delta, \theta) = (1 - \delta/\theta)^2$, или $L(\delta, \theta) = (\delta^{-1} - \theta^{-1})^2$ и т. д.

Примеры модально несмещенных оценок

Пример 4. Пусть $X = (X_1, \quad, X_n)$ — выборка из нормального распределения $\mathcal{N}(\theta, \sigma^2)$. В качестве оценки для θ возьмем оптимальную несмещенную оценку $\delta(X) = \overline{X}$. Поскольку $\mathcal{L}_\theta(\overline{X}) = \mathcal{N}(\theta, \sigma^2/n)$, а нормальное распределение является унимодальным, причем его мода $\varkappa(\theta)$ совпадает с математическим ожиданием θ, то в данном случае $\varkappa(\theta) = \theta$ при всех θ и потому

обычная несмещенная оценка $\overline{X}$ является здесь одновременно и модально несмещенной оценкой θ. •

Пример 5. Пусть $X = (X_1, \ldots, X_n)$ — выборка из равномерного распределения $\mathcal{U}(0, \theta)$. Покажем, что статистика $\delta(X) = X_{(n)}$ — максимальное значение выборки — является модально несмещенной оценкой параметра θ. Плотность статистики $X_{(n)}$ следует из общего соотношения (13) § 3.4 при $a = 0$, $Q_1(\theta) = 1/\theta$, $M_1(x) \equiv 1$ (см. пример 3 в § 3.4) и имеет вид

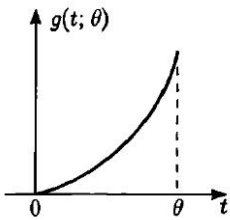


Рис. 1

$$g(t; \theta) = \frac{nt^{n-1}}{\theta^n} I(0 \leqslant t \leqslant \theta).$$

Так как (см. рис. 1) при любом $\theta > 0$ функция $g(t; \theta)$ имеет единственный максимум в точке $\varkappa(\theta) = \theta$, то статистика $X_{(n)}$ является модально несмещенной оценкой для θ. Для сравнения напомним (см. пример 3 в § 3.4), что оптимальной несмещенной оценкой для θ является статистика $\theta^* = ((n+1)/n)X_{(n)}$. •

Пример 6 (Показательное распределение, модально несмещенная оценка параметра). Пусть $X = (X_1, \ldots, X_n)$, $n \geqslant 3$ — выборка из показательного распределения $\Gamma(\theta, 1)$ (см. п. 3 § 1.2). Рассмотрим статистику

$$\delta_c(X) = c \sum_{i=1}^{n} X_i, \quad c > 0.$$

Из свойств гамма-распределения имеем, что

$$\mathcal{L}_\theta(\delta_c(X)) = \Gamma(c\theta, n),$$

т. е. плотность статистики $\delta_c(X)$ есть (см. (18) § 1.2)

$$f(x|c\theta, n) = \frac{x^{n-1}e^{-x/c\theta}}{\Gamma(n)(c\theta)^n}, \quad x \geqslant 0.$$

Здесь функция $f(x|c\theta, n)$ имеет при любом θ единственный максимум в точке $x = \varkappa(\theta) = c\theta(n-1)$, следовательно, при $c = 1/(n-1)$ статистика

$$\delta_c(X) = V_n = \frac{1}{n-1} \sum_{i=1}^{n} X_i$$

будет модально несмещенной оценкой θ.

Мы знаем, что при квадратичной функции потерь несмещенная в среднем оценка $\overline{X}$ является оптимальной несмещенной оценкой для θ (см., например, табл. 2 в § 3.2). Сравним теперь риски этих двух оценок, V_n и $\overline{X}$, относительно функции потерь $L(\delta, \theta) = (\delta^{-1} - \theta^{-1})^2$ Используя формулы для моментов гамма-распределения (см. п. 3 § 1.2), нетрудно получить (это мы оставляем читателю в качестве упражнения), что

$$R(\overline{X}, \theta) = \mathbf{E}_\theta \left(\frac{1}{\overline{X}} - \frac{1}{\theta} \right)^2 = \frac{n+2}{(n-1)(n-2)\theta^2},$$

$$R(V_n, \theta) = \mathbf{E}_\theta \left(\frac{1}{V_n} - \frac{1}{\theta} \right)^2 = \frac{1}{(n-2)\theta}.$$

Отсюда имеем $R(V_n, \theta) < R(\overline{X}, \theta)$, $\forall\, \theta > 0$, т. е. при такой функции потерь модально несмещенная оценка имеет равномерно меньший риск, нежели оценка, несмещенная в среднем. •

4. Медианно несмещенные оценки

Определение медианы для произвольного распределения F было дано в п. 1 § 2.4 — это p-квантиль ζ_p при $p = 1/2$. Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения с плотностью $f(x; \theta)$, $\theta \in \Theta$, и пусть $\delta(X)$ — некоторая оценка параметра θ. Обозначим ее медиану $m(\theta)$.

В таком случае, если

$$m(\theta) = \theta \quad \text{при всех} \quad \theta \in \Theta,$$

то статистика $\delta(X)$ называется *медианно несмещенной оценкой* θ.

Пример 7. В условиях примера 6 рассмотрим статистику $X_{(1)}$ — минимальное значение выборки. Ее плотность по формуле (6) § 2.4 имеет вид

Показательное распределение: оценивание параметра

$$g_1(x; \theta) = \frac{n}{\theta} e^{-nx/\theta}, \quad x \geqslant 0,$$

поэтому, решая уравнение

$$\frac{1}{2} = \int_0^x g_1(t; \theta)\, dt = 1 - e^{-nx/\theta}$$

найдем медиану

$$m(\theta) = \frac{1}{n} \theta \ln 2.$$

Следовательно, статистика $\delta(X) = nX_{(1)}/(\ln 2)$ является медианно несмещенной оценкой θ. Найдем распределение этой статистики. Так как (см. (18) § 1.2)

$$\mathcal{L}_\theta(X_{(1)}) = \Gamma\left(\frac{\theta}{n}, 1\right), \quad \text{то} \quad \mathcal{L}_\theta(\delta(X)) = \Gamma\left(\frac{\theta}{\ln 2}, 1\right),$$

т. е. распределение $\delta(X)$ от n вообще не зависит и потому ее с. к. о. не сходится к 0 при $n \to \infty$, более того, $\delta(X)$ вообще не является состоятельной оценкой. •

Разницу между различными понятиями несмещенности можно проиллюстрировать на примере гамма-модели $\Gamma(\theta, \lambda)$ (см. п. 3 § 1.2). Если рассмотреть статистику $T(X) = X_1 + \ldots + X_n$, то, так как $\mathcal{L}_\theta(T(X)) = \Gamma(\theta, n\lambda)$, ее плотность есть

Гамма-модель оценивания параметра

$$g(t; \theta) = \frac{t^{n\lambda - 1}}{\Gamma(n\lambda)\theta^{n\lambda}} e^{-t/\theta}, \quad t \geqslant 0.$$

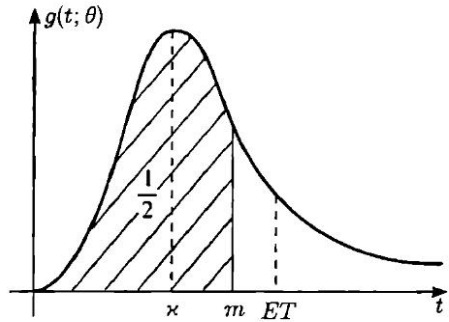


Рис. 2

График этой плотности, а также мода $\varkappa$, медиана m и среднее значение ET изображены на рис. 2. Здесь $\mathbf{E}_\theta T = n\lambda\theta$, т. е. $T/(n\lambda)$ — несмещенная в среднем оценка θ. Мода $\varkappa$ статистики T находится из уравнения $\dfrac{\partial g(t;\theta)}{\partial t} = 0$ и равна $\theta(n\lambda - 1)$, откуда следует, что модально несмещенной оценкой параметра θ будет статистика $\delta(X) = T/(n\lambda - 1)$. Медиана же m статистики T является корнем уравнения

$$\frac{1}{\Gamma(n\lambda)\theta^{n\lambda}} \int_0^m t^{n\lambda-1} e^{-t/\theta}\, dt = \frac{1}{2}, \quad \text{или} \quad \frac{1}{\Gamma(n\lambda)} \int_0^{m/\theta} t^{n\lambda-1} e^{-t}\, dt = \frac{1}{2}.$$

Отсюда следует, что m пропорционально θ: $m = \theta A(n\lambda)$, где $A(n\lambda)$ — константа, не зависящая от θ, следовательно, медианно несмещенной оценкой θ будет статистика $\delta_1(X) = T/A(n\lambda)$.

5. Эквивариантные оценки

В некоторых случаях класс рассматриваемых оценок параметра θ разумно ограничить заранее некоторым условием инвариантности относительно того или иного преобразования данных. Рассмотрим, например, задачу оценивания *параметра сдвига*, т. е. когда плотность наблюдаемой случайной величины ξ имеет вид $f(x;\theta) = f(x-\theta)$, где f — известная функция (здесь $x \in R^1$ $\theta \in R^1$). В этом случае для любой константы c

Параметр сдвига: эквивариантная оценка

$$\mathbf{P}_\theta\{X_i + c \leqslant x\} = \mathbf{P}_\theta\{X_i \leqslant x - c\} =$$

$$\int_{-\infty}^{x-c} f(t-\theta)\, dt = \int_{-\infty}^{x} f(u-c-\theta)\, du = \mathbf{P}_{\theta+c}\{X_i \leqslant x\}.$$

Поэтому разумно рассмотреть в качестве оценок параметра θ статистики $T = T(X_1, \ldots, X_n)$, инвариантные относительно сдвига, т. е. удовлетворя-

ющие условию

$$T(X_1 + c, \quad , X_n + c) = T(X) + c. \tag{5}$$

Оценки параметра сдвига θ, удовлетворяющие условию (5), называются *эквивариантными оценками*.

Легко видеть, что статистики $T_1(X) = \overline{X}$ и $T_2(X) = 1/2(X_{(1)} + X_{(n)})$ являются эквивариантными оценками параметра сдвига θ. Действительно, для любого числа c

$$T_1(X_1 + c, \ldots, X_n + c) = \frac{1}{n}\sum_{i=1}^{n}(X_i + c) = \overline{X} + c = T_1(X) + c,$$

$$T_2(X_1 + c, \quad , X_n + c) = \frac{1}{2}\big((X_{(1)} + c) + (X_{(n)} + c)\big) = T_2(X) + c.$$

Статистики же $T_3(X) = S^2$ и $T_4(X) = X_{(n)} - X_{(1)}$ условию (5) не удовлетворяют.

Покажем теперь, что с. к. о. $R(T, \theta) = \mathbf{E}_\theta(T(X) - \theta)^2$ любой эквивариантной оценки $T = T(X)$ не зависит от оцениваемого параметра сдвига θ. Действительно,

$$R(T,\theta) = \int (T(\underline{x}) - \theta)^2 \prod_{i=1}^{n} f(x_i - \theta)\, d\underline{x} =$$

$$= \int T^2(x_1 - \theta, \ldots, x_n - \theta) \prod_{i=1}^{n} f(x_i - \theta)\, d\underline{x} = \int T^2(\underline{x}) \prod_{i=1}^{n} f(x_i)\, d\underline{x} = R(T, 0).$$

В свете этого представляется естественным в качестве оценки параметра сдвига выбрать такую эквивариантную оценку T, которая минимизирует с. к. о. $R(T, 0)$ в классе всех эквивариантных оценок параметра сдвига θ. Такая оценка, называемая *оценкой Питмена*, существует и имеет вид

Оценка Питмена

$$\theta_n^* = \frac{\displaystyle\int \theta \prod_{i=1}^{n} f(X_i - \theta)\, d\theta}{\displaystyle\int \prod_{i=1}^{n} f(X_i - \theta)\, d\theta}. \tag{6}$$

Таким образом, оценка Питмена (6) является наилучшей в смысле минимума квадратичного риска (с. к. о.) несмещенной эквивариантной оценкой параметра сдвига θ.

Пример 8 (Оценка Питмена для среднего нормального распределения). Пусть $\mathcal{L}(\xi) \in \mathcal{N}(\theta, 1)$, т. е. $f(x; \theta) = \dfrac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}$ Здесь неизвестный параметр (среднее) является параметром сдвига, поэтому оценка Питмена (6) для

него есть

$$\theta_n^* = \frac{\int\limits_{-\infty}^{\infty} \theta \exp\left\{-\frac{1}{2}\sum\limits_{i=1}^{n}(X_i-\theta)^2\right\} d\theta}{\int\limits_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\sum\limits_{i=1}^{n}(X_i-\theta)^2\right\} d\theta} = \frac{\int\limits_{-\infty}^{\infty} \theta \exp\left\{-\frac{n\theta^2}{2}+n\theta\overline{X}\right\} d\theta}{\int\limits_{-\infty}^{\infty} \exp\left\{-\frac{n\theta^2}{2}+n\theta\overline{X}\right\} d\theta} =$$

$$= \frac{\int\limits_{-\infty}^{\infty} \theta \exp\left\{-\frac{n}{2}(\theta-\overline{X})^2\right\} d\theta}{\int\limits_{-\infty}^{\infty} \exp\left\{-\frac{n}{2}(\theta-\overline{X})^2\right\} d\theta} = \overline{X},$$

т. е. θ_n^* совпадает с выборочным средним $\overline{X}$, которое является эффективной оценкой θ (см. табл. 2 § 3.2). •

Пример 9 (Оценка Питмена для параметра сдвига равномерного распределения). Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ — выборка из равномерного распределения $\mathcal{U}(\theta - 1/2, \theta + 1/2)$, т. е. здесь $f(x;\theta) = I(\theta - 1/2 \leqslant x \leqslant \theta + 1/2) = I(-1/2 \leqslant x - \theta \leqslant 1/2)$, следовательно, θ является параметром сдвига. Поскольку

$$\prod_{i=1}^{n} f(X_i;\theta) = I\left(-\frac{1}{2} \leqslant X_{(1)} - \theta \leqslant X_{(n)} - \theta \leqslant \frac{1}{2}\right) =$$

$$= I\left(X_{(n)} - \frac{1}{2} \leqslant \theta \leqslant X_{(1)} + \frac{1}{2}\right), \tag{7}$$

то оценка Питмана (6) принимает здесь вид

$$\theta_n^* = \frac{\int\limits_{X_{(n)}-1/2}^{X_{(1)}+1/2} \theta\, d\theta}{\int\limits_{X_{(n)}-1/2}^{X_{(1)}+1/2} d\theta} = \frac{1}{2}\frac{\left(X_{(1)}+\frac{1}{2}\right)^2 - \left(X_{(n)}-\frac{1}{2}\right)^2}{\left(X_{(1)}+\frac{1}{2}\right) - \left(X_{(n)}-\frac{1}{2}\right)} = \frac{1}{2}(X_{(1)} + X_{(n)}), \tag{8}$$

т. е. совпадает с несмещенной оценкой максимального правдоподобия $\widehat{\theta}$ (из (7) следует, что любое $\theta \in [X_{(n)} - 1/2, X_{(1)} + 1/2]$ максимизирует функцию правдоподобия, а средняя точка этого интервала (8) в силу упр. 46 к гл. 1 является несмещенной оценкой θ). •

Замечание. Можно показать, что оценка Питмена (6) может быть получена по формуле

$$\theta_n^* = T(X) - \mathbf{E}_0(T(X)|Y),$$

где $T(X)$ — произвольная эквивариантная оценка, $Y = (X_1 - X_n, \ldots, X_{n-1} - X_n)$ и $\mathbf{E}_0$ — символ математического ожидания $\mathbf{E}_\theta$ при $\theta = 0$; при этом вид оценки θ_n^* от выбора $T(X)$ не зависит.

Оценивание параметра масштаба

Рассмотрим теперь аналогичный подход в задаче оценивания *параметра масштаба*, т. е. когда теоретическая плотность имеет вид

$$f(x;\theta) = \frac{1}{\theta} f\left(\frac{x}{\theta}\right),$$

$\theta > 0$ — неизвестный параметр масштаба, а f — известная функция. В этом случае для любой константы $c > 0$

$$\mathbf{P}_\theta\{cX_i \leqslant x\} = \mathbf{P}_\theta\left\{X_i \leqslant \frac{x}{c}\right\} = \int\limits_{-\infty}^{x/c} f(t;\theta)\,dt =$$

$$= \int\limits_{-\infty}^{x/c} \frac{1}{\theta} f\left(\frac{t}{\theta}\right) dt = \int\limits_{-\infty}^{x} \frac{1}{c\theta} f\left(\frac{u}{c\theta}\right) du = \mathbf{P}_{c\theta}\{X_i \leqslant x\},$$

т. е. изменение масштаба измерения данных в c раз эквивалентно в данном случае изменению параметра в c раз. Поэтому в задаче оценивания параметра θ для таких моделей разумно рассматривать в качестве оценок статистики, инвариантные относительно изменения масштаба, т. е. удовлетворяющие условию

$$T(cX) = T(cX_1, \ldots, cX_n) = cT(X). \tag{9}$$

Статистические оценки параметра масштаба θ, удовлетворяющие условию (9), также называются *эквивариантными оценками.*

Примерами таких статистик являются $\overline{X}$, S, $(X_{(1)} + X_{(2)})/2$. В задачах оценивания параметра масштаба естественной функцией потерь является

$$L(T,\theta) = \frac{1}{\theta^2}(T-\theta)^2 = \left(\frac{T}{\theta} - 1\right)^2 \tag{10}$$

Таким образом, если измерять качество оценок с помощью функции риска

$$R(T,\theta) = \left(\frac{1}{\theta^2}\right)\mathbf{E}_\theta(T-\theta)^2$$

которая представляет собой нормированную с. к. о., то оптимальная (т. е. минимизирующая эту меру точности) эквивариантная оценка параметра

Оценка Питмена параметра масштаба

масштаба θ, называемая также *оценкой Питмена*, существует и имеет вид

$$\theta_n^* = \frac{\int\limits_0^\infty \frac{1}{\theta^2} \prod\limits_{i=1}^n f(X_i;\theta)\,d\theta}{\int\limits_0^\infty \frac{1}{\theta^3} \prod\limits_{i=1}^n f(X_i;\theta)\,d\theta} = \frac{\int\limits_0^\infty \frac{1}{\theta^{n+2}} \prod\limits_{i=1}^n f\left(\frac{X_i}{\theta}\right)d\theta}{\int\limits_0^\infty \frac{1}{\theta^{n+3}} \prod\limits_{i=1}^n f\left(\frac{X_i}{\theta}\right)d\theta}. \qquad (11)$$

Пример 10 (Оценка Питмена для параметра масштаба равномерного распределения). Пусть $X \quad (X_1, \quad , X_n)$ — выборка из распределения $\mathcal{U}(0,\theta)$, $\theta > 0$, т. е.

$$f(x;\theta) = \frac{1}{\theta} I(0 \leqslant x \leqslant \theta) = \frac{1}{\theta} I\left(0 \leqslant \frac{x}{\theta} \leqslant 1\right)$$

Здесь $f(x) = I(0 \leqslant x \leqslant 1)$ — равномерная плотность на $[0, 1]$ и

$$\prod_{i=1}^n f\left(\frac{X_i}{\theta}\right) = I(X_{(n)} \leqslant \theta).$$

Поэтому по формуле (11) находим, что оценка Питмена для параметра θ есть

$$\theta_n^* = \int\limits_{X_{(n)}}^\infty \frac{d\theta}{\theta^{n+2}} \left(\int\limits_{X_{(n)}}^\infty \frac{d\theta}{\theta^{n+3}}\right) = \frac{n+2}{n+1} X_{(n)}. \qquad (12)$$

Мы знаем (см. пример 5), что наилучшей несмещенной оценкой θ относительно квадратичной функции потерь $L(T,\theta) = (T-\theta)^2$ является статистика $((n+1)/n)X_{(n)}$. Поэтому оценка Питмена (12) является смещенной, но она — наилучшая в классе эквивариантных оценок параметра θ относительно функции потерь (10). •

6. Байесовские и минимаксные оценки

В этом пункте мы познакомим читателя с основными элементами весьма популярного и в то же время неоднозначно воспринимаемого специалистами так называемого *байесовского подхода* (*метода*) в статистике. Оценки неизвестных параметров распределений, получаемые этим методом, носят название *байесовских оценок*. При байесовском подходе предполагается, что параметр θ изучаемой модели

Байесовский подход

$$\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$$

— это случайная величина с некоторым распределением на параметрическом множестве Θ, задаваемом плотностью $\pi(\theta)$. (В таких случаях будем использовать одно и то же обозначение θ как для случайной величины, так и для

принимаемого ею значения, полагая, что это не вызовет затруднений при чтении). Это распределение $\mathcal{L}(\theta)$ называют *априорным распределением параметра* и считают известным. Итак, можно представлять себе, что до начала наблюдений над случайной величиной ξ с $\mathcal{L}(\xi) \in \mathcal{F}$ производится дополнительный эксперимент, в котором разыгрывается (моделируется) значение случайного параметра с заданной плотностью $\pi(\theta)$, $\theta \in \Theta$, и затем уже с полученным значением θ моделируется выборка $\boldsymbol{X} = (X_1, \quad, X_n)$ из распределения $F(x; \theta)$.

Априорное распределение параметра

Если мы исходим при оценивании параметра из заданной функции потерь $L(T, \theta)$, то мы можем усреднить соответствующий риск $R(T, \theta) = \mathbf{E}_\theta L(T, \theta)$ оценки $T = T(\boldsymbol{X})$ по априорному распределению параметра и тем самым вычислить полную среднюю потерю (усредненный риск)

Байесовский риск

$$r(T) = \int R(T, \theta)\pi(\theta)\, d\theta, \tag{13}$$

которую мы несем при использовании статистики T в качестве оценки неизвестного параметра θ. Величина $r(T)$ в (13) называется *байесовским риском*. Таким образом, при таком подходе оценка характеризуется одним числом — байесовским риском и, следовательно, все оценки могут быть упорядочены в соответствии со значением этой характеристики. Оптимальной является оценка T^* минимизирующая байесовский риск $r(T)$:

Байесовская оценка параметра

$$T^* = \arg\min_T r(T); \tag{14}$$

T^* и называется *байесовской оценкой* для θ. Подчеркнем, что T^* зависит от выбора априорного распределения π, поэтому для различных априорных распределений параметра соответствующие байесовские оценки будут, вообще говоря, различны, — в этом состоит основной недостаток байесовского подхода.

Рассмотрим теперь вопрос о том, как практически находить байесовские оценки T^*. Основой для этого служит *теорема Байеса*, согласно которой *апостериорное распределение параметра* при наблюдении $\boldsymbol{X} = \underline{x}$ задается условной плотностью

Апостериарное распределение

$$\pi(\theta|\underline{x}) = \frac{f(\underline{x}; \theta)\pi(\theta)}{f(\underline{x})}, \tag{15}$$

где

$$f(\underline{x}) = \mathbf{E}f(\underline{x}; \theta) = \int f(\underline{x}; \theta)\pi(\theta)\, d\theta, \quad \text{или} \quad f(\underline{x}) = \sum_i f(\underline{x}; \theta_i)\pi(\theta_i),$$

если θ принимает дискретные значения $\{\theta_i\}$. Тогда байесовская оценка T^* — это оценка, минимизирующая *апостериорный риск*

$$\mathbf{E}\big(L(T, \theta)|\boldsymbol{X} = \underline{x}\big) = \int L(T, \theta)\pi(\theta|\underline{x})\, d\theta. \tag{16}$$

Действительно, если $\mathbf{E}\big(L(T^*,\theta)|X=\underline{x}\big) \leqslant \mathbf{E}\big(L(T,\theta)|X=\underline{x}\big)$, то по формуле полного математического ожидания

$$r(T^*)=\mathbf{E}L(T^*,\theta)=\mathbf{E}\big[\mathbf{E}\big(L(T^*,\theta)|X\big)\big] \leqslant \mathbf{E}\big[\mathbf{E}\big(L(T,\theta)|X\big)\big]=\mathbf{E}L(T,\theta)=r(T).$$

Апостериорное среднее параметра

Для случая квадратичной функции потерь

$$L(T,\theta)=(T-\theta)^2$$

байесовская оценка T^* вычисляется особенно просто:

$$T^*=\mathbf{E}(\theta|X)=\int \theta\pi(\theta|X)\,d\theta, \tag{17}$$

т. е. она совпадает с *апостериорным средним параметра*.

Это легко следует из цепочки соотношений

$$\mathbf{E}\left[(\theta-T(X))^2|X\right]=\mathbf{E}\left[\big(\theta-T^*(X)+T^*(X)-T(X)\big)^2|X\right]=$$
$$=\mathbf{D}(\theta|X)+\big(T^*(X)-T(X)\big)^2 \geqslant \mathbf{D}(\theta|X);$$

минимум здесь достигается при $T=T^*$, и он равен апостериорной дисперсии параметра $\mathbf{D}(\theta|X)$. Отсюда также следует, что $r(T^*)=\mathbf{E}\mathbf{D}(\theta|X)$.

Сопряженные априорные распределения

Важную роль в теории байесовских оценок играют *сопряженные априорные распределения* $\mathcal{L}(\theta)$, т. е. когда апостериорное распределение параметра $\mathcal{L}(\theta|X=\underline{x})$ принадлежит тому же семейству распределений, что и $\mathcal{L}(\theta)$. Примеры сопряженных семейств приведены в упр. 64 к гл. 1 (см. также замечание к упр. 63 гл. 1). Для сопряженных априорных распределений задача вычисления байесовской оценки (17) значительно упрощается.

Пример 11 (Байесовская оценка вероятности успеха в схеме Бернулли). Пусть $X=(X_1,\ \ ,X_n)$ — выборка из распределения Бернулли $\mathrm{Bi}(1,\theta)$ и $X=X_1+\ldots+X_n$ — наблюдавшееся, число «успехов». Тогда $\mathcal{L}_\theta(X)=\mathrm{Bi}(n,\theta)$ и, как мы знаем, оптимальной несмещенной оценкой θ при квадратичной функции потерь является статистика $T=X/n=\overline{X}$ (см. табл. 2 § 3.2). Предположим теперь, что параметр θ — случайная величина с априорным распределением типа бета-распределения $\mathrm{Be}(a,b)$ (см. п. 4 § 1.2), при этом параметры a и b считаются известными. Согласно упр. 64 к гл. 1, бета-распределение $\mathrm{Be}(a,b)$ является сопряженным к семейству бернуллиевских распределений $\mathrm{Bi}(1,\theta)$, и при этом для апостериорного распределения параметра θ справедливо соотношение

$$\mathcal{L}(\theta|X=\underline{x})=\mathrm{Be}(a+x,b+n-x),$$

где $x=x_1+\ldots+x_n$. Отсюда по формуле (24) § 1.2 имеем, что апостериорное среднее (17), а тем самым и байесовская оценка параметра θ, есть в данном случае

$$T^*=\frac{a+X}{a+b+n}. \tag{18}$$

•

Если априорная информация об оцениваемом параметре θ отсутствует или байесовский подход вообще неприменим, для построения оценки можно использовать так называемый *минимаксный подход*, при котором за меру точности оценки T принимается максимальное значение ее риска, или *максимальный риск*,

$$m(T) = \max_{\theta} R(T, \theta).$$

Минимаксная оценка

Оценка, минимизирующая максимальный риск, называется *минимаксной оценкой* θ и обозначается $\widetilde{T}$:

$$\widetilde{T} = \arg\min_{T}(T). \tag{19}$$

В общем случае вопрос об отыскании такой оценки весьма трудный, но в ряде случаев она может быть найдена так: предположим, что существует априорное распределение $\pi(\theta) > 0$, для которого соответствующая байесовская оценка T_π^* имеет постоянный риск: $R(T_\pi^*, \theta) = a = \text{const}$. Тогда $\widetilde{T} = T_\pi^*$.

Действительно, предположим, что существует оценка T_1, для которой $m(T_1) < m(T_\pi^*)$. Очевидно, что для любой оценки T и любого априорного распределения π параметра θ справедливо неравенство $r(T) \leqslant m(T)$ (см. (13)), а для оценки T_π^* здесь имеет место равенство. Отсюда имеем

$$r(T_1) \leqslant m(T_1) < m(T_\pi^*) = r(T_\pi^*),$$

что противоречит байесовости оценки T_π^*. Следовательно, $m(T_\pi^*) \leqslant m(T)$ для любой оценки T, и потому T_π^* — минимаксная оценка.

Наименее благоприятное априорное распределение

Добавим к этому, что такое распределение $\pi(\theta)$ называется *наименее благоприятным априорным распределение* параметра, и оно часто находится в сопряженном семействе.

Замечание. Этот термин отражает следующую особенность такого априорного распределения. Пусть π' — любое другое априорное распределение и $T_{\pi'}^*$ — соответствующая ему байесовская оценка. Тогда

$$r(T_{\pi'}^*) = \int R(T_{\pi'}^*, \theta)\pi'(\theta)\,d\theta \leqslant \int R(T_\pi^*, \theta)\pi'(\theta)\,d\theta = a = r(T_\pi^*),$$

т. е. π — такое априорное распределение, которое причиняет статистику *наибольшие неизбежные средние потери*.

Пример 12 (Минимаксная оценка вероятности успеха в схеме Бернулли). В ситуации примера 11 вычислим функцию риска, т. е. с. к. о. байесовской оценки (18). Поскольку эта оценка смещенная (несмещенной является оценка X/n), то, согласно формуле (5) § 3.1, ее с. к. о. равна сумме квадрата смещения

$$b(\theta) = \mathbf{E}_\theta T^* - \theta = \frac{a + n\theta}{a + b + n} - \theta = \frac{a - \theta(a + b)}{a + b + n}$$

и дисперсии

$$\mathbf{D}_\theta T^* = \frac{n\theta(1 - \theta)}{(a + b + n)^2},$$

т. е.

$$R(T^*, \theta) = \mathbf{E}_\theta (T^* - \theta)^2 = \frac{[a - \theta(a+b)]^2 + n\theta(1-\theta)}{(a+b+n)^2}. \tag{20}$$

Чтобы получить минимаксную оценку, надо найти такое априорное распределение $\mathrm{Be}(a, b)$, для которого эта функция риска не зависит от θ. Из (20) легко находим, что это имеет место при $(a+b)^2 = n$ и $2a(a+b) = n$, т. е. при $a = b = \sqrt{n}/2$. Следовательно, минимаксная оценка θ в данной задаче существует и имеет вид

$$\widetilde{T} = \frac{X + \sqrt{n}/2}{n + \sqrt{n}}, \tag{21}$$

а ее риск равен

$$R(\widetilde{T}, \theta) = m(\widetilde{T}) = \frac{a^2}{(a+b+n)^2} = \frac{1}{4n(1 + 1/\sqrt{n})^2}. \tag{22}$$

Для сравнения рассмотрим риск оптимальной несмещенной оценки $T = X/n$. Имеем

$$R(T, \theta) = \mathbf{D}_\theta T = \frac{\theta(1-\theta)}{n},$$

а

$$m(T) = \max_\theta \frac{\theta(1-\theta)}{n} = \frac{1}{4n} > m(\widetilde{T}).$$

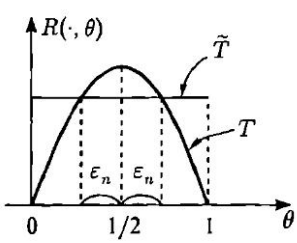


Рис. 3

График функций риска оценок T и $\widetilde{T}$ приведены на рис. 3. Из рисунка видно, что зона предпочтения минимаксной оценки $\widetilde{T}$ перед оценкой T имеет вид

$$\{\theta \quad R(\widetilde{T}, \theta) < R(T, \theta)\} = \left\{\theta \in \left(\frac{1}{2} \mp \varepsilon_n\right)\right\},$$

где

$$\varepsilon_n = 1/2\sqrt{1 - \left(1 + \frac{1}{\sqrt{n}}\right)^{-2}}$$

Если n велико $(n \to \infty)$, то $\varepsilon_n \to 0$, и эта зона предпочтения сужается к точке $\theta = 1/2$. Таким образом, при большом числе наблюдений n оценка $\widetilde{T}$ точнее оценки T лишь для незначительной зоны значений параметра θ. Для большинства же значений параметра θ более точной является несмещенная оценка $T = X/n$. •

Изложенная выше теория байесовского и минимаксного оценивания относится к случаю скалярного параметра, но все ее основные положения переносятся и на случай векторного параметра $\theta = (\theta_1, \ldots, \theta_r)$. Если для оценки

$\underline{T}$ $(T_1, \quad , T_r)$ параметра θ ее точность измеряется с помощью квадратичной функции потерь

Квадратичная функция потерь

$$L(\underline{T}, \theta) = |\underline{T} - \theta|^2 = \sum_{i=1}^{r} (T_i - \theta_i)^2,$$

т. е. квадратичным риском $R(\underline{T}, \theta) \quad \mathbf{E}_\theta |\underline{T} - \theta|^2$, то байесовская оценка $\underline{T}^* = \arg\min_{\underline{T}} r(\underline{T})$, где байесовский риск $r(\underline{T}) = \mathbf{E}R(\underline{T}, \theta)$, есть вектор апостериорных средних $\underline{T}^* = \mathbf{E}(\theta | \boldsymbol{X})$, а минимаксная оценка $\widetilde{\underline{T}} = \arg\min_{\underline{T}} m(\underline{T})$, где $m(\underline{T}) = \max_{\theta} R(\underline{T}, \theta)$, есть такая байесовская оценка $\underline{T}^*$, для которой риск постоянен: $R(\underline{T}^*, \theta) \equiv \text{const}$.

Пример 13 (Байесовская и минимаисная оценки параметров полиномиального распределения). Пусть наблюдается целочисленный случайный вектор $\underline{\nu} = (\nu_1, \quad , \nu_N)$, имеющий полиномиальное распределение $M(n; \underline{p} = (p_1, \quad , p_N))$ (см. п. 6 § 1.1), параметры $\underline{p}$ которого неизвестны, т. е. здесь

$$\underline{p} = \theta \in \Theta = \{\theta = (\theta_1, \quad , \theta_N) \quad \theta_i \geqslant 0, \ i = 1, \ldots , N, \ \theta_1 + \quad + \theta_N = 1\}.$$

Предположим, что вектор $\underline{p}$ случаен и его априорное распределение есть распределение Дирихле $\mathbf{D}(\underline{\alpha})$ с известными параметрами $\underline{\alpha} = (\alpha_1, \ldots , \alpha_N)$ (см. п. 10 § 1.2). Распределение Дирихле сопряжено к полиномиальному распределению и при этом для апостериорного распределения параметра $\underline{p}$ справедливо соотношение $\mathcal{L}(\underline{p} | \underline{\nu} = \underline{x}) = \mathbf{D}(\underline{\alpha} + \underline{x})$ (см. упр. 63 к гл. 1). Отсюда с учетом формул (41) § 1.2 имеем, что байесовская оценка параметра $\underline{p}$ есть вектор

$$\underline{T}^* = \mathbf{E}(\underline{p} | \underline{\nu}) = \frac{\underline{\alpha} + \underline{\nu}}{\alpha + n} \tag{23}$$

(здесь $\alpha = \alpha_1 + \quad + \alpha_N$, $n = \nu_1 + \quad + \nu_N$). Напомним, что наилучшей несмещенной оценкой для $\underline{p}$ является статистика $\underline{p}^* = \underline{\nu}/n$ (см., например, пример 10 в § 3.2).

Вычислим функцию риска оценки (23). Для этого удобно переписать представление (23) в виде

$$\underline{T}^* = w\underline{p}^* + (1 - w)\underline{\lambda}, \quad \text{где} \quad \underline{\lambda} = \frac{\underline{\alpha}}{\alpha}, \quad w = \frac{n}{\alpha + n}, \tag{24}$$

показывающем, что точка $\underline{T}^* \in \Theta$ лежит на отрезке, соединяющем точки $\underline{p}^*$ и $\underline{\lambda}$. Из (24) имеем

$$R(\underline{T}^*, \theta) = \mathbf{E}_\theta |\underline{T}^* - \theta|^2 = \mathbf{E}_\theta \left| w(\underline{p}^* - \theta) + (1 - w)(\underline{\lambda} - \theta) \right|^2 =$$

$$= w^2 \mathbf{E}_\theta |\underline{p}^* - \theta|^2 + (1 - w)^2 |\underline{\lambda} - \theta|^2 = \frac{w^2}{n}(1 - |\theta|^2) + (1 - w)^2 |\underline{\lambda} - \theta|^2$$

Если здесь $\underline{\lambda} = \underline{e} = (1/N, \ldots, 1/N)$, то получим

$$R(\underline{T}^*, \theta) = \frac{\alpha^2 - n}{(\alpha + n)^2}|\theta|^2 + \frac{nN - \alpha^2}{N(\alpha + n)^2}.$$

Отсюда следует, что при $\alpha = \sqrt{n}$ риск $R(\underline{T}^*, \theta)$ будет постоянным и равным

$$\frac{N - 1}{[N(\sqrt{n} + 1)]^2}.$$

Но это означает, что $\mathbf{D}(\sqrt{n}\underline{e})$ — наименее благоприятное априорное распределение, а минимаксной оценкой является статистика

$$\widetilde{\underline{T}} = \frac{\sqrt{n}}{\sqrt{n} + 1}\,\underline{p}^* + \frac{1}{\sqrt{n} + 1}\,\underline{e} = \left(\frac{\nu_j + \sqrt{n}/N}{n + \sqrt{n}}, \quad j = 1, \quad , N\right). \qquad (25)$$

•

В предыдущих примерах 12 и 13 удалось найти наименее благоприятное априорное распределение параметра модели и тем самым легко построить минимаксную оценку. Но такое распределение не всегда существует, и в таких случаях для отыскания минимаксной оценки может быть полезным следующий критерий.

Лемма. Пусть $\{\pi_k\}$ — некоторая последовательность априорных распределений параметра и $\{T_k^*\}$ — последовательность соответствующих байесовских оценок. Если существует оценка T_0, для которой функция риска удовлетворяет условию

$$R(T_0, \theta) \leqslant \lim_{k \to \infty} r(T_k^*), \quad \forall\, \theta,$$

то T_0 есть минимаксная оценка: $T_0 = \widetilde{T}$

Доказательство. Для любой оценки T выполняется цепочка соотношений

$$m(T) = \sup_{\theta} R(T, \theta) \geqslant \int R(T, \theta)\pi_k(\theta)\, d\theta \geqslant r(T_k^*).$$

Отсюда следует, что

$$m(T) \geqslant \lim_{k \to \infty} r(T_k^*) \geqslant R(T_0, \theta), \quad \forall\, \theta, \quad \text{т.е.} \quad m(T) \geqslant m(T_0). \ \blacksquare$$

| **Замечание.** Очевидно, что в этих распределениях lim может быть заменен на lim sup.

Пример 14 (Минимаксность выборочного среднего для нормальной модели). Пусть $X = (X_1, \quad , X_n)$ — выборка из распределения $\mathcal{N}(\theta, b^2)$ (b^2 известно). Покажем, что статистика $\overline{X}$, являющаяся оптимальной несмещенной оценкой θ относительно квадратичной функции потерь $L(T, \theta) = (T - \theta)^2$ (см. табл. 2 § 3.2), есть также и минимаксная оценка для θ. Для этого мы сначала найдем байесовскую оценку T^* для θ относительно априорного распределения $\mathcal{L}(\theta) = \mathcal{N}(\mu, \sigma^2)$, являющегося сопряженным к семейству $\mathcal{N}(\theta, b^2)$

(см. упр. 64(5) к гл. 1). Используя результат этого упражнения, сразу находим

$$T^* = \mathbf{E}(\theta|\boldsymbol{X}) = \sigma_1^2\left(\frac{\mu}{\sigma^2} + \frac{n}{b^2}\overline{X}\right), \quad \sigma_1^2 = \left(\frac{1}{\sigma^2} + \frac{n}{b^2}\right) \tag{26}$$

при этом байесовский риск

$$r(T^*) = \mathbf{ED}(\theta|\boldsymbol{X}) = \sigma_1^2.$$

Можно проверить (это мы оставляем читателю в качестве упражнения), что функция риска оценки T^* имеет вид

$$R(T^*, \theta) = \mathbf{E}_\theta(T^* - \theta)^2 = \frac{n\sigma_1^4}{b^2} + \left[\left(\frac{n\sigma_1^2}{b^2} - 1\right)\theta + \frac{\mu\sigma_1^2}{\sigma^2}\right]^2 \tag{27}$$

и здесь нет априорного распределения (в сопряженном семействе), для которого $R(T^*, \theta) \equiv \text{const}$. Поэтому применим критерий минимаксности, сформулированный в лемме. Для этого рассмотрим последовательность априорных распределений $\{\mathcal{N}(\mu, \sigma_k^2)\}$, где $\sigma_k^2 \to \infty$ при $k \to \infty$. Соответствующие байесовские оценки в силу (26) есть $T_k^* = \sigma_{1k}^2(\mu\sigma_k^{-2} + nb^{-2}\overline{X})$ и их байесовские риски удовлетворяют соотношению

$$r(T_k^*) = \sigma_{1k}^2 = \left(\frac{1}{\sigma_k^2} + \frac{n}{b^2}\right) \xrightarrow[k\to\infty]{} \frac{b^2}{n}.$$

Подчеркнем также, что $T_k^* \to \overline{X}$ при $k \to \infty$. Но для оценки $T_0 = \overline{X}$ функция риска есть

$$R(\overline{X}, \theta) = \mathbf{E}_\theta(\overline{X} - \theta)^2 = \mathbf{D}_\theta\overline{X} = \frac{b^2}{n} = \lim_{k\to\infty} r(T_k^*),$$

следовательно $\overline{X} = \widetilde{T}$ — минимаксная оценка θ, и она является пределом при $k \to \infty$ байесовских оценок T_k^*. •

Замечание. Таким образом, выборочное среднее $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ является «образцовой» оценкой неизвестного теоретического среднего θ в нормальной модели $\mathcal{N}(\theta, b^2)$: оно является несмещенной оценкой с равномерно минимальной дисперсией, а также минимаксной оценкой при квадратичной функции потерь. Естественно поставить вопрос, сохраняется ли свойство оптимальности выборочного среднего в многомерном случае. Как показал К. Стейн (1956), ответ на этот вопрос отрицателен. Точнее, рассмотрим k-мерную нормальную модель $\mathcal{N}\big(\theta = (\theta_1, \ldots, \theta_k), b^2\mathbb{1}_k\big)$ (т. е. (см. п. 2 § 1.2) наблюдается k-мерный вектор $\underline{\xi} = (\xi_1, \quad, \xi_k)$ с независимыми (для простоты) нормальными $\mathcal{N}(\theta_j, b^2)$, $j = 1, \quad, k$, компонентами) и предположим, что по соответствующей выборке $X_l = (X_{l1}, \ldots, X_{lk})$, $l = 1, \ldots, n$, оценивается неизвестный вектор средних $\theta = (\theta_1, \quad, \theta_k)$ при квадратичной функции потерь

$$L(\underline{T}, \theta) = |\underline{T} - \theta|^2 = \sum_{j=1}^{k}(T_j - \theta_j)^2$$

Тогда (см. п. 2 § 2.3) вектор выборочных средних $\overline{X} = (\overline{X}_1, \ldots, \overline{X}_k)$, где

$$\overline{X}_j = \frac{1}{n}\sum_{l=1}^{n} X_{lj}, \quad j = 1, \ldots, k,$$

будет несмещенной оценкой для вектора θ и ее функция риска равна

$$R(\overline{X}, \theta) = \mathbf{E}_\theta |\overline{X} - \theta|^2 = \sum_{j=1}^{k} \mathbf{D}_\theta \overline{X}_j = \frac{k}{n} b^2,$$

т. е. не зависит от θ. Постоянность функции риска подсказывает, что рассматриваемая оценка является минимаксной при любой размерности k. Но, оказывается, что при $k > 2$ она может быть строго улучшена. Именно, в 1961 г. Джеймс и Стейн предложили следующую простую оценку для математического ожидания многомерного нормального распределения (при $k \geqslant 3$)

$$\theta^* = \left(1 - \frac{(k-2)b^2}{|\overline{X}|^2}\right)\overline{X}$$

и показали, что ее риск $R(\theta^*, \theta) < R(\overline{X}, \theta)$, т. е. оценка θ^* равномерно лучше, чем $\overline{X}$ (говорят также, что оценка $\overline{X}$ является *недопустимой*, см. п. 2 § 7.1). Открытие Стейна, явившееся сюрпризом, показывает, что даже тогда, когда рассматривается классическая проблема оценивания (т. е. оценивание среднего значения нормального распределения), выборочное среднее — это не единственная оценка, которую следует приниимать во внимание.

7. Оценивание по цензурированным данным

До сих пор рассматривались методы оценивания, использующие информацию, доставляемую полной выборкой $X = (X_1, \ldots, X_n)$. Однако, на практике возникают задачи оценивания и по неполной выборке, т. е. когда некоторые наблюдения либо отсутствуют, либо о них имеется лишь частичная информация, — в таких случаях говорят о *цензурированных* данных. Типичными примерами цензурирования являются следующие планы испытаний на надежность: берется контрольная партия из n однотипных изделий, «времена жизни» которых — независимые одинаково распределенные случайные величины, и наблюдаются либо моменты отказов за заданное время t (I тип цензурирования), либо моменты первых $r (< n)$ отказов (II тип цензурирования). В обоих случаях достигается экономия времени на проведение эксперимента (получение исходных данных), что бывает важным фактором в реальных условиях. В общем случае I тип цензурирования определяется заданием такого интервала (t_1, t_2), что наблюдаются лишь значения $X_i \in (t_1, t_2)$, а II тип — заданием двух целых чисел $r_1, r_2 \geqslant 0$, таких что наблюдаются лишь значения порядковых статистик $X_{(k)}$ при $r_1 + 1 \leqslant k \leqslant n - r_2$ (если соответствующее ограничение с какой-нибудь одной стороны отсутствует, то говорят о *простом* цензурировании).

Рассмотренные выше методы могут быть применены и к цензурированным данным. Проиллюстрируем это на примере оценивания параметров $\theta = (\theta_1, \theta_2)$ нормальной модели $\mathcal{N}(\theta_1, \theta_2^2)$ при простом цензурировании.

Пример 15 (I тип цензурирования). Пусть из n наблюдений фиксируются лишь те, которые принадлежат интервалу (t, ∞). Пусть таких наблюдений оказалось N, а их самих обозначим $Y = (Y_1, \ldots, Y_N)$ (таким образом, $n - N \geqslant 0$ наблюдений потеряно). Тогда функция правдоподобия данных, очевидно, имеет вид (далее $u = (t - \theta_1)/\theta_2$)

$$L(\underline{y}; \theta) = C_n^N \Phi^{n-N}(u)\left(\sqrt{2\pi}\theta_2\right)^{-N} \exp\left\{-\frac{1}{2\theta_2^2}\sum_{i=1}^{N}(y_i - \theta_1)^2\right\} =$$

$$= C_n^N \Phi^{n-N}(u)\left(\sqrt{2\pi}\theta_2\right)^{-N} \exp\left\{-\frac{N}{2\theta_2^2}s^2 - \frac{N}{2\theta_2^2}(\bar{y} - \theta_1)^2\right\},$$

где

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2$$

Отсюда следует, что в данном случае достаточной статистикой является тройка $(N, \overline{Y}, S^2(Y))$.

Будем искать оценки параметров $\theta = (\theta_1, \theta_2)$ по методу максимального правдоподобия. Для этого запишем уравнения правдоподобия

$$\frac{\partial \ln L(\underline{y}; \theta)}{\partial \theta_1} = \frac{N}{\theta_2^2}(\bar{y} - \theta_1) - \frac{n - N}{\theta_2}\frac{\Phi'(u)}{\Phi(u)} = 0,$$

$$\frac{\partial \ln L(\underline{y}; \theta)}{\partial \theta_2} = \frac{N}{\theta_2^3}\left(s^2 + (\bar{y} - \theta_1)^2\right) - \frac{N}{\theta_2} - \frac{(n - N)u}{\theta_2}\frac{\Phi'(u)}{\Phi(u)} = 0.$$

Если обозначить $A(u) = (u - N)\Phi'(u)/(N\Phi(u))$, то эти уравнения приводятся к виду

$$\bar{y} - \theta_1 = \theta_2 A(u), \quad s^2 + (\bar{y} - \theta_1)^2 = \theta_2^2(1 + uA(u)). \tag{28}$$

Учитывая, что $\theta_2 u = t - \theta_1 = t - \bar{y} + \theta_2 A(u)$, т. е. $\theta_2 = (\bar{y} - t)/(A(u) - u)$, из первого уравнения (28) имеем

$$\theta_1 = \bar{y} - \lambda(\bar{y} - t), \quad \lambda = \frac{A(u)}{A(u) - u}. \tag{29}$$

Далее, так как

$$(\bar{y} - \theta_1)^2 - \theta_2^2 uA(u) = \theta_2^2 A(u)(A(u) - u) = \lambda(\bar{y} - t)^2,$$

из второго уравнения (28) следует, что

$$\theta_2^2 = s^2 + \lambda(\bar{y} - t)^2 \tag{30}$$

Но также $\theta_2^2 = (\bar{y} - t)^2/(A(u) - u)^2$, поэтому

$$s^2 = (\bar{y} - t)^2\left[\frac{1}{(A(u) - u)^2} - \frac{A(u)}{A(u) - u}\right] = (\bar{y} - t)^2\frac{1 - A(u)(A(u) - u)}{(A(u) - u)^2}.$$

Отсюда

$$\gamma \equiv \frac{s^2}{(\bar{y}-t)^2} = \frac{1 - A(u)(A(u)-u)}{(A(u)-u)^2}. \tag{31}$$

Теперь, вычислив по данным величину γ, из уравнения (31) находим значение u, затем $\lambda = \lambda(u)$ в (29) и, наконец, из (29) и (30) значения θ_1 и θ_2. Тем самым однозначно определяются о. м. п. $\widehat{\theta}_1, \widehat{\theta}_2$.

Отметим, что если $N = n$, т. е. наблюдается полная выборка, то $A(u) = 0$, следовательно, $\lambda = 0$, и мы приходим к обычным о. м. п. $\widehat{\theta}_1 = \overline{Y}$, $\widehat{\theta}_2 = S(Y)$ (см. пример 1 в § 3.5). •

Пример 16 (II тип цензурирования). Пусть наблюдению доступны лишь порядковые статистики $X_{(k)}$ при $k > r$ для заданного $r \geqslant 1$. Обозначим $Y = (Y_1, \ldots, Y_N) = (X_{(r+1)}, \ldots, X_{(n)})$, $N = n - r$, тогда из п. 2 § 2.4 имеем, что плотность распределения вектора Y (функция правдоподобия данных) есть

$$L(\underline{y}; \theta) = \frac{n!}{r!}\Phi^r\left(\frac{y_1 - \theta_1}{\theta_2}\right)\left(\sqrt{2\pi}\theta_2\right)^{-N} \exp\left\{-\frac{1}{2\theta_2^2}\sum_{i=1}^{N}(y_i - \theta_1)^2\right\}$$

Отсюда следует, что достаточной статистикой является в данном случае тройка $(X_{(r+1)}, \overline{Y}, S^2(Y))$, а метод максимального правдоподобия приводит к тем же уравнениям (28) с заменой t на $y_1 = x_{(r+1)}$ и u — на $u' = (y_1 - \theta_1)/\theta_2$. •

Рассмотрим еще один практически важный пример.

Пример 17 (Оценивание параметра масштаба показательного распределения по цензурированным данным). Пусть для выборки $X = (X_1, \ldots, X_n)$ из показательного распределения $\mathcal{L}(\xi) \in \Gamma(\theta, 1)$ (см. п. 3 § 1.2) доступны наблюдению лишь первые r порядковые статистики $X_{(1)}, \ldots, X_{(r)}$. Рассмотрим класс линейных оценок

$$T_r = \sum_{k=1}^{r} \lambda X_{(k)}$$

для параметра θ и найдем в нем оптимальную несмещенную оценку (т. е. н. о. р. м. д.). Для этого прежде всего заметим, что $\mathcal{L}_\theta(\xi/\theta) = \Gamma(1, 1)$ — стандартное показательное распределение, поэтому, воспользовавшись результатами примера 1 § 2.4, запишем представление

$$X_{(k)} = \theta \sum_{i=1}^{k} \frac{Y_i}{n - i + 1},$$

где $\{Y_i\}$ — независимые случайные величины с распределением $\Gamma(1, 1)$.

Статистика T_r при этом примет вид

$$T_r = \theta \sum_{k=1}^{r} \lambda_k \sum_{i=1}^{k} \frac{Y_i}{n - i + 1} = \theta \sum_{i=1}^{r} \frac{Y_i \Lambda_i}{n - i + 1},$$

где $\Lambda_i = \sum_{k=i}^{r} \lambda_k$, $i = 1, \quad , r$. Отсюда, учитывая, что $\mathbf{E}Y_i = \mathbf{D}Y_i = 1$, имеем

$$\mathbf{E}_\theta T_r = \theta \sum_{i=1}^{r} \frac{\Lambda_i}{n-i+1}, \quad \mathbf{D}_\theta T_r = \theta^2 \sum_{i=1}^{r} \frac{\Lambda_i^2}{(n-i+1)^2}.$$

Условие несмещенности теперь принимает вид

$$\sum_{i=1}^{r} \frac{\Lambda_i}{n-i+1} = 1,$$

и при этом условии надо минимизировать по $\Lambda_1, \ldots, \Lambda_r$ сумму

$$F(\Lambda_1, \ldots, \Lambda_r) = \sum_{i=1}^{r} \frac{\Lambda_i^2}{(n-i+1)^2}.$$

Так как сумма $\sum_{i=1}^{r} x_i^2$ при условии $\sum_{i=1}^{r} x_i = 1$ обращается в минимум при $x_1 = \quad = x_r = 1/r$ (показать!), то в нашем случае оптимальный выбор Λ_i таков: $\Lambda_i^* = (n-i+1)/r$, $i = 1, \quad , r$ Тем самым оптимальная оценка существует и имеет вид (далее учитывается, что $Y_i = (n-i+1)(X_{(i)} - X_{(i-1)})/\theta$, $i = 1, \quad , r$, $X_{(0)} = 0$)

$$T_r^* = \frac{1}{r} \sum_{i=1}^{r} (n-i+1)(X_{(i)} - X_{(i-1)}) = \frac{1}{r}(X_{(1)} + \ldots + X_{(r)}) + \frac{n-r}{r} X_{(r)}$$

и

$$\mathbf{D}_\theta T_r^* = \frac{\theta^2}{r}.$$

Подчеркнем, что при $r = n$ (наблюдается полная выборка) $T_n^* = \overline{X}$, т. е. мы приходим к эффективной оценке для θ (см. табл. 2 § 3.2). •

Дальнейший материал — для тех, кто хочет расширить и углубить свои представления об обсуждаемой проблематике. Мы дадим общие определения современных моделей неполных данных и приведем основные результаты о соответствующих оценках неизвестной функции распределения наблюдений.

Модели неполных данных

Итак, пусть $\{X_i\}$ — независимые одинаково распределенные случайные величины (н. о. р. с. в.) с общей функцией распределения $F(t) = \mathbf{P}\{X_1 \leqslant t\}$, которая неизвестна и должна быть оценена по соответствующей выборке с неполными данными. Опишем сначала наиболее распространенные модели неполных данных.

Модель 1 (цензурированные данные). Задана последовательность н. о. р. с. в. $\{T_i\}$, независимая от $\{X_i\}$, и предполагается, что исходные данные (выборка)

имеют вид $(\widetilde{X}_i, \delta_i)$, $i = 1, \ldots, n$, где

$$\widetilde{X}_i = X_i \wedge T_i \quad (\wedge \equiv \min), \quad \delta_i = I(X_i \leqslant T_i). \tag{32}$$

Таким образом, пара $(\widetilde{X}_i, 1)$ означает, что наблюдается $\widetilde{X}_i = X_i$, если же дано $(\widetilde{X}_i, 0)$, то это означает, что наблюдается $\widetilde{X}_i = T_i$, а об X_i известно лишь, что $X_i > T_i$. Можно также сказать, что в этой модели о результатах некоторых наблюдений известна лишь частичная информация типа, что $X_i > T_i$ (о такой модели иногда говорят как о *правоцензурированных* данных).

С. в. $\{T_i\}$ называются *цензурирующими* переменными. •

Модель 2 (усеченные данные). Здесь задается последовательность н. о. р. с. в. $\{t_i\}$, не зависящая от $\{X_i\}$, и предполагается, что X_i наблюдается лишь в случае, когда $X_i \geqslant t_i$. Таким образом, в данной модели выборка представляет собой n пар (X_i°, t_i°) с $X_i^\circ \geqslant t_i^\circ$, i 1, , n. Эта выборка порождается большей выборкой (X_i, t_i), $i = 1,$, $m(n)$, где

$$m(n) = \min\left\{m \ \sum_{i=1}^{m} I(X_i \geqslant t_i) = n\right\} \tag{33}$$

Здесь речь идет о *левоусеченных* данных (случай усечения справа: X_i наблюдается лишь в случае $X_i \leqslant t_i$, сводится к предыдущему умножением X_i и t_i на -1). Таким образом, здесь о наблюдениях с $X_i < t_i$ вообще нет информации. С. в. $\{t_i\}$ называются *усекающими* переменными. •

Модель 3 (смешанная). В этом случае данные подвергаются как цензурированию, так и усечению. Имеются последовательности $\{T_i\}$ (цензурирующие переменные) и $\{t_i\}$ (усекающие переменные), не зависящие от $\{X_i\}$ и между собой и состоящие из независимых с. в., которые с положительной вероятностью могут принимать значения $\pm\infty$ (расширенные с. в.). Здесь исходные данные (выборка) имеют вид $(\widetilde{X}_i^\circ, \delta_i^\circ, t_i^\circ)$ с $\widetilde{X}_i^\circ \geqslant t_i$, $i = 1, \ldots n$, которые порождаются большей выборкой (X_i, T_i, t_i), $i = 1,$, $m(n)$, где

$$m(n) = \min\left\{m \ \sum_{i=1}^{m} I(X_i \wedge T_i \geqslant t_i) = n\right\} \tag{34}$$

Если здесь положить все T_i ∞, то получим усеченную модель, а при $t_i \equiv -\infty$ — цензурированную. •

Оценивание функции распределения

Оценки теоретической функции распределения $F(t)$ в таких моделях данных носят название *product–limit* (*p.-l.*) оценок и их конструкции имеют следующий вид.

Оценка Каплана—Мейера

Для модели 1 *p.-l.* оценка $\widetilde{F}_n$ (Каплан и Мейер, 1958) определяется равенством

$$1 - \widetilde{F}_n(t) = \prod_{s \leqslant t}\left\{1 - \frac{\Delta N_n(s)}{Y_n(s)}\right\}, \tag{35}$$

где $0/0 = 0$, $\Delta N_n(s)$ $N_n(s) - N_n(s-0)$ — величина скачка в точке s (так что в (35) произведение берется по точкам скачков функции $N_n(s)$),

$$N_n(s) = \sum_{i=1}^{n} I(X_i \leqslant s \wedge T_i) = \sum_{i=1}^{n} I(\widetilde{X}_i \leqslant s,\ \delta_i = 1),$$

$$Y_n(s) = \sum_{i=1}^{n} I(X_i \wedge T_i \geqslant s) = \sum_{i=1}^{n} I(\widetilde{X}_i \geqslant s).$$

Асимптотические (при $n \to \infty$) свойства оценки $\widetilde{F}_n(t)$ аналогичны свойствам классической эмпирической функции распределения (см. п. 2 и 3 § 2.1), например, имеет место аналог теоремы Гливенко

$$\mathbf{P}\left\{\lim_{n\to\infty} \sup_t \left|\widetilde{F}_n(t) - F(t)\right| = 0\right\} = 1.$$

Для модели 2 *р.-l.* оценка F_n^* (Линден-Бэлл, 1971) определяется равенством

Оценка Линден-Бэлла

$$1 - F_n^*(t) = \prod_{s \leqslant t}\left\{1 - \frac{\Delta L_n(s)}{R_n(s)}\right\}, \tag{36}$$

где

$$L_n(s) = \sum_{i=1}^{n} I(X_i^\circ \leqslant s), \quad R_n(s) = \sum_{i=1}^{n} I(X_i^\circ \geqslant s \geqslant t_i^\circ).$$

Оценка F_n^* обладает хорошими асимптотическими свойствами, в частности, свойством состоятельности, когда с. в. t_i имеют непрерывную функцию распределения. Для случаев, когда t_i не являются н. о. р. с. в., а также для дискретных с. в., для обеспечения состоятельности оценки Лей и Йинг (1991) предложили модифицированную *р.-l.* оценку $\widehat{F}_n$:

$$1 - \widehat{F}_n(t) = \prod_{s \leqslant t}\left\{1 - \frac{\Delta L_n(s)}{R_n(s)}\, I(R_n(s) \geqslant cn^\alpha)\right\}, \tag{37}$$

где $c > 0$ и $0 < \alpha < 1$ — заданные числа.

Для модели 3 ими же предложен модифицированный вариант оценки $\widehat{F}_n$:

$$1 - \widehat{F}_n'(t) = \prod_{s \leqslant t}\left\{1 - \frac{\Delta L_n'(s)}{R_n'(s)}\, I(R_n'(s) \geqslant cn^\alpha)\right\}, \tag{38}$$

где

$$L_n'(s) = \sum_{i=1}^{n} I(\widetilde{X}_i^\circ \leqslant s,\ \delta_i^\circ = 1), \quad R_n'(s) = \sum_{i=1}^{n} I(\widetilde{X}_i^\circ \geqslant s \geqslant t_i^\circ).$$

Асимптотические свойства этих *р.-l.* оценок и дальнейшие детали их теории можно найти в работах: *Lai T. L., Ying Z.* Estimating a distribution function

with truncated and censored data // Ann. Statist. 1991. V. 19. № 1. P 417–442; *Gill R. D.* Glivenko-Cantelly for Kaplan-Meier // Math. Meth. Statist. 1994. V. 3. № 1. P. 76–87.

§ 3.7. Объединение и улучшение оценок

Настоящий параграф посвящен важным вопросам построения на базе уже имеющихся оценок (для некоторой теоретической характеристики или параметра) новой оценки, обладающей бóльшей точностью, нежели исходные оценки. Здесь можно выделить два аспекта:

1) проблема объединения информации, получаемой в разных экспериментах, и построение итоговой оценки с бóльшей точностью, нежели оценки, порождаемые соответствующими экспериментами, и
2) проблема «улучшения» имеющейся оценки с целью построения на ее основе новой оценки с бóльшей точностью. Ниже будут изложены некоторые общие подходы в решении этих проблем и даны практически полезные иллюстрации их применения в конкретных задачах.

1. Объединение оценок

В качестве введения в общую проблематику рассмотрим простейшую ситуацию, когда имеются две независимые и несмещенные оценки T_1 и T_2 некоторого параметра θ и их дисперсии $\mathbf{D}T_i = \sigma_i^2$, $i = 1, 2$, известны. Как на их основе построить более точную новую несмещенную оценку? Рассмотрим класс линейных комбинаций

$$\Delta = \{T_\alpha \quad \alpha T_1 + (1-\alpha)\, T_2, 0 \leqslant \alpha \leqslant 1\}.$$

Все они — несмещенные оценки:

$$\mathbf{E}T_\alpha = \alpha \mathbf{E}T_1 + (1-\alpha)\, \mathbf{E}T_2 = \alpha\theta + (1-\alpha)\,\theta = \theta$$

и

$$\mathbf{D}T_\alpha = \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2 \equiv \psi(\alpha).$$

Найдем α из условия обращения дисперсии $\mathbf{D}T_\alpha$ в минимум, т. е. решая уравнение $\psi'(\alpha) = 0$. Отсюда находим точку минимума

$$\alpha^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

Объединение независимых оценок

и, следовательно, оценку с минимальной дисперсией (в классе Δ)

$$T^* = T_{\alpha^*} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\, T_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\, T_2; \tag{1}$$

ее дисперсия

$$\mathbf{D}T^* = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} < \min(\sigma_1^2, \sigma_2^2),$$

т. е. это более точная несмещенная оценка, нежели исходные оценки T_1, T_2.

Пример 1 (Нормальное распределение, объединение оценок). Пусть $X = (X_1, \quad, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ — две независимые выборки из одного и того же нормального распределения $\mathcal{N}(\theta_1, \theta_2^2)$ (оба параметра неизвестны). Мы знаем (см. теорему 2 § 2.2), что

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{и} \quad S_0^2(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

— несмещенные оценки соответственно для θ_1 и θ_2^2 и аналогично $\overline{Y}$ и $S_0^2(Y)$. Мы также знаем (см. теорему Фишера и соотношение (7) в § 2.5), что эти четыре статистики независимы и

$$\mathbf{D}\overline{X} = \frac{\theta_2^2}{n}, \quad \mathbf{D}\overline{Y} = \frac{\theta_2^2}{m}, \quad \mathbf{D}S_0^2(X) = \frac{2\theta_2^4}{n-1}, \quad \mathbf{D}S_0^2(Y) = \frac{2\theta_2^4}{m-1}.$$

Отсюда и из (1) получаем, что объединенная оценка для θ_1 есть

$$\theta_1^* = \frac{\mathbf{D}\overline{Y}}{\mathbf{D}\overline{X}+\mathbf{D}\overline{Y}}\,\overline{X} + \frac{\mathbf{D}\overline{X}}{\mathbf{D}\overline{X}+\mathbf{D}\overline{Y}}\,\overline{Y} =$$

$$= \frac{n}{n+m}\,\overline{X} + \frac{m}{n+m}\,\overline{Y} = \frac{1}{n+m}\left(\sum_{i=1}^{n} X_i + \sum_{i=1}^{m} Y_i\right), \qquad (2)$$

т. е. среднее арифметическое всех данных, и при этом $\mathbf{D}\theta_1^* = \theta_2^2/(n+m)$. Аналогично имеем, что объединенная оценка для теоретической дисперсии θ_2^2 есть

$$(\theta_2^2)^* = \frac{n-1}{n+m-2}\,S_0^2(X) + \frac{m-1}{n+m-2}\,S_0^2(Y) =$$

$$= \frac{1}{n+m-2}\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{i=1}^{m}(Y_i - \overline{Y})^2\right], \qquad (3)$$

и при этом $\mathbf{D}(\theta_2^2)^* = 2\theta_2^4/(n+m-2)$. •

Объединение коррелированных оценок

В общем случае, когда оценки коррелированы: $\operatorname{corr}(T_1, T_2) = \rho$, для дисперсии их линейной комбинации T_α имеет место соотношение

$$\mathbf{D}T_\alpha = \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2 + 2\alpha(1-\alpha)\rho\sigma_1\sigma_2 \equiv \psi(\alpha),$$

и решение уравнения $\psi'(\alpha) = 0$ есть

$$\alpha^* = \frac{\sigma_2(\sigma_2 - \rho\sigma_1)}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}.$$

В итоге имеем, что в классе Δ существует единственная оценка с минимальной дисперсией

$$T^* = T_{\alpha^*} = \frac{\sigma_2(\sigma_2 - \rho\sigma_1)\,T_1 + \sigma_1(\sigma_1 - \rho\sigma_2)T_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2} \tag{4}$$

и ее дисперсия есть

$$\mathbf{D}T^* = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2},$$

что сводится к результату (1) при $\rho = 0$.

Перейдем к общей постановке проблемы объединения оценок. Пусть $T_1, \quad , T_k$ — несмещенные оценки некоторой величины g, полученные в k экспериментах. Обычно на практике это оценки, построенные по соответствующим независимым выборкам $X_j = (X_{j1}, \quad , X_{jn_j})$, $j = 1, \quad , k$, т. е. $T_j = T_j(X_j)$, $j = 1, \quad , k$, и тогда они представляют собой независимые случайные величины с $\mathbf{E}T_j = g$ и некоторыми дисперсиями, которые мы будем обозначать $\mathbf{D}T_j = \sigma_j^2/n_j$, $j = 1, \quad , k$. Такой вид дисперсии имеют в случае, когда $T_j = \overline{X}_j = (X_{j1} + \quad + X_{jn_j})/n_j$ и $\mathbf{D}X_{ji} = \sigma_j^2$, $i = 1, \quad , n_j$, но такое обозначение мы сохраним для них и для случаев, когда оценки T_j могут задаваться и каким-либо другим образом. Требуется по оценкам T_j, $j = 1, \quad , k$, построить более точную оценку для g, нежели любая из этих исходных оценок, учитывая при этом различные специфические условия, могущие иметь место на практике, как-то: дисперсии известны либо нет. В более общем случае исходные оценки могут быть коррелированными, и тогда могут возникать дополнительные предположения об их смешанных вторых моментах (ковариациях). Возможны и дальнейшие различного типа детализации общих ситуаций, связанные, например, с условием несмещенности оценок T_j (в принципе, это исходное условие может также нарушаться, т. е. оценки T_j могут быть неоднородными и в смысле математического ожидания, а не только в смысле дисперсии). Таким образом, эта проблематика весьма многоплановая, и мы далее дадим описание оптимальных алгоритмов построения итоговой оценки лишь при простейших естественных предположениях.

1) *Некоррелированные оценки с известными дисперсиями.* Предположим, что дисперсии $\mathbf{D}T_j$ (т. е. величины σ_j^2, $j = 1, \quad , k$) исходных оценок известны, и эти оценки некоррелированы. Итоговую оценку величины g будем искать в классе линейных несмещенных оценок вида

$$T_{\underline{\alpha}} = \alpha_1 T_1 + \quad + \alpha_k T_k = \underline{\alpha}'\underline{T}, \tag{5}$$

где

$$\underline{\alpha} = (\alpha_1, \ldots, \alpha_k), \quad \alpha_1 + \quad + \alpha_k = 1, \quad \underline{T} = (T_1, \quad , T_k).$$

При этом в качестве критерия точности оценки $T_{\underline{\alpha}}$ будем считать ее дисперсию. Таким образом, нашей целью будет построение несмещенной оценки с минимальной дисперсией (н. о. м. д.) в классе оценок (5).

В принятых обозначениях и при сделанных предположениях дисперсия произвольной оценки (5) равна

$$\mathbf{D}T_{\underline{\alpha}} = \alpha_1^2 \frac{\sigma_1^2}{n_1} + \quad + \alpha_k^2 \frac{\sigma_k^2}{n_k}. \tag{6}$$

Минимум правой части (6) по $\underline{\alpha}$ при условии

$$\sum_{j=1}^{k} \alpha_j = 1$$

находится методом неопределенных множителей Лагранжа[6], согласно которому надо найти абсолютный минимум функции

$$F(\underline{\alpha}, \lambda) = \sum_{j=1}^{k} \frac{\alpha_j^2 \sigma_j^2}{n_j} - \lambda\left(\sum_{j=1}^{k} \alpha_j - 1\right).$$

Решая систему уравнений

$$\frac{\partial F(\underline{\alpha}, \lambda)}{\partial \alpha_j} = 0, \quad j = 1, \ldots, k, \quad \frac{\partial F(\underline{\alpha}, \lambda)}{\partial \lambda} = 0,$$

легко находим, что минимум достигается при

$$\alpha_j = \alpha_j^* = \frac{n_j}{\sigma_j^2} B^2, \quad j = 1, \ldots, k, \quad B^2 = \left(\sum_{j=1}^{k} \frac{n_j}{\sigma_j^2}\right) \tag{7}$$

Таким образом, в рассматриваемом случае н. о. м. д. T^* существует и имеет вид

Объединенная оценка при отсутствии корреляций

$$T^* = T_{\underline{\alpha}^*} = \alpha_1^* T_1 + \quad + \alpha_k^* T_k = B^2 \sum_{j=1}^{k} \frac{n_j}{\sigma_j^2} T_j; \tag{8}$$

при этом из (6) и (7) следует, что

$$\mathbf{D}T^* = B^2 < \min_j \frac{\sigma_j^2}{n_j} = \min_j \mathbf{D}T_j \tag{9}$$

(как и должно быть: ведь $\mathbf{D}T^*$ есть минимум среди всех линейных комбинаций (5), включая и сами исходные оценки T_j). Отметим также очевидное неравенство:

$$B^2 \leqslant \frac{1}{k} \max_j \frac{\sigma_j^2}{n_j}.$$

2) *Объединение оценок с неизвестными дисперсиями.* Более реалистичной является ситуация, когда дисперсии объединяемых оценок неизвестны. Для

[6] Лагранж Жозеф Луи (1736–1813) — французский математик, механик и астроном, почетный член Петербургской АН (1776).

этого случая мы будем предполагать, что оценки T_j порождены независимыми выборками $X_j = (X_{j1}, \ldots, X_{jn_j})$: $T_j = \overline{X}_j$, и при этом нам известны также выборочные дисперсии

$$S_j^2(X) = \frac{1}{n_j} \sum_{i=1}^{n_j} (X_{ji} - \overline{X}_j)^2, \quad j = 1, \ldots, k.$$

В рассматриваемом случае естественно поступить следующим образом: заменить в формулах (7) и (8) неизвестные дисперсии σ_j^2 их оценками S_j^2 и в итоге взять в качестве объединенной оценки следующую оценку со случайными «весами»:

Объединенная оценка в случае неизвестных дисперсий

$$\widetilde{T} = \widetilde{\alpha}_1 T_1 + \quad + \widetilde{\alpha}_k T_k, \tag{10}$$

где веса

$$\widetilde{\alpha}_j = \frac{n_j}{S_j^2} \widetilde{B}^2 \quad j = 1, \quad , k, \quad \widetilde{B}^2 = \left(\sum_{j=1}^{k} \frac{n_j}{S_j^2} \right) \tag{11}$$

Однако исследование свойств оценок вида (10) значительно усложняется. Так, если в случае известных дисперсий σ_j^2 дисперсия объединенной оценки (8) не превосходит дисперсии любой исходной оценки T_j, то такое утверждение для оценки (10), вообще говоря, уже перестает быть верным. Но, как показывают соответствующие исследования, такое явление может происходить лишь при небольших объемах выборок. В случае же больших выборок, т.е. при $\min_j n_j \to \infty$, свойства оценок (10) будут «достаточно хорошими». Приведем соответствующую предельную теорему, дающую условия, при которых оценка (10) является состоятельной и асимптотически нормальной.

Теорема 1. Пусть объемы выборок удовлетворяют условиям:

$$n_j = \gamma_j(n) n, \quad \gamma_j(n) \to \gamma_j > 0 \quad \text{при} \quad n \to \infty, \quad j = 1, \quad , k, \tag{12}$$

k фиксировано и, кроме того, $\mathrm{E}(X_{ji} - g)^8 \leqslant c < \infty$, $0 < \underline{\sigma}^2 \leqslant \sigma_j^2 \leqslant \bar{\sigma}^2 < \infty$, $j = 1, \ldots, k$. Тогда величина $(\widetilde{T} - g)/\widetilde{B}$ при $n \to \infty$ асимптотически нормальна с параметрами (0, 1): $\mathcal{L}((\widetilde{T} - g)/\widetilde{B}) \to \mathcal{N}(0, 1)$.

Из этой теоремы следует, что для больших выборок (12) при достаточно широких условиях оценка (10) обладает хорошими асимптотическими свойствами: она является асимптотически несмещенной, состоятельной (так как $\widetilde{B}^2 \to 0$ при $n \to \infty$) и, более того, она позволяет построить для величины g асимптотический γ-доверительный интервал. Действительно, так как при $n \to \infty$

$$\mathbf{P}\left\{ \frac{|\widetilde{T} - g|}{\widetilde{B}} < c_\gamma \right\} \to \Phi(c_\gamma) - \Phi(-c_\gamma) = 2\Phi(c_\gamma) - 1 = \gamma,$$

где $\gamma \in (0, 1)$ — заданный доверительный уровень и $c_\gamma = \Phi^{-1}((1+\gamma)/2)$, то отсюда следует, что интервал

$$\Delta_\gamma(n) = \left(\widetilde{T} \mp c_\gamma \widetilde{B}\right) \tag{13}$$

является асимптотическим γ-доверительным интервалом для g:

$$\mathbf{P}\{g \in \Delta_\gamma(n)\} \to \gamma \quad \text{при} \quad n \to \infty.$$

Объединенная оценка в случае нормальных выборок

Дальнейшие уточнения свойств оценок (10) можно сделать, если дополнительно предположить, что выборки X_j извлечены из нормальных совокупностей $\mathcal{N}(g, \sigma_j^2)$, j 1, , k. Вместо смещенных оценок S_j^2 теоретических дисперсий σ_j^2 будем использовать несмещенные оценки

$$S_{j0}^2 = \frac{n_j S_j^2}{n_j - 1}.$$

Тогда формулы (10)–(11) заменяются на формулы

$$\widetilde{T}_0 = \widetilde{\alpha}_{10} T_1 + \quad + \widetilde{\alpha}_{k0} T_k, \tag{14}$$

где (в условиях (12))

$$\widetilde{\alpha}_{j0} = \frac{\gamma_j(n)}{S_{j0}^2} \widetilde{B}_0^2, \quad j = 1, \quad ., k, \quad \widetilde{B}_0^2 = \left(\sum_{j=1}^{k} \frac{\gamma_j(n)}{S_{j0}^2}\right) \tag{15}$$

Покажем, что в данном случае оценка $\widetilde{T}_0$ является несмещенной. Действительно, так как для нормальных выборок по теореме Фишера (см. теорему 2 § 2.5) статистики $\overline{X}_j$ и S_{j0}^2 независимы, то независимы T_j и $\widetilde{\alpha}_{j0}$. Но тогда

$$\mathbf{E}\widetilde{T}_0 = \sum_{j=1}^{k} \mathbf{E}\widetilde{\alpha}_{j0} \mathbf{E} T_j = g \sum_{j=1}^{k} \mathbf{E}\widetilde{\alpha}_{j0} = g\mathbf{E}\left(\sum_{j=1}^{k} \widetilde{\alpha}_{j0}\right) = g, \tag{16}$$

поскольку $\sum_{j=1}^{k} \widetilde{\alpha}_{j0} = 1$.

Вычисление дисперсии оценки (14) более сложно, и мы приведем для нее лишь асимптотическую формулу в условиях (12).

Теорема 2. Если для нормальных выборок выполняются условия (12) и

$$0 < \underline{\sigma}^2 \leqslant \sigma_j^2 \leqslant \overline{\sigma}^2 < \infty,$$

то

$$\mathbf{D}\widetilde{T}_0 = \frac{B_0^2}{n}\left[1 + \frac{2}{n}\sum_{j=1}^{k} \frac{\alpha_{j0}(1-\alpha_{j0})}{\gamma_j}\right] + O\left(\frac{1}{n^{5/2}}\right), \tag{17}$$

где

$$B_0^2 = \left(\sum_{j=1}^{k} \frac{\gamma_j}{\sigma_j^2}\right) \qquad \alpha_{j0} = \frac{\gamma_j}{\sigma_j^2} B_0^2, \quad j = 1, \quad , k.$$

Таким образом, по крайней мере для нормальных выборок точность несмещенной объединенной оценки (14) асимптотически (в условиях (12)) такая же, как и для оптимальной объединенной оценки (8) в случае известных дисперсий.

3) *Объединение коррелированных оценок.* В общем случае наличия корреляций $\rho_{ij} = \operatorname{corr}(T_i, T_j)$ между исходными оценками $T_1, \quad , T_k$ дисперсия оценки (5) имеет вид

$$\mathbf{D}T_{\underline{\alpha}} = \sum_{,j=1}^{k} \alpha_i \alpha_j \operatorname{cov}(T_i, T_j) = \frac{1}{n} \sum_{i,j=1}^{k} \mathbf{D}_{ij} \alpha_i \alpha_j = \frac{1}{n} \underline{\alpha}^t \mathbf{D} \underline{\alpha} \equiv \frac{1}{n} Q(\underline{\alpha}). \quad (18)$$

где

$$\mathbf{D}_{ij} = n \operatorname{cov}(T_i, T_j) = \frac{\rho_{ij} \sigma_i \sigma_j}{\sqrt{\gamma_i \gamma_j}}, \quad n_j = \gamma_j n, \quad i, j = 1, \ldots, k.$$

Считая матрицу $\mathbf{D} = \|\mathbf{D}_{ij}\|_1^k$ известной и невырожденной, будем минимизировать квадратичную форму $Q(\underline{\alpha})$ при ограничении

$$\sum_{j=1}^{k} \alpha_j = 1.$$

Снова применяя метод неопределенных множителей Лагранжа, введем функцию

$$F(\underline{\alpha}, \lambda) = Q(\underline{\alpha}) - \lambda\left(\sum_{j=1}^{k} \alpha_j - 1\right)$$

и найдем ее точку минимума, решая систему уравнений

$$\frac{\partial F(\underline{\alpha}, \lambda)}{\partial \alpha_i} = 0, \quad i = 1, \ldots, k, \quad \sum_{j=1}^{k} \alpha_j = 1.$$

Эта система имеет вид

$$\frac{\partial Q(\underline{\alpha})}{\partial \alpha_i} = 2 \sum_{j=1}^{k} \mathbf{D}_{ij} \alpha_j = \lambda, \quad i = 1, \quad , k, \quad \sum_{j=1}^{k} \alpha_j = 1,$$

или в матричной форме

$$\mathbf{D}\underline{\alpha} = \frac{\lambda}{2} \underline{1}, \quad \sum_{j=1}^{k} \alpha_j = 1,$$

где $\underline{1}$ — вектор-столбец из единиц. Отсюда

$$\underline{\alpha} = \frac{\lambda}{2}\mathbf{D}^{-1}\underline{1}, \qquad \underline{1}'\underline{\alpha} = \frac{\lambda}{2}\underline{1}'\mathbf{D}^{-1}\underline{1} = 1.$$

Таким образом,

$$\frac{\lambda}{2} = \left(\underline{1}'\,\mathbf{D}^{-1}1\right)^{-1} \quad \text{и} \quad \underline{\alpha} = \underline{\alpha}^* = \left(\underline{1}'\,\mathbf{D}^{-1}\underline{1}\right)^{-1}\mathbf{D}^{-1}\underline{1}.$$

Это и есть оптимальный выбор $\underline{\alpha}^*$ коэффициентов линейной комбинации (5) с учетом наличия корреляций между исходными оценками $\{T_j\}$. Запишем этот результат в более явном виде. Для этого обозначим

$$\begin{gathered}\mathbf{D}^{-1} = \left\|\mathbf{D}^{ij}\right\|_1^k, \quad W_i = \sum_{j=1}^{k}\mathbf{D}^{ij}, \\ W = \sum_{i=1}^{k} W_i = \sum_{i,j=1}^{k}\mathbf{D}^{ij}, \quad H_i = \frac{W_i}{W}.\end{gathered} \tag{19}$$

Тогда

$$W = \underline{1}'\mathbf{D}^{-1}\underline{1}, \quad \underline{\alpha}^{*\prime} = \frac{\underline{1}'\mathbf{D}^{-1}}{W} = (H_1, \ldots, H_k)$$

Объединенная оценка в случае наличия корреляций

и н. о. м. д.

$$T^* = T_{\underline{\alpha}^*} = \underline{\alpha}^{*\prime}\underline{T} = \sum_{j=1}^{k} H_j T_j = \frac{1}{W}\sum_{j=1}^{k} W_j T_j. \tag{20}$$

При этом (см. (18))

$$\mathbf{D}T^* = \frac{1}{n}\underline{\alpha}^{*\prime}\mathbf{D}\underline{\alpha}^* = \frac{1}{nW^2}\,\underline{1}'\mathbf{D}^{-1}\mathbf{D}\mathbf{D}^{-1}\underline{1} = \frac{1}{nW^2}\,\underline{1}'\mathbf{D}^{-1}\underline{1} = \frac{1}{nW}. \tag{21}$$

Оценка (20) точнее оценивает g, нежели любая из исходных оценок T_j.

2. Улучшение оценок

Выше мы акцентировали внимание на использовании несмещенных (или асимптотически несмещенных) оценок и искали среди них оценку с наименьшей дисперсией, как наиболее точную. Однако в реальных ситуациях в ряде случаев требование несмещенности итоговой оценки не является «естественным», более того, использование смещенных оценок может оказаться даже предпочтительным, если в качестве критерия точности оценок использовать среднеквадратичную ошибку (с. к. о.). Один пример такого рода мы уже встречали ранее в связи с задачей оценивания дисперсии общей нормальной модели (см. пример 4 в § 3.1). В общем случае, пусть T есть несмещенная оценка с равномерно минимальной дисперсией для некоторой параметрической функции $\tau = \tau(\theta)$, $\theta \in \Theta$, но нас не устраивает ее точность, хотя оценка T и является наилучшей

среди несмещенных оценок. Спрашивается, можно ли с помощью T построить оценку с меньшей с. к. о.? Ясно, что такая оценка, если она существует, будет смещенной, тем самым, допуская смещение оценки, мы стремимся увеличить точность (уменьшить квадратичный риск $\equiv$ с. к. о.). Итак, в этой проблеме платой за более высокую точность оценки является отказ от условия ее несмещенности. Рассмотрим, когда и как этого удается достичь.

С помощью оценки T введем класс оценок

$$\mathcal{T} = \{T_\lambda = \lambda T,\ \lambda \in R^1\}.$$

Тогда с. к. о. произвольной оценки T_λ из этого класса есть

$$\mathbf{E}_\theta(T_\lambda - \tau(\theta))^2 = \mathbf{E}_\theta\big[\lambda(T - \tau(\theta)) + (\lambda - 1)\tau(\theta)\big]^2 = \lambda^2 \mathbf{D}_\theta T + (\lambda - 1)^2 \tau^2(\theta)$$

и минимум этого риска достигается при

$$\lambda = \lambda_T(\theta) = \frac{\tau^2(\theta)}{\mathbf{D}_\theta T + \tau^2(\theta)} = \frac{(\mathbf{E}_\theta T)^2}{\mathbf{E}_\theta T^2} \in (0, 1). \tag{22}$$

Если $\lambda_T(\theta) = \lambda^* \equiv \text{const}$ (не зависит от θ), то при всех $\theta \in \Theta$

$$\mathbf{E}_\theta(T_{\lambda^*} - \tau(\theta))^2 < \mathbf{E}_\theta(T_\lambda - \tau(\theta))^2 = \mathbf{D}_\theta T,$$

т. е. T_{λ^*} — искомая оценка, причем она является смещенной (именно так было в примере 4 в § 3.1)).

В общем случае, когда $\lambda_T(\theta)$ не является тождественной константой, введем два числа

$$\lambda_T^{(1)} = \inf_\theta \lambda_T(\theta) \quad \text{и} \quad \lambda_T^{(2)} = \sup_\theta \lambda_T(\theta). \tag{23}$$

Тогда, если $\lambda_{T-c}^{(1)} > 0$ при некотором $c \in R^1$ то оценка $T^{(1)} = \lambda_{T-c}^{(1)}(T - c) + c$ будет иметь меньшую с. к. о., чем оценка T; если же $\lambda_{T-c}^{(2)} < 1$, то лучшей (более точной) оценкой будет $T^{(2)} = \lambda_{T-c}^{(2)}(T - c) + c$. Достаточным условием существования какой-то одной из оценок $T^{(1)}$ или $T^{(2)}$ является выполнение неравенства

$$\frac{\mathbf{D}_\theta T}{[\mathbf{E}_\theta(T - c)]^2} \geqslant \varepsilon, \quad \forall\, \theta \in \Theta, \tag{24}$$

при некоторых $c \in R^1$ и $\varepsilon > 0$ (c и ε не зависят от θ).

Доказательства. Докажем это, например, для sup. Имеем (θ для сокращения записи опускаем)

$$\mathbf{D}T - \mathbf{E}(T^{(2)} - \tau)^2 = \mathbf{D}T - \mathbf{E}\big[\lambda^{(2)}(T - \tau) + (c - \tau)(1 - \lambda^{(2)})\big]^2 =$$

$$= \big(1 - \lambda^{(2)^2}\big)\mathbf{D}T - (1 - \lambda^{(2)})^2(c - \tau)^2 > 0 \iff$$

(так как $\lambda^{(2)} < 1$ и $\mathbf{D}T = \mathbf{D}(T - c)$)

$$(1 + \lambda^{(2)})\big[\mathbf{E}(T - c)^2 - (\mathbf{E}(T - c))^2\big] > (1 - \lambda^{(2)})[\mathbf{E}(T - c)]^2 \iff$$

$$\Longleftrightarrow (1+\lambda^{(2)}_{T-c})(1-\lambda_{T-c}(\theta)) > (1-\lambda^{(2)}_{T-c})\lambda_{T-c}(\theta) \Longleftrightarrow 2\lambda_{T-c}(\theta) < 1+\lambda^{(2)}_{T-c},$$

что очевидно, так как $\lambda_{T-c}(\theta) \leqslant \lambda^{(2)}_{T-c} < 1$. ■

Пример 2. Пусть наблюдения производятся над геометрической случайной величиной $\xi' = \xi + 1$ с $\mathcal{L}_\theta(\xi) = \overline{\mathrm{Bi}}(1, 1-1/\theta)$ (см. п. 2 § 1.1), т. е.

$$\mathbf{P}_\theta\{\xi' = x\} = \frac{1}{\theta}\left(1-\frac{1}{\theta}\right)^{x-} \qquad x = 1, 2, \ldots, \quad \theta > 1.$$

Тогда, в соответствии с формулами (7) § 1.1,

$$\mathbf{E}_\theta\xi' = \mathbf{E}_\theta\xi + 1 = \theta\left(1-\frac{1}{\theta}\right) + 1 = \theta, \quad \mathbf{D}_\theta\xi' = \mathbf{D}_\theta\xi = \theta(\theta-1).$$

Пусть $X = (X_1, \quad, X_n)$ — соответствующая выборка. Поскольку рассматриваемая модель относится к семейству моделей степенного ряда, то (см. п. 1 § 3.4) сумма наблюдений $\sum_{i=1}^{n} X_i$ является здесь полной достаточной статистикой и потому выборочное среднее $T = \overline{X}$ есть н. о. р. м. д. для θ. В данном примере условие (24) принимает вид

$$\frac{\mathbf{D}_\theta T}{[\mathbf{E}_\theta(T-c)]^2} = \frac{\theta(\theta-1)}{n(\theta-c)^2}\bigg|_{c=1} \geqslant \varepsilon = \frac{1}{n}, \quad \forall\,\theta > 1,$$

и, следовательно, оценка, имеющая меньшую с. к. о. чем T, существует. Так как (см. (22))

$$\lambda_{T-1}(\theta) = \frac{[\mathbf{E}_\theta(T-1)]^2}{\mathbf{E}_\theta(T-1)^2} = \frac{(\theta-1)^2}{\mathbf{D}_\theta T + (\theta-1)^2} = \frac{1}{1+\dfrac{\theta}{n(\theta-1)}} < 1,$$

а (см. (23))

$$\lambda^{(2)}_{T-1} = \sup_{\theta>1}\lambda_{T-1}(\theta) = \frac{1}{1+1/n} = \frac{n}{n+1},$$

то искомая оценка есть

$$T^{(2)} = \lambda^{(2)}_{T-1}(T-1) + 1 = \frac{n\overline{X}+1}{n+1}. \tag{25}$$

Конечно, эта оценка уже не будет несмещенной, но ее с. к. о.

$$\mathbf{E}_\theta(T^{(2)}-\theta)^2 = \frac{\theta(\theta-1)}{n+1} - \frac{\theta-1}{(n+1)^2} < \mathbf{D}_\theta T = \frac{\theta(\theta-1)}{n}. \tag{26}$$

•

На практике иногда сталкиваются с противоположной ситуацией, когда имеется смещенная оценка $\delta_n(X)$, $X = (X_1, \ldots, X_n)$, для параметра θ, а условия задачи требуют использования несмещенной оценки или хотя бы оценки

с меньшим смещением, чем у $\delta_n(X)$. При этом экспериментатор даже готов пойти на увеличение риска оценки, лишь бы уменьшить ее смещение. Общий прием, с помощью которого можно уменьшать смещение исходной смещенной оценки, носит название метода *«складного ножа»* (jackknife method (англ.)). Суть его состоит в следующем.

Метод «складного ножа»

Рассмотрим оценки $\delta_{n(i)}(X)$, основанные на выборках объема $n-1$, полученных из X в результате исключения i-го наблюдения:

$$\delta_{n(i)}(X) = \delta_{n-1}(X_1, \quad , X_{i-1}, X_{i+1}, \quad , X_n), \quad i = 1, \quad n.$$

Построим оценку

$$\widehat{\theta}_1 = n\delta_n(X) - (n-1)\delta_{(1)}(X), \quad \delta_{(1)}(X) = \frac{1}{n}\sum_{i=1}^{n}\delta_{n(i)}(X). \tag{27}$$

Можно показать, что если смещение исходной оценки $\delta_{(n)}(X)$ имеет вид

$$\mathbf{E}\delta_n(X) - \theta = \frac{a_1}{n} + \frac{a_2}{n^2} +$$

то смещение $\mathbf{E}\widehat{\theta}_1 - \theta = O(1/n^2)$ оценки $\widehat{\theta}_1$ уже не содержит членов порядка $O(1/n)$, т. е. оценка $\widehat{\theta}_1$ устраняет смещение порядка $1/n$. При этом во многих случаях дисперсии оценок $\widehat{\theta}_1$ и $\delta_n(X)$ приблизительно одинаковы, а дисперсию $\mathbf{D}\widehat{\theta}_1$ можно оценить по формуле

$$\frac{n-1}{n}\sum_{i=1}^{n}\left(\delta_{n(i)}(X) - \delta_{(1)}(X)\right)^2$$

Оценка $\widehat{\theta}_1$ называется *оценкой «складного ножа» первого порядка*. Метод «складного ножа» дает хорошие результаты при устранении смещения оценок максимального правдоподобия (о. м. п.), которые, как правило, оказываются смещенными; при этом оценка $\widehat{\theta}_1$ имеет такое же асимптотическое распределение, как и исходная о. м. п. (см. соотношение (23) § 3.5).

Пример 3. Пусть $X = (X_1, \quad , X_n)$ — выборка из нормального распределения $\mathcal{N}(\theta_1, \theta_2^2)$ и надо оценить θ_2^2. В качестве оценки можно использовать выборочную дисперсию

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

которая является о. м. п. для θ_2^2 (см. пример 1 в § 3.5). Но эта оценка смещена (несмещенной оценкой является статистика $S_0^2 = \dfrac{n}{n-1}S^2$, см. пример 4 в § 3.1) и ее смещение есть

$$\mathbf{E}_\theta S^2 - \theta_2^2 = \frac{n-1}{n}\mathbf{E}_\theta S^2 - \theta_2^2 = -\frac{\theta_2^2}{n}.$$

Чтобы уменьшить смещение, применим формулу (27). Так как

$$\delta_n(X) = S^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2$$

то

$$\delta_{n(i)}(X) = \frac{1}{n-1}\left(X_1^2 + \quad + X_{i-1}^2 + X_{i+1}^2 + \ldots + X_n^2\right) -$$

$$- \frac{1}{(n-1)^2}\left(X_1 + \quad + X_{i-1} + X_{i+1} + \ldots + X_n\right)^2 =$$

$$= \frac{1}{n-1}\left(\sum_{j=1}^{n} X_j^2 - X_i^2\right) - \frac{1}{(n-1)^2}\left[\left(\sum_{j=1}^{n} X_j\right)^2 - 2X_i\sum_{j=1}^{n} X_j + X_i^2\right],$$

откуда

$$\delta_{(1)}(X) = \frac{1}{n}\sum_{i=1}^{n}\delta_{n(i)}(X) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n(n-1)^2}\left[\sum_{i=1}^{n} X_i^2 + (n-2)\left(\sum_{i=1}^{n} X_i\right)^2\right].$$

Подставляя $\delta_n(X)$ и $\delta_{(1)}(X)$ в формулу (27), имеем

$$\widehat{\theta}_1 = \sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2 - \frac{n-1}{n}\sum_{i=1}^{n} X_i^2 + \frac{1}{n(n-1)}\sum_{i=1}^{n} X_i^2 +$$

$$+ \frac{n-2}{n(n-1)}\left(\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n} X_i\right)^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S_0^2,$$

т. е. мы устранили смещение порядка $1/n$ и пришли к наилучшей несмещенной оценке S_0^2 (см. пример 9 в § 3.2). •

Обобщенная оценка «складного ножа»

В случае, когда необходимо устранять смещение и более высоких порядков, чем $O(1/n)$, используют *обобщенную оценку «складного ножа» k-го порядка*

$$\widehat{\theta}_k = \frac{n^k}{k!}\sum_{i=0}^{k}(-1)^i C_k^i\left(1 - \frac{i}{n}\right)^k \delta_{(i)}(X),$$

где $\delta_{(0)}(X) = \delta_n(X)$, $\delta_{(1)}(X)$ — среднее значение оценок $\delta_{n(i)}(X)$, построенных по выборкам объема $n-1$ с пропущенным i-м наблюдением (как выше в (27)), $\delta_{(2)}(X)$ — среднее от оценок $\delta_{n(ij)}(X)$ при выборках объемами $n-2$ (пропущены i-е и j-е наблюдения) и т. д. Оценка $\widehat{\theta}_k$ устраняет смещение вплоть до членов порядка $1/n^k$, $k = 1, 2, \ldots$.

§ 3.8. Доверительное оценивание

До сих пор мы акцентировали внимание на точечных оценках параметра θ исследуемой модели $\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$ и функций от него. Любая точечная оценка представляет собой функцию $T = T(X)$ выборки $X = (X_1, \quad , X_n)$, т. е. является случайной величиной, и при каждой реализации $\underline{x}$ выборки X эта функция определяет единственное значение $t = T(\underline{x})$ оценки, принимаемое за приближаемое значение оцениваемой характеристики. При этом надо принимать во внимание, что в каждом конкретном случае значение оценки может отличаться от значения параметра, поэтому желательно было бы знать и возможную погрешность, возникающую при использовании предлагаемой оценки, например, указывая такой интервал (или область в случае векторного параметра), внутри которого с высокой (т. е. близкой к 1) вероятностью γ находится истинное значение оцениваемого параметра. При таком подходе говорят об *интервальном* или *доверительном оценивании*. Основная цель при этом состоит в том, чтобы при заданном *доверительном уровне* γ построить кратчайший интервал, обеспечивающий наиболее точную локализацию оцениваемой характеристики.

С подобного рода задачами мы уже встречались ранее (см. п. 2 § 2.1, п. 5 § 2.2, п. 4 § 2.3, п. 3 § 2.5, п. 5 § 3.5). Общее определение γ-доверительного интервала дано в п. 5 § 2.2 (соотношение (26)), в случае параметрических моделей — в п. 5 § 3.5 (соотношение (28)), там же (соотношение (34)) дано определение и γ-доверительной области для векторного параметра. Для удобства дальнейшей работы воспроизведем это определение здесь еще раз в случае оценивания скалярного параметра θ При интервальном оценивании ищут две такие статистики $T_i = T_i(X)$, $i = 1, 2$, что $T_1 < T_2$, для которых при заданном доверительном уровне γ выполняется условие

$$\mathbf{P}_\theta\{T_1(X) < \theta < T_2(X)\} \geqslant \gamma, \quad \forall\, \theta \in \Theta. \tag{1}$$

Иногда рассматривают *односторонние доверительные интервалы*, соответственно *верхний* (вида $\theta < T_2(X)$) и *нижний* (вида $T_1(X) < \theta$), определяемые условиями, аналогичными (1), в которых опускают соответствующую вторую границу. Вместо (1) иногда также пишут

$$\mathbf{P}_\theta\{\theta \in (T_1(X), T_2(X))\} \geqslant \gamma.$$

Аналогично определяется доверительный интервал для отдельной компоненты (например, θ_1) в случае многомерного параметра $\theta = (\theta_1, \quad , \theta_r)$:

$$\mathbf{P}_\theta\{T_1(X) < \theta_1 < T_2(X)\} \geqslant \gamma, \quad \forall\, \theta \in \Theta.$$

Перейдем теперь к изложению общих приемов построения доверительных интервалов в произвольных параметрических моделях (во встречавшихся нам ранее примерах мы имели дело лишь с нормальными либо асимптотически нормальными распределениями).

1. Построение доверительного интервала с помощью центральной статистики

Центральная статистика

Пусть модель $\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$ абсолютно непрерывна и существует случайная величина $G(X;\theta)$, зависящая от выборки $X = (X_1, \quad , X_n)$ и от параметра θ и такая, что: 1) ее распределение не зависит от θ и 2) при каждом $\underline{x} \in \mathfrak{X}$ функция $G(\underline{x};\theta)$ непрерывна и строго монотонна по θ — такую случайную величину называют *центральной статистикой* (для θ). Аналогично определяется центральная статистика для отдельной компоненты (например, θ_1) многомерного параметра θ — $G(X;\theta_1)$, а также для скалярной параметрической функции $\tau = \tau(\theta)$ (θ может быть как скаляром, так и вектором) — $G(X;\tau)$. Далее будем рассматривать для краткости лишь случай скалярного параметра (для векторного параметра все рассуждения аналогичны).

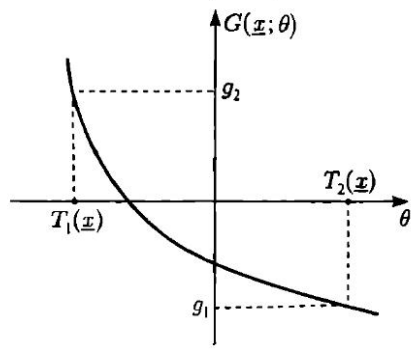


Рис. 1

Пусть для модели $\mathcal{F}$ построена центральная статистика $G(X;\theta)$ и $f_G(g)$ — ее плотность распределения. Функция $f_G(g)$ от параметра θ не зависит [условие 1)], поэтому для любого $\gamma \in (0, 1)$ можно выбрать величины $g_1 < g_2$ (многими способами) так, чтобы

$$\mathbf{P}_\theta\{g_1 < G(X;\theta) < g_2\} = \int_{g_1}^{g_2} f_G(g)\,dg = \gamma. \tag{2}$$

Определим теперь два числа $T_i(\underline{x})$, $i = 1, 2$, $\underline{x} \in \mathfrak{X}$, где $T_1(\underline{x}) < T_2(\underline{x})$, как решения относительно θ уравнений

$$G(\underline{x};\theta) = g_1, g_2 \tag{3}$$

(однозначность определения этих чисел обеспечивается условием 2), наложенным на функцию $G(\underline{x};\theta)$). Тогда неравенства $g_1 < G(X;\theta) < g_2$ эквивалентны неравенствам $T_1(\underline{x}) < \theta < T_2(\underline{x})$ (см. рис. 1) и, следовательно, (2) можно переписать в виде

$$\mathbf{P}_\theta\{T_1(X) < \theta < T_2(X)\} = \gamma, \quad \forall\,\theta.$$

Таким образом, построенный интервал $(T_1(X), T_2(X))$ в соответствии с (1) является γ-доверительным интервалом для θ.

На практике, чтобы устранить неопределенность в выборе чисел g_1 и g_2 в (2), рекомендуется выбирать их так, чтобы выполнялись равенства

$$\int_{-\infty}^{g_1} f_G(g)\,dg = \int_{g_2}^{\infty} f_G(g)\,dg = \frac{1-\gamma}{2}; \tag{4}$$

Центральный доверительный интервал

в этом случае соответствующий доверительный интервал называют *центральным*.

С применением изложенной методики мы уже встречались ранее в § 2.5 (п. 3) в связи с обсуждением теоремы Фишера. Так в примере 1 в § 2.5 была построена центральная статистика $G(X;\theta_2^2) = nS^2/\theta_2^2$ для дисперсии θ_2^2 нормальной модели $\mathcal{N}(\theta_1\theta_2^2)$, с помощью которой был получен центральный γ-доверительный интервал (10) для θ_2^2. Аналогичный интервал для среднего θ_1, указанный в (12), был построен с помощью центральной статистики t_{n-1} в (11). Другие иллюстрации этой методики дают примеры 2 и 3 § 2.5. Но все это было осуществлено с учетом специфики нормальной модели, т. е. в достаточно частном (хотя и важном) случае.

В то же время можно выделить целый класс моделей, для которых центральная статистика всегда существует и имеет простой вид. Именно: если функция распределения $F(x;\theta)$ наблюдаемой случайной величины ξ непрерывна и монотонна по параметру θ, то можно положить

$$G(X;\theta) = -\sum_{i=1}^{n} \ln F(X_i,\theta). \tag{5}$$

Действительно, условие 2) здесь обеспечено предположением о свойствах функции $F(x;\theta)$, а так как преобразование Смирнова $\eta = F(\xi;\theta)$ порождает случайную величину с равномерным распределением $\mathcal{U}(0,1)$ (см. упр. 17 к гл. 2), т. е. $\mathcal{L}_\theta(\eta) = \mathcal{U}(0,1)$ — не зависит от θ, и, в свою очередь, $\mathcal{L}(-\ln\eta) = \Gamma(1,1)$ (см. пример 1 в § 1.2), то слагаемые в (5) независимы, и каждое из них имеет стандартное показательное распределение $\Gamma(1,1)$. Отсюда по воспроизводящему свойству гамма-распределения (см. п. 3 § 1.2) имеем $\mathcal{L}_\theta(G(X;\theta)) = \Gamma(1,n)$, т. е. $G(X;\theta)$ — центральная статистика с плотностью (см. (18) § 1.2) $f_G(g) = g^{n-1}e^{-g}/\Gamma(n)$, $g > 0$. Следовательно, согласно изложенной выше общей методике, для построения γ-доверительного интервала для θ мы должны определить числа $g_1 < g_2$ из условия

$$\frac{1}{\Gamma(n)}\int_0^{g_1} g^{n-1}e^{-g}\,dg = \frac{1}{\Gamma(n)}\int_{g_2}^{\infty} g^{n-1}e^{-g}\,dg = \frac{1-\gamma}{2} \tag{6}$$

и, решая относительно θ уравнения

$$\sum_{i=1}^{n} \ln F(X_i;\theta) = g_i \quad (i = 1,2), \tag{7}$$

найти корни $T_1(X) < T_2(X)$. В итоге мы построим центральный γ-доверительный интервал $(T_1(X), T_2(X))$ для θ.

Добавим некоторые комментарии к изложенному алгоритму. Обозначим $\Gamma_{p,n}$ p-квантиль распределения $\Gamma(1,n)$, тогда условия (5) означают, что

$g_1 = \Gamma_{(1-\gamma)/2,n}$, $g_2 = \Gamma_{(1+\gamma)/2,n}$ и, следовательно, для определения g_1 и g_2 можно воспользоваться известными таблицами квантилей гамма-распределения $\Gamma(1, n)$. Очевидно, что наибольшая трудность в применении методики к конкретным задачам заключается в нахождении корней уравнений (7).

Рассмотрим несколько важных примеров использования центральных статистик при построении доверительных интервалов.

Пример 1 (Доверительный интервал для параметра равномерной модели). Пусть по выборке $\boldsymbol{X} = (X_1, \quad , X_n)$ из распределения $\mathcal{U}(0, \theta)$ требуется построить доверительный интервал для параметра θ Здесь $F(x; \theta) = x/\theta$, $0 \leqslant x \leqslant \theta$, т. е. функция распределения непрерывна и монотонна по θ при $\theta > 0$, и центральная статистика (5) принимает вид

$$G(\boldsymbol{X}; \theta) = -\sum_{i=1}^{n} \ln X_i + n \ln \theta.$$

Легко видеть, что решениями уравнений (7) являются здесь

$$T_j(\boldsymbol{X}) = \exp\left\{\frac{g_j}{n} + \frac{1}{n}\sum_{i=1}^{n} \ln X_i\right\}, \quad j = 1, 2,$$

где $g_1 = \Gamma_{(1-\gamma)/2,n}$, $g_2 = \Gamma_{(1+\gamma)/2,n}$, а искомый интервал есть $(T_1(\underline{x}), T_2(\underline{x}))$.

Центральная статистика может быть не единственна. Так, в обсуждаемом примере можно указать еще одну центральную статистику, а именно, из п. 2 § 3.4 (соотношение (13)) следует, что плотность статистики $\boldsymbol{X}_{(n)} = \max\limits_{1 \leqslant i \leqslant n} \boldsymbol{X}_i$ имеет здесь вид

$$g_1(t; \theta) = \frac{n}{\theta^n} t^{n-1} \qquad t \in [0, \theta].$$

Отсюда по формуле (45) § 1.2 легко получить (задача!), что

$$\mathcal{L}_\theta\left(\left(\frac{\boldsymbol{X}_{(n)}}{\theta}\right)^n\right) = \mathcal{U}(0, 1), \quad \forall\, \theta > 0.$$

Таким образом, случайную величину $(\boldsymbol{X}_{(n)}/\theta)^n$ также можно использовать в качестве центральной статистики для оценивания θ. Согласно общей методике (соотношения (2)–(3)), γ-доверительным для θ является любой интервал вида

$$\left(\frac{\boldsymbol{X}_{(n)}}{(g+\gamma)^{1/n}}, \frac{\boldsymbol{X}_{(n)}}{g^{1/n}}\right), \quad 0 < g \leqslant 1 - \gamma;$$

центральным среди них является интервал, соответствующий значению $g = (1 - \gamma)/2$, наименьшую же длину имеет интервал при $g = 1 - \gamma$, т. е. интервал $(\boldsymbol{X}_{(n)}, \boldsymbol{X}_{(n)}/(1-\gamma)^{1/n})$ (показать!). •

Пример 2 (Доверительный интервал для среднего нормальной модели). Пусть по выборке $\boldsymbol{X} = (X_1, \quad , X_n)$ из распределения $\mathcal{N}(\theta, \sigma^2)$ (σ^2 известно) требуется построить доверительный интервал для неизвестного среднего θ. Поскольку $\mathcal{L}_\theta(\overline{\boldsymbol{X}}) = \mathcal{N}(\theta, \sigma^2/n)$ или $\mathcal{L}_\theta(\sqrt{n}(\overline{\boldsymbol{X}} - \theta)/\sigma) = \mathcal{N}(0, 1)$, то

Таблица 1

Таблица доверительных интервалов для параметров нормальных моделей
(неизвестные параметры обозначены символом θ)

Модель	Что оценивается	Центральный γ-доверительный интервал (T_1, T_2)	Центральная статистика и ее распределение
$\mathcal{N}(\theta, \sigma^2)$	среднее θ	$T_{1,2} = \overline{X} \mp \frac{\sigma c_\gamma}{\sqrt{n}}$, $c_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$	$G(X;\theta) = \frac{(\overline{X}-\theta)\sqrt{n}}{\sigma}$; $\mathcal{N}(0,1)$
$\mathcal{N}(\theta_1, \theta_2^2)$	среднее θ_1	$T_{1,2} = \overline{X} \mp \frac{S}{\sqrt{n-1}} t_{(1+\gamma)/2,\, n-1}$	$G(X;\theta_1) = (\overline{X}-\theta_1)\frac{\sqrt{n-1}}{S}$; $S(n-1)$
$\mathcal{N}(\mu, \theta^2)$	дисперсия θ^2	$T_{1,2} = \frac{T^2}{\chi^2_{(1\pm\gamma)/2,\, n}}$, $T^2 = \sum_{i=1}^{n}(X_i-\mu)^2$	$G(X;\theta^2) = \frac{T^2}{\theta^2}$; $\chi^2(n)$
$\mathcal{N}(\theta_1, \theta_2^2)$	дисперсия θ_2^2	$T_{1,2} = \frac{nS^2}{\chi^2_{(1\pm\gamma)/2,\, n-1}}$	$G(X;\theta_2^2) = \frac{nS^2}{\theta_2^2}$; $\chi^2(n-1)$
$\mathcal{N}(\theta^{(1)}, \sigma_1^2)$ $\mathcal{N}(\theta^{(2)}, \sigma_2^2)$	разность средних $\tau = \theta^{(1)} - \theta^{(2)}$	$T_{1,2} = \overline{X} - \overline{Y} \mp \sigma c_\gamma$, $\sigma^2 = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$	$G(\boldsymbol{X}, \underline{Y}; \tau) = \frac{\overline{X}-\overline{Y}-\tau}{\sigma}$; $\mathcal{N}(0,1)$
$\mathcal{N}(\theta_1^{(1)}, \theta_2^2)$ $\mathcal{N}(\theta_1^{(2)}, \theta_2^2)$	разность средних $\tau = \theta_1^{(1)} - \theta_1^{(2)}$	$T_{1,2} = \overline{X} - \overline{Y} \mp t_{(1+\gamma)/2,\, m+n-2} \times \left[\frac{m+n}{mn(m+n-2)}\left(nS^2(X) + mS^2(\underline{Y})\right)\right]^{1/2}$	$G(X, \underline{Y}; \tau) = \sqrt{\frac{mn(m+n-2)}{m+n}} \times \frac{\overline{X}-\overline{Y}-\tau}{[nS^2(X)+mS^2(\underline{Y})]^{1/2}}$; $S(m+n-2)$
$\mathcal{N}(\theta_{11}, \theta_{12}^2)$ $\mathcal{N}(\theta_{21}, \theta_{22}^2)$	отношение дисперсий $\tau = \theta_{12}^2/\theta_{22}^2$	$T_{1,2} = \frac{n(m-1)}{m(n-1)}\frac{S^2(X)}{S^2(Y)} \Big/ F_{(1\pm\gamma)/2,\, n-1,\, m-1}$	$G(\boldsymbol{X}, Y\ \tau) = \frac{n(m-1)}{m(n-1)}\frac{S^2(\boldsymbol{X})}{S^2(Y)} \Big/ \tau$; $S(n-1, m-1)$

Примечание. Здесь $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_m)$, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S^2 = S^2(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ $\overline{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$, $S^2(Y) = \frac{1}{m}\sum_{i=1}^{m}(Y_i - \overline{Y})^2$

центральной статистикой для θ является здесь $G(X;\theta)=\sqrt{n}(\overline{X}-\theta)/\sigma$. Уравнения (4) принимают в данном случае вид $\Phi(g_1)=(1-\gamma)/2$, $\Phi(g_2)=(1+\gamma)/2$ и при этом $g_1=-g_2$ (так как $\Phi(-x)=1-\Phi(x)$). Будем использовать привычное нам обозначение $c_\gamma=\Phi^{-1}((1+\gamma)/2)$ (см. (27) § 2.2), тогда $g_2=-g_1=c_\gamma$, и, решая уравнения (3), легко находим центральный γ-доверительный интервал для θ: $(\overline{X}\mp c_\gamma\sigma/\sqrt{n})$. Это симметричный относительно случайной точки $\overline{X}$ интервал длины $2c_\gamma\sigma/\sqrt{n}$. В частности, для доверительного уровня $\gamma=0{,}99$ величина $c_\gamma=2{,}5758$, для $\gamma=0{,}998$ $c_\gamma=3{,}0902$ и т. д. Значения c_γ для различных γ можно извлечь из таблиц квантилей стандартного нормального распределения $\mathcal{N}(0,1)$. Доверительный интервал для среднего при неизвестной дисперсии σ^2 построен нами раньше в примере 1 в § 2.5 (соотношение (12)).•

Пример 3 (Доверительный интервал для дисперсии нормальной модели). Для модели $\mathcal{N}(\mu,\theta^2)$ (μ известно) также легко найти центральную статистику для $\tau=\tau(\theta)=\theta^2$:

$$G(X;\tau)=\frac{1}{\tau}\sum_{i=1}^{n}(X_i-\mu)^2$$

Действительно, так как $\mathcal{L}_\theta((X_i-\mu)/\theta)=\mathcal{N}(0,1)$, то $\mathcal{L}_\theta((X_i-\mu)^2/\tau)=\chi^2(1)$ (см. упр. 38 к гл. 1) и потому $\mathcal{L}_\theta(G(X;\theta))=\chi^2(n)$. Если $\chi^2_{p,n}$ есть p-квантиль распределения $\chi^2(n)$, то решениями уравнений (4) являются $g_1=\chi^2_{(1-\gamma)/2,n}$ и $g_2=\chi^2_{(1+\gamma)/2,n}$. Разрешая относительно τ уравнения $G(X;\tau)=g_i$ $(i=1,2)$, получаем центральный γ-доверительный интервал для дисперсии $\tau=\theta^2$:

$$\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(1+\gamma)/2,n}},\ \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{(1-\gamma)/2,n}}\right)$$

(сравни с решением (10) § 2.5 для случая неизвестного среднего μ). •

Учитывая важность нормальной модели, особую роль ее в приложениях, приведем сводную таблицу (см. табл. 1) доверительных интервалов для ее параметров (в различных постановках), объединив результаты примеров 2, 3 и полученные ранее в § 2.5 (примеры 1–3).

2. Построение доверительного интервала с использованием распределения точечной оценки параметра

Если уже имеется некоторая точечная оценка $T=T(X)$ для параметра θ и известна ее функция распределения $F_T(t;\theta)$, то доверительный интервал

Непрерывная модель

для θ можно построить, основываясь на этой функции. Предположим сначала, что распределение оценки T непрерывно и функция $F_T(t;\theta)$ непрерывна и монотонна по θ. Пусть задан доверительный уровень γ Определим при каждом $\theta \in \Theta$ числа $t_i = t_i(\theta)$, $i = 1, 2$, где $t_1 < t_2$ и

$$\mathbf{P}_\theta\{t_1 < T(X) < t_2\} = F_T(t_2;\theta) - F_T(t_1;\theta) = \gamma. \qquad (8)$$

Чтобы данная процедура была однозначной, рекомендуется выбирать эти числа так, чтобы выполнялись условия

$$F_T(t_1;\theta) = 1 - F_T(t_2;\theta) = \frac{1-\gamma}{2}, \qquad (9)$$

т. е. речь идет о построении *центрального* доверительного интервала. Обозначим через $\mathcal{D}_\gamma$ подмножество $\Theta \times \Theta$ $\mathcal{D}_\gamma$ $\{(\theta,\theta')$ $t_1(\theta) < \theta' < t_2(\theta)\}$ (рис. 2). Тогда $\mathbf{P}_\theta\{(\theta, T(X)) \in \mathcal{D}_\gamma\} = \gamma, \forall\, \theta \in \Theta$. Определим теперь при фиксированном θ' сечение $\mathcal{D}_\gamma(\theta')$ множества $\mathcal{D}_\gamma$ $\mathcal{D}_\gamma(\theta') = \{\theta$ $(\theta,\theta') \in \mathcal{D}_\gamma\}$ — и рассмотрим случайное множество $\mathcal{D}_\gamma(T(X)) \subset \Theta$. Событие $\theta \in \mathcal{D}_\gamma(T(X))$ происходит тогда и только тогда, когда $T(X) \in (t_1(\theta), t_2(\theta))$ и, следовательно, при каждом θ имеет вероятность γ Таким образом, построено случайное множество $\mathcal{D}_\gamma(T(X))$, которое накрывает истинное значение параметра θ с вероятностью γ. Если это множество является интервалом, то тем самым построен (центральный) γ-доверительный интервал для θ. Это будет так, если кривые $\theta' = t_i(\theta)$, $i = 1, 2$, являются монотонными одного типа (т. е. одновременно либо возрастают, либо убывают), что обеспечивается условием непрерывности и монотонности функции $F_T(t;\theta)$ по θ (рис. 3).

Таким образом, при сделанных предположениях множество $\mathcal{D}_\gamma(\theta')$ при каждом θ' представляет собой интервал; следовательно, определены его концы $\theta_1(\theta') < \theta_2(\theta')$, а тем самым и соответствующий γ-доверительный интервал $(T_1(X), T_2(X))$ для θ, где $T_i(X) = \theta_i(T(X))$, $i = 1, 2$.

Построение диаграммы придает наглядный смысл этой методике. Диаграмму строят *по вертикали* (для любой абсциссы θ находят из (9) соответствующие

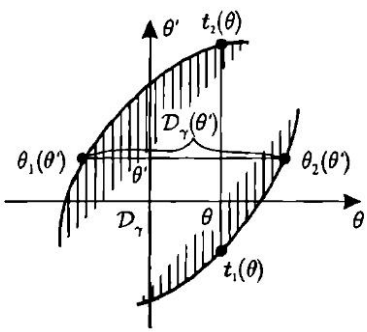


Рис. 2

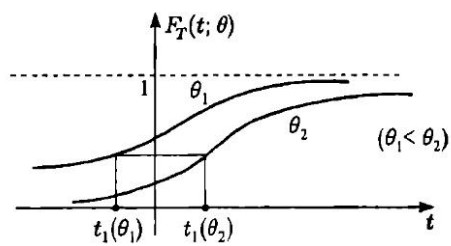


Рис. 3

ординаты t_1 и t_2); «читают» же ее *по горизонтали*, т. е. для наблюдавшейся ординаты $t = T(\underline{x})$ ($\underline{x}$ — раелизация выборки X) «считывают» две величины θ_1 и θ_2 и утверждают, что $\theta \in (\theta_1, \theta_2)$. Таким образом, если «читать» диаграмму по вертикали, то границы области $\mathcal{D}_\gamma$ описываются парой переменных точек (θ, t_1) и (θ, t_2), таких что выполняются условия (9). Если же «читать» диаграмму по горизонтали, то границы $\mathcal{D}_\gamma$ можно эквивалентно описать парой переменных точек (t, θ_1) и (t, θ_2), взяв за независимую переменную наблюдаемое значение t оценки $T(X)$. Тогда соотношения (9) можно переписать в терминах этих величин так: если функции $t_i(\theta)$ возрастают (см. рис. 2), что соответствует убыванию по θ функции $F_T(t; \theta)$, то

$$1 - F_T(t; \theta_1) = F_T(t; \theta_2) = \frac{1-\gamma}{2},$$

если же функции $t_i(\theta)$ убывают, что соответствует возрастанию по θ функции $F_T(t; \theta)$, то

$$F_T(t; \theta_1) = 1 - F_T(t; \theta_2) = \frac{1-\gamma}{2}.$$

Итак, алгоритм построения центрального γ-доверительного интервала для θ в случае, когда функция распределения $F_T(t; \theta)$ оценки $T = T(X)$ непрерывна и монотонна по θ, состоит в следующем. Пусть $t = T(\underline{x})$ — наблюдавшееся значение оценки. Решая относительно θ уравнения

$$F_T(t; \theta) = \frac{1-\gamma}{2}, \frac{1+\gamma}{2}, \tag{10}$$

находим два числа $\theta_1 < \theta_2$. Далее утверждаем, что $\theta \in (\theta_1, \theta_2)$. Приведенная теория гарантирует, что при γ, близком к 1, вероятность ошибки в случае применения этого правила равна $1 - \gamma$

Дискретная модель

Аналогичные рассуждения можно провести и для случая дискретной модели. Единственное отличие состоит в том, что из-за ступенчатости функции распределения $F_T(t; \theta)$ в соотношении (8) можно, вообще говоря, обеспечить лишь выполнение неравенства

$$\mathrm{P}_\theta\{t_1 < T(X) < t_2\} = F_T(t_2 - 0; \theta) - F_T(t_1; \theta) \geqslant \gamma, \tag{11}$$

поэтому вместо условий (9) следует ввести условия

$$F_T(t_1; \theta) \leqslant \frac{1-\gamma}{2}, \quad 1 - F_T(t_2 - 0; \theta) \leqslant \frac{1-\gamma}{2}, \tag{12}$$

где t_1 — наибольшее, а t_2 — наименьшее значения T, удовлетворяющие этим неравенствам. Кривые $\theta' = t_i(\theta)$, $i = 1, 2$, в данном случае являются ступенчатыми (рис. 4) и при «считывании» величин $\theta_i(\theta')$, $i = 1, 2$, следует брать крайнюю правую точку отрезка пересечения горизонтальной прямой с левой кривой и крайнюю левую точку — с правой кривой (когда линия пересечения попадает на горизонтальные «ступеньки» кривых).

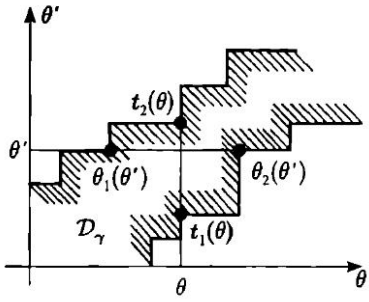


Рис. 4

Алгоритм построения центрального γ-доверительного интервала для θ в дискретном случае в основном тот же, что и для непрерывного, только вместо уравнений (10) надо решать относительно θ уравнения

$$F_T(t;\theta) = \frac{1-\gamma}{2}, \qquad 1 - F_T(t-0;\theta) = \frac{1-\gamma}{2}; \tag{13}$$

t — наблюдавшееся значение оценки T.

Замечание. Часто вместо центрального доверительного интервала в дискретном случае рекомендуется строить множество $\mathcal{D}_\gamma$ по принципу включения в него значений статистики T с наибольшими (при каждом θ) вероятностями их появления. В случае применения этого метода вместо условий (11)–(12) границы $t_i = t_i(\theta)$, $i = 1, 2$, определяются из условия

$$\mathbf{P}_\theta\{t_1 < T < t_2\} = \sum_{t_1 < t < t_2} \mathbf{P}_\theta\{T = t\} \geqslant \gamma; \tag{14}$$

при этом в сумму должны быть включены значения t с наибольшими вероятностями их появления, а интервал (t_1, t_2) должен быть наименьшим, удовлетворяющим этому условию. В ряде случаев, когда распределение статистики T асимметрично, такой способ позволяет получить более короткие доверительные интервалы.

3. Асимптотические доверительные интервалы

В ряде случаев, когда имеется состоятельная и асимптотически нормальная оценка $T_n = T_n(X)$, $X = (X_1, \ldots, X_n)$, для параметра θ, можно построить асимптотический (при больших n) доверительный интервал. С примерами построения таких интервалов мы встречались в п. 5 § 2.2, п. 4 § 2.3 и п. 5 § 3.5. В общем случае, пусть при $n \to \infty$ имеет место соотношение

$$\mathcal{L}_\theta\big(\sqrt{n}(T_n - \theta)\big) \longrightarrow \mathcal{N}(0, \sigma^2(\theta)), \quad \forall\, \theta \in \Theta,$$

причем асимптотическая дисперсия $\sigma^2(\theta)$ непрерывна по θ. Тогда в силу (17) § 3.5

$$\mathcal{L}_\theta\left(\frac{\sqrt{n}(T_n - \theta)}{\sigma(T_n)}\right) \longrightarrow \mathcal{N}(0, 1), \quad \forall\, \theta,$$

откуда следует, что при $n \to \infty$ и всех θ

$$\mathbf{P}_\theta\left\{\frac{\sqrt{n}|T_n - \theta|}{\sigma(T_n)} < c_\gamma\right\} = 2\Phi(c_\gamma) - 1 = \gamma,$$

если $c_\gamma = \Phi^{-1}((1+\gamma)/2)$. Переписав это соотношение в виде

$$\mathbf{P}_\theta\left\{T_n - \frac{c_\gamma \sigma(T_n)}{\sqrt{n}} < \theta < T_n + \frac{c_\gamma \sigma(T_n)}{\sqrt{n}}\right\} \longrightarrow \gamma,$$

получаем, что $(T_n \mp c_\gamma \sigma(T_n)/\sqrt{n})$ — асимптотический γ-доверительный интервал для θ. Это симметричный относительно центральной точки T_n интервал шириной $2c_\gamma\sigma(T_n)/\sqrt{n}$, поэтому он будет тем уже, чем меньше асимптотическая дисперсия оценки T_n, т.е. чем выше асимптотическая эффективность T_n (в смысле п. 5 § 3.2). Следовательно, асимптотически кратчайший доверительный интервал будет порождаться асимптотически эффективной оценкой. Если исходная модель $\mathcal{F} \quad \{F(x;\theta),\ \theta \in \Theta\}$ регулярна, то перечисленными свойствами обладают оценки максимального правдоподобия (см. п. 4 § 3.5). Поэтому построенные в п. 5 § 3.5 доверительные интервалы являются асимптотически оптимальными (кратчайшими).

Проиллюстрируем изложенные общие принципы примерами.

Пример 4 (Доверительный интервал для параметра бернуллиевской модели). Пусть требуется построить доверительный интервал для параметра θ модели $\mathrm{Bi}(1,\theta)$. Если имеется соответствующая выборка $X = (X_1, \quad, X_n)$, то, как мы хорошо знаем, оптимальной несмещенной оценкой для θ является выборочное среднее $T = \overline{X}$. Статистика T принимает значения вида k/n, $k = 0, 1, \quad, n$, при этом (см. п. 1 § 1.1)

$$F_T\left(\frac{k}{n};\theta\right) = \mathbf{P}_\theta\left\{T \leqslant \frac{k}{n}\right\} = \mathbf{P}_\theta\left\{\sum_{i=1}^{n} X_i \leqslant k\right\} = \sum_{r=0}^{k} C_n^r \theta^r (1-\theta)^{n-r}$$

Здесь

$$\frac{d}{d\theta} F_T\left(\frac{k}{n};\theta\right) = -n\, C_{n-1}^k \theta^k (1-\theta)^{n-k-1} < 0 \quad \text{при} \quad k < n,$$

так что при $k < n$ функция $F_T(k/n;\theta)$ монотонно убывает по θ. Следовательно, применима описанная в п. 2 методика, согласно которой при $T = k/n$ центральным γ-доверительным интервалом для θ является интервал (θ_1, θ_2), где θ_1 и θ_2 определяются в соответствии с уравнениями (13), которые в данном случае принимают вид

$$\begin{aligned} 1 - F_T\left(\frac{k-1}{n};\theta_1\right) &= \sum_{r=k}^{n} C_n^r \theta_1^r (1-\theta_1)^{n-r} = \frac{1-\gamma}{2}, \\ F_T\left(\frac{k}{n};\theta_2\right) &= \sum_{r=0}^{k} C_n^r \theta_2^r (1-\theta_2)^{n-r} = \frac{1-\gamma}{2}; \end{aligned} \tag{15}$$

при этом, если наблюдалось $T = 1$ (т.е. $k = n$), то принимают $\theta_2 = 1$, а при $T = 0$ (т.е. $k = 0$) полагают $\theta_1 = 0$.

Практически границы θ_1 и θ_2 вычисляют, используя их связь с квантилями бета-распределения (см. п. 4 § 1.2). Именно, по формулам (4)–(4′) § 2.4:

$$\sum_{r=k}^{n} C_n^r \theta_1^r (1-\theta_1)^{n-r} = B(\theta_1; k, n-k+1),$$

$$\sum_{r=0}^{k} C_n^r \theta_2^r (1-\theta_2)^{n-r} = 1 - \sum_{r=k+1}^{n} C_n^r \theta_2^r (1-\theta_2)^{n-r} = 1 - B(\theta_2; k+1, n-k).$$

Поэтому уравнения (15) можно переписать в виде

$$B(\theta_1; k, n-k+1) = \frac{1-\gamma}{2}, \quad B(\theta_2; k+1, n-k) = \frac{1+\gamma}{2} \tag{16}$$

Доверительный интервал для параметра модели Bi(1, θ)

Следовательно, если обозначить $z(p; a, b)$ p-квантиль бета-распределения Be(a, b), т. е. решение уравнения $B(z; a, b) = p$, $0 < p < 1$, то из (16) имеем (заменяя k на nT)

$$\theta_1 = z\left(\frac{1-\gamma}{2}; nT, n-nT+1\right), \quad \theta_2 = z\left(\frac{1+\gamma}{2}; nT+1, n-nT\right) \tag{17}$$

Доверительные интервалы (θ_1, θ_2) рассчитаны для широкого диапазона значений n при $\gamma = 0{,}9;\ 0{,}95;\ 0{,}99$ (см. [3]).

Если число наблюдений n велико, то для быстрого нахождения приближенного доверительного интервала для θ можно воспользоваться асимптотической теорией, изложенной в примере 17 в § 3.5. •

Пример 5 (Доверительный интервал для параметра пуассоновской модели). Как и в предыдущем примере, здесь оптимальной несмещенной оценкой для θ является статистика $T = \overline{X}$ (см. пример 1 в § 3.4), принимающая значения k/n, $k = 0, 1, 2, \ldots$; при этом (см. п. 3 § 1.1) $\mathcal{L}_\theta(nT) = \Pi(n\theta)$, поэтому

$$F_T\left(\frac{k}{n}; \theta\right) = \mathbf{P}_\theta\left\{T \leqslant \frac{k}{n}\right\} = \sum_{r=0}^{k} e^{-n\theta} \frac{(n\theta)^r}{r!}.$$

Здесь

$$\frac{d}{d\theta} F_T\left(\frac{k}{n}; \theta\right) = -ne^{-n\theta} \frac{(n\theta)^k}{k!} < 0,$$

т. е. функция $F_T(k/n; \theta)$ монотонно убывает по θ. Согласно общей методике, если наблюдалось $T = k/n$, то центральным γ-доверительным интервалом для θ является интервал (θ_1, θ_2), где θ_1 и θ_2 определяются соответственно уравнениями

$$\begin{aligned} 1 - F_T\left(\frac{k-1}{n}; \theta_1\right) &= \sum_{r=k}^{\infty} e^{-n\theta_1} \frac{(n\theta_1)^r}{r!} = \frac{1-\gamma}{2}, \\ F_T\left(\frac{k}{n}; \theta_2\right) &= \sum_{r=0}^{k} e^{-n\theta_2} \frac{(n\theta_2)^r}{r!} = \frac{1-\gamma}{2}. \end{aligned} \tag{18}$$

В данном случае границы θ_1 и θ_2 можно выразить через квантили хи-квадрат распределения. Для этого получим сначала вид функции распределения $G_n(x)$

закона $\chi^2(n)$ (см. п. 3 § 1.2) при $n = 2k$. Обозначая для краткости $\pi_r(t) = e^{-t}t^r/r!$, $r = 0, 1, \ldots$, имеем

$$1 - G_{2k}(2\lambda) = \int_{2\lambda}^{\infty} \frac{x^{k-1}e^{-x/2}}{(k-1)!2^k}\,dx = \int_{\lambda}^{\infty} \pi_{k-1}(t)\,dt = \sum_{r=0}^{k-1} \pi_r(\lambda)$$

(интеграл легко вычисляется интегрированием по частям). Отсюда находим, что

$$G_{2k}(2\lambda) = 1 - \sum_{r=0}^{k-1} \pi_r(\lambda) = \sum_{r=k}^{\infty} \pi_r(\lambda). \tag{19}$$

С учетом (19) уравнения (18) можно переписать теперь в виде

$$G_{2k}(2n\theta_1) = \frac{1-\gamma}{2}, \quad G_{2k+2}(2n\theta_2) = \frac{1+\gamma}{2}. \tag{20}$$

Доверительный интервал для параметра модели $\Pi(\theta)$

Следовательно, если $\chi^2_{p,n}$ обозначает p-квантиль распределения $\chi^2(n)$, т. е. решение уравнения $G_n(x) = p$, то из (20) имеем (заменяя k на nT)

$$\theta_1 = \frac{1}{2n}\chi^2_{(1-\gamma)/2,\,2nT}, \quad \theta_2 = \frac{1}{2n}\chi^2_{(1+\gamma)/2,\,2nT+2} \tag{21}$$

Как и в предыдущем примере, для больших объемов выборки можно просто рассчитать приближенный доверительный интервал, воспользовавшись асимптотической теорией оценок максимального правдоподобия, что проделано уже в примере 15 § 3.5.

В данной модели можно построить асимптотический доверительный интервал для θ и другим способом, а именно, воспользовавшись нормальной аппроксимацией для выборочного среднего: $\mathcal{L}_\theta(\overline{X}) \sim \mathcal{N}(\theta, \theta/n)$ (см. теорему 3 § 2.2 с учетом того, что для пуассоновского распределения $\Pi(\theta)$ среднее и дисперсия равны θ, см. п. 3 § 1.1). Отсюда можем записать, что при больших n

$$\gamma \approx \mathbf{P}_\theta\left\{\frac{\sqrt{n}|\overline{X} - \theta|}{\sqrt{\theta}} < c_\gamma\right\} = \mathbf{P}_\theta\left\{(\overline{X} - \theta)^2 < \frac{c_\gamma^2\theta}{n}\right\} =$$

$$= \mathbf{P}_\theta\left\{\theta^2 - 2\theta\left(\overline{X} + \frac{c_\gamma^2}{2n}\right) + \overline{X}^2 < 0\right\} = \mathbf{P}_\theta\{T_1 < \theta < T_2\},$$

где

$$T_{1,2} = \overline{X} + \frac{c_\gamma^2}{2n} \mp \left(\overline{X}\frac{c_\gamma^2}{n} + \frac{c_\gamma^4}{4n^2}\right)^{1/2}$$

Таким образом, (T_1, T_2) — другой приближенный γ-доверительный интервал для θ. С точностью до членов порядка $1/n$ этот интервал эквивалентен интервалу $(\overline{X} \mp c_\gamma\sqrt{\overline{X}/n})$, полученному в примере 15 в § 3.5. •

Пример 6 (Доверительное оценивание квантилей). Пусть по выборке $X = (X_1, \quad , X_n)$ из непрерывного распределения с неизвестной функцией распределения $F(x)$ требуется построить доверительный интервал для p-квантили ζ_p при заданном $p \in (0, 1)$ (см. п. 1 § 2.4). Точные доверительные интервалы в этой задаче можно строить, основываясь на формуле (5) § 2.4. Именно, подставляя в эту формулу $x = \zeta_p$ и учитывая, что $F(\zeta_p) = p$, получим

$$\mathbf{P}\{X_{(k)} \leqslant \zeta_p\} = B(p; k, n-k+1) \equiv \gamma_1,$$

т. е. порядковая статистика $X_{(k)}$ является нижней γ_1-доверительной границей для ζ_p. Аналогично,

$$\mathbf{P}\{X_{(l)} > \zeta_p\} = 1 - B(p; l, n-l+1) \equiv \gamma_2,$$

т. е. $X_{(l)}$ — верхняя γ_2-доверительная граница для ζ_p. Наконец, так как $X_{(k)} < X_{(l)}$ при $k < l$, то

$$\begin{aligned}\mathbf{P}\{X_{(k)} < \zeta_p < X_{(l)}\} &= 1 - \mathbf{P}\{\zeta_p \leqslant X_{(k)}\} - \mathbf{P}\{X_{(l)} \leqslant \zeta_p\} = \\ &= B(p; k, n-k+1) - B(p; l, n-l+1) = \gamma_1 + \gamma_2 - 1 \equiv \gamma\end{aligned}$$

Это означает, что $(X_{(k)}, X_{(l)})$ — γ-доверительный интервал для ζ_p. •

§ 3.9. Оценивание при выборе из конечной совокупности

1. Оценивание среднего совокупности

В прикладных исследованиях часто приходится иметь дело со следующей ситуацией. Имеется конечная совокупность $U = \{u_1, \quad , u_n\}$ из N объектов, каждый из которых характеризуется некоторой величиной $x(u)$, $u \in U$. Значения $x_i = x(u_i)$, $i = 1, \quad , N$, неизвестны, и требуется оценить их сумму

$$T = T(\underline{x}) = \sum_{i=1}^{N} x_i, \quad \underline{x} = (x_1, \quad , x_N), \tag{1}$$

или, что то же, величину $\mu = T(\underline{x})/N$, называемую *средним совокупности* U (можно рассматривать и другие функции от $\underline{x}$, подлежащие оценке).

При этом предполагается, что в силу объективных причин (большой размер совокупности или большая стоимость наблюдения и т. п.) мы не можем наблюдать все объекты с их соответствующими x-значениями, а обследованию доступна лишь некоторая часть из них (выборка), выбираемая по некоторому случайному закону. Математическая формализация сказанного состоит в следующем.

Назовем выборкой s любую конечную упорядоченную последовательность объектов из U, т. е.

$$s = (u_{i_1}, \quad , u_{i_{n(s)}}), \quad i \in \{1, \quad , N\}.$$

Здесь $n(s)$ есть объем выборки s, который может быть как фиксированным, так и меняться от выборки к выборке. В первом случае говорят о вы-

борках фиксированного объема n. Если все элементы выборки различны, то говорят о *выборе без возвращения* (в этом случае $n(s) \leqslant N$), в противном случае (допускаются повторения объектов в выборке) — о *выборе с возвращением*. Множество всех выборок обозначается через S и называется *выборочным пространством*. Если на S задана вероятностная мера $\mathbf{P}$, приписывающая каждой выборке $s \in S$ вероятность $p(s)$ ее извлечения $\left(\sum_{s \in S} p(s) = 1\right)$, то говорят, что задан *выборочный план* (в. п.), подразумевая под этим тройку $\mathcal{A} = (U, S, \mathbf{P})$. Итак, мы можем наблюдать лишь некоторую совокупность объектов из U, выбираемую в соответствии с заданным в. п. $\mathcal{A}$.

Выборочный план

Далее, мы рассматриваем $\underline{x} = (x_1, \quad , x_N)$ как параметрическую точку в R^N и определяем оценку для $T(\underline{x})$ как функцию $e(s, \underline{x})$ на $S \times R^N$, которая зависит от $\underline{x}$ только через те x_i, для которых $u_i \in s$ (таким образом, $e(s, \underline{x}) = e(s, \underline{x}')$, если $x_i = x_i'$ для всех $u_i \in s$). Другими словами, оценка есть функция выбранных объектов и их наблюденных x-значений. Мы будем рассматривать класс $\mathcal{E} = \mathcal{E}(\mathcal{A})$ всех несмещенных оценок:

$$\mathcal{E} = \left\{ e(s, \underline{x}) \quad \sum_{s \in S} p(s) e(s, \underline{x}) = T(\underline{x}), \ \forall \underline{x} \in R^N \right\}, \tag{2}$$

и искать в нем наилучшую (*оптимальную*) оценку, т. е. оценку, для которой дисперсия

$$\mathbf{D}e(s, \underline{x}) = \sum_{s \in S} p(s) \left[e(s, \underline{x}) - T(\underline{x}) \right]^2$$

минимальна для $\forall \underline{x} \in R^N$. Оптимальная оценка, вообще говоря, не всегда существует, и в таких случаях для сравнения различных оценок используется понятие допустимой оценки. По определению, оценка $e(s, \underline{x})$ (для $T(\underline{x})$) называется *допустимой* (в классе $\mathcal{E}(\mathcal{A})$), если не существует другой оценки $e_1(s, \underline{x}) \in \mathcal{E}(\mathcal{A})$, для которой

Допустимые оценки

$$\mathbf{D}e_1(s, \underline{x}) \leqslant \mathbf{D}e(s, \underline{x}), \quad \forall \underline{x} \in R^N$$

со строгим неравенством хотя бы в одной точке $\underline{x}$.

Ясно, что недопустимые оценки должны быть отвергнуты. С другой стороны, обычно класс допустимых оценок является весьма широким и никакие две допустимые оценки e_1 и e_2 уже несравнимы (по значению дисперсии), так как для них существуют непустые множества $X_1, X_2 \subset R^N$ с условием

$$\mathbf{D}e_1 < \mathbf{D}e_2, \quad \underline{x} \in X_1, \qquad \text{и} \qquad \mathbf{D}e_2 < \mathbf{D}e_1, \quad \underline{x} \in X_2.$$

Поэтому выбор оценки из класса допустимых должен проводиться с учетом дополнительных соображений.

Оценка Горвица—Томпсона

Наиболее употребительными оценками для $T(\underline{x})$ являются *оценки Горвица—Томпсона* $\bar{e}(s, \underline{x})$, которые определяются следующим образом. Обозначим

$$\pi(u) = \mathbf{P}(u \in s) = \sum_{s\,:\,s \ni u} p(s) \tag{3}$$

вероятность включения в выборку объекта $u \in U$ (сумму вероятностей всех выборок, содержащих u), и предположим, что в. п. $\mathcal{A}$ таков, что $\pi(u) > 0$, $\forall\, u \in U$. Тогда, по определению,

$$\bar{e}(s, \underline{x}) = \sum_{u \in s} \frac{x(u)}{\pi(u)} = \sum_{u} I(u \in s)\frac{x(u)}{\pi(u)}. \tag{4}$$

Так как

$$\mathbf{E}I(u \in s) = \mathbf{P}(u \in s) = \pi(u),$$

то

$$\mathbf{E}\bar{e}(s, \underline{x}) = \sum_{u} x(u) = T(\underline{x}), \quad \text{т. е.} \quad \bar{e}(s, \underline{x}) \in \mathcal{E}.$$

Нетрудно заметить, что оценка Горвица—Томпсона является *единственной линейной несмещенной оценкой* для T, т. е. оценкой вида

$$\sum_{u \in s} a(u)x(u) = \sum_{u} I(u \in s)a(u)x(u).$$

Действительно, условие несмещенности означает, что

$$\sum_{u} x(u)a(u)\pi(u) = \sum_{u} x(u), \quad \forall\, \underline{x} \in R^N$$

Полагая здесь $\underline{x} = (0 \ldots 010 \ldots 0)$, где 1 стоит на i-м месте, имеем

$$a(u_i)\pi(u_i) = 1, \quad \text{т. е.} \quad a(u_i) = \frac{1}{\pi(u_i)}, \quad i = 1, \quad , N.$$

Замечание. Если рассматривать класс всех несмещенных оценок $\mathcal{E}(\mathcal{A})$ для в. п. $\mathcal{A}$ с условием $\pi(u) > 0$, $\forall\, u \in U$, то можно лишь доказать, что $\bar{e}(s, \underline{x})$ является допустимой оценкой.

Выбор без возвращения

На практике чаще всего используются *выборочные планы фиксированного объема* и для *выбора без возвращения с равномерной мерой* на S (*выборочный план* $\mathcal{A}^*$). В этом случае множество S состоит из $(N)_n = N(N-1) \quad (N-n+1)$ всех упорядоченных комбинаций длины n из различных элементов U и $p(s) = 1/(N)_n$, $\forall\, s \in S$. Нетрудно видеть, что для любого фиксированного элемента u существует $n(N-1)_{n-1}$ различных выборок из S, содержащих этот элемент, поэтому

$$\pi(u) = \frac{n(N-1)_{n-1}}{(N)_n} = \frac{n}{N}.$$

Следовательно, для в. п. $\mathcal{A}^*$ оценка Горвица—Томпсона принимает вид

$$\bar{e}(s, \underline{x}) = \frac{N}{n} \sum_{u \in s} x(u), \tag{5}$$

т. е. $\frac{1}{N}\bar{e}(s, \underline{x})$ есть обычное выборочное среднее $\bar{x}$ наблюденных x-значений. Из предыдущего следует, что для в. п. $\mathcal{A}^*$ $\bar{x}$ *есть оптимальная линейная оценка для среднего совокупности* μ. Вычислим дисперсию этой оценки. Так как

$$\bar{x}^2 = \frac{1}{n^2}\left(\sum_{i=1}^{N} I(u_i \in s)x_i\right)^2 = \frac{1}{n^2}\sum_{i \neq j} I\big((u_i, u_j) \in s\big)\, x_i x_j + \frac{1}{n^2}\sum_{i=1}^{N} I(u_i \in s)x_i^2$$

и

$$\mathrm{E}I\big((u_i, u_j) \in s\big) = \mathrm{P}\big((u_i, u_j) \in s\big) = \frac{n(n-1)(N-2)_{n-2}}{(N)_n} = \frac{n(n-1)}{N(N-1)}, \quad i \neq j,$$

то

$$\mathrm{D}\bar{x} = \mathrm{E}\bar{x}^2 - \mu^2 = \frac{n-1}{nN(N-1)}\sum_{i \neq j} x_i x_j + \frac{1}{nN}\sum_{i=1}^{N} x_i^2 - \mu^2,$$

что легко приводится к виду

$$\mathrm{D}\bar{x} = \left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2, \quad \text{где} \quad \sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{6}$$

Величина σ^2 называется *дисперсией совокупности* U.

Иногда требуется оценить не только среднее μ, но и дисперсию σ^2 совокупности U. По аналогии со случаем бесконечной совокупности рассмотрим в качестве оценки для σ^2 величину

Оценка дисперсии совокупности

$$\hat{\sigma}^2 = \hat{\sigma}^2(s, \underline{x}) = \frac{1}{n-1}\sum_{u \in s}(x(u) - \bar{x})^2 \tag{7}$$

и найдем ее математическое ожидание. Перепишем $\hat{\sigma}^2$ в виде

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{N} I(u_i \in s)\big((x_i - \mu) - (\bar{x} - \mu)\big)^2 =$$

$$= \frac{1}{n-1}\sum_{i=1}^{N} I(u_i \in s)(x_i - \mu)^2 - \frac{n}{n-1}(\bar{x} - \mu)^2$$

Тогда с учетом предыдущих соотношений имеем

$$\mathrm{E}\hat{\sigma}^2 = \frac{n}{(n-1)N}\sum_{i=1}^{N}(x_i - \mu)^2 - \frac{n}{n-1}\mathrm{D}\bar{x} = \sigma^2$$

Таким образом, $\hat{\sigma}^2$ есть несмещенная оценка для σ^2.

Замечание. Справедлив более сильный результат: $\widehat{\sigma}^2$ есть оптимальная оценка σ^2 в классе всех несмещенных квадратичных оценок, т. е. оценок вида

$$\sum_{u,\,v\in s} a(u,v)(x(u)-\bar{x})(x(v)-\bar{x}).$$

2. Оценивание состава совокупности

Часто составляющие совокупность U объекты классифицируются по некоторому признаку (например, полу, возрасту, если U — биологическая популяция, уровню качества, если U — партия промышленных изделий и т. д.) на заданное число k классов (слоев) $U_1, \ldots, U_k$ ($U = U_1 \cup \ldots \cup U_k$, $U_i \cap U_j = \varnothing$, $i \neq j$) — в таких случаях говорят о *расслоенной совокупности*. Размеры классов $N_1, \ldots, N_k$ (состав совокупности) являются неизвестными параметрами, которые должны быть оценены по результатам соответствующего выборочного эксперимента. Что касается размера $N = N_1 + \ldots + N_k$ всей совокупности U, то в одних случаях он может быть известен, в других же также представляет собой неизвестный параметр. Например, в задачах статистического (выборочного) контроля качества продукции U представляет собой партию некоторых изделий, каждое из которых может классифицироваться по альтернативному признаку «исправно-дефектно»; в этом случае число классов $k = 2$, размер N всей партии известен и, следовательно, речь идет об оценивании размера лишь одного класса, например, числа дефектных изделий в партии. При изучении же биологических популяций (например, популяции рыб в некотором водоеме), в социологических исследованиях, анализе текстов в лингвистике и т. д., как правило, общий размер совокупности неизвестен и также подлежит оцениванию по результатам выборочного обследования. Задачи, связанные с оцениванием состава конечных совокупностей, — обширная и многообразная область прикладных статистических исследований, достаточно полно отраженная в обзоре Ивченко и Хонова[7] Здесь же мы в качестве иллюстрации лишь кратко коснемся некоторых из этого круга вопросов.

Расслоенная совокупность

Выборочный контроль

1) Рассмотрим сначала *задачу выборочного контроля*. Пусть имеется партия из N изделий, содержащая некоторое (неизвестное) число D дефектных изделий, $D \in \{0, 1, \ldots, N\}$. Чтобы оценить параметр D или, более общо, некоторую заданную функцию от него $\tau(D)$, случайным образом без возвращения из всей партии извлекается $n (< N)$ изделий, каждое из которых проверяется на доброкачественность. Пусть $X_i = 1$, если i-е проверяемое изделие дефектно, и $X_i = 0$ в противном случае, $i = 1, \ldots, n$. Положим $d = X_1 + \ldots + X_n$ — общее число обнаруженных в выборке $X = (X_1, \ldots, X_n)$ дефектных изделий. Тогда d — *полная достаточная статистика*.

[7] *Ивченко Г. И. Хонов С. А.* Статистическое оценивание состава конечной совокупности // Дискретная математика. 1996. Т. 8. № 1. С. 3–40.

Доказательство. Вычислим функцию правдоподобия

$$\mathbf{P}_D(X = \underline{x}), \quad \underline{x} = (x_1, \ldots, x_n), \quad x_\cdot = 0, 1,$$

наших данных. По формуле умножения вероятностей можем записать

$$\mathbf{P}_D(X = \underline{x}) = \mathbf{P}_D(X_1 = x_1)\mathbf{P}_D(X_2 = x_2 | X_1 = x_1) \quad \times$$
$$\times \mathbf{P}_D\big(X_n = x_n | X_i = x_i,\ i = 1, \quad , n-1\big).$$

Легко видеть, что здесь

$$\mathbf{P}_D\big(X_j = x_j | X_i = x_i,\ i = 1, \quad , j-1\big) =$$
$$= \left\{ \begin{array}{ll} \dfrac{D - x_1 - \quad - x_{j-1}}{N - j + 1} & \text{при} \quad x_j = 1, \\ 1 - \dfrac{D - x_1 - \quad - x_{j-1}}{N - j + 1} & \text{при} \quad x_j = 0 \end{array} \right\} =$$
$$= \frac{(D - x_1 - \quad - x_{j-1})^{x_j}}{N - j + 1}(N - D - j + 1 + x_1 + \quad + x_{j-1})^{1-x_j}$$

Отсюда

$$\mathbf{P}_D(X = \underline{x}) = D^{x_1}(D - x_1)^{x_2} \quad (D - x_1 - \quad - x_{n-1})^{x_n}(N - D)^{1-x_1} \times$$
$$\times \frac{\big(N - D - (1 - x_1)\big)^{1-x_2} \quad \big(N - D - (n - 1 - x_1 - \quad - x_{n-1})\big)^{1-x_n}}{(N)_n}.$$

Наконец, воспользовавшись формулой

$$l^{\varepsilon_1}(l - \varepsilon_1)^{\varepsilon_2} \quad (l - \varepsilon_1 - \quad - \varepsilon_{n-1})^{\varepsilon_n} = \frac{l!}{(l - \varepsilon_1 - \quad - \varepsilon_n)!}, \quad \varepsilon_\cdot = 0, 1,$$

которая легко устанавливается, например, индукцией по n, окончательно находим

$$\mathbf{P}_D(X = \underline{x}) = \frac{D!(N - D)!(N - n)!}{(D - d)!(N - D - n + d)!N!} = \frac{C_{N-n}^{D-d}}{C_N^D}. \tag{8}$$

Согласно критерию факторизации (1) § 3.3, это означает, что d — достаточная статистика для параметра D.

Чтобы убедиться в ее полноте, требуется доказать, что если $\mathbf{E}_D\varphi(d) = 0$ при $D = 0, 1, \quad , N$, то $\varphi(k) = 0$ при $k = 0, 1, \quad . \quad \min(n, D) \equiv n \wedge D$ при всех возможных D, т. е. при $k = 0, 1, \quad , n$. Но статистика d имеет известное гипергеометрическое распределение (см. п. 4 § 1.1):

$$\mathbf{P}_D(d = k) \equiv f(k; D) = \frac{C_D^k C_{N-D}^{n-k}}{C_N^n}, \quad k = 0, 1, \quad , n \wedge D, \tag{9}$$

поэтому предыдущее условие имеет вид

$$\sum_{k=0}^{n \wedge D} \varphi(k) f(k; D) = 0, \quad D = 0, 1, \ldots, N.$$

Полагая здесь последовательно $D=0$, $D=1$ и т.д., получим

$$\varphi(k)=0, \quad k=0,1, \quad ,n. \ \blacksquare$$

Пусть теперь $\tau(D)$ — подлежащая оцениванию функция параметра D, тогда оптимальная несмещенная оценка для нее — это решение уравнения несмещенности

$$\mathbf{E}_D T(d)=\tau(D), \quad D=0,1, \quad ,N$$

Но при любой функции T левая часть здесь представляет собой полином от D степени $\leqslant n$, поэтому, если $\tau(D)$ не есть полином степени $\leqslant n$, то это уравнение решения не имеет, и, следовательно, для таких функций $\tau(D)$ несмещенных оценок не существует. Если же $\tau_j(D)=(D)_j$, $j\leqslant n$, то воспользовавшись формулой для факториальных моментов гипергеометрического распределения: $\mathbf{E}_D(d)_j=(D)_j(n)_j/(N)_j$, сразу же получаем вид оптимальной оценки факториальной степени $(D)_j$: $\tau_j^*=(d)_j(N)_j/(n)_j$. Оценки для произвольных полиномов

$$\tau(D)=\sum_{j=0}^{n} a_j(D)_j$$

можно строить с помощью линейных комбинаций статистик τ_j^* на основании теоремы 3 § 3.1. В частности, $\tau_1^*=dN/n$ — оптимальная оценка самого параметра D, а $\tau^*=d(n-d)(N)_2/(n)_2$ — оптимальная оценка функции $\tau(D)=D(N-D)$ при $n\geqslant 2$.

Для сравнения приведем вид оценки максимального правдоподобия параметра D, которая получается максимилизацией по D правой части (8): $\widehat{D}=[d(N+1)/n]$ (здесь $[a]$ — целая часть числа a).

2) *Задача выборочного контроля* (*продолжение*). В некоторых системах статистического контроля качества продукции поступают следующим образом. Если число d обнаруженных в выборке дефектных изделий удовлетворяет неравенству $d\leqslant k_0$, где k_0 — задаваемый заранее некоторый уровень ($k_0<n$), то они заменяются на исправные, после чего вся партия принимается. Если же $d>k_0$, то контролю подвергаются все N изделий, и все дефектные изделия заменяются на исправные. Пусть случайная величина ξ есть число обнаруженных при описанном способе действия дефектных изделий (подчеркнем, что это более «богатая» статистика, чем d). Тогда для любой (!) параметрической функции $\tau(D)$ можно построить несмещенную оценку вида $T(\xi)$, т.е. являющуюся функцией от ξ, причем такая оценка единственна.

Доказательство. Так как распределение статистики ξ имеет, очевидно, вид (см. (9))

$$\mathbf{P}_D(\xi=k)=f(k;D), \quad k=0,1,\ldots,k_0,$$

$$\mathbf{P}_D(\xi=D)=\sum_{k=k_0+1}^{n} f(k;D)\equiv F(k_0;D) \quad (\text{при } D>k_0),$$

то условие несмещенности в данном случае выглядит так:

$$\mathbf{E}_D T(\xi) = \sum_{k=0}^{k_0} T(k) f(k; D) + T(D) F(k_0; D) = \tau(D), \quad D = 0, 1, \ldots, N.$$

Но при $D \leqslant k_0$ вероятности $f(k; D) = 0$ для $k > D$, поэтому отсюда получаем систему уравнений

$$\sum_{k=0}^{D} T(k) f(k; D) = \tau(D), \quad D = 0, 1, \ldots, k_0.$$

Это треугольная система, в которой диагональные коэффициенты $f(D; D) \neq 0$. Следовательно, при любой правой части значения $T(k)$, $k = 0, 1, \ldots, k_0$, отсюда однозначно определяются. Для значений же $m > k_0$ полагаем

$$T(m) = \frac{\tau(m) - \sum_{k=0}^{k_0} T(k) f(k; m)}{F(k_0; m)},$$

что возможно, так как $F(k_0; m) \neq 0$ при $m > k_0$. ■

Если $\tau(D) = D$, то (так как $\mathbf{E}_D d = Dn/N$) функция T имеет вид:

$$T(k) = \frac{kN}{n} \quad \text{при} \quad k = 0, 1, \ldots, k_0 \quad \text{и}$$

$$T(m) = \frac{m - \frac{N}{n} \sum_{k=0}^{k_0} k f(k; m)}{F(k_0; m)} \quad \text{при} \quad m > k_0. \tag{10}$$

В частности, если положить здесь $k_0 = n$ (контроль всей партии не предусматривается), то это решение сводится к полученному в предыдущей задаче: $T(\xi) = \tau_1^*$.

Рассмотренный пример иллюстрирует тот факт, что наличие дополнительной информации позволяет получать более сильные статистические выводы.

3) *Схема повторного выбора.* Пусть из совокупности U, размер которой N и размеры ее классов $N_1, \ldots, N_k$ неизвестны, по схеме случайного выбора с возвращением извлекается n объектов и при этом фиксируется, к какому классу принадлежит наблюдавшийся объект. В результате подсчитывается число μ_{jr} объектов класса U_j, появившихся в выборке по r раз каждый, и тем самым исходная статистическая информация об исследуемой совокупности представляется набором данных $\{\mu_{jr},\ j = 1, \ldots, k;\ r = 1, \ldots, n\}$. По этой информации требуется оценить неизвестный параметрический вектор $\underline{N} = (N_1, \ldots, N_k)$ и, вообще говоря, произвольную функцию $\tau(\underline{N})$ от него. Показано[8], что в дан-

Повторный выбор

[8] *Ивченко Г. И.* О некоторых задачах статистического вывода для расслоенных совокупностей // Теория вероятностей и ее применения. 1984. Т. 29. № 1. С. 113–118.

ной схеме полной достаточной статистикой является $\underline{\eta} = (\eta_1, \ldots, \eta_k)$, где

$$\eta_j = \sum_{r=1}^{n} \mu_{jr}$$

— число различных наблюдавшихся объектов класса U_j, $j = 1, \ldots, k$. Таким образом, задача построения оптимальной несмещенной оценки для $\tau(\underline{N})$ сводится к выяснению условий разрешимости и отысканию при этих условиях явного вида решения уравнения несмещенности

$$\mathbf{E}_{\underline{N}} T(\underline{\eta}) = \tau(\underline{N}), \quad \forall\, \underline{N} \neq \underline{0}. \tag{11}$$

Обозначим через A_j оператор сдвига на единицу j-й переменной:

$$A_j f(\underline{x}) = f(\underline{x} + \underline{e}_j), \quad \underline{x} = (x_1, \ldots, x_k), \quad \underline{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0),$$

где 1 стоит на j-м месте, через I — тождественный оператор: $If(\underline{x}) = f(\underline{x})$ и положим $\Delta_j = A_j - I$ (Δ_j называется *оператором разности по j-й переменной*). Тогда, если возможные значения вектора $\underline{N}$ ограничены условием $N = N_1 + \ldots + N_k \leqslant n$, уравнение (11) разрешимо для любой функции $\tau(\underline{N})$ и при этом оптимальной оценкой для $\tau(\underline{N})$ является статистика

$$\tau^* = \frac{\left(\Delta_1^{\eta_1} \ldots \Delta_k^{\eta_k} - (-1)^{\eta} I\right) f(\underline{0})}{\Delta^{\eta} 0^n}, \tag{12}$$

где

$$\eta = \eta_1 + \ldots + \eta_k, \quad f(\underline{x}) = (x_1 + \ldots + x_k)^n \tau(\underline{x})$$

и

$$\Delta^r t^n = \sum_{j=0}^{r} (-1)^{r-j} C_r^j (t + j)^n$$

Оптимальная оценка размера совокупности

В частности, оптимальная оценка для размера N всей совокупности имеет вид

$$N^* = \frac{\Delta^{\eta} 0^{n+1}}{\Delta^{\eta} 0^n}. \tag{13}$$

Если же $\underline{N} \neq \underline{0}$ априори может быть любым целочисленным вектором, то уравнение (11) разрешимо лишь для функций вида $\tau(\underline{N}) = f(\underline{N})/N^n$ где $f(\underline{x})$ — полином степени $\leqslant n$, удовлетворяющий условию $f(\underline{0}) = 0$. Для любой такой функции оптимальная оценка имеет вид (12). В частности, несмещенные оценки существуют лишь для полиномов степени $\leqslant n$ от долей $\beta_j = N_j/N$, $j = 1, \ldots, k$, а оптимальной оценкой для величины $1/N$ (которая всегда существует при $n \geqslant 2$) является статистика

$$\left(\frac{1}{N}\right)^* = \frac{\Delta^{\eta} 0^{n-1}}{\Delta^{\eta} 0^n}. \tag{14}$$

Упражнения

Я занимался до сих пор решением ряда задач, ибо при изучении наук примеры полезнее правил.

Исаак Ньютон

1 Пусть в урне находится a шаров, из которых a_1 — белые и $a_2 = a - a_1$ — черные. Общее число шаров a известно, а неизвестным и подлежащим оцениванию является число a_1 белых шаров или, что то же самое, их доля $p = a_1/a$. Опыт заключается в извлечении без возвращения n, шаров $n < a$, причем предполагается, что любая из C_a^n комбинаций шаров может быть извлечена с равной вероятностью. Пусть X обозначает число белых шаров в выборке и $\widehat{p}_n = X/n$. Найти $\mathbf{E}\,\widehat{p}_n$ и $\mathbf{D}\,\widehat{p}_n$, и обосновать использование случайной величины $\widehat{p}_n$ в качестве оценки для p. Сравнить с результатом примера 6 в § 3.1.

◄ *Указание.* Использовать тот факт, что X имеет гипергеометрическое распределение $H(a_1, a_2, n)$ (см. п. 4 § 1.1). ►

2 (*Продолжение*). Рассматривается модель Маркова—Пойа извлечения шаров из урны (см. п. 5 § 1.1). Пусть, по-прежнему, X есть число белых шаров в выборке объема n, т. е. $\mathcal{L}(X) = \mathrm{МП}(n; a_1, a_2, c)$, и пусть $c > 0$. Убедиться в том, что доля белых шаров в выборке $\widehat{p}_n = X/n$ является несмещенной, но не состоятельной оценкой для $p = a_1/a$.

◄ *Указание.* Воспользоваться формулами (13) § 1.1. ►

3 Рассматривается ситуация, описанная в п. 7 § 1.1 (многомерная модель Маркова—Пойа для урны с N цветами шаров). Используя формулы (19) § 1.1, убедиться в том, что вектор относительных наблюдавшихся частот цветов в n извлечениях $\underline{\nu}/n = (\nu_1/n, \ldots, \nu_N/n)$ является несмещенной оценкой для вектора $\underline{p} = (a_1/a, \ldots, a_N/a)$ долей цветов в урне в стартовый момент.

4 Пусть $X = (X_1, \quad, X_n)$ — выборка из некоторого распределения $\mathcal{L}(\xi)$, для которого среднее $\alpha_1 = \mathbf{E}\,\xi$ известно. Требуется оценить неизвестную дисперсию $\mu_2 = \mathbf{D}\xi$. Убедиться в том, что статистика $T_n(X) = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \alpha_1)^2$ является несмещенной и состоятельной оценкой μ_2, если $\mathbf{E}\,\xi^4 < \infty$.

5 В каких случаях статистика $T_n(X) = \sqrt{\widehat{\alpha}_{n2}/2}$ является состоятельной оценкой теоретического среднего α_1? Убедиться в том, что это имеет место, в частности, для моделей $\Gamma(\theta, 1)$ и $\mathcal{N}(\theta, \theta^2)$, $\theta > 0$.

◄ *Указание.* Воспользоваться состоятельностью выборочного момента $\widehat{\alpha}_{n2}$. ►

6 По n независимым наблюдениям $X_1, \ldots, X_n$ над случайной величиной ξ требуется оценить неизвестную вероятность $p = \mathrm{P}\{\xi \in \Delta\}$, где интервал Δ задан. Положим

$$\nu_n = \sum_{i=1}^{n} I(X_i \in \Delta)$$

— число попаданий в интервал Δ. Показать, что статистика $T_n = \nu_n/n$ — несмещенная и состоятельная оценка p. При каком значении p дисперсия $\mathbf{D}T_n$ будет максимальной?

7 Пусть случайная величина ξ принимает значения $1, 2, \ldots, N$ с некоторыми (неизвестными) вероятностями $p_1, \ldots, p_N$ соответственно. Над ξ произведено n независимых наблюдений и пусть ν_r означает число тех из них, когда наблюдался исход «r» ($r = 1, \ldots, N$, $\nu_1 + \quad + \nu_N = n$). Доказать, что относительная частота $\nu_r^* = \nu_r/n$ является несмещенной и состоятельной оценкой для вероятности p_r.

◀ *Указание.* Воспользоваться тем, что случайная величина ν_r имеет биномиальное распределение $\mathrm{Bi}(n, p_r)$; применить неравенство Чебышева. ▶

8 Пусть $\widehat{g}$ — оценка для теоретической характеристики g и $0 < \mathrm{D}\widehat{g} < \infty$. Убедиться в том, что статистика $\widehat{g}^2$ не является несмещенной оценкой для g^2 и вычислить ее смещение.

9 По выборке $\boldsymbol{X} = (X_1, \ldots, X_n)$ из распределения $\mathcal{L}(\xi)$ построить несмещенную оценку его характеристической функции (х. ф.) $\psi(t) = \mathrm{E}\, e^{it\xi}$

◀ *Указание.* Рассмотреть эмпирическую х. ф. ▶

10 По результатам n испытаний оценивается неизвестная вероятность «успеха» θ в схеме Бернулли $\mathrm{Bi}(1, \theta)$. Обозначим r_n число успехов в этих испытаниях и рассмотрим класс оценок вида $T = (r_n + \alpha)/(n + \beta)$. Вычислить среднеквадратичную ошибку T, убедиться в том, что при $\alpha = \sqrt{n}/2$, $\beta = \sqrt{n}$ она равна $\left[4(\sqrt{n} + 1)^2\right]^{-1}$ и сравнить эту ошибку с риском «обычной» оценки r_n/n.

11 (*Продолжение*). Убедиться в том, что в классе оценок вида $H(r_n)$ несмещенная оценка для параметрической функции $\tau(\theta) = \theta^s(1-\theta)^t$ при целых $s, t \geqslant 0$ существует лишь при $s + t \leqslant n$ и при этом условии она единственна и имеет вид

$$H(r_n) = \frac{(r_n)_s (n - r_n)_t}{(n)_{s+t}}, \quad \text{где} \quad (a)_r = a(a-1) \quad (a - r + 1).$$

◀ *Указание.* Использовать тот факт, что $\mathcal{L}_\theta(r_n) = \mathrm{Bi}(n, \theta)$. ▶

12 Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ — выборка из биномиального распределения $\mathrm{Bi}(k, \theta)$ и $T = X_1 + \ldots + X_n$. Описать класс оцениваемых параметрических функций, т. е. таких, для которых существуют несмещенные оценки вида $H(T)$; построить несмещенную оценку для $\tau_j(\theta) = \theta^j$

◀ *Указание.* Воспользоваться свойством воспроизводимости распределения $\mathrm{Bi}(k, \theta)$ (см. п. 1 § 1.1) и результатом примера 1 в § 3.1. ▶

13 Пусть произведено одно наблюдение X над пуассоновской случайной величиной с неизвестным параметром θ. Показать, что $T(X) = (X)_j$ — несмещенная оценка для $\tau_j(\theta) = \theta^j$ ($j = 1, 2, \ldots$), а для функций $\tau(\theta) = \theta^{-a}$ при $a > 0$ несмещенных оценок не существует. Построить несмещенную оценку для $\tau(\theta) = (\theta + 1)^{-1}$

14 По одному наблюдению над дискретной случайной величиной ξ с распределением

$$f(x; \theta) = \frac{e^{-\theta}}{1 - e^{-\theta}} \frac{\theta^x}{x!}, \quad x = 1, 2, \ldots$$

(*урезанное в нуле пуассоновское распределение*), требуется оценить функцию $\tau(\theta) = 1 - e^{-\theta}$ Убедиться в том, что статистика

$$T(\xi) = \begin{cases} 2 & \text{при } \xi \text{ четном,} \\ 0 & \text{в противном случае} \end{cases}$$

— единственная несмещенная оценка $\tau(\theta)$, но она практически бесполезна.

15 Показать, что $T(X) = \sum_{i=1}^{X} \frac{1}{i}$ — единственная несмещенная оценка функции $\tau(\theta) = \ln(1-\theta)$ в модели $\overline{\mathrm{Bi}}(1, \theta)$ при одном наблюдении.

16 В модели $\mathcal{N}(\mu, \theta^2)$ требуется оценить $\tau(\theta) = \theta^2$ по выборке объема n. Показать, что выборочная дисперсия S^2 имеет меньшую с. к. о., чем несмещенная оценка T_n, указанная в упр. 4, но в классе несмещенных оценок T_n точнее, чем $S_0^2 = (n/(n-1))S^2$.

17 Доказать, что статистика

$$T_n(X) = \sqrt{\frac{\pi}{2}} \frac{1}{n} \sum_{i=1}^{n} |X_i - \mu|$$

— несмещенная и состоятельная оценка параметра θ в модели $\mathcal{N}(\mu, \theta^2)$.

18 Дана выборка $X = (X_1, X_2, X_3)$ из нормального распределения $\mathcal{N}(0, \theta^2)$. Рассмотрим статистику

$$p_T(x) = \frac{1}{2T} I(|x| \leqslant T),$$

где $T = T(X) = (X_1^2 + X_2^2 + X_3^2)^{1/2}$, которая как функция переменного x представляет собой плотность равномерного распределения $U(-T, T)$. Убедиться в том, что $p_T(x)$ при любом x является несмещенной оценкой теоретической плотности.

◀ *Указание.* Учесть, что $\mathcal{L}_\theta(T^2/\theta^2) = \chi^2(3)$. ▶

19 Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения $\mathcal{N}(\mu, \theta^2)$ и

$$T = \sum_{i=1}^{n} (X_i - \mu)^2$$

Проверить, что несмещенной оценкой для функции $\tau_k(\theta) = \theta^k$ при любом $k > -n$ является статистика (см. (45) § 3.4)

$$\tau_k^* = \left(\frac{T}{2}\right)^{k/2} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+k}{2}\right)}.$$

Сравнить оценку τ_1^* с оценкой, указанной в упр. 17.

◀ *Указание.* Использовать тот факт, что $\mathcal{L}_\theta(T/\theta^2) = \chi^2(n)$, и формулу для моментов гамма-распределения (см. п. 3 § 1.1). ▶

20 Доказать, что в общей нормальной модели $\mathcal{N}(\theta_1, \theta_2^2)$ несмещенной оценкой для функции $\tau_k(\theta) = \theta_2^k$ по выборке объема n при $k > -n+1$ является статистика (см. (46) § 3.4)

$$\tau_k^* = \left(\frac{nS^2}{2}\right)^{k/2} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n+k-1}{2}\right)},$$

где S^2 — выборочная дисперсия. Рассмотреть случай $n = 2$ и вычислить смещение статистики $|X_1 - X_2|$ как оценки θ_2.

◀ *Указание.* Учесть, что по теореме Фишера $\mathcal{L}_\theta(nS^2/\theta_2^2) = \chi^2(n-1)$. ▶

21 Пусть $X = (X_1, \quad, X_n)$ — выборка из гамма-распределения $\Gamma(\theta, \lambda)$ и $T = X_1 + \quad + X_n$. Убедиться в том, что статистика

$$\tau_r^* = T^r \frac{\Gamma(\lambda n)}{\Gamma(\lambda n + r)}$$

— несмещенная оценка функции $\tau_r(\theta) = \theta^r$ при любом $r > -\lambda n$ (см. (39) § 3.4).

◀ *Указание.* Воспользоваться свойствами гамма-распределения (см. п. 3 § 1.2). ▶

22 По выборке $X = (X_1, \quad, X_n)$ из равномерного распределения $U(\theta, 2\theta)$ требуется оценить параметр θ. Доказать, что в классе оценок вида $T = \alpha X_{(1)} + \beta X_{(n)}$, $\alpha, \beta \geqslant 0$, оптимальная несмещенная оценка есть

$$T^* = \frac{n+1}{5n+4}(X_{(1)} + 2X_{(n)});$$

вычислить ее дисперсию.

◀ *Указание.* Воспользоваться результатами упр. 46 к гл. 1. ▶

23 Оценивается параметр θ распределения $U(0, \theta)$ по выборке $X = (X_1, \quad, X_n)$. Убедиться в том, что статистики $T_1 = (n+1)/nX_{(n)}$, $T_2 = 2\overline{X}$ и $T_3 = (n+1)X_{(1)}$ несмещенные. Какая из них предпочтительнее?

24 Пусть $X = (X_1, \quad, X_n)$ — выборка из $U(\theta_1, \theta_2)$. Доказать, что статистики $T_1 = (X_{(1)} + X_{(n)})/2$ и $T_2 = (n+1)/(n-1)(X_{(n)} - X_{(1)})$ — несмещенные и состоятельные оценки функций $\tau_1(\theta) = (\theta_1 + \theta_2)/2$ и $\tau_2(\theta) = \theta_2 - \theta_1$ соответственно.

25 Показать, что для *логистического распределения*, задаваемого плотностью

$$f(x; \theta) = e^{-x+\theta}(1 + e^{-x+\theta})^{-2}, \quad x, \theta \in R^1$$

несмещенной и состоятельной оценкой параметра θ является выборочное среднее $\overline{X}$.

26 Является ли выборочное среднее состоятельной оценкой параметра θ модели Коши $\mathcal{K}(\theta)$?

◀ *Указание.* Использовать упр. 48(б) к гл. 1. ▶

27 Проверить, что количество информации $i(\theta)$ для соответствующих моделей имеет указанный в табл. 1 § 3.2 вид.

28 Установить результаты, указанные в табл. 2 § 3.2.

29 Рассматривается задача оценивания функции $\tau(\theta) = \theta^2$ в модели $\Gamma(\theta, \lambda)$ по выборке $X = (X_1, \ldots, X_n)$. Воспользовавшись критерием Бхаттачария, показать, что статистика $T^* = n/(\lambda(\lambda n + 1))\overline{X}^2$ является оптимальной, но не эффективной оценкой $\tau(\theta)$.

30 Пусть $X = (X_1, \ldots, X_n)$ — выборка из *обратного гауссовского распределения*, задаваемого плотностью

$$f(x; \lambda, \mu) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0, \quad \lambda > 0, \quad \mu \neq 0.$$

1) Убедиться в том, что $\overline{X}$ — оптимальная несмещенная оценка параметра μ в любом случае, известен или нет параметр λ. Получить отсюда, что $\mathbf{E}X_1 = \mu$, $\mathbf{D}X_1 = \mu^3/\lambda$.

2) Найти эффективную оценку λ^{-1} при известном μ.

◀ *Указание.* Воспользоваться общими результатами для экспоненциальной модели и критерием Бхаттачария. ▶

31 Доказать непосредственно полноту достаточной статистики $T = X_1 + \ldots + X_n$ для биномиальной модели $\mathrm{Bi}(k, \theta)$. Получить отсюда, что в данном случае несмещенные оценки существуют лишь для полиномов

$$\tau(\theta) = \sum_{j=0}^{r} a_j \theta^j$$

при $r \leqslant kn$, при этом оптимальная оценка имеет вид

$$\tau^* = \sum_{j=0}^{r} \frac{a_j (T)_j}{(kn)_j}.$$

◀ *Указание.* Воспользоваться свойством воспроизводимости распределения $\mathrm{Bi}(k, \theta)$. ▶

32 Доказать полноту достаточной статистики $T = X_1 + \ldots + X_n$ для пуассоновской модели $\Pi(\theta)$. Показать, что оптимальной несмещенной оценкой для сходящегося при всех $\theta > 0$ степенного ряда $\tau(\theta) = \sum_{j \geqslant 0} a_j \theta^j$ является статистика $\tau^* = \sum_j \frac{a_j (T)_j}{n^j}$

◀ *Указание.* Воспользоваться свойством воспроизводимости распределения $\Pi(\theta)$. ▶

33 Показать, что оптимальной оценкой параметра θ, урезанного в нуле пуассоновского распределения (упр. 14) по соответствующей выборке $\boldsymbol{X} = (X_1, \ldots, X_n)$ является статистика $\tau^* = T\Delta^n 0^{T-1} / (\Delta^n 0^T)$ при $T = X_1 + \ldots + X_n \geqslant n + 1$ и $\tau^* = 0$ при $T = n$, где

$$\Delta^n 0^k = \sum_{r=0}^{n} (-1)^{n-r} C_n^r r^k$$

◀ *Указание.* Воспользоваться общим результатом для модели степенного ряда. ▶

34 Доказать, что оценки, указанные в упр. 19, 20 и 21 — оптимальные.

35 Показать, что в модели $\mathcal{N}(\theta, \gamma\theta^2)$ достаточной является статистика $T = (\overline{X}, S^2)$, но она не полна.

◀ *Указание.* Вычислить среднее функции

$$\varphi(T) = \frac{S^2(n+\gamma)}{(n-1)\gamma} - \overline{X}^2 \quad \blacktriangleright$$

36 По результатам $n \geqslant 2$ независимых измерений диаметра θ_1 круга построить оптимальную несмещенную оценку

$$\tau^* = \frac{\pi}{4}\left(\overline{X}^2 - \frac{1}{n-1} S^2\right)$$

его площади $\tau(\theta) = \frac{\pi}{4}\theta_1^2$, считая погрешности измерений нормальными $\mathcal{N}(0, \theta_2^2)$ случайными величинами.

37 Доказать, что в упр. 23 статистика T_1 — оптимальная несмещенная оценка θ. Рассмотреть класс оценок $T_\lambda = \lambda T_1$ и убедиться в том, что в нем имеются оценки с меньшей с. к. о., чем у оценки T_1, оптимальным же по критерию минимума с. к. о. является выбор $\lambda^* = n(n+2)/(n+1)^2$.

38 Доказать, что $T = (X_{(1)}, X_{(n)})$ — полная достаточная статистика для модели $U(\theta_1, \theta_2)$ и обосновать оптимальность оценок в упр. 24. Построить оптимальные оценки для θ_1 и θ_2.

◀ *Указание.* Воспользоваться упр. 46 к гл. 1. ▶

39 Рассматривается двухпараметрическая показательная модель $W(\theta_1, 1, \theta_2)$ (см. пример 2 в § 3.3). Основываясь на том факте, что здесь достаточная статистика

$$T = (T_1, T_2) = (X_{(1)}, \overline{X} - X_{(1)})$$

полна, убедиться в том, что оптимальными несмещенными оценками параметров модели являются соответственно

$$\theta_1^* = T_1 - \frac{T_2}{(n-1)} = \frac{1}{n-1}(nX_{(1)} - \overline{X}) \qquad \theta_2^* = \frac{n}{n-1}T_2;$$

при этом

$$\mathrm{D}_\theta\theta_1^* = \frac{\theta_2^2}{n(n-1)}, \quad \mathrm{D}_\theta\theta_2^* = \frac{\theta_2^2}{n-1}$$

◀ *Указание.* Используя результаты примера 1 в § 2.4, представить T_1 и T_2 через показательные спейсинги $Y_1, \quad , Y_n$: $Y_1 = n(T_1 - \theta_1)/\theta_2$, $Y_2 + \ldots + Y_n = nT_2/\theta_2$, где $Y_r = (n - r + 1)\big(X_{(r)} - X_{(r-1)}\big)/\theta_2$, $r \geqslant 2$. Отсюда, в частности, следует, что T_1 и T_2 независимы. ▶

40 Убедиться в том, что оптимальная несмещенная оценка параметра θ распределения Вейбулла $W(0, \alpha, \theta^{1/\alpha})$ по выборке $X = (X_1, \ldots, X_n)$ есть T/n, где $T = \sum_{i=1}^{n} X_i^\alpha$.

◀ *Указание.* См. пример 15 § 3.4. ▶

41 Доказать непосредственно, вычислив $\mathrm{E}_\theta\widehat{\theta}_n$ и $\mathrm{D}_\theta\widehat{\theta}_n$, асимптотическую несмещенность и состоятельность о. м. п. $\widehat{\theta}_n$ параметра θ нормальной модели $\mathcal{N}(\mu, \theta^2)$ и найти ее предельный при $n \to \infty$ закон распределения; вычислить асимптотическую эффективность оценки T_n в упр. 17.

◀ *Указание.* Выразить $\widehat{\theta}_n = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right]^{1/2}$ через оптимальную несмещенную оценку τ_1^* в упр. 19; применить к оценке T_n центральную предельную теорему. ▶

42 Убедиться в том, что о. м. п. $\widehat{\theta}_n$ параметра θ в модели $\mathcal{N}(\theta, 2\theta)$ имеет вид $\widehat{\theta}_n = \sqrt{1 + T_n} - 1$, где $T_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$, и доказать ее состоятельность.

◀ *Указание.* Применить к T_n закон больших чисел. ▶

43 Имеется выборка (X_i, Y_i), $i = 1, \ldots, n$, из двумерного нормального распределения

$$\mathcal{N}\left((0,0), \begin{Vmatrix} 1 & \theta \\ \theta & 1 \end{Vmatrix}\right), \quad \theta \in (-1, 1).$$

Составить уравнение правдоподобия для о. м. п. $\widehat{\theta}_n$ и убедиться в том, что ее асимптотическая дисперсия равна $(1-\theta^2)^2/\big(n(1+\theta^2)\big)$. Рассмотреть в качестве альтернативной оценки коэффициента корреляции θ выборочный коэффициент корреляции

$$T_n = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$$

и доказать, что его асимптотическая эффективность есть

$$e_0(T_n;\theta)=\left(\frac{1-\theta^2}{1+\theta^2}\right)^2$$

◄ *Указание.* При вычислении моментов использовать характеристическую функцию (см. указание к упр. 29 гл. 1). ►

44 Пусть $\boldsymbol{X}=(X_1,\ldots,X_n)$ — выборка из *распределения Кептайна*, задаваемого плотностью

$$f(x;\theta)=\frac{g'(x)}{\sqrt{2\pi}\theta_2}\exp\left\{-\frac{1}{2\theta_2^2}(g(x)-\theta_1)^2\right\},\quad \theta=(\theta_1,\theta_2),$$

где $g(x)$ — некоторая дифференцируемая монотонно возрастающая фуикция (при $g(x)\equiv x$ имеем нормальное распределение $\mathcal{N}(\theta_1,\theta_2^2)$). Убедиться в том, что справедливо следующее обобщение результата примера 1 в § 3.5: о. м. п. $\widehat{\theta}=(\overline{g},T)$, где

$$\overline{g}=\frac{1}{n}\sum_{i=1}^n g(X_i),\quad T^2=\frac{1}{n}\sum_{i=1}^n(g(X_i)-\overline{g})^2$$

Является ли $\overline{g}$ эффективной оценкой θ_1? Показать, что при известном значеиии $\theta_1=a$ эффективной оценкой θ_2^2 является статистика

$$T_1^2=\frac{1}{n}\sum_{i=1}^n(g(X_i)-a)^2$$

(ср. с соответствующими результатами для нормальной модели, см. примеры 1, 2 и 8 § 3.2).

45 Найти о. м. п. $\widehat{\theta}_n$ параметра θ отрицательной биномиальной модели $\overline{\mathrm{Bi}}(r,\theta)$ и вычислить ее асимптотическую дисперсию.

◄ *Указание.* Использовать общие результаты примера 9 в § 3.5 для модели степенного ряда. ►

46 Пусть в полиномиальном распределении $M(n;p_1,\ldots,p_N)$ вероятности исходов имеют вид $p_i=p_i(\theta)$, $i=1,\ldots,N$, где θ — общий неизвестный (скалярный) параметр. Составить уравнения метода накопления для приближенного вычисления о. м. п. $\widehat{\theta}$.

47 Показать, что асимптотическая эффективность выборочной медианы в модели Коши $\mathcal{K}(\theta)$ (см. пример 8 в § 3.5) равна $8/\pi^2=0{,}8\ldots$

◄ *Указание.* Воспользоваться теоремой 1 § 2.4 (соотношение (15)). ►

48 Убедиться в асимптотической несмещенности и состоятельностн о. м. п. $\widehat{\theta}_n=X_{(n)}$ параметра θ равномерной модели $U(0,\theta)$ (см. примеры 3 и 11 § 3.5). Доказать непосредственно, что при $n\to\infty$

$$\mathrm{P}_\theta\left\{\frac{n(\theta-\widehat{\theta}_n)}{\theta}\leqslant t\right\}\longrightarrow 1-e^{-t}\qquad t\geqslant 0.$$

◄ *Указание.* Использовать упр. 23, а также 46 к гл. 1. ►

49 Показать, что в случае модели $U(\theta - 1/2,\ \theta + 1/2)$ любое значение

$$\theta \in \left[X_{(n)} - \frac{1}{2}, X_{(1)} + \frac{1}{2}\right]$$

является о. м. п. $\widehat{\theta}_n$. Какая точка этого интервала является несмещенной оценкой θ?

◄ *Указание.* Записать о. м. п. в виде

$$\widehat{\theta}_n = \alpha\left(X_{(n)} - \frac{1}{2}\right) + (1-\alpha)\left(X_{(1)} + \frac{1}{2}\right), \quad \alpha \in [0, 1],$$

и найти α из уравнения несмещенности. ►

50 Рассматривается семейство распределений Вейбулла $W(\theta, \alpha, b)$ с неизвестным параметром сдвига. Убедиться в том, что при $0 < \alpha \leqslant 1$ о. м. п. $\widehat{\theta}_n = X_{(1)}$ асимптотически несмещена и состоятельна, а ее закон распределения не является нормальным при $n \to \infty$.

◄ *Указание.* Использовать результаты примера 6 § 3.2 и упр. 56 к гл. 1. ►

51 Случайная величина ξ, характеризующая срок службы элементов электронной аппаратуры, имеет распределение Релея $W(0, 2, \sqrt{\theta})$ (см. упр. 40). Убедиться в том, что о. м. п. $\widehat{\theta}_n = T/n$, т. е. совпадает с оптимальной несмещенной оценкой θ

52 По выборке $\boldsymbol{X} = (X_1, \ldots, X_n)$ из гамма-распределения $\Gamma(\theta, \lambda)$ требуется оценить функцию $\tau(\theta) = \theta^{-1}$ Показать, что о. м. п. $\widehat{\tau}_n = \lambda/\overline{X}$, убедиться в ее асимптотической несмещенности и состоятельности, а также установить, что при $n \to \infty$ $\mathcal{L}_\theta(\widehat{\tau}_n) \sim \mathcal{N}(1/\theta, 1/(\lambda n \theta^2))$.

◄ *Указание.* Использовать упр. 21. ►

53 Доказать, что для *распределения Лапласа с параметром сдвига*, задаваемого плотностью $f(x; \theta) = (1/2)e^{-|x-\theta|}$, $x, \theta \in R^1$, о. м. п. $\widehat{\theta}_n$ совпадает с выборочной медианой и установить ее асимптотическую нормальность: $\mathcal{L}_\theta(\widehat{\theta}_n) \sim \mathcal{N}(\theta, n^{-1})$.

54 Убедиться в асимптотической несмещенности и состоятельности о. м. п. $\widehat{\theta}_n = (\widehat{\theta}_{1n}, \widehat{\theta}_{2n})$ параметров двухпараметрической показательной модели (см. пример 4 в § 3.5).

55 *Преобразования, стабилизирующие дисперсию.* Установить следующие, полезные аппроксимации для распределений специальных функций от о. м. п.:
для модели $\mathrm{Bi}(k, \theta)$:

$$\mathcal{L}_\theta\left(\arcsin\sqrt{\widehat{\theta}_n}\right) \sim \mathcal{N}\left(\arcsin\sqrt{\theta}, \frac{1}{4kn}\right), \quad \widehat{\theta}_n = \frac{\overline{X}}{k};$$

для модели $\Pi(\theta)$:

$$\mathcal{L}_\theta\left(\sqrt{\widehat{\theta}_n}\right) \sim \mathcal{N}\left(\sqrt{\theta}, \frac{1}{4n}\right), \quad \widehat{\theta}_n = \overline{X};$$

для модели $\mathcal{N}(\mu, \theta^2)$:

$$\mathcal{L}_\theta(\ln\widehat{\theta}_n) \sim \mathcal{N}\left(\ln\theta, \frac{1}{2n}\right), \quad \widehat{\theta}_n = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right]^{1/2}$$

для модели $\Gamma(\theta, \lambda)$:

$$\mathcal{L}_\theta(\ln\widehat{\theta}_n) \sim \mathcal{N}\left(\ln\theta, \frac{1}{\lambda n}\right), \quad \widehat{\theta}_n = \frac{\overline{X}}{\lambda}.$$

◀ *Указание.* См. п. 5 (соотношение (31)) § 3.5. ▶

56 *Объединение статистической информации.* Пусть $X_j = (X_{j1}, \quad, X_{jn_j})$, $j = 1, \quad, k$, — независимые выборки из распределений $\mathcal{N}(\theta_{j1}, \theta_2^2)$, $j = 1, \ldots, k$, соответственно и $\overline{X}_j$, S_j^2 — соответствующие выборочные средние и дисперсии. Доказать, что $\widehat{\theta} = (\overline{X}_1, \ldots, \overline{X}_k, \widehat{\theta}_2)$ — о. м. п. для $\theta = (\theta_{11}, \ldots, \theta_{k1}, \theta_2)$, где

$$\widehat{\theta}_2^2 = \frac{1}{n}\sum_{j=1}^{k} n_j S_j^2, \quad n = n_1 + \quad + n_k;$$

несмещенной же оценкой для общей дисперсии θ_2^2 является статистика

$$\widetilde{\theta}_2^2 = \left(\frac{n}{n-k}\right)\widehat{\theta}_2^2$$

◀ *Указание.* Использовать пример 1 в § 3.5, записав функцию правдоподобия для всех данных в виде

$$L = (2\pi e t^2)^{-n/2} \exp\left\{-\frac{1}{2\theta_2^2}\sum_{j=1}^{k} n_j(\overline{x}_j - \theta_{j1})^2 - n\left[\frac{1}{2}\left(\frac{t^2}{\theta_2^2} - 1\right) - \ln\frac{t}{\theta_2}\right]\right\},$$

где

$$t^2 = \frac{1}{n}\sum_{j=1}^{k} n_j s_j^2. \ ▶$$

57 Показать, что γ-доверительный интервал для параметра θ модели $\mathcal{N}(\theta, \theta^2)$, $\theta > 0$, по выборке $\boldsymbol{X} = (X_1, \ldots, X_n)$ имеет вид

$$\left(\frac{\overline{X}}{1 + c_\gamma/\sqrt{n}}, \ \frac{\overline{X}}{1 - c_\gamma/\sqrt{n}}\right), \quad c_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right).$$

Получить аналогичный результат для случая $\theta < 0$.

◀ *Указание.* Использовать тот факт, что $\mathcal{L}_\theta(\sqrt{n}(\overline{X} - \theta)/\theta) = \mathcal{N}(0, 1)$. ▶

58 Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ — выборка из распределения $\mathcal{N}(\theta, \sigma^2)$.

1) Убедиться в том, что любой интервал вида

$$\Delta_\gamma(\boldsymbol{X}) = \left(\overline{X} - \frac{\sigma g_2}{\sqrt{n}}, \ \overline{X} - \frac{\sigma g_1}{\sqrt{n}}\right),$$

где $g_1 < g_2$ — любые числа, удовлетворяющие условию $\Phi(g_2) - \Phi(g_1) = \gamma$, является γ-доверительным интервалом для параметра θ. Доказать, что кратчайший из этих интервалов есть $\Delta_\gamma^*(\boldsymbol{X}) = (\overline{X} \mp \sigma c_\gamma/\sqrt{n})$.

2) Сколько необходимо произвести наблюдений $n = n(l, \gamma)$, чтобы точность локализации параметра, т. е. длина интервала $\Delta_\gamma^*(\boldsymbol{X})$, была равна заданной величине l? Проверить, что при заданных n и l доверительный уровень $\gamma = \gamma(n, l) = 2\Phi(l\sqrt{n}/(2\sigma)) - 1$.

◀ *Указание.* Использовать результаты примера 2 в § 3.8. ▶

59 Доказать, что γ-доверительным интервалом для среднеквадратичного отклонения θ модели $\mathcal{N}(\mu, \theta^2)$ является любой интервал $\delta_\gamma(\underline{x}) = (T/a_2, T/a_1)$, где

$$T^2 = \sum_{i=1}^{n} (X_i - \mu)^2$$

а числа $a_1 < a_2$ выбираются из условия

$$\int_{a_1}^{a_2} x k_n(x^2)\, dx = \frac{\gamma}{2},$$

где $k_n(t)$ — плотность распределения $\chi^2(n)$. Определить кратчайший в этом классе интервал $\delta_\gamma^*(X)$.

◀ *Указание.* Использовать то, что $\mathcal{L}_\theta(T^2/\theta^2) = \chi^2(n)$. ▶

60 Доказать, что нижний и верхний γ-доверительные интервалы для среднего θ_1 модели $\mathcal{N}(\theta_1, \theta_2^2)$ по выборке $X = (X_1, \ldots, X_n)$ имеют соответственно вид

$$\left(\overline{X} - \frac{t_{\gamma, n-1} S}{\sqrt{n-1}} < \theta_1\right) \qquad \left(\theta_1 < \overline{X} + \frac{t_{\gamma, n-1} S}{\sqrt{n-1}}\right)$$

◀ *Указание.* См. пример 1 в § 2.5, где построен двусторонний интервал для θ_1 ▶

61 (*Продолжение*). Построить односторонние γ-доверительные интервалы для дисперсии θ_2^2:

$$\left(\frac{nS^2}{\chi^2_{\gamma, n-}} < \theta_2^2\right) \quad \text{и} \quad \left(\theta_2^2 < \frac{nS^2}{\chi^2_{1-\gamma, n-1}}\right)$$

(двусторонний центральный γ-доверительный интервал указан в (10) § 2.5).

62 Пусть $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ — две независимые выборки, причем первая из распределения $\mathcal{N}(\theta^{(1)}, \sigma_1^2)$, а вторая — из $\mathcal{N}(\theta^{(2)}, \sigma_2^2)$. Показать, что центральный γ-доверительный интервал для разности средних $\tau = \theta^{(1)} - \theta^{(2)}$ имеет вид, указанный в табл. 1 § 3.8.

63 Пусть $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ — независимые выборки из распределений $\Gamma(\theta_1, 1)$ и $\Gamma(\theta_2, 1)$ соответственно. Доказать, что центральный γ-доверительный интервал для отношения $\tau = \theta_2/\theta_1$ имеет вид

$$\left(\frac{F_{(1-\gamma)/2,\, 2n,\, 2m}\, \overline{Y}}{\overline{X}},\ \frac{F_{(1+\gamma)/2,\, 2n,\, 2m}\, \overline{Y}}{\overline{X}}\right),$$

где $F_{p, r, s}$ — p-квантиль распределения Снедекора $S(r, s)$.

◀ *Указание.* Использовать тот факт, что $\mathcal{L}_\theta(\tau \overline{X}/\overline{Y}) = S(2n, 2m)$ (см. упр. 54 к гл. 1). ▶

64 Убедиться в том, что γ-доверительный интервал для параметра сдвига показательного распределения: $f(x; \theta) = e^{-x+\theta}$, $x \geqslant \theta$, по n наблюдениям имеет вид

$$\left(X_{(1)} + \left(\frac{1}{n}\right) \ln(1-\gamma),\ X_{(1)}\right)$$

Какой вид будет иметь здесь центральный γ-доверительный интервал?

◀ *Указание.* Найти распределение статистики $X_{(1)} = \min\limits_{1 \leqslant i \leqslant n} X_i$ и учесть, что событие $\{X_{(1)} \geqslant \theta\}$ является достоверным. ▶

65 Рассматривается модель упр. 40. Убедиться в том, что центральный γ-доверительный интервал для параметра θ имеет вид

$$\left(\frac{2T}{\chi^2_{(1+\gamma)/2,\,2n}},\ \frac{2T}{\chi^2_{(1-\gamma)/2,\,2n}}\right).$$

В частности, при $\alpha = 1$ имеем соответствующее решение для модели $\Gamma(\theta, 1)$.

◀ *Указание.* Учесть, что $\mathcal{L}_\theta(2\xi^\alpha/\theta) = \chi^2(2)$ (см. пример 15 в § 3.4 и п. 3 § 1.2). ▶

66 Рассматривается модель упр. 39. Учитывая, что

$$\mathcal{L}_\theta\left(\frac{2nT_2}{\theta_2}\right) = \Gamma(2, n-1) = \chi^2(2(n-1)),$$

т. е. $2nT_2/\theta_2$ — центральная статистика для θ_2 (см. указание к упр. 39), убедиться в том, что

$$\left(\frac{2nT_2}{\chi^2_{(1+\gamma)/2,\,2(n-1)}},\ \frac{2nT_2}{\chi^2_{(1-\gamma)/2,\,2(n-1)}}\right)$$

есть центральный γ-доверительный интервал для θ_2. Далее, так как

$$\mathcal{L}_\theta\left(\frac{2n(T_1-\theta_1)}{\theta_2}\right) = \Gamma(2, 1) = \chi^2(2)$$

и статистики T_1 и T_2 независимы, то

$$\frac{2n(T_1-\theta_1)}{2\theta_2}\ \frac{2nT_2}{2(n-1)\theta_2} = \frac{(n-1)(T_1-\theta_1)}{T_2}$$

— центральная статистика для θ_1, имеющая распределение Снедекора $S(2, 2(n-1))$. Следовательно, центральный γ-доверительный интервал для θ_1 есть

$$\left(T_1 - \frac{T_2}{n-1}F_{(1+\gamma)/2,\,2,\,2(n-1)},\ T_1 - \frac{T_2}{n-1}F_{(1-\gamma)/2,\,2,\,2(n-1)}\right).$$

67 *Прогнозирование будущих наблюдений.* Пусть n, $\overline{X}$ и S^2 — объем, выборочные среднее и дисперсия выборки из распределения $\mathcal{N}(\theta_1, \theta_2^2)$. Показать, что с вероятностью γ результат следующего, $(n+1)$-го испытания X_{n+1} попадет в интервал

$$\left(\overline{X} \mp t_{\frac{(1+\gamma)}{2},\,n-1}S\sqrt{\frac{(n+1)}{(n-1)}}\right),$$

где $t_{p,\,n}$ — p-квантиль распределения Стьюдента $S(n)$.

◀ *Указание.* Учесть, что

$$\mathcal{L}\left(\sqrt{\frac{(n-1)}{(n+1)}}\,\frac{(\overline{X}-X_{n+1})}{S}\right) = S(n-1)$$

(следствие теоремы 2 § 2.5). ▶

68 Пусть $X = (X_1, \quad, X_n)$ — выборка из нормального распределения $\mathcal{N}(\mu, \theta^2)$. Показать, что приближенные (при больших n) γ-доверительные интервалы для параметра θ и функции $\tau(\theta) = \ln \theta$ имеют вид соответственно

$$\widehat{\theta}_n\left(1 \mp \frac{c_\gamma}{\sqrt{2n}}\right) \qquad \left(\ln \widehat{\theta}_n \mp \frac{c_\gamma}{\sqrt{2n}}\right),$$

где

$$\widehat{\theta}_n = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right]^{1/2} \quad \text{и} \quad c_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right).$$

Получить из второго результата другой интервал для θ.

◀ *Указание.* Использовать результат, указанный в упр. 55. ▶

69 Доказать, что в случае больших выборок асимптотически кратчайший γ-доверительный интервал для параметра θ модели $\Gamma(\theta, \lambda)$ имеет вид $\overline{X}\lambda^{-1}(1 \mp c_\gamma/\sqrt{\lambda n})$, а аналогичный интервал для функции $\tau(\theta) = \ln \theta$ есть $(\ln \widehat{\theta}_n \mp c_\gamma/\sqrt{\lambda n})$ (см. упр. 55). Получить из второго результата другой интервал для θ.

70 По выборке $X = (X_1, \ldots, X_n)$ из распределения $\mathrm{Bi}(1, \theta)$ построить асимптотический (при $n \to \infty$) γ-доверительный интервал для θ основываясь на нормальной аппроксимации

$$\mathcal{L}_\theta\left(\frac{\sqrt{n}(\overline{X} - \theta)}{\sqrt{\theta(1-\theta)}}\right) \sim \mathcal{N}(0, 1)$$

(теорема Муавра—Лапласа). Сравнить полученное решение с решением, указанным в примере 17 в § 3.5.

71 Показать, что асимптотический γ-доверительный интервал для параметра θ отрицательной биномиальной модели $\overline{\mathrm{Bi}}(r, \theta)$ имеет вид

$$\left(\frac{\overline{X}}{r + \overline{X}} \mp c_\gamma\left[\frac{r\overline{X}}{n(r+\overline{X})^3}\right]^{1/2}\right)$$

(ср. с результатом примера 16 § 3.5).

◀ *Указание.* Воспользоваться общим результатом (33) § 3.5 для модели степенного ряда и формулами (7) § 1.1. ▶

72 Показать, что оценкой по методу моментов параметра θ равномерного распределения $U(0, \theta)$ является статистика $\widetilde{\theta} = 2\overline{X}$ и исследовать ее на несмещенность, состоятельность и эффективность (см. упр. 23).

73 Убедиться в том, что оценка по методу моментов параметра θ распределения $U(-\theta, \theta)$, $\theta > 0$, есть $\widetilde{\theta} = \sqrt{3\widehat{\alpha}_2}$.

74 Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения Лапласа с плотностью

$$f(x; \theta) = \left(\frac{\theta}{2}\right)e^{-\theta|x|}, \quad \theta > 0.$$

Показать, что оценка θ по методу моментов есть $\widetilde{\theta} = \sqrt{2/\widehat{\alpha}_2}$.

75 Найти методом моментов оценки параметров θ_1 и θ_2 «двойного» распределения Пуассона, задаваемого вероятностями

$$\mathbf{P}_\theta\{\xi = x\} = \frac{1}{2}\left(e^{-\theta_1}\frac{\theta_1^x}{x!} + e^{-\theta_2}\frac{\theta_2^x}{x!}\right), \qquad x = 0, 1, 2, \ldots, \quad 0 < \theta_1 < \theta_2.$$

76 Пусть $X_{(1)} < \quad < X_{(r)}$ — цензурированные данные для n наблюдений ($r \leqslant n$) в модели упр. 39 и $Y_i = (n-i+1)(X_{(i)} - X_{(i-1)})$, $i = 2, \ldots, r$. Показать, что

$$\delta_2 = \overline{Y} = \frac{Y_2 + \quad + Y_r}{r-1} = \frac{1}{r-1}\left(\sum_{i=1}^{r} X_{(i)} - nX_{(1)} - (n-r)X_{(r)}\right)$$

и $\delta_1 = X_{(1)} - \overline{Y}/n$ — несмещенные оценки соответственно для θ_2 и θ_1 и вычислить их дисперсии

$$\mathbf{D}_\theta \delta_2 = \frac{\theta_2^2}{(r-1)}, \quad \mathbf{D}_\theta \delta_1 = \frac{\theta_2^2 r}{(r-1)n^2}$$

(при $r = n$ имеем результат упр. 39).

◀ *Указание.* Перейти к величине $\xi_0 = (\xi - \theta_1)/\theta_2$ и использовать пример 1 в § 2.4. ▶

77 Рассматривается задача оценивания неизвестной вероятности «успеха» θ по наблюдению числа «успехов» X в n испытаниях Бернулли. Пусть функция потерь имеет вид

$$L(\delta, \theta) = \frac{(\delta - \theta)^2}{\theta(1-\theta)},$$

а априорное распределение параметра $\mathcal{L}(\theta) = U(0, 1)$. Доказать, что байессовская оценка есть $\delta^*(X) = X/n$ — относительная частота «успеха», и она же является минимаксной; при этом байессовский риск $r(\delta^*) = 1/n$.

◀ *Указание.* Вычислить апостериорный риск (16) § 3.6 и убедиться в том, что он минимизируется при $\delta(X) = X/n$, при этом функция риска $R(\delta^*, \theta) = 1/n$ — не зависит от θ. ▶

78 Пусть испытания Бернулли продолжаются до получения r-го «неуспеха» и X — число «успехов» в этих испытаниях, т. е. $\mathcal{L}_\theta(X) = \overline{\mathrm{Bi}}(r, \theta)$. Убедиться в том, что если функция потерь $L(\delta, \theta) = (\delta - \theta)^2$, а априорное распределение $\mathcal{L}(\theta) = \mathrm{Be}(a, b)$, то байессовская оценка вероятности «успеха» θ есть

$$\delta^*(X) = \mathbf{E}(\theta | X) = \frac{a + X}{a + b + r + X}.$$

◀ *Указание.* Использовать упр. 64 (2) к гл. 1 и формулы (24) § 1.2. ▶

79 По выборке $X = (X_1, \ldots, X_n)$ из пуассоновского распределения $\Pi(\theta)$ оценивается параметр θ при квадратичной функции потерь $L(\delta, \theta) = (\delta - \theta)^2$ и априорном распределении $\mathcal{L}(\theta) = \Gamma(a, \lambda)$ Показать, что байессовская оценка имеет вид

$$\delta^*(X) = \frac{a(\lambda + X)}{na + 1}, \quad X = X_1 + \ldots + X_n,$$

и ее риск есть $r(\delta^*) = \lambda a^2/(na + 1)$. Определить оптимальный объем выборки при цене $c > 0$ одного наблюдения (т. е. минимизирующий общие потери $r(\delta^*) + cn$).

◀ *Указание.* Использовать упр. 64(3) к гл. 1 и формулы (21) § 1.2. ▶

80 (*Продолжение*). Убедиться в том, что если функция потерь $L(\delta, \theta) = (\delta - \theta)^2/\theta$, то байессовская оценка при $\lambda + X > 1$ имеет вид

$$\delta^*(X) = \frac{a}{na + 1}(X + \lambda - 1)$$

и ее риск $r(\delta^*) = a/(na + 1)$.

81 Рассматривается задача оценивания параметра θ показательного распределения с плотностью $f(x;\theta) = \theta e^{-\theta x}$ $x \geqslant 0$, $\theta > 0$, по выборке $\boldsymbol{X} = (X_1, \ldots, X_n)$. Пусть функция потерь $L(\delta, \theta) = (\delta - \theta^{-1})^2$ и априорное распределение $\mathcal{L}(\theta) = \Gamma(a, \lambda)$, $\lambda > 2$. Доказать, что байесовская оценка имеет вид

$$\delta^*(\boldsymbol{X}) = \frac{1}{a(\lambda + n - 1)}\left(a\sum_{i=1}^{n} X_i + 1\right)$$

и ее риск $r(\delta^*) = [a^2(\lambda + n - 1)(\lambda - 1)(\lambda - 2)]^{-1}$. Убедиться в том, что оптимальное число наблюдений при цене $c > 0$ одного наблюдения равно

$$n^* = \frac{1}{a\sqrt{c(\lambda - 1)(\lambda - 2)}} - \lambda + 1.$$

◀ *Указание.* Использовать упр. 64 (4) к гл. 1 и формулы для моментов гамма-распределения (п. 3 § 1.2). ▶

82 Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ — выборка из равномерного распределения $U(0, \theta)$, где априорное распределение параметра θ есть распределение Парето с параметрами x_0 и $\alpha > 2$ (см. п. 9 § 1.2). Убедиться в том, что при квадратичной функции потерь байесовская оценка θ имеет вид

$$\delta^*(\boldsymbol{X}) = \frac{n + \alpha}{n + \alpha - 1}\max(x_0, X_{(n)}), \quad X_{(n)} = \max_{1 \leqslant i \leqslant n} X_i,$$

и ее риск $r(\delta^*) = \alpha x_0^2/[(\alpha - 2)(n + \alpha - 1)^2]$. Определить оптимальный объем выборки при цене $c > 0$ одного наблюдения.

◀ *Указание.* Использовать упр. 64 (6) к гл. 1 и формулы (39) § 1.2. ▶

83 Убедиться в том, что байесовская оценка параметра θ нормального распределения $\mathcal{N}(\theta, b^2)$ при квадратичной функции потерь и априорном распределении $\mathcal{L}(\theta) = \mathcal{N}(\mu, \sigma^2)$ есть $\delta^*(\boldsymbol{X}) = \mu_1$ и ее риск $r(\delta^*) = \sigma_1^2$ (обозначения упр. 64 (5) к гл. 1), а оптимальное число наблюдений при цене $c > 0$ одного наблюдения равно

$$n^* = b(c^{-1/2} - b\sigma^{-2}).$$

◀ *Указание.* Использовать упр. 64 (5) к гл. 1. ▶

Глава 4

Проверка статистических гипотез

...Узловым вопросом математической статистики является вопрос: как далеко могут отклоняться величины, вычисленные по выборке, от соответствующих идеальных значений?

Б. Л. Ван дер Варден [1)]

Эта глава представляет собой введение в теорию проверки статистических гипотез. Здесь вводятся основные понятия этой теории (статистической гипотезы, статистического критерия, критической области, функции мощности и др.) и на примерах решения задач проверки типичных и наиболее распространенных в приложениях статистических гипотез излагаются общие принципы построения и исследования критериев согласия. Значительное внимание уделяется методу группировки наблюдений с последующим применением классического критерия хи-квадрат.

§ 4.1. Основные понятия и общие принципы теории проверки гипотез

Как отмечалось в общем Введении, математическая статистика, помимо оценивания неизвестных параметров распределений, занимается еще и проверкой статистических гипотез, т. е. различных предположений о виде или свойствах распределений наблюдаемых случайных величин. Общее понятие статистической гипотезы (далее будем говорить для краткости просто гипотезы) и примеры типичных и наиболее распространенных гипотез приведены во Введении. Поэтому здесь мы ограничимся лишь кратким напоминанием основных фактов, необходимых нам для развития этой темы.

Пусть $\boldsymbol{X}$ — данные (выборка), $\mathfrak{X} = \{\boldsymbol{x}\}$ — выборочное пространство, $\mathcal{F} = \{F\}$ — совокупность априори допустимых распределений $\boldsymbol{X}$, $F_{\boldsymbol{X}}$ — неизвестное истинное распределение данных $\boldsymbol{X}$, $F_{\boldsymbol{X}} \in \mathcal{F}$. В общем случае задача проверки гипотез ставится так: выделяется некоторое подмножество $\mathcal{F}_0 \subset \mathcal{F}$ допустимых распределений и требуется по данным $\boldsymbol{X}$ проверить, справедливо ли утверждение H_0: $F_{\boldsymbol{X}} \in \mathcal{F}_0$ или же оно ложно. В этом случае H_0 называется *основной* (*нулевой*) *гипотезой*. Распределения $F \in \mathcal{F}_1 = \mathcal{F} \setminus \mathcal{F}_0$ называются *альтернативными*, а утверждение H_1: $F_{\boldsymbol{X}} \in \mathcal{F}_1$ — *альтернативной*

[1)] Ван дер Варден Бартел Лендерт (1903–1996) — голландский математик — статистик и алгебраист.

гипотезой. В этом случае речь идет о проверке гипотезы H_0 против альтернативы H_1 или о задаче (H_0, H_1); иногда также говорят, что гипотеза H_0 проверяется внутри общей гипотезы $H \quad F_X \in \mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$. Если подмножество $\mathcal{F}_0(\mathcal{F}_1)$ состоит из одного элемента, то гипотеза H_0 (альтернатива H_1) называется *простой*, в противном случае — *сложной*.

Правило, согласно которому мы, наблюдая X, принимаем решение принять гипотезу H_0 как истинную либо отклонить ее как ложную (т. е. принять альтернативную гипотезу H_1) называется *статистическим критерием* (или просто *критерием*).

Статистические гипотезы могут быть самые разные. В простейших случаях проверяемая гипотеза может однозначно фиксировать распределение F_X, но типично она указывает лишь тот класс допустимых распределений, которому, по предположению, принадлежит истинное распределение данных. В некоторых случаях класс $\mathcal{F}$ — это все распределения на выборочном пространстве $\mathfrak{X}$, тогда альтернативная гипотеза $H_1 = \overline{H}_0$ («не H_0») никак не конкретизируется, и здесь речь идет просто о *согласии* данных X с нулевой гипотезой H_0 (согласуются ли данные X с гипотезой H_0 или же они ее опровергают) — соответствующие критерии называются *критериями согласия*. В других случаях класс всех допустимых распределений данных является заданным параметрическим семейством $\mathcal{F} = \{F(x; \theta),\ \theta \in \Theta\}$; здесь гипотезы имеют вид $H_0 : \theta \in \Theta_0 \subset \Theta$, а $H_1 : \theta \in \Theta_1 = \Theta \backslash \Theta_0$, и называются *параметрическими*. О соответствующих критериях также говорят как о *параметрических критериях*. В любом случае «хороший» критерий должен учитывать специфику проверяемой гипотезы: в непараметрическом случае он должен быть универсального типа, т. е. обнаруживать любые отклонения от H_0, в параметрическом же случае он должен быть направлен на обнаружение конкретных (в рамках рассматриваемой параметрической модели $\mathcal{F}$) отклонений от H_0. Но любой статистический критерий строится по следующей схеме.

Поскольку критерий — это правило, которое для каждой реализации x выборки X должно приводить к одному из двух решений: принять гипотезу H_0 или отклонить ее (принять альтернативу H_1), то каждому критерию соответствует некоторое разбиение выборочного пространства $\mathfrak{X}$ на два взаимно дополнительных множества $\mathfrak{X}_0$ и $\mathfrak{X}_1$ ($\mathfrak{X}_0 \cap \mathfrak{X}_1 = \varnothing$, $\mathfrak{X}_0 \cup \mathfrak{X}_1 = \mathfrak{X}$), где $\mathfrak{X}_0$ состоит из тех точек x, для которых H_0 принимается, а $\mathfrak{X}_1$ — из тех, для которых H_0 отвергается (принимается H_1). Таким образом, $\mathfrak{X}_0$ — это область принятия гипотезы H_0, а $\mathfrak{X}_1$ — область ее отклонения, которую принято называть *критической областью*. Тем самым, любой критерий проверки гипотезы H_0 однозначно задается соответствующей критической областью $\mathfrak{X}_1$, и о таком критерии часто говорят как о «критерии $\mathfrak{X}_1$». Итак, критерий $\mathfrak{X}_1$ имеет вид

$$\boxed{H_0 \quad \text{отвергается} \quad \Longleftrightarrow \quad X \in \mathfrak{X}_1.}$$

Как в конкретной задаче (H_0, H_1) выбирать критическую область $\mathfrak{X}_1$? Это делается на основе постулируемого в математической статистике следующего

Общий принцип принятия решений

общего принципа принятия решения: если в эксперименте наблюдается маловероятное при справедливости гипотезы H_0 событие, то считается, что гипотеза H_0 не согласуется с данными (или противоречит им), и в этом случае она отвергается (отклоняется); в противном случае считается, что данные не противоречат H_0 (или согласуются с ней), и H_0 принимается.

В соответствии с этим принципом критическая область $\mathfrak{X}_1$ должна быть выбрана так, чтобы была мала вероятность $\mathbf{P}\{X \in \mathfrak{X}_1 | H_0\}$, т. е. условная (при условии, что H_0 справедлива) вероятность попадания значения выборки X в область $\mathfrak{X}_1$. Поэтому при построении критерия задаются заранее некоторым малым числом α (например, $\alpha = 0{,}001; 0{,}01; 0{,}05$ и т. д.) и налагают условие

$$\mathbf{P}\{X \in \mathfrak{X}_1 | H_0\} \leqslant \alpha. \tag{1}$$

Если выполнено (1), то говорят, что критерий $\mathfrak{X}_1$ имеет *уровень значимости* α и подчеркивают это обозначением $\mathfrak{X}_1 = \mathfrak{X}_{1\alpha}$. Ясно, что условием (1) критическая область $\mathfrak{X}_{1\alpha}$ определяется неоднозначно, и чтобы устранить эту неопределенность, надо ввести дополнительное понятие *ошибок критерия*.

Следуя любому критерию $\mathfrak{X}_{1\alpha}$, мы можем принять правильное решение либо совершить одну из двух ошибок: ошибку 1-го рода, отвергнув H_0, когда она верна, или ошибку 2-го рода, приняв H_0, когда она ложна (ошибка 1-го рода совершается, когда наблюдение X попадает в критическую область $\mathfrak{X}_{1\alpha}$, в то время как верна гипотеза H_0: вероятность этого есть как раз левая часть (1), а ошибка 2-го рода — когда $X \in \overline{\mathfrak{X}}_{1\alpha} = \mathfrak{X}_{0\alpha}$, но гипотеза H_0 не верна). Введем теперь фундаментальное понятие теории проверки статистических гипотез — понятие *функции мощности* критерия. По определению, функцией мощности критерия $\mathfrak{X}_{1\alpha}$ называется следующий функционал на множестве всех допустимых распределений $\mathcal{F} = \{F\}$ выборки X

Функция мощности критерия

$$W(F) = W(F; \mathfrak{X}_{1\alpha}) = \mathbf{P}\{X \in \mathfrak{X}_{1\alpha} | F\}, \quad F \in \mathcal{F} \tag{2}$$

Другими словами, $W(F; \mathfrak{X}_{1\alpha})$ — это вероятность попадания значения выборки X в критическую область $\mathfrak{X}_{1\alpha}$, когда F — ее истинное распределение. Через функцию мощности легко выразить вероятности обоих типов ошибок, свойственных критерию $\mathfrak{X}_{1\alpha}$. Именно, вероятность ошибки 1-го рода есть $W(F)$ при $F \in \mathcal{F}_0$, а 2-го рода — $1 - W(F)$ при $F \in \mathcal{F}_1$. Иногда удобно записывать эти вероятности в символическом виде: $\mathbf{P}\{H_1 | H_0\}$ (вероятность ошибки 1-го рода) и $\mathbf{P}\{H_0 | H_1\}$ (вероятность ошибки 2-го рода).

Естественно стремиться построить критерий так, чтобы свести к минимуму вероятности обоих типов ошибок. Однако при фиксированном объеме выборки сумма вероятностей ошибок обоих типов не может быть сделана как угодно малой (это своего рода *принцип неопределенности* при проверке гипотез). Поэтому **рациональный принцип выбора критической области формулируется следующим образом**: из всех критических областей $\mathfrak{X}_{1\alpha}$ (т. е. удовлетворяющих

условию (1)) выбирается та, для которой вероятность ошибки 2-го рода минимальна, т. е. при условии

$$W(F; \mathfrak{X}_{1\alpha}) \leqslant \alpha, \quad \forall F \in \mathcal{F}_0, \tag{3}$$

надо минимизировать (за счет выбора $\mathfrak{X}_{1\alpha}$) величину $1 - W(F; \mathfrak{X}_{1\alpha})$, $\forall F \in \mathcal{F}_1$, или, что эквивалентно, максимизировать мощность $W(F; \mathfrak{X}_{1\alpha})$, $\forall F \in \mathcal{F}_1$.

Итак, при заданном уровне значимости α наилучшим среди всех критериев $\mathfrak{X}_{1\alpha}$ является тот, который обладает наибольшей мощностью при альтернативах. Если для двух критериев $\mathfrak{X}^*_{1\alpha}$ и $\mathfrak{X}_{1\alpha}$ имеют место соотношения

$$\begin{aligned} &W(F; \mathfrak{X}^*_{1\alpha}) \leqslant W(F; \mathfrak{X}_{1\alpha}), \quad \forall F \in \mathcal{F}_0, \quad \text{и} \\ &W(F; \mathfrak{X}^*_{1\alpha}) \geqslant W(F; \mathfrak{X}_{1\alpha}), \quad \forall F \in \mathcal{F}_1, \end{aligned} \tag{4}$$

то говорят, что критерий $\mathfrak{X}^*_{1\alpha}$ *равномерно мощнее* критерия $\mathfrak{X}_{1\alpha}$; если (4) выполняется для любого критерия $\mathfrak{X}_{1\alpha}$, то $\mathfrak{X}^*_{1\alpha}$ называется *равномерно наиболее мощным критерием* (р. н. м. к.) в задаче (H_0, H_1). Р. н. м. к. идеал, но он не всегда достижим. Поэтому в конкретных задачах зачастую приходится ограничиваться более умеренными требованиями к критериям. Минимальным таким требованием является свойство *несмещенности*, которое означает, что одновременно с (3) должно выполняться условие

Р. н. м. к.

Несмещенность критерия

$$W(F; \mathfrak{X}_{1\alpha}) > \alpha, \quad \forall F \in \mathcal{F}_1, \tag{5}$$

т. е. при любом альтернативном распределении данных мы должны попадать в критическую область с бо́льшей вероятностью, нежели при нулевой гипотезе.

В случае «больших выборок», т. е. когда $X = (X_1, \quad , X_n)$ есть случайная выборка объема n из некоторого распределения $\mathcal{L}(\xi)$ и $n \to \infty$, критерий должен обладать свойством *состоятельности* (чтобы подчеркнуть зависимость от n пишем W_n):

Состоятельность критерия

$$W_n(F; \mathfrak{X}_{1\alpha}) \longrightarrow 1, \quad \forall F \in \mathcal{F}_1 \tag{6}$$

Это свойство очень важно. Оно означает, что, если истинной является некоторая альтернатива, то при большом числе наблюдений мы будем попадать в критическую область с вероятностью, близкой к 1, т. е. отклоняя нулевую гипотезу H_0, мы будем принимать правильное решение. Таким образом, можно сказать, что состоятельный критерий «улавливает» любые отклонения от нулевой гипотезы, когда они имеют место, если число наблюдений (объем выборки) неограниченно возрастает.

«Близкие» альтернативы

Более тонкие свойства состоятельного критерия выражаются в терминах «близких» альтернатив, т. е. когда вместо фиксированной альтернативы $F \in \mathcal{F}_1$, рассматривается последовательность альтернативных распределений $\{F_n\}$, «сближающихся» (в том или ином смысле) с $\mathcal{F}_0$, и исследуется поведение мощности $W_n(F_n; \mathfrak{X}_{1\alpha})$ при $n \to \infty$. В этом случае условие (6), вообще говоря, уже

не будет иметь места, и оно заменяется на условие

$$W_n(F_n; \mathfrak{X}_{1\alpha}) \longrightarrow \gamma, \quad \alpha < \gamma < 1. \tag{7}$$

Если последовательность «близких» альтернатив $\{F_n\}$ такова, что при некотором $\gamma \in (\alpha, 1)$ выполняется (7), то такие альтернативы называются «пороговыми»; они определяют тот «порог чувствительности» критерия $\mathfrak{X}_{1\alpha}$ к отклонениям от нулевой гипотезы H_0, когда критерий еще «реагирует» на такие отклонения: он способен различать распределения $F \in \mathcal{F}_0$ и $F_n \in \mathcal{F}_1$ с ошибками $W_n(F) \leqslant \alpha$ и $1 - W_n(F_n) \approx 1 - \gamma$. Однако в конкретных задачах такое полное исследование асимптотических свойств критериев не всегда осуществимо ввиду аналитических трудностей.

Рандомизированные критерии

В теории часто оказывается полезным рассмотрение так называемых *рандомизированных критериев*, когда при наблюдении $X = x$ гипотеза H_0 отклоняется с некоторой вероятностью $\varphi(x)$ и принимается с дополнительной вероятностью $1 - \varphi(x)$. В этом случае функция $\varphi(x) \in [0, 1]$ называется *критической функцией*, а функция мощности «критерия φ» определяется равенством

$$W(F; \varphi) = \mathbf{E}\big(\varphi(X)|F\big) = \int \varphi(x)\, dF(x).$$

Если функция φ принимает лишь два значения 0 и 1, т. е. является индикаторной функцией некоторого подмножества выборочного пространства, то получаем предыдущую схему *нерандомизированного критерия* с критической областью $\mathfrak{X}_1 = \{x \;\; \varphi(x) = 1\}$ — они обычно и применяются на практике.

Тестовая статистика

В приложениях обычно критическая область задается с помощью некоторой статистики $T = T(X)$, которая «измеряет» отклонение эмпирических данных от соответствующих (гипотезе H_0) гипотетических значений и распределение которой при справедливости гипотезы H_0 должно быть известно (точно или хотя бы приближенно, т. е. для больших выборок). Если $\mathcal{T} = \{t = T(x),\ x \in \mathfrak{X}\}$ — пространство значений этой статистики, то критическая область критерия может быть задана непосредственно в этом пространстве как некоторое подмножество $\mathcal{T}_{1\alpha} \subset \mathcal{T}$, которое должно удовлетворять условию

$$\mathbf{P}\{T(X) \in \mathcal{T}_{1\alpha}|F\} \leqslant \alpha, \quad \forall F \in \mathcal{F}_0, \tag{8}$$

т. е. включать все маловероятные при гипотезе H_0 значения T. Решающим моментом для расчета критерия (т. е. обеспечения (8)) является проблема отыскания распределения статистики $T(X)$ при гипотезе H_0. В частности, если H_0 — сложная гипотеза, то желательно, чтобы распределение $T(X)$ было одним и тем же для всех $F \in \mathcal{F}_0$. Чтобы вычислить полностью функцию мощности $W(F; T) = \mathbf{P}\{T(X) \in \mathcal{T}_{1\alpha}|F\}$, $F \in \mathcal{F}$, а тем самым исследовать и вероятность ошибки 2-го рода, надо знать распределение $T(X)$ также и при альтернативах $F \in \mathcal{F}_1$, что представляет собой, как правило, весьма трудную задачу. Итак, в терминах выбранной статистики $T = T(X)$ критерий имеет вид:

$$\boxed{H_0 \quad \text{отвергается} \quad \Longleftrightarrow \quad T(X) \in \mathcal{T}_{1\alpha},}$$

где при заданном уровне значимости α критическая область $\mathcal{T}_{1\alpha}$ удовлетворяет условию (8). Статистика T называется *статистикой критерия* или *тестовой статистикой*, а сам критерий называется «критерием $\mathcal{T}_{1\alpha}$».

Параметрическая модель

Наконец, отметим, что для параметрических задач, когда допустимые распределения данных X идентифицируются параметром θ, для функции мощности используется обозначение

$$W(\theta) = W(\theta; \mathfrak{X}_{1\alpha}) = \mathbf{P}_\theta\{X \in \mathfrak{X}_{1\alpha}\}, \quad \theta \in \Theta, \tag{9}$$

а проблема построения наилучшего критерия в задаче

$$H_0 \quad \theta \in \Theta_0, \quad H_1 \quad \theta \in \Theta_1 = \Theta \backslash \Theta_0$$

при уровне значимости α формулируется так:

$$W(\theta; \mathfrak{X}_{1\alpha}) \leqslant \alpha, \quad \forall\, \theta \in \Theta_0, \quad W(\theta; \mathfrak{X}_{1\alpha}) \longrightarrow \max, \quad \forall\, \theta \in \Theta_1. \tag{10}$$

Общие замечания

В заключение этого введения сделаем несколько общих замечаний. В конкретных задачах выбор уровня значимости критерия до некоторой степени произволен и связан с практической стороной вопроса. Так, часто ошибочное принятие или отклонение гипотезы H_0 связано с материальными затратами. Если принятие гипотезы H_0 в то время, когда она неверна (ошибка 2-го рода), приводит к большим затратам, тогда как отклонение истинной гипотезы H_0 (ошибка 1-го рода) приводит к небольшим потерям, то ясно, что желательно сделать как можно меньшей вероятность ошибки 2-го рода, допуская сравнительно большие значения α. Обычно для α выбирают одно из следующих стандартных значений: 0,005; 0,01; 0,05; для этих значений рассчитывают соответствующие таблицы, используемые при проведении различных испытаний.

Далее, в отличие от теории оценивания неизвестных параметров распределений, когда для характеризации качества соответствующих оценок достаточно было уметь вычислять, как правило, лишь первые два момента используемых статистик, в теории проверки гипотез требуется знание всего распределения статистики критерия: как минимум — при нулевой гипотезе (чтобы рассчитать критическую область), а вообще же говоря — и при альтернативах (чтобы оценить мощность). В этой связи отметим, что для выборок фиксированного объема (или, как говорят, случая «малых выборок») эта проблема не всегда поддается практическому решению, поэтому при проверке

Большие выборки

гипотез часто применяется *асимптотический подход*, предполагающий неограниченное возрастание объема n выборки $X = (X_1, \quad , X_n)$, т. е. проблема решается для случая «больших выборок». Асимптотический подход широко используется при расчете статистических критериев, часто позволяя исследовать единым методом различные задачи и получать весьма общие результаты. При этом финальные результаты, как правило, имеют простую и наглядную форму, удобную для практических расчетов. Эти результаты в дальнейшем могут использоваться и для выборок

конечного объема, давая приближенное решение задачи. Асимптотический подход — «зачастую единственная возможность, открытая статистику в суровой действительности» (Дж. Русас) и с его применением мы неоднократно будем встречаться в дальнейшем.

Наконец, отметим, что важным показателем каждого критерия является трудоемкость практической реализации соответствующего алгоритма. На практике, когда требуется быстро получить ответ, предпочтение нередко отдается просто реализуемому критерию, даже если он не является оптимальным в теоретическом смысле. При этом типичной является ситуация, когда для одной и той же задачи имеется несколько разработанных критериев, каждый из которых обладает определенными специфическими достоинствами, среди которых не последнее место занимает простота его практической реализации.

Далее изложенные принципы применяются при решении описанных во Введении задач проверки типичных статистических гипотез.



Из истории (по книге Г. Секея [23]). Трудно сказать что-то определенное о том, когда предпринимались первые попытки проверять статистические гипотезы. Б. В. Гнеденко в своей книге [8] отмечает, что учет населения, проведенный в древнем Китае в 2238 г. до н. э., показал, что доля родившихся мальчиков составляла 50 %. Джон Арбутнот (1667–1735), английский математик, врач и писатель, был первым, кто (в 1710 г.) заметил, что гипотеза о равном соотношении родившихся мальчиков и девочек должна быть опровергнута, так как согласно демографическим данным за 82 года (доступным в то время) мальчиков каждый год рождалось больше, чем девочек. Этот факт заинтересовал П. Лапласа. В 1784 г. он с удивлением обнаружил, что в нескольких различных районах доля родившихся мальчиков приблизительно равнялась 22/43, а в Париже это отношение было равно 25/49, что меньше 22/43. Лаплас был заинтригован таким различием, но вскоре нашел для него разумное объяснение: в общее число родившихся в Париже включались также все подкидыши, а население из пригородов предпочитало подкидывать младенцев одного пола. Когда Лаплас исключил подкидышей из общего числа родившихся, доля новорожденных мальчиков стала близкой к 22/43.

В 1734 г. французская академия присудила Даниилу Бернулли (1700–1782) — швейцарскому физику и математику — премию за исследование по орбитам планет. С помощью некоторого критерия проверки гипотез Бернулли пытался показать, что схожесть орбит планет является далеко не случайной (в его модели каждая орбита соответствует некоторой точке на единичной сфере, и Бернулли проверял гипотезу о равномерном распределении этих точек). В 1812 г. Лаплас исследовал похожую проблему. Он пытался применить статистические методы для решения вопроса о том, какую из гипотез следует принять: являются ли кометы обычными элементами Солнечной системы или они всего лишь «незванные гости». В последнем случае углы между орбитами комет и эклиптикой были бы равномерно распределены на интервале от 0 до $\pi/2$, что как раз совпадает с математической записью предположения Лапласа (он обнаружил, что кометы не являются обычными элементами Солнечной системы). Основоположниками современной теории проверки статистических гипотез были К. Пирсон, Э. Пирсон, Р. Фишер и Ю. Нейман. Большой вклад в ее становление и развитие внесли Г. Крамер, Р. фон Мизес, А. Н. Колмогоров, Н. В. Смирнов и другие ученые, работавшие позднее.

Хотя это может показаться парадоксом,
вся наука основана на идее аппроксимации.

Бертран Рассел [2)]

§ 4.2. Проверка гипотезы о виде распределения

Классика — это то, что все хотели бы прочитать, но никто читать не хочет.

Марк Твен [3)]

Мы начинаем изложение основных результатов теории проверки гипотез с рассмотрения классической задачи о виде распределения наблюдений (см. пример 4 Введения). В простейшем ее варианте задача ставится так: пусть $\boldsymbol{X} = (X_1, \quad, X_n)$ — выборка из распределения $\mathcal{L}(\xi)$ с неизвестной функцией распределения $F_\xi(x)$, о которой выдвинута простая гипотеза $H_0 \quad F_\xi(x) = F(x)$, где функция $F(x)$ полностью задана. Здесь альтернативные распределения никак не конкретизируются, т. е. допустимыми распределениями данных (выборки $\boldsymbol{X}$) являются любые распределения на выборочном пространстве $\mathfrak{X} = \{\underline{x} = (x_1, \quad, x_n)\}$ (априори может быть ясен лишь тип распределения F_ξ — абсолютно непрерывный либо дискретный, что определяется обычно физической природой наблюдаемой случайной величины ξ). Следовательно, в данном случае альтернатива $H_1 = \overline{H}_0$ («не H_0»), и речь идет просто о согласии данных и гипотезы H_0. Наиболее известными (классическими) критериями согласия в данной задаче являются критерий Колмогорова и критерий хи-квадрат.

1. Критерий согласия Колмогорова

Статистика Колмогорова

Этот критерий применяют в тех случаях, когда функция $F(x)$ непрерывна. Статистика критерия определяется формулой

$$D_n = D_n(X) = \sup_{-\infty<x<\infty} \left|\widehat{F}_n(x) - F(x)\right|, \tag{1}$$

т. е. представляет собой максимальное отклонение эмпирической функции распределения $\widehat{F}_n(x)$, построенной по выборке $\boldsymbol{X}$, от гипотетической (т. е. определяемой гипотезой H_0) функции распределения $F(x)$. То, что это «разумная» статистика для проверки гипотезы H_0, следует из известных нам свойств э. ф. р. $\widehat{F}_n(x)$. Так, мы знаем (см. теорему 2 § 3.1 и комментарий

[2)] Рассел Бертран Артур Уильям (1872–1970) — английский математик, философ, логик, социолог, общественный деятель.

[3)] Твен Марк (Сэмюэль Ленгхорн Клеменс) (1835–1910) — знаменитый американский писатель.

к ней), что при каждом x величина $\widehat{F}_n(x)$ является оптимальной несмещенной оценкой для $F(x)$, и эта оценка состоятельна (см. комментарий к формуле (23) § 2.2), т. е. с увеличением объема выборки n происходит сближение $\widehat{F}_n(x)$ с $F(x)$ (если, конечно, гипотеза H_0 справедлива). Поэтому, по крайней мере при больших n, в тех случаях, когда гипотеза H_0 истинна, значение D_n не должно существенно отклоняться от нуля. Отсюда следует, что критическую область критерия, основанного на статистике $T = D_n$, следует задавать в виде $\mathcal{T}_{1\alpha} = \{t \geqslant t_\alpha\}$, т. е. большие значения D_n надо интерпретировать как свидетельство против проверяемой гипотезы H_0. Критическая граница t_α при заданном уровне значимости α рассчитывается при этом на основании теоремы Колмогорова (см. теорему 2 § 2.1). Именно, если n достаточно велико (уже при $n \geqslant 20$), то положив $t_\alpha = \lambda_\alpha/\sqrt{n}$, где (см. (11) § 2.1) $K(\lambda_\alpha) = 1 - \alpha$, будем иметь

$$\mathbf{P}\{D_n \in \mathcal{T}_{1\alpha} | H_0\} = \mathbf{P}\{\sqrt{n}D_n \geqslant \lambda_\alpha | H_0\} \approx 1 - K(\lambda_\alpha) = \alpha, \qquad (2)$$

т. е. по крайней мере для больших выборок условие (8) § 4.1 будет приближенно выполняться.

Критерий Колмогорова

Тем самым *критерий согласия Колмогорова* формулируется следующим образом: если $n \geqslant 20$ и при выбранном уровне значимости α число λ_α определено соотношением $K(\lambda_\alpha) = 1 - \alpha$, то

$$\boxed{H_0 \quad \text{отвергается} \quad \Longleftrightarrow \quad \sqrt{n}D_n \geqslant \lambda_\alpha.} \qquad (3)$$

Следуя этому правилу, можно ошибочно отклонить гипотезу H_0, когда она верна, с вероятностью, приблизительно равной α. Этот критерий состоятелен (альтернативой здесь является любое непрерывное распределение $F_\xi \not\equiv F$).

Комментарии

Дополним сказанное некоторыми комментариями. Прежде всего подчеркнем, что распределение статистики D_n при гипотезе H_0 не зависит от вида функции $F(x)$. Этот факт имеет принципиальное значение, так как имея таблицы значений функции Колмогорова $K(t)$, мы имеем возможность рассчитывать критерий (3) для проверки гипотезы относительно произвольной непрерывной функции распределения $F(x)$. Такие подобные таблицы имеются в книге [3], где также даны детальные практические рекомендации по их использованию.

Далее, по своей сути критерий Колмогорова (3) является асимптотическим (т. е. он «работает» лишь для больших выборок), поскольку в основе его лежит предельная теорема 2 § 2.1. Но помимо вывода предельного результата (см. (11) § 2.1) Колмогоров (1933) дал рекуррентные соотношения для конечных n, которые затем были использованы для табулирования точного распределения статистики D_n. Поэтому для расчета точных значений критической границы t_α в случае малых выборок также можно воспользоваться соответствующими таблицами, информация о которых имеется в [12, с. 611].

Наконец, отметим, что для практических вычислений статистики D_n полезна эквивалентная (1) формула $D_n = \max(D_n^+, D_n^-)$, где

$$D_n^+ = \max_{1 \leqslant k \leqslant n} \left(\frac{k}{n} - F(X_{(k)}) \right), \quad D_n^- = \max_{1 \leqslant k \leqslant n} \left(F(X_{(k)}) - \frac{k-1}{n} \right)$$

и $X_{(1)} \leqslant X_{(2)} \leqslant \quad \leqslant X_{(n)}$ — вариационный ряд выборки X.

Замечание. Описанную методику проверки гипотезы о виде распределения наблюдаемой случайной величины можно распространить и на случай проверки сложной гипотезы $H_0 \quad F_\xi(x) \in \mathcal{F}_0 = \{F(x;\theta),\ \theta \in \Theta\}$, где $\mathcal{F}_0$ — заданное параметрическое семейство распределений (комментарии см. в примере 4 Введения). В этом случае вместо (1) используют тестовую статистику

$$\widehat{D}_n = \sup_{-\infty < x < \infty} \left| \widehat{F}_n(x) - F(x; \widehat{\theta}_n) \right|,$$

где $\widehat{\theta}_n$ — оценка максимального правдоподобия параметра θ. Если семейство $\mathcal{F}_0$ регулярно (см. п. 4 § 3.5), то известно асимптотическое (при $n \to \infty$) распределение статистики $\widehat{D}_n$, которым можно воспользоваться для расчета соответствующей критической границы критерия. Однако это распределение имеет уже отличный от (11) § 2.1 и гораздо более сложный вид, отражающий специфические свойства семейства $\mathcal{F}_0$; тем самым расчет соответствующего критерия значительно усложняется [4)]

2. Критерий согласия хи-квадрат К. Пирсона

Одним из наиболее универсальных методов проверки различных статистических гипотез является *метод хи-квадрат* (χ^2). В этом методе работают с дискретными данными, но поскольку любые исходные данные можно свести к дискретным методом группировки наблюдений, как это описано в п. 2 § 3.6, т. е. перейти от исходной выборки $X \quad (X_1, \quad , X_n)$ к частотам $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ попадания ее элементов в соответствующие подмножества группировки $\mathcal{E}_1, \quad , \mathcal{E}_N$, то тем самым метод χ^2 можно применять к данным любой природы, в том числе и многомерным (подчеркнем, что описанный в п. 1 критерий Колмогорова «работает» лишь с выборками из непрерывного одномерного распределения).

Итак, пусть в эксперименте наблюдается дискретная случайная величина ξ, принимающая значения $1, 2, \quad , N$ с некоторыми вероятностями $p_1, \quad , p_N$ $(p_1 + \quad + p_N = 1)$. Если произведено n независимых испытаний над ξ, т. е. имеется выборка $(\xi_1, \quad , \xi_n)$, то пусть

$$\nu_j = \sum_{i=1}^{n} I(\xi_i = j), \quad j = 1, \quad , N,$$

— соответствующие частоты исходов $(\nu_1 + \quad + \nu_N = n)$. Тогда вектор $\underline{\nu} = (\nu_1, \quad , \nu_N)$ имеет полиномиальное распределение $M(n; p_1, \quad , p_N)$

[4)] *Тюрин Ю. Н.* О предельном распределении статистик Колмогорова—Смирнова для сложной гипотезы // Изв. АН СССР. Сер. Матем. 1984. Т. 48. № 6. С. 1314–1343.

(см. п. 6 § 1.1). В методе χ^2 работают именно с данными $\underline{\nu}$, а проверяемые гипотезы формулируются в терминах вектора вероятностей $\underline{p} = (p_1, \quad , p_N)$. По сути — это параметрическая модель, так как она определяется конечным числом параметров $(p_1, \quad , p_{N-1})$. Рассмотрим сначала задачу проверки простой гипотезы в рамках этой модели.

Итак, пусть по наблюдению вектора частот $\underline{\nu} = (\nu_1, \quad , \nu_N)$ требуется проверить простую гипотезу $H_0 \quad \underline{p} = \underline{p}^\circ$ где $\underline{p}^\circ = (p_1^\circ, \quad , p_N^\circ)$ — заданный вероятностный вектор $(0 < p_j^\circ < 1, j = 1, \quad , N, p_1^\circ + \ldots + p_N^\circ = 1)$. К. Пирсон в 1900 г. предложил использовать в качестве меры отклонения эмпирических данных (относительных частот $\underline{\nu}/n$) от гипотетических значений $\underline{p}^\circ$ меру хи-квадрат (см. (2) § 3.6)

Мера χ^2

$$\mathring{X}_n^2 = \mathring{X}_n^2(\underline{\nu}) = \sum_{j=1}^{N} \frac{(\nu_j - np_j^\circ)^2}{np_j^\circ} = \sum_{j=1}^{N} \frac{\nu_j^2}{np_j^\circ} - n. \tag{4}$$

На этой тестовой статистике и основывается знаменитый *критерий* χ^2 *К. Пирсона*. В основе этого критерия лежат следующие очевидные соображения. Если гипотеза H_0 справедлива, то, поскольку относительная частота ν_j/n события $\{\xi = j\}$ является состоятельной оценкой его вероятности p_j° $(j = 1, \ldots, n)$, при больших n разности $|\nu_j/n - p_j^\circ|$ должны быть малы, следовательно, и значение статистики $\mathring{X}_n^2$ не должно быть слишком большим. Поэтому естественно задать критическую область для гипотезы H_0 в виде $\mathcal{T}_{1\alpha} = \{\mathring{X}_n^2 > t_\alpha\}$, где критическая граница t_α при заданном уровне значимости α должна быть выбрана из условия

$$\mathrm{P}\{\mathring{X}_n^2 > t_\alpha | H_0\} = \alpha. \tag{5}$$

Такой критической областью и определяется критерий χ^2 Главная проблема здесь — вычисление границы t_α в (5), для чего надо знать распределение статистики $\mathring{X}_n^2$ при нулевой гипотезе H_0. Точное распределение $\mathcal{L}(\mathring{X}_n^2 | H_0)$ неудобно для расчета критерия, но для больших объемов выборок n статистика $\mathring{X}_n^2$ имеет при гипотезе H_0 простое предельное распределение, не зависящее от гипотезы (т. е. от $\underline{p}^\circ$). Именно, справедливо следующее утверждение.

Теорема 1. Если $0 < p_j^\circ < 1$, $j = 1, \quad , N$, то при $n \to \infty$

$$\mathcal{L}(\mathring{X}_n^2 | H_0) \to \chi^2(N-1).$$

Доказательство. Как показано в примере 12 § 3.5, при сформулированных условиях вектор нормированных частот $\underline{\nu}^* = (\nu_1^*, \quad , \nu_{N-1}^*)$, где

$$\nu_j^* = \frac{\nu_j - np_j^\circ}{\sqrt{n}}, \quad j = 1, \ldots, N,$$

имеет в пределе невырожденное нормальное распределение $\mathcal{N}(\underline{0}, \mathring{\Sigma}_{N-1})$, где

$$\mathring{\Sigma}_{N-1} = \left\|\sigma_{ij}^{\circ} = p_i^{\circ}\delta_{ij} - p_i^{\circ}p_j^{\circ}\right\|_1^{N-1}$$

δ_{ij} — символ Кронеккера. Отсюда на основании соотношения (34) § 3.5 можно заключить, что при $n \to \infty$

$$\mathcal{L}\left(Q_n \equiv \underline{\nu}_n^{*\prime}\mathring{\Sigma}_{N-1}^{-1}\underline{\nu}_n^{*} | H_0\right) \to \chi^2(N-1).$$

С другой стороны, из (4) имеем

$$\mathring{X}_n^2 = \sum_{j=1}^{N-1}\frac{(\nu_j^*)^2}{p_j^{\circ}} + \frac{(\nu_1^* + \quad + \nu_{N-1}^*)^2}{p_N^{\circ}} = \underline{\nu}_n^{*\prime}\mathring{\mathcal{I}}\underline{\nu}_n^{*},$$

где матрица

$$\mathring{\mathcal{I}} = \left\|g_{ij}^{\circ}\right\|_1^{N-1} = \mathring{\Sigma}_{N-1}^{-1}$$

(см. пример 10 в § 3.2). Таким образом, $\mathring{X}_n^2 = Q_n$. ■

На практике предельное распределение $\chi^2(N-1)$ можно использовать для расчетов с хорошим приближением уже при $n \geqslant 50$ и $\nu_j \geqslant 5$, $j = 1, \quad , N$. При выполнении этих условий можно записать, что

$$\mathbf{P}\{\mathring{X}_n^2 > t_\alpha | H_0\} \approx 1 - F_{N-1}(t_\alpha),$$

где $F_{N-1}(t)$ — функция распределения закона $\chi^2(N-1)$. Следовательно, полагая

$$1 - F_{N-1}(t_\alpha) = \alpha,$$

т. е.

$$t_\alpha = \chi^2_{1-\alpha, N-}$$

(напомним, что $\chi^2_{p,r}$ обозначает p-квантиль распределения $\chi^2(r)$), получаем асимптотический вариант критерия χ^2 задаваемый критической областью $\{\mathring{X}_n^2 > \chi^2_{1-\alpha,N-1}\}$.

Таким образом, классический критерий согласия χ^2 имеет следующий вид: пусть объем выборки n и наблюдавшиеся значения вектора частот $\underline{\nu} = (\nu_1, \quad , \nu_N)$ удовлетворяют условиям $n \geqslant 50$, $\nu_j \geqslant 5$, $\forall j$; тогда при заданном уровне значимости α

$$\boxed{H_0 \quad \text{отвергается} \iff \{\mathring{X}_n^2 > \chi^2_{1-\alpha,N-1}\},} \tag{6}$$

Критерий χ^2

где статистика $\mathring{X}_n^2$ определена в (4).

Комментарии

Сделаем несколько общих замечаний. Критерий согласия χ^2 применяется в тех случаях, когда в каждом испытании наблюдается одно из N несовместных событий $A_1, \ldots, A_N$ и заданы частоты появления этих событий в n (независимых) испытаниях (можно говорить также, что наблюдается дискретная случайная

величина ≡ номер реализуемого в испытании события). Если же исходные данные представляют собой выборку из некоторого непрерывного распределения, то, применяя предварительно метод группировки наблюдений, приходят к рассмотрению дискретной схемы, в которой в качестве событий A_j рассматриваются события $\{\xi \in \mathcal{E}_j\}$, где $\mathcal{E}_1, \quad , \mathcal{E}_N$ — интервалы группировки. Недостатком метода является то, что при группировке данных происходит некоторая потеря информации. Кроме того, остается еще вопрос о выборе числа интервалов N и их виде (более детально эти вопросы освещены в [12, гл. 30]). Однако критерий χ^2 имеет и свои достоинства: при его применении нет необходимости учитывать точные значения наблюдений (бывают случаи, когда исходные данные имеют не числовой характер, см. пример 2 ниже); несомненными преимуществами критерия являются его простота, наглядность и универсальность.

Приведем несколько примеров практического применения критерия χ^2

Пример 1. При $n = 4040$ бросаниях монеты Бюффон получил $\nu_1 = 2048$ выпадений герба и $\nu_2 = n - \nu_1 = 1992$ выпадений решки. Проверим, используя критерий χ^2 совместимы ли эти данные с гипотезой H_0 о симметричности монеты, т. е. что вероятность выпадения герба $p = 1/2$. Здесь $N = 2$, $p_1^\circ = p_2^\circ \quad 1/2$ и по формуле (4) имеем

Эксперимент Бюффона

$$\mathring{\chi}_n^2 = \frac{(\nu_1 - np_1^\circ)^2}{np_1^\circ} + \frac{(\nu_2 - np_2^\circ)^2}{np_2^\circ} = \frac{(\nu_1 - np_1^\circ)^2}{np_1^\circ p_2^\circ} = 0{,}776.$$

Пусть уровень значимости $\alpha = 0{,}05$. Из таблиц распределения χ^2 находим $\chi^2_{0{,}95;\,1} = 3{,}841$. Следовательно, данные не противоречат гипотезе H_0. •

Пример 2 [12, с. 563]. В экспериментах с селекцией гороха Мендель (см. ниже) наблюдал частоты различных видов семян, получаемых при скрещивании растений с круглыми желтыми семенами и растений с морщинистыми зелеными семенами. Эти данные и значения теоретических вероятностей, определяемые в соответствии с теорией наследственности Менделя, приведены в следующей таблице:

Эксперимент Менделя

Семена	Частоты ν_j	Вероятности p_j°
Круглые и желтые	315	9/16
Морщинистые и желтые	101	3/16
Круглые и зеленые	108	3/16
Морщинистые и зеленые	32	1/16
Σ	$n = 556$	1

Следует проверить гипотезу H_0 о согласовании частотных данных с теоретическими вероятностями. Вычисления по формуле (4) дают здесь $\mathring{X}_n^2 = 0{,}47$ Из таблиц распределения $\chi^2(3)$ следует, что при любом уровне значимости $\alpha \leqslant 0{,}90$ критерий (6) не отвергает гипотезу, так что между наблюдениями и гипотезой имеется очень хорошее согласие. •



Из истории науки. С именем монаха августинского монастыря Грегора Менделя [5] связан поучительный эпизод из истории научных открытий. В 1866 г. Мендель открыл законы классической генетики. Но открыть мало — их нужно доказать. И он провел серию опытов на растении с названием ястребинка. Почему из тысячи видов он остановился именно на этом растении, сказать трудно. Факт в том, что опыты не удались, и монах сделал вывод, что его законы неверны. Впоследствии выяснилось, что Мендель выбрал неудачный объект исследований. Но дело было сделано — имя монаха и его законы предали забвению. В 1900 г. эти же законы были заново открыты сразу тремя ботаниками — Хуго де Фризом, Карлом Корренсом и Эрихом Чермаком. Казалось бы, им ничего не мешало закрепить за собой лавры первооткрывателей. Но с непомерным усердием они принялись рыться в научных архивах и нашли забытые статьи о «правилах Менделя». Ученые не прекратили поиск, а, наоборот, углубились в него. И пришли к неожиданному выводу: «их» законы за 30 лет до них открыл тот самый монах из августинского монастыря. Что ботаники делают дальше? Все просто: они признают за собой «приоритет вторичного открытия», а законы нарекают именем выдающегося монаха.

> *Предмет математики настолько серьезен, что полезно не упускать случая сделать его немного занимательным.*
>
> **Б. Паскаль**

Рассмотрим, наконец, пример, когда гипотетическое распределение является непрерывным.

Пример 3. Используя данные, приведенные в упр. 10 к гл. 2, проверим согласуются ли они с гипотезой H_0 о том, что показания часов равномерно распределены на интервале $(0, 12)$. Здесь $N = 12$ и при гипотезе H_0 $p_j^\circ = 1/12$, $j = 0, \quad 11$. Значение статистики (4) есть 10,000, а, например, $\chi^2_{0{,}95;\,11} = 19{,}675$, поэтому следует признать, что согласие предположения с опытными данными хорошее. •

Состоятельность критерия χ^2 Для критерия χ^2 можно провести полное исследование предельного при $n \to \infty$ поведения его мощности при альтернативах. В рассматриваемой методике гипотезы характеризуется вектором $\underline{p} = (p_1, \quad , p_N)$ вероятностей, с которыми появляются в каждом испытании события $A_1, \quad , A_N$, поэтому для

[5] Мендель Грегор Иоганн (1822–1884) — австрийский естествоиспытатель, основоположник учения о наследственности.

функции мощности критерия (6) будем использовать обозначение

$$W_n(\underline{p}) = \mathbf{P}\{\mathring{X}_n^2 > \chi^2_{1-\alpha, N-1} | \underline{p}\}, \tag{7}$$

а о соответствующей гипотезе будем говорить для краткости как о гипотезе $\underline{p}$. В этих обозначениях полученный выше результат записывается в виде:

$$W_n(\underline{p}^\circ) \to \alpha \quad \text{при} \quad n \to \infty.$$

Установим теперь следующий результат.

Теорема 2. Для любого вектора $\underline{p} \neq \underline{p}^\circ$ выполняется условие $W_n(\underline{p}) \to 1$ при $n \to \infty$, т. е. критерий χ^2 является состоятельным.

Доказательство. Вычислим среднее и дисперсию статистики $\mathring{X}_n^2$ при произвольной гипотезе $\underline{p}$. Для этого перепишем формулу (4) в виде

$$\mathring{X}_n^2 = \sum_{j=1}^N \frac{(\nu_j - np_j)^2}{np_j^\circ} + 2\sum_{j=1}^N \frac{(\nu_j - np_j)(p_j - p_j^\circ)}{p_j^\circ} + n\sum_{j=1}^N \frac{(p_j - p_j^\circ)^2}{p_j^\circ}.$$

Так как

$$\mathbf{E}(\nu_j | \underline{p}) = np_j, \quad \mathbf{E}\big[(\nu_j - np_j)^2 | \underline{p}\big] = \mathbf{D}(\nu_j | \underline{p}) = np_j(1 - p_j),$$

то

$$\mathbf{E}(\mathring{X}_n^2 | \underline{p}) = n\sum_{j=1}^N \frac{(p_j - p_j^\circ)^2}{p_j^\circ} + \sum_{j=1}^N \frac{p_j(1 - p_j)}{p_j^\circ}. \tag{8}$$

Отсюда, в частности, имеем $\mathbf{E}(\mathring{X}_n^2 | \underline{p}^\circ) = N - 1$. Этот результат согласуется с асимптотическим результатом теоремы 1, поскольку среднее предельного распределения $\chi^2(N-1)$ равно $N - 1$ (см п. 3 § 1.2). Аналогично можно вывести и формулу для дисперсии:

$$\mathbf{D}(\mathring{X}_n^2 | \underline{p}) = 4\frac{(n-1)(n-2)}{n}(R_{32} - R_{21}^2) + \\ + 2\frac{n-1}{n}(3R_{22} - 2R_{21}R_{11} - R_{21}^2) + \frac{1}{n}(R_{12} - R_{11}^2), \tag{9}$$

где

$$R_{ks} = \sum_{j=1}^N \frac{p_j^k}{p_j^{\circ s}}.$$

Если $\underline{p} = \underline{p}^\circ$ то $R_{ks} = \sum\limits_{j=1}^N (p_j^\circ)^{k-s}$ и из (9) имеем

$$\mathbf{D}(\mathring{X}_n^2 | \underline{p}^\circ) = 2(N-1) + \frac{1}{n}\left(\sum_{j=1}^N \frac{1}{p_j^\circ} - N^2 - 2N + 2\right) \to 2(N-1)$$

при $n \to \infty$, что также согласуется с теоремой 1. Если же $\underline{p} \neq \underline{p}^\circ$, то из (8) и (9) следует, что при $n \to \infty$ среднее и дисперсия статистики $\mathring{X}_n^2$ при гипотезе $\underline{p}$ имеют порядок роста n. С учетом этого из (7) по неравенству Чебышева (при $t = \chi^2_{1-\alpha, N-1}$) имеем

$$1 - W_n(\underline{p}) = \mathbf{P}\{\mathring{X}_n^2 \leqslant t|\underline{p}\} = \mathbf{P}\left\{\left|\mathring{X}_n^2 - \mathbf{E}(\mathring{X}_n^2|\underline{p})\right| \geqslant \left|\mathbf{E}(\mathring{X}_n^2|\underline{p}) - t\right| \,\middle|\, \underline{p}\right\} \leqslant$$

$$\leqslant \frac{\mathbf{D}(\mathring{X}_n^2|\underline{p})}{\left(\mathbf{E}(\mathring{X}_n^2|\underline{p}) - t\right)^2} = 0\left(\frac{1}{n}\right) \quad \blacksquare$$

«Близкие» альтернативы и мощность критерия χ^2

Тем самым, если гипотеза H_0 неверна, а истинной является некоторая альтернатива $\underline{p}$, то критерий (6) обнаружит это с вероятностью, близкой к 1, когда n велико. Но что будет, если альтернатива $\underline{p}$ «мало отличается» от $\underline{p}^\circ$, т. е. если $\underline{p} = \underline{p}(n) \to \underline{p}^\circ$ при $n \to \infty$? Что собой представляют в данном случае «пороговые» альтернативы, т. е. для которых выполняется условие (7) § 4.1, которое в наших обозначениях запишется в виде

$$W_n(\underline{p}(n)) \to \gamma, \quad \alpha < \gamma < 1?$$

Ответ на эти вопросы дает следующее утверждение.

Теорема 3. Пусть альтернатива имеет вид

$$\underline{p}(n) = \underline{p}^\circ + \frac{\underline{\beta}}{\sqrt{n}},$$

где $\underline{\beta} = (\beta_1, \quad , \beta_N) \neq \underline{0}$ — фиксированный вектор, задающий отклонение от гипотезы H_0 $\left(\sum_{j=1}^{N} \beta_j = 0\right)$ Тогда при $n \to \infty$

$$W_n(\underline{p}(n)) \to 1 - F_{N-1}(\chi^2_{1-\alpha, N-1}; \lambda^2), \quad \lambda^2 = \sum_{j=1}^{N} \frac{\beta_j^2}{p_j^\circ}, \tag{10}$$

где $F_{N-1}(t; \lambda^2)$ — функция нецентрального распределения хи-квадрат

$$\chi^2(N-1; \lambda^2)$$

(см п. 2 § 2.5).

Доказательство. Как и при доказательстве теоремы 1, можно показать, что при гипотезе $\underline{p}(n)$ распределение вектора нормированных частот

$$\left(\frac{\nu_j - np_j(n)}{\sqrt{n}}, \; j = 1, \ldots, N-1\right)$$

сходится при $n \to \infty$ к нормальному распределению $\mathcal{N}(\underline{0}, \mathring{\Sigma}_{N-1})$. Но

$$\nu_j^* = \frac{\nu_j - np_j^\circ}{\sqrt{n}} = \frac{\nu_j - np_j(n)}{\sqrt{n}} + \beta_j,$$

поэтому

$$\mathcal{L}(\underline{\nu}^* | \underline{p}(n)) \to \mathcal{N}(\underline{\beta}_{N-1}, \mathring{\Sigma}_{N-1}), \quad \underline{\beta}_{N-1} = (\beta_1, \ldots, \beta_{N-1}).$$

А так как

$$\mathring{X}_n^2 = \underline{\nu}^{*\prime} \mathring{\Sigma}_{N-1}^{-1} \underline{\nu}^*,$$

то по лемме 3 § 2.5 (точнее, на основании асимптотического варианта этой леммы)

$$\mathcal{L}(\mathring{X}_n^2 | \underline{p}(n)) \to \chi^2(N-1; \lambda^2),$$

где

$$\lambda^2 = \underline{\beta}'_{N-1} \mathring{\Sigma}_{N-1}^{-1} \underline{\beta}_{N-1} = \sum_{j=1}^{N} \frac{\beta_j^2}{p_j^\circ}$$

(проверяется непосредственно). Отсюда и из (7) следует (10). ■

Таким образом, критерий χ^2 «реагирует» на отклонения от нулевой гипотезы, имеющие при больших n порядок $n^{-1/2}$ — это и есть «пороговые» альтернативы для данного критерия; более близкие альтернативы, т. е. отклоняющиеся от H_0 на величину $o(n^{-1/2})$ (в этом случае $\lambda^2 \to 0$), критерий не отличает от H_0, так как в таких случаях $W_n(\underline{p}(n)) \to \alpha$ (вероятность попадания в критическую область при альтернативе асимптотически такая же, как и при основной гипотезе); для более же удаленных альтернатив (т. е. при $\lambda^2 \to \infty$) критерий остается состоятельным: $W_n(\underline{p}(n)) \to 1$.

Тем самым при простой гипотезе H_0 удается провести исчерпывающее исследование асимптотических свойств критерия χ^2 для больших выборок.

3. Критерий хи-квадрат для сложной гипотезы

Сложные гипотезы

Сложные гипотезы для полиномиального распределения $M(n; \underline{p})$ в общем случае имеют следующий вид:

$$H_0 \quad \underline{p} = \underline{p}(\theta), \quad \theta = (\theta_1, \ldots, \theta_r) \in \Theta, \quad r < N - 1.$$

Таким образом, при гипотезе H_0 вероятности исходов являются некоторыми функциями от неизвестного параметра θ с множеством возможных значений Θ. Чтобы построить критерий проверки такой гипотезы по наблюдению вектора частот $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ можно, по аналогии с предыдущим случаем, поступить следующим образом. Построим статистику, аналогичную (4),

$$X_n^2(\theta) = \sum_{j=1}^{N} \frac{(\nu_j - np_j(\theta))^2}{np_j(\theta)}.$$

Эта статистика зависит от неизвестного параметра θ, поэтому непосредственно использовать ее нельзя — требуется предварительно исключить неопределенность в θ. Для этого заменяют θ некоторой оценкой $\widetilde{\theta}_n = \widetilde{\theta}_n(\underline{\nu})$ и получают в итоге статистику $\widetilde{X}_n^2 = X_n^2(\widetilde{\theta}_n)$. Это уже функция только от наблюдений $\underline{\nu}$, следовательно, ее значение можно однозначно вычислить для каждой реализации вектора $\underline{\nu}$. Однако, чтобы рассчитать соответствующий критерий, надо знать распределение $\widetilde{X}_n^2$ при гипотезе H_0 (хотя бы приближенно). Положение осложняется тем, что теперь величины $p_j(\widetilde{\theta}_n)$ уже не постоянные, а представляют собой функции от наблюдений $\underline{\nu}$, поэтому теорема 1 к статистике $\widetilde{X}_n^2$ неприменима. Более того, следует ожидать, что распределение $\widetilde{X}_n^2$ будет, вообще говоря, зависеть от способа построения оценки $\widetilde{\theta}_n$.

Проблема нахождения предельного (при $n \to \infty$) распределения для $\widetilde{X}_n^2$ при этих усложненных условиях была решена американским статистиком Р. Фишером в 1924 г. Он показал, что существуют методы оценивания параметра θ, при которых это предельное распределение имеет простой вид, а именно, является распределением $\chi^2(N - 1 - r)$. В частности, это будет иметь место при использовании оценки максимального правдоподобия

$$\widehat{\theta}_n = \arg\max_{\theta} \prod_{j=1}^{N} (p_j(\theta))^{\nu_j}$$

или, что то же самое, оценки по видоизмененному методу минимума хи-квадрат (см п. 2 § 3.6), которая является решением системы уравнений (3) § 3.6. В этом случае для построения критерия используется тестовая статистика

$$\widehat{X}_n^2 = X_n^2(\widehat{\theta}_n) = \sum_{j=1}^{N} \frac{(\nu_j - np_j(\widehat{\theta}_n))^2}{np_j(\widehat{\theta}_n)}. \tag{11}$$

Сам критерий рассчитывается на основании следующего утверждения.

Теорема 4. Пусть функции $p_j(\theta)$, $j = 1, \quad , N$, удовлетворяют условиям:

1) $\sum_{j=1}^{N} p_j(\theta) = 1, \forall\, \theta \in \Theta$;

2) $p_j(\theta) \geqslant c > 0, \forall\, j$, и существуют непрерывные производные

$$\frac{\partial p_j(\theta)}{\partial \theta_k}, \quad \frac{\partial^2 p_j(\theta)}{\partial \theta_k \partial \theta_l}, \quad k, l = 1, \quad , r;$$

3) $(N \times r)$-матрица $\left\| \partial p_j(\theta)/\partial \theta_k \right\|$ имеет ранг r для всех $\theta \in \Theta$. Тогда

$$\mathcal{L}(\widehat{X}_n^2 | H_0) \to \chi^2(N - 1 - r),$$

где статистика $\widehat{X}_n^2$ определена в (11).

Доказательство этой теоремы можно найти в [15, с. 462–470]. При выполнении условий теоремы 4 критерий имеет вид:

Критерий χ^2 для сложной гипотезы

$$\boxed{H_0 \text{ отвергается} \iff \widehat{X}_n^2 > \chi^2_{1-\alpha, N-1-r}} \tag{12}$$

(здесь α — заданный уровень значимости и предполагается, что $n \geqslant 50$ и $\nu_j \geqslant 5, \forall j$).

Этот критерий можно применять также в следующей общей ситуации. Пусть задана выборка $X = (X_1, \ldots, X_n)$ из некоторого распределения $\mathcal{L}(\xi)$, о котором требуется проверить гипотезу H_0 $\mathcal{L}(\xi) \in \mathcal{F}_0$, где $\mathcal{F}_0 = \{F(x;\theta), \theta \in \Theta\}$ — заданное параметрическое семейство. Проведем группировку наблюдений, основываясь на интервалах $\mathcal{E}_1, \quad, \mathcal{E}_N$ ($\mathcal{E}_i \cap \mathcal{E}_j \quad \varnothing$, $i \neq j$, $\mathcal{E}_1 \cup \quad \cup \mathcal{E}_N = \mathcal{E}$ — множество возможных значений ξ), т. е. подсчитаем частоты

$$\nu_j = \sum_{i=1}^{n} I(X_i \in \mathcal{E}_j), \quad j = 1, \ldots, N,$$

и гипотетические вероятности

$$p_j(\theta) = \mathbf{P}\{\xi \in \mathcal{E}_j | H_0\} = \int_{\mathcal{E}_j} dF(x;\theta), \quad j = 1, \quad, N$$

Решив далее систему (3) § 3.6, найдем оценку $\widehat{\theta}_n$ и вычислим значение статистики $\widehat{X}_n^2$ в (11). Если выполнены условия теоремы 4, то применяется критерий (12). Следуя этому критерию, можно ошибочно отклонить гипотезу H_0, когда она верна, с вероятностью, приближенно равной α.

Пример 4. Среди 2020 семей, имеющих двух детей, 527 семей, в которых два мальчика, и 476 — две девочки (в остальных 1017 семьях дети разного пола). Можно ли с уровнем значимости $\alpha = 0{,}05$ считать, что количество мальчиков в семье с двумя детьми (ξ) — биномиальная случайная величина? Здесь речь идет о проверке гипотезы H_0 $\mathcal{L}(\xi) \in \mathrm{Bi}(2, \theta)$, поэтому гипотетические вероятности исходов имеют вид

$$p_0(\theta) = \mathbf{P}\{\xi = 0\} = (1-\theta)^2,$$
$$p_1(\theta) = \mathbf{P}\{\xi = 1\} = 2\theta(1-\theta),$$
$$p_2(\theta) = \mathbf{P}\{\xi = 2\} = \theta^2$$

Отсюда в соответствии с (3) § 3.6 получаем уравнение для нахождения оценки параметра θ

$$\sum_{j=0}^{2} \frac{\nu_j p_j'(\theta)}{p_j(\theta)} = -\frac{2\nu_0}{1-\theta} + \frac{1-2\theta}{\theta(1-\theta)}\nu_1 + \frac{2\nu_2}{\theta} = 0,$$

решая которое, находим

$$\widehat{\theta}_n = \frac{\nu_1 + 2\nu_2}{2n}.$$

В данном случае $\nu_0 = 476$, $\nu_1 = 1017$, $\nu_2 = 527$, $n = 2020$, поэтому $\widehat{\theta}_n = 0{,}513$. Далее вычисляем значение статистики (11):

$$\widehat{X}_n^2 = \sum_{j=0}^{2} \frac{(\nu_j - np_j(\widehat{\theta}_n))^2}{np_j(\widehat{\theta}_n)} = 0{,}116,$$

и этот результат надо сравнить с $\chi^2_{1-\alpha,\,1}$. Поскольку $\chi^2_{0{,}3;\,1} = 0{,}148$, при любом уровне значимости $\alpha \leqslant 0{,}7$ гипотез принимается. •

Пример 5 (Пуассоновская модель, критерий χ^2 для нее). Пусть производится n независимых наблюдений над неотрицательной целочисленной случайной величиной ξ. Требуется проверить гипотезу $H_0 \quad \mathcal{L}(\xi) \in \Pi(\theta)$. Здесь множество $\mathcal{E}$ возможных значений ξ есть $\{0, 1, 2, \ .\}$, т. е. бесконечно, поэтому надо предварительно провести группировку наблюдений. В качестве соответствующих «интервалов» выберем $\mathcal{E}_j = \{j-1\}$, $j = 1, \ , N \ \ 1$, $\mathcal{E}_N = \{N-1, N, \ldots\}$ при некотором $N \geqslant 3$ (тогда $N - 1 - 1 > 0$). Гипотетические вероятности при таком способе группировки будут равны (см. (8) § 1.1)

$$p_j(\theta) = f(j-1|\theta), \quad j = 1, \quad , N-1, \quad p_N(\theta) = \sum_{x=N-1}^{\infty} f(x|\theta),$$

где

$$f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}.$$

Уравнение для нахождения оценки $\widehat{\theta}_n$ в данном случае имеет вид (см. (3) § 3.6)

$$\sum_{j=1}^{N-1}\left(\frac{j-1}{\theta} - 1\right)\nu_j + \frac{\nu_N \sum\limits_{x=N-1}^{\infty}\left(\frac{x}{\theta} - 1\right) f(x|\theta)}{\sum\limits_{x=N-1}^{\infty} f(x|\theta)} = 0.$$

Отсюда

$$\theta = \frac{1}{n}\left[\sum_{j=1}^{N-1}(j-1)\nu_j + \frac{\nu_N \sum\limits_{x=N-1}^{\infty} x f(x|\theta)}{\sum\limits_{x=N-1}^{\infty} f(x|\theta)}\right]$$

Здесь первый член в квадратных скобках равен сумме всех наблюдений X_i таких, что $X_i \leqslant N-2$, а последний член приближенно равен сумме наблюдений, оказавшихся бóльшими $N-2$. Таким образом, в качестве приближенной оценки для θ можно взять выборочное среднее: $\widehat{\theta}_n = \overline{X}$ (напомним, что обычная оценка максимального правдоподобия для параметра θ в модели $\Pi(\theta)$ точно равна $\overline{X}$, см. п. 1 § 3.5).

Теперь находим

$$\widehat{p}_j = p_j(\widehat{\theta}_n) = \begin{cases} \dfrac{e^{-\overline{X}}\overline{X}^{j-1}}{(j-1)!} & \text{при} \quad j = 1, \quad , N-1, \\ e^{-\overline{X}} \displaystyle\sum_{k=N-1}^{\infty} \dfrac{\overline{X}^k}{k!} & \text{при} \quad j = N, \end{cases}$$

и (при больших выборках) гипотеза H_0 отвергается на уровне значимости α тогда и только тогда, когда

$$\sum_{j=1}^{N} \frac{(\nu_j - n\widehat{p}_j)^2}{n\widehat{p}_j} > \chi^2_{1-\alpha,\, N-2}. \qquad \bullet$$

Пример 6 (Нормальная модель, критерий χ^2 для нее). Пусть H_0 — гипотеза, состоящая в том, что выборка $\boldsymbol{X} = (X_1, \quad , X_n)$ получена из нормального распределения $\mathcal{N}(\theta_1, \theta_2^2)$ (здесь $\theta = (\theta_1, \theta_2)$). Зададим интервалы группировки в виде $\mathcal{E}_j = (a_{j-1}, a_j]$, $j \quad 1, \quad , N$, $N > 3$, где $a_0 = -\infty$, $a_N = \infty$, $a_j = a_1 + (j-1)h$, $j = 1, \quad , N-1$, a_1, h — некоторые фиксированные величины. В таком случае, обозначив для краткости

$$g(x; \theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta_1)^2}{2\theta_2^2}\right\},$$

имеем

$$p_j(\theta) = \frac{1}{\theta_2} \int_{\mathcal{E}_j} g(x; \theta)\, dx, \quad j = 1, \quad , N$$

Система (3) § 3.6 в данном случае состоит из следующих двух уравнений:

$$\sum_{j=1}^{N} \frac{\nu_j}{p_j(\theta)} \int_{\mathcal{E}_j} (x - \theta_1) g(x; \theta)\, dx = 0,$$

$$\sum_{j=1}^{N} \frac{\nu_j}{p_j(\theta)} \left[\frac{1}{\theta_2^2} \int_{\mathcal{E}_j} (x-\theta_1)^2 g(x; \theta)\, dx - \int_{\mathcal{E}_j} g(x; \theta)\, dx\right] = 0.$$

Поскольку $\sum_{j=1}^{N} \nu_j = n$, эти уравнения можно привести к виду

$$\theta_1 = \frac{1}{n} \sum_{j=1}^{N} \nu_j \frac{\int_{\mathcal{E}_j} x g(x; \theta)\, dx}{\int_{\mathcal{E}_j} g(x; \theta)\, dx}, \qquad \theta_2^2 = \frac{1}{n} \sum_{j=1}^{N} \nu_j \frac{\int_{\mathcal{E}_j} (x-\theta_1)^2 g(x; \theta)\, dx}{\int_{\mathcal{E}_j} g(x; \theta)\, dx}.$$

Разрешая эту систему относительно θ_1, θ_2^2, можно найти оценки $\widehat{\theta}_1$ и $\widehat{\theta}_2$.

При малом h можно получить практически приемлемые приближенные значения для этих оценок. Предположим сначала, что $\nu_1 = \nu_N = 0$ (левее точки a_1 и правее точки $a_{N-1} = a_1 + (N-2)h$ нет выборочных точек). Заменяя интегралы по областям $\mathcal{E}_j$ произведением подынтегральной функции в средней точке z_j на длину $|\mathcal{E}_j| = h$ для $j = 2, \quad , N-1$ получим следующие приближения:

$$\widehat{\theta}_1 \approx \frac{1}{n}\sum_{j=2}^{N-1} \nu_j z_j, \quad \widehat{\theta}_2^2 \approx \frac{1}{n}\sum_{j=2}^{N-1} \nu_j (z_j - \widehat{\theta}_1)^2$$

Если крайние интервалы не пусты, но содержат малую часть выборки, то можно использовать приближения

$$\widehat{\theta}_1 \approx \frac{1}{n}\sum_{j=1}^{N} \nu_j z_j, \quad \widehat{\theta}_2^2 \approx \frac{1}{n}\sum_{j=1}^{N} \nu_j (z_j - \widehat{\theta}_1)^2$$

где $z_1 = a_1$, $z_N = a_{N-1}$.

С учетом этих допущений можно считать, что если проверяемая гипотез H_0 верна, то статистика (11) распределена приближенно (при больших выборках) по закону $\chi^2(N-3)$, и на этой основе рассчитывать соответствующий критерий (12). •

4. Критерий квантилей

Метод хи-квадрат можно применять для построения критериев согласия и в задачах другого типа, когда нулевая гипотеза фиксирует некоторые частные характеристики функции распределения $F_\xi(x)$ наблюдаемой случайной величины ξ. Рассмотрим типичную ситуацию такого рода, когда гипотеза имеет вид $H_0 \quad F_\xi(\zeta_j) = p_j$, $j = 1, \quad , N-1$, где $-\infty < \zeta_1 < \quad < \zeta_{N-1} < \infty$ и $0 < p_1 < \quad < p_{N-1} < 1$ — заданные числа, т. е. H_0 — это гипотеза о том, что распределение $\mathcal{L}(\xi)$ имеет заданные квантили ζ_j для заданных уровней p_j Это сложная гипотеза: соответствующий ей класс распределений $\mathcal{F}_0$ включает все распределения с указанными квантилями.

Пусть имеется выборка $X = (X_1, \quad , X_n)$ из некоторого непрерывного распределения $\mathcal{L}(\xi)$, о котором выдвинута такая гипотез H_0. Построим критерий типа χ^2 для проверки этой гипотезы. Для этого положим $\mathcal{E}_j = (\zeta_{j-1}, \zeta_j]$, $j = 1, \ldots, N$, $\zeta_0 = -\infty$, $\zeta_N = \infty$, и пусть ν_j — число выборочных точек X_i, принадлежащих интервалу $\mathcal{E}_j$. Тем самым мы имеем схему группировки данных с естественными (т. е. порождаемыми самой проверяемой гипотезой) интервалами. Тогда

$$\mathcal{L}\big(\underline{\nu} = (\nu_1, \ldots, \nu_N) | H_0\big) = M\big(n; \underline{p}^\circ = (p_1^\circ \quad , p_N^\circ)\big),$$

где $p_j^\circ = p_j - p_{j-1}$, $j = 1, \quad , N$, $p_0 = 0$, $p_N = 1$, и гипотеза H_0 эквивалентна утверждению, что вероятности исходов в построенной полиномиальной схеме равны заданным числам $p_1^\circ \ldots, p_N^\circ$. Следовательно, для проверки этой

гипотезы можно использовать построенный в п. 2 критерий χ^2 (6). В данном контексте этот критерий называют *критерием квантилей.*

При $N = 2$ и $p_1 = 1/2$ соответствующий критерий называют *критерием знаков*. В этом случае H_0 — это гипотеза о том, что выборка X извлечена из распределения $\mathcal{L}(\xi)$ с медианой ζ_1, а тестовая статистика (4) принимает вид

$$\hat{X}_n^2 = \frac{2}{n}\left(\nu_1 - \frac{n}{2}\right)^2 + \frac{2}{n}\left(\nu_2 - \frac{n}{2}\right)^2 = \frac{4}{n}\left(\nu_1 - \frac{n}{2}\right)^2 \tag{13}$$

где ν_1 — число выборочных точек X_i в интервале $(-\infty, \zeta_1]$, или, что то же самое, число отрицательных разностей $X_i - \zeta_1$, $i = 1, \ldots, n$ (отсюда и название критерия).

На практике критерий знаков чаще всего применяют в следующей ситуации. Пусть (X_i, Y_i), $i = 1, \ldots, n$, — независимые наблюдения над двумерной случайной величиной $\underline{\xi} = (\xi_1, \xi_2)$. Требуется проверить гипотезу H_0 о том, что компоненты ξ_1 и ξ_2 независимы и одинаково распределены, т. е. что $F_{\underline{\xi}}(x, y) = F(x)F(y)$, где $F(x)$ — некоторая (неизвестная) одномерная функция распределения. Чтобы построить соответствующий критерий, поступают следующим образом. Составим разности $Z_i = X_i - Y_i$, $i = 1, \ldots, n$. Если гипотеза H_0 верна, то $\mathrm{P}\{Z_i < 0\} = \mathrm{P}\{Z_i > 0\} = 1/2$, и тем самым гипотеза сводится к утверждению, что выборка $(Z_1, \ldots, Z_n)$ извлечена из распределения, медиана которого равна 0. В этом случае ν_1 равно числу отрицательных Z_i, а статистика (13) для больших выборок распределена при гипотезе H_0 приближенно по закону $\chi^2(1)$. Тем самым при уровне значимости α асимптотический вариант критерия знаков имеет вид

Критерий знаков

$$\boxed{H_0 \quad \text{отвергается} \quad \Longleftrightarrow \quad \frac{4}{n}\left(\nu_1 - \frac{n}{2}\right)^2 > \chi^2_{1-\alpha},} \tag{14}$$

Замечание. Из (13) видно, что критерий знаков по существу основывается на статистике ν_1, которая имеет при гипотезе H_0 биномиальное распределение $\mathrm{Bi}(n, 1/2)$, поэтому на самом деле здесь речь идет о проверке простой параметрической гипотезы о параметре биномиального распределения $\mathrm{Bi}(n, \theta)$, именно, гипотезы H_0 $\theta = 1/2$. В основе критерия (14) лежит нормальная аппроксимация $\mathcal{L}(\nu_1 | H_0) \sim \mathcal{N}(n/2, n/4)$ (теорема Муавра—Лапласа), из которой и следует, что

$$\mathcal{L}\left(\frac{4}{n}\left(\nu_1 - \frac{n}{2}\right)^2 \middle| H_0\right) \to \chi^2(1) \quad \text{при} \quad n \to \infty.$$

При малых выборках критерий знаков как частный случай критериев для общих параметрических гипотез в биномиальной модели $\mathrm{Bi}(n, \theta)$ будет рассчитан позже (в гл. 5) при изложении общей теории проверки параметрических гипотез. Вообще же критерий знаков — «старый и знаменитый» (Я. Гаек и З. Шидак), он применяется уже очень давно, по крайней мере не позже, чем с 1925 г. (Р. Фишер); детальное обсуждение его практического использования можно найти в книге [26, с. 58–65], где приведена и соответствующая обширная библиография

5. Критерий пустых ящиков

В этом пункте мы кратко коснемся еще одного современного подхода к задаче проверки простой гипотезы H_0 о виде распределения, связанного с методом группировки наблюдений. Как уже отмечалось в п. 2, при группировке наблюдений (в случае выборки из непрерывного распределения) происходит неизбежная потеря информация, так как вместо самих наблюдений используются лишь частоты попадания в интервалы группировки. Этот недостаток метода можно уменьшить, устраивая «более мелкое» разбиение носителя распределения наблюдаемой случайной величины, т. е. увеличивая число интервалов группировки N. Однако, тогда критерий χ^2 становится трудоемким, так как подсчет и обработка большого (при больших N) числа частот $\nu_1, \ldots, \nu_N$ требует больших вычислительных затрат. Следует также иметь в виду и то обстоятельство, что теорема 1 о предельном распределении статистики $\hat{X}_n^2$ справедлива лишь при фиксированных значениях параметров N и p_j°, $j = 1, \quad , N$, и ее уже нельзя использовать для расчета критической границы критерия, если $N \to \infty$. Выход из этой противоречивой ситуации можно найти, сконструировав более простые, нежели $\hat{X}_n^2$, тестовые статистики и доказав для них предельные теоремы, учитывающие одновременное неограниченное возрастание параметров n и N. Одной из таких простых статистик является число пустых (т. е. не содержащих ни одной выборочной точки) интервалов $\mu_0 = \mu_0(n, N)$ (мю-ноль). Критерий, основанный на этой статистике, называют *критерием пустых ящиков*, его мы и опишем. (Если интерпретировать интервалы группировки как ящики, а испытания как бросание в них дробинок, то $\mu_0(n, N)$ есть число пустых ящиков после бросания n дробинок — отсюда и название критерия.)

Для некоторого упрощения мы конкретизируем нулевую гипотезу

$$H_0 \quad F_\xi(x) = F(x),$$

считая, что $F(x) = x$, $0 \leqslant x \leqslant 1$; таким образом, H_0 — это гипотеза о том, что наблюдаемая случайная величина ξ имеет равномерное распределение на отрезке $[0, 1]$. (На самом деле это не есть сужение проблемы, так как с помощью преобразования Смирнова (см. упр. 17 к гл. 2) $\eta = F(\xi)$ мы всегда можем перейти от случайной величины ξ с абсолютно непрерывным распределением F к равномерно распределенной на отрезке $[0, 1]$ величине η.) Далее, обычно в методе группировки рекомендуется интервалы $\mathcal{E}_1, \quad , \mathcal{E}_N$ выбирать *равновероятными* при нулевой гипотезе, т. е. чтобы все $p_j^\circ \quad 1/N$, поэтому далее мы полагаем $\mathcal{E}_j = ((j-1)/N, j/N]$, $j = 1, \ldots, N$.

Итак, пусть $\nu_1, \quad , \nu_N$ — частоты попадания элементов выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$ из некоторого абсолютно непрерывного распределения на $[0, 1]$ в эти интервалы. Определим индикаторные случайные величины $\eta_j = I(\nu_j = 0)$, $j = 1, \quad , N$, тогда, очевидно, $\mu_0 = \eta_1 + \quad + \eta_N$ Вычислим первые два момента этой статистики для произвольного распределения наблюдений F_ξ, для которого

$$F_\xi(j/N) - F_\xi((j-1)/N) = p_j, \quad j = 1, \ldots, N.$$

Имеем

$$\mathbf{E}\mu_0 = \sum_{j=1}^{N} \mathbf{E}\eta_j = \sum_{j=1}^{N} \mathbf{P}\{\eta_j = 1\},$$

$$\mathbf{D}\mu_0 = \sum_{j=1}^{N} \mathbf{D}\eta_j + 2\sum_{i<j} \operatorname{cov}(\eta_i, \eta_j) = \sum_{j=1}^{N} \mathbf{P}\{\eta_j = 1\}(1 - \mathbf{P}\{\eta_j = 1\}) +$$

$$+ 2\sum_{i<j} \left[\mathbf{P}\{\eta_i = \eta_j = 1\} - \mathbf{P}\{\eta_i = 1\}\mathbf{P}\{\eta_j = 1\}\right]$$

Испытания независимы, поэтому

$$\mathbf{P}\{\eta_j = 1\} = \mathbf{P}\{\nu_j = 0\} = (1 - p_j)^n$$

$$\mathbf{P}\{\eta_i = \eta_j = 1\} = \mathbf{P}\{\nu_i = \nu_j = 0\} = (1 - p_i - p_j)^n$$

$$i \neq j$$

Следовательно,

$$\mathbf{E}\mu_0 = \sum_{j=1}^{N}(1 - p_j)^n \qquad \mathbf{D}\mu_0 = 2\sum_{i<j}(1 - p_i - p_j)^n + \mathbf{E}\mu_0 - (\mathbf{E}\mu_0)^2 \tag{15}$$

С помощью метода неопределенных множителей Лагранжа (см. о нем в п. 1 (1) § 3.7) легко проверить, что среднее $\mathbf{E}\mu_0$ как функция параметров $p_1, \ldots, p_N$ достигает минимума при $p_j = 1/N$, $j = 1, \ldots, N$, т. е. при нулевой гипотезе. В этом случае формулы (15) принимают вид

$$\mathbf{E}\mu_0 = N\left(1 - \frac{1}{N}\right)^n$$

$$\mathbf{D}\mu_0 = N(N-1)\left(1 - \frac{2}{N}\right)^n + N\left(1 - \frac{1}{N}\right)^n - N^2\left(1 - \frac{1}{N}\right)^{2n} \tag{16}$$

Таким образом, при любом отклонении от нулевой гипотезы, таком, что вероятности попадания в интервалы группировки не все равны $1/N$, статистика μ_0 имеет тенденцию увеличиваться, т. е. «слишком большие» значения μ_0 «свидетельствуют» против гипотезы H_0. Отсюда следует, что критическую область критерия следует задавать в виде

$$\{\mu_0(n, N) > t_\alpha(n, N)\}. \tag{17}$$

Для нахождения критической границы $t_\alpha(n, N)$ надо знать распределение статистики μ_0 при гипотезе H_0. Точное ее распределение известно и имеет вид [14, с. 10],

$$\mathbf{P}\{\mu_0(n, N) = k | H_0\} = C_N^k \sum_{j=0}^{N-k}(-1)^j C_{N-k}^j \left(1 - \frac{k+j}{N}\right)^n$$

однако использовать его для практических расчетов при больших n и N (что и представляет наибольший интерес для приложений) неудобно. Поэтому при больших значениях n и N для расчета критерия (17) естественно воспользоваться предельными теоремами. Одна из необходимых нам таких теорем, дающая вид асимптотического распределения μ_0 при любом абсолютно непрерывном распределении наблюдений F_ξ (а не только для нулевой гипотезы $H_0 : F_\xi(x) = x$, $0 \leqslant x \leqslant 1$) формулируется следующим образом.

Обозначим через

$$p_j = \int_{(j-1)/N}^{j/N} f(x)\,dx, \quad j = 1, \quad , N,$$

вероятности попадания наблюдений в интервалы группировки для распределения с плотностью $f(x)$ и будем предполагать, что все допустимые плотности $f(x) > 0$, $x \in [0, 1]$, и непрерывны.

Теорема 5. Пусть $n, N \to \infty$ так, что $n/N \to \rho > 0$. Тогда

$$\mathcal{L}\big(\mu_0(n, N)|f\big) \sim \mathcal{N}\big(Nm(f),\ N\sigma^2(f)\big),$$

где

$$m(f) = \int_0^1 e^{-\rho f(x)}\,dx,$$

$$\sigma^2(f) = \int_0^1 e^{-\rho f(x)}(1 - e^{-\rho f(x)})\,dx - \rho\left[\int_0^1 f(x)e^{-\rho f(x)}\,dx\right]^2 \tag{18}$$

Для нулевой гипотезы $f(x) \equiv 1$, $x \in [0, 1]$, поэтому утверждение теоремы 5 принимает вид

$$\mathcal{L}\left(\frac{\mu_0(n, N) - Ne^{-\rho}}{\sqrt{Ne^{-\rho}\big(1 - (1 + \rho)e^{-\rho}\big)}}\,\middle|\,H_0\right) \to \mathcal{N}(0, 1). \tag{19}$$

Этот результат уже позволяет приближенно рассчитать критическую границу в (17) при больших n и N и заданном уровне значимости α. В самом деле, из (19) следует, что если в (17) положить

$$t_\alpha(n, N) = Ne^{-\rho} + t_\alpha\sqrt{Ne^{-\rho}\big(1 - (1 + \rho)e^{-\rho}\big)}, \quad \Phi(-t_\alpha) = \alpha, \tag{20}$$

то при $n, N \to \infty$, $n/N \to \rho > 0$ критерий (17) будет иметь в пределе уровень значимости α.

Критерий пустых ящиков

Итак, асимптотический (при больших n, N и $n/N = \rho > 0$) вариант критерия пустых ящиков имеет следующий вид:

$$\boxed{H_0 \quad \text{отвергается} \iff \mu_0(n, N) > t_\alpha(n, N),} \tag{21}$$

где $t_\alpha(n, N)$ дано в (20). Следуя этому правилу, можно ошибочно отвергнуть истинную гипотезу H_0 $F_\xi(x) = x$, $0 \leqslant x \leqslant 1$, с вероятностью, приближенно равной α.

Теорема 5 позволяет также исследовать предельное поведение мощности $W_N(f)$ критерия (21) и при произвольной альтернативе, задаваемой плотностью $f(x) \not\equiv 1$, $x \in [0, 1]$. Действительно, в условиях этой теоремы

$$W_N(f) = \mathbf{P}\{\mu_0(n, N) > t_\alpha(n, N) | f\} =$$
$$= \mathbf{P}\left\{ \frac{\mu_0(n, N) - Nm(f)}{\sqrt{N}\,\sigma(f)} > -z_\alpha(N) \middle| f \right\} \approx \Phi(z_\alpha(N)), \tag{22}$$

где

$$z_\alpha(N) = \sqrt{N}\,\frac{m(f) - e^{-\rho}}{\sigma(f)} - t_\alpha \frac{\sigma_0}{\sigma(f)}, \qquad \sigma_0^2 = e^{-\rho}\bigl(1 - (1+\rho)e^{-\rho}\bigr).$$

Воспользуемся, далее, известным в теории вероятностей *неравенством Йенсена*[6], справедливым для любой случайной величины η и любой выпуклой вниз функции $g(x)$ $g(\mathbf{E}\eta) \leqslant \mathbf{E}g(\eta)$ с равенством только в случае, когда $\eta \equiv \text{const}$ (предполагается, что указанные математические ожидания существуют, для выпуклой вверх функции $g(x)$ неравенство заменяется на противоположное); для неотрицательной величины η и $g(x) = \ln x$ это неравенство принимает вид $\ln \mathbf{E}\eta \geqslant \mathbf{E}\ln\eta$ или $\mathbf{E}\eta \geqslant \exp\{\mathbf{E}\ln\eta\}$. Полагая здесь $\eta = e^{-\rho f(\xi)}$, где $\mathcal{L}(\xi) = \mathcal{U}(0, 1)$, получим неравенство (см. (18))

$$m(f) \geqslant \exp\left\{-\rho \int_0^1 f(x)\,dx\right\} = e^{-\rho}$$

так как $f(x)$ — функция плотности на $[0, 1]$; равенство здесь имеет место только в случае $f(x) \equiv \text{const} = 1$, т. е. при нулевой гипотезе. Итак, для любой альтернативной плотности $f(x) \not\equiv 1$ разность $m(f) - e^{-\rho} > 0$, и потому величина $z_\alpha(N)$ в (22) при $N \to \infty$ неограниченно возрастает. Это, в свою очередь, в силу (22) означает, что $\lim\limits_{N\to\infty} W_N(f) = 1$, т. е. критерий пустых ящиков является состоятельным.

Состоятельность критерия пустых ящиков

Итак, если выполнены условия теоремы 5, то мощность критерия пустых ящиков при любой фиксированной альтернативе стремится к 1. Однако, если при изменении n и N изменяется также и альтернатива, сближаясь с нулевой гипотезой H_0, то мощность критерия уже необязательно будет сходиться к 1 — это зависит от скорости сближения альтернативы с нулевой гипотезой. Теорема 5 дает возможность ответить и на вопрос о том, с какой скоростью

[6] Йенсен Иоганн Людвиг (1859–1925) — датский математик, труды по математическому анализу, геометрии.

может происходить это сближение, чтобы критерий (21) обладал еще способностью «реагировать» на альтернативу. В данном случае альтернативами являются любые абсолютно непрерывные распределения на $[0, 1]$, отличающиеся от равномерного распределения $U(0, 1)$; следовательно, «близкую» альтернативу F_n можно задать плотностью

«Близкие» альтернативы

$$f_n(x) = 1 + a(n)b(x), \tag{23}$$

где $a(n) \to 0$ при $n \to \infty$, а $b(x)$ — произвольная непрерывная функция, удовлетворяющая условию

$$\int_0^1 b(x)\,dx = 0.$$

Здесь $a(n)$ и определяет скорость сближения (в смысле поточечной сходимости) альтернативной плотности f_n с плотностью при нулевой гипотезе.

Следующее утверждение является простым следствием теоремы 5.

Мощность критерия пустых ящиков

Теорема 6. Если в соотношении (23) положить

$$a(n) = n^{-1/4},$$

то при $n, N \to \infty$, $n/N \to \rho > 0$ мощность $W_N(f_n)$ удовлетворяет предельному соотношению

$$W_N(f_n) \to \Phi\big(\Delta(b^2, \rho) - t_\alpha\big), \tag{24}$$

где

$$\Delta(b^2, \rho) = \frac{b^2}{2}\frac{\rho^{3/2}}{\sqrt{e^\rho - 1 - \rho}}, \qquad b^2 = \int_0^1 b^2(x)\,dx.$$

Из этой теоремы следует, что критерий пустых ящиков позволяет отличать от гипотезы H_0 альтернативы, сближающиеся с ней со скоростью $n^{-1/4}$. Более близкие альтернативы вида (23) (т. е. при $a(n)n^{1/4} \to 0$) этот критерий асимптотически (в условиях теоремы 6) не отличает от H_0 (т. е. $W_N(f_n) \to \alpha$), а для более удаленных альтернатив (т. е. при $a(n)n^{1/4} \to \infty$) он сохраняет свойство состоятельности (т. е. $W_N(f_n) \to 1$).

Дальнейшие результаты

Критерий пустых ящиков привлекателен своей простотой и достаточно широко известен (по крайней мере с работы Ф. Дейвид 1950 г.). Подробное изложение относящихся к нему результатов и историю вопроса можно найти в книге [14, гл. 5], где также рассмотрены и обобщения этого критерия. «Критерий пустых ящиков может быть полезен при проверке гипотезы о равномерности распределения каких-либо объектов в ограниченной части пространства,

например, звезд в части космического пространства, особей какой-либо биологической популяции в ее ареале и т. п.» [14, с. 10]. В то же время очевидно, что статистика μ_0 включает далеко не всю информацию, содержащуюся в выборке $X = (X_1, \quad , X_n)$. Более полно эта информация может быть учтена, если конструировать тестовые статистики вида

Симметрические статистики

$$\sum_{r=0}^{n} c_r \mu_r,$$

где

$$\mu_r = \mu_r(n, N) = \sum_{j=1}^{N} I \quad (\nu_j = r)$$

— число частот $\nu_1, \quad , \nu_N$, принявших значение $r = 0, 1, \quad , n$, а c_0, c_1, c_n — некоторые «веса», выбираемые из тех или иных условий оптимальности. Статистики такого вида называются *симметрическими статистиками в схеме группировки наблюдений*, а порождаемые ими критерии — *симметрическими критериями*. Общие результаты для симметрических критериев, обобщающих критерий пустых ящиков, изложены в работах авторов[7] Здесь же мы лишь отметим, что в условиях теоремы 6 наибольшую асимптотическую мощность среди всех симметрических критериев имеет известный нам критерий χ^2, тестовая статистика которого (см. (4))

Критерий χ^2

в рассматриваемом случае принимает вид

$$\mathring{X}_n^2 = \frac{N}{n} \sum_{j=1}^{N} \nu_j^2 - n = \frac{N}{n} \sum_{r=0}^{n} r^2 \mu_r - n.$$

Критическая граница для этого критерия, превышение которой приводит к отклонению гипотезы H_0, имеет асимптотически вид $t_\alpha(n, N) = N + t_\alpha \sqrt{2N}$, $\Phi(-t_\alpha) = \alpha$, а предельная мощность при «пороговых» альтернативах равна $\Phi\left(b^2 \sqrt{\rho/2} - t_\alpha\right)$

Интересно сравнить, насколько уступает по мощности критерий пустых ящиков — простейший симметрический критерий — оптимальному (но трудоемкому) критерию χ^2 Из (24) следует, что

$$\Delta(b^2, \rho) = b^2 \sqrt{\frac{\rho}{2}}\, e(\rho), \quad \text{где} \quad e(\rho) = \frac{\rho}{\sqrt{2(e^\rho - 1 - \rho)}}.$$

Как функция от ρ величина $e(\rho)$ монотонно убывает от 1 при $\rho = 0$ до 0 при $\rho = \infty$, т. е. при малых значениях $\rho = n/N$ критерий пустых ящиков по мощности мало уступает критерию χ^2 Некоторые численные значения функции $e(\rho)$ указаны в следующей таблице:

[7] *Ивченко Г. И. Медведев Ю. И.* Разделимые статистики и проверка гипотез. Случай малых выборок // Теория вероятностей и ее применения. 1978. Т. 23. № 4. С. 796–806; *Ивченко Г. И., Медведев Ю. И.* Разделимые статистики и проверка гипотез для группированных данных // Теория вероятностей и ее применения. 1980. Т. 25. № 4. С. 549–560.

ρ	0,05	0,10	0,15	0,20	0,30	0,40	0,50	0,60	0,70	0,80
$e(\rho)$	0,992	0,983	0,975	0,967	0,950	0,933	0,917	0,900	0,884	0,865
ρ	0,90	1,00	1,10	1,20	1,30	1,40	1,50	1,60	1,80	2,00
$e(\rho)$	0,851	0,834	0,818	0,802	0,786	0,769	0,753	0,738	0,706	0,675

§ 4.3. Гипотеза и критерии однородности

Задача о двух выборках

О гипотезе однородности и ее прикладном значении подробно рассказано в примере 5 Введения. Ее простейший вариант, который часто называют также *задачей о двух выборках*, формулируется следующим образом. Пусть $\boldsymbol{X} = (X_1, \quad , X_n)$ — выборка из распределения $\mathcal{L}(\xi)$ с некоторой (неизвестной) функцией распределения $F_1(x)$, а $Y = (Y_1, \quad , Y_m)$ — выборка из распределения $\mathcal{L}(\eta)$ с неизвестной функцией распределения $F_2(x)$. Гипотезой однородности является утверждение $H_0 \quad F_1(x) \equiv F_2(x)$. Это — одна из основных проблем непараметрической статистики, для решения которой разработано большое число соответствующих *критериев однородности*. Некоторые из них будут описаны и обсуждены в этом параграфе.

1. Критерий однородности Смирнова

Статистика Смирнова

Этот критерий является одним из первых критериев однородности в задаче о двух выборках из непрерывных распределений (1939 г.). Его тестовой статистикой является

$$D_{n,m} = \sup_{-\infty < x < \infty} \left| \widehat{F}_{1n}(x) - \widehat{F}_{2m}(x) \right|,$$

где $\widehat{F}_{1n}(x)$ и $\widehat{F}_{2m}(x)$ — эмпирические функции распределения, построенные по выборкам X и Y соответственно. Статистика $D_{n,m}$ «сама напрашивается» в качестве разумной тестовой статистики для гипотезы однородности H_0. Ведь, как мы знаем (см., например, п. 1 § 4.2), эмпирическая функция распределения является оптимальной оценкой для теоретической функции распределения, и с увеличением объема выборки они сближаются, поэтому в случае справедливости гипотезы H_0 функции $\widehat{F}_{1n}(x)$ и $\widehat{F}_{2m}(x)$ оценивают одну и ту же неизвестную функцию распределения. Тем самым в этом случае (по крайней мере при больших n и m) статистика $D_{n,m}$ не должна существенно отклоняться от нуля. Отсюда следует, что слишком большие значения этой статистики следует расценивать как свидетельство против гипотезы H_0 и потому разумно выбрать критическую область в виде $T_{1\alpha} = \{D_{n,m} > t_\alpha(n,m)\}$. Критическую границу $t_\alpha(n,m)$ при заданном уровне значимости α находят на основании известного при гипотезе H_0 предельного (при $n, m \to \infty$) распределения статистики $D_{n,m}$, даваемого теоремой Смирнова (см. теорему 4 § 2.1). На основании этой теоремы при больших n и m можно положить

$t_\alpha(n, m) = \sqrt{1/n + 1/m}\,\lambda_\alpha$, где $K(\lambda_\alpha) = 1-\alpha$. Действительно, в этом случае

$$\mathbf{P}\left\{D_{n,m} > \sqrt{\frac{1}{n}+\frac{1}{m}}\lambda_\alpha \,\middle|\, H_0\right\} = \mathbf{P}\left\{\sqrt{\frac{nm}{n+m}}D_{n,m} > \lambda_\alpha \,\middle|\, H_0\right\} \approx 1 - K(\lambda_\alpha) = \alpha.$$

Итак асимптотический (при больших n и m) вариант критерия однородности Смирнова имеет вид:

Критерий Смирнова

$$\boxed{H_0 \quad \text{отвергается} \quad \Longleftrightarrow \quad D_{n,m} > \lambda_\alpha\sqrt{\frac{1}{n}+\frac{1}{m}},} \tag{1}$$

где $K(\lambda_\alpha) = 1 - \alpha$, $K(t)$ — функция распределения Колмогорова (см. (11) § 2.1). Вероятность ошибочно отвергнуть при этом истинную гипотезу H_0 приблизительно равна α. Отметим также, что этот критерий состоятелен (альтернативой здесь является любая пара непрерывных распределений (F_1, F_2) такая, что $F_1(x) \not\equiv F_2(x)$).

Подчеркнем, что критерий (1) проверки неизменности функции распределения не зависит от конкретного вида функции. Для приложений это имеет важное значение, так как истинное распределение наблюдений, как правило, бывает неизвестно, а интерес представляет вопрос о том, не изменялось ли это распределение от выборки к выборке. Для применения критерия Смирнова необходимо лишь выполнение условия непрерывности, которое обычно бывает ясно из физической природы изучаемого явления и не требует специальной проверки.

Отметим также, что практически значение статистики $D_{n,m}$ рекомендуется вычислять по формуле $D_{n,m} = \max(D^+_{n,m}, D^-_{n,m})$, где

$$D^+_{n,m} = \max_{1\leqslant r\leqslant m}\left(\frac{r}{m} - \widehat{F}_{1n}(Y_{(r)})\right) = \max_{1\leqslant s\leqslant n}\left(\widehat{F}_{2m}(X_{(s)}) - \frac{s-1}{n}\right),$$

$$D^-_{n,m} = \max_{1\leqslant r\leqslant m}\left(\widehat{F}_{1n}(Y_{(r)}) - \frac{r-1}{m}\right) = \max_{1\leqslant s\leqslant n}\left(\frac{s}{n} - \widehat{F}_{2m}(X_{(s)})\right)$$

и $X_{(1)} \leqslant X_{(2)} \leqslant \quad \leqslant X_{(n)}$, $Y_{(1)} \leqslant Y_{(2)} \leqslant \quad \leqslant Y_{(m)}$ — порядковые статистики выборок X и Y соответственно. В книге [3, с. 88–89]. приводятся детальные практические рекомендации по применению (расчету) критерия Смирнова в том числе и в случае малых выборок, где имеются и соответствующие таблицы для численных расчетов.

2. Критерий однородности хи-квадрат

Этот критерий применяют к анализу данных, имеющих дискретную структуру, конкретнее, когда в опытах наблюдается некоторый переменный признак, принимающий конечное число $N \geqslant 2$ различных значений. Но к такой схеме, как мы знаем (см. п. 2 § 4.2), можно свести любую другую модель, применяя предварительно метод группировки данных, поэтому метод хи-квадрат применим на самом деле к анализу данных любой природы, т. е. является в этом

смысле универсальным. Кроме того, с помощью этого метода можно анализировать одновременно любое конечное число выборок.

Итак, пусть осуществлено k серий независимых наблюдений, состоящих из $n_1, \ldots, n_k$ наблюдений соответственно, и пусть $\underline{\nu}_i = (\nu_{i1}, \ldots, \nu_{iN})$ — частоты исходов i-й серии, а $\underline{p}_i = (p_{i1}, \ldots, p_{iN})$ — их вероятности ($i = 1, \ldots, k$). Тогда *гипотеза однородности* означает утверждение, что вероятности исходов не менялись от серии к серии, т. е.

$$H_0 : \underline{p}_1 = \quad = \underline{p}_k = \underline{p} = (p_1, \quad , p_N), \tag{2}$$

где $\underline{p}$ — некоторый (неизвестный) вектор вероятностей ($p_1 + \quad + p_N = 1$).

Чтобы построить критерий проверки этой гипотезы, поступают следующим образом. Поскольку при гипотезе H_0 каждый из векторов частот $\underline{\nu}_i$, $i \quad 1, \ldots, k$, распределен по полиномиальному закону $M(n_i; \underline{p})$, то $\mathbf{E}(\nu_{ij} | H_0) = n_i p_j$, $j = 1, \quad , N$, и в качестве меры отклонения эмпирических данных от их гипотетических (при гипотезе H_0) значений можно выбрать статистику (см. п. 3 § 4.2)

$$X^2_{n_1 \ldots n_k}(\underline{p}) = \sum_{i=1}^{k} X^2_{n_i}(\underline{p}), \quad \text{где} \quad X^2_{n_i}(\underline{p}) = \sum_{j=1}^{N} \frac{(\nu_{ij} - n_i p_j)^2}{n_i p_j},$$

заменив здесь неизвестные параметры $\underline{p}$ их оценкой максимального правдоподобия, которая вычисляется в предположении истинности проверяемой гипотезы H_0:

$$\widehat{\underline{p}} = \arg\max_{\underline{p}} \prod_{i,j} p_j^{\nu_{ij}} = \arg\max_{\underline{p}} \prod_{j} p_j^{\nu_{\bullet j}} \quad \text{где} \quad \nu_{\bullet j} = \sum_{i=1}^{k} \nu_{ij}.$$

Таким образом, в этом случае (при гипотезе H_0) все данные можно объединить в одну выборку объема $n \quad n_1 + \quad + n_k$ с частотами исходов $\nu_{\bullet 1}, \quad , \nu_{\bullet N}$, следовательно (см. пример 12 § 3.5), $\widehat{p}_j = \nu_{\bullet j}/n$, $j = 1, \quad , N$. В итоге мы получаем тестовую статистику критерия однородности хи-квадрат

Тестовая статистика в задачах однородности

$$\widehat{X}^2_{n_1 \ldots n_k} = X^2_{n_1 \ldots n_k}(\widehat{\underline{p}}) = n \sum_{i=1}^{k} \sum_{j=1}^{N} \frac{1}{n_i \nu_{\bullet j}} \left(\nu_{ij} - \frac{n_i \nu_{\bullet j}}{n} \right)^2 \tag{3}$$

с помощью которой и строится критическая область критерия: $\{\widehat{X}^2_{n_1 \ldots n_k} > t_\alpha\}$. Для нахождения критической границы t_α при заданном уровне значимости α используется следующий предельный результат [15, с. 483], аналогичный теореме 4 § 4.2: при $n_i \to \infty$ ($i = 1, \quad , k$)

$$\mathcal{L}\left(\widehat{X}^2_{n_1 \ldots n_k} | H_0\right) \to \chi^2\big((k-1)(N-1)\big).$$

В силу этого результата при больших n_i можно полагать $t_\alpha = \chi^2_{1-\alpha,\,(k-1)(N-1)}$.

Итак, асимптотический (при больших n_i) вариант критерия имеет вид: Критерий однородности χ^2

$$\boxed{H_0 \quad \text{отвергается} \quad \Longleftrightarrow \quad \widehat{X}^2_{n_1 \ldots n_k} > \chi^2_{1-\alpha,\,(k-1)(N-1)}} \tag{4}$$

где статистика $\widehat{X}^2_{n_1 \ldots n_k}$ вычисляется по формуле (3). Вероятность ошибочно отклонить при этом истинную гипотезу (2) приблизительно равна α, если n достаточно велико.

Параметрические гипотезы

Эту же методику можно использовать и для проверки гипотезы о том, что гипотетические вероятности p_j имеют заданную параметрическую форму: $p_j = p_j(\theta)$, $j = 1, \ldots, N$. Такая ситуация возникает, например, когда требуется проверить гипотезу о том, что k серий наблюдений произведены над одной и той же случайной величиной с распределением из заданного параметрического семейства $\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$. В этом случае предварительно проводится группировка наблюдений с N интервалами и вычисляются оценки $\widehat{p}_j = p_j(\widehat{\theta})$, где

$$\widehat{\theta} = \arg\max_{\theta} \prod_j \left(p_j(\theta)\right)^{\nu_{\cdot j}}$$

Число степеней свободы в предельном (для статистики $X^2_{n_1 \ldots n_k}(\widehat{p})$ при гипотезе H_0) распределении χ^2 заменяется при этом на $k(N-1) - r$, где $r = \dim \Theta$ (dimention — размерность) (предыдущая ситуация соответствует случаю $r = N - 1$).

Критерий однородности χ^2 (4) является состоятельным, т. е. с вероятностью, стремящейся к 1 при $n \to \infty$, он «улавливает» любые отклонения от нулевой гипотезы (2), при которых вероятности появления исходов от серии к серии не сохраняют постоянного значения.

Для случая двух выборок ($k = 2$) статистика (3) принимает, как нетрудно проверить, следующий вид

$$\widehat{X}^2_{n_1,n_2} = n_1 n_2 \sum_{j=1}^{N} \frac{1}{\nu_{1j} + \nu_{2j}} \left(\frac{\nu_{1j}}{n_1} - \frac{\nu_{2j}}{n_2} \right)^2$$

Если положить $\omega_j = \nu_{1j}/(\nu_{1j}+\nu_{2j})$, $\omega = n_1/(n_1+n_2)$, то последнее выражение преобразуется к виду, более удобному для практических вычислений:

$$\widehat{X}^2_{n_1,n_2} = \frac{1}{\omega(1-\omega)} \left(\sum_{j=1}^{N} \omega_j \nu_{1j} - n_1 \omega \right) \tag{5}$$

Выделим также важный для приложений частный случай двух исходов ($N = 2$), которые можно обозначить A и $\bar{A}$. Гипотеза однородности (2) в этом случае есть утверждение, что событие A имеет во всех испытаниях одну и ту же (хотя и неизвестную) вероятность реализации p. Здесь оценкой для p является

относительная частота появления события A во всей совокупности данных:

$$\widehat{p} = 1 - \widehat{q} = \frac{1}{n}\sum_{i=1}^{k} \nu_{i1},$$

где ν_{i1} — число появлений события A в испытаниях i-й серии, а статистика (3) принимает вид

$$\widehat{X}^2_{n_1 \dots n_k} = \frac{1}{\widehat{p}\widehat{q}} \sum_{i=1}^{k} \frac{(\nu_{i1} - n_i \widehat{p})^2}{n_i} = \frac{1}{\widehat{p}\widehat{q}} \sum_{i=1}^{k} \frac{\nu_{i1}^2}{n_i} - n\frac{\widehat{p}}{\widehat{q}}. \tag{6}$$

Таблица «2 × 2» Рассмотрим здесь детальнее случай двух выборок ($n = 2$). Для наглядности в этом случае данные сводят в так называемую таблицу «2 × 2» (два-на-два):

	A	$\bar{A}$	Σ
(1)	ν_{11}	ν_{12}	$\nu_{1\bullet} = n_1$
(2)	ν_{21}	ν_{22}	$\nu_{2\bullet} = n_2$
Σ	$\nu_{\bullet 1}$	$\nu_{\bullet 2}$	n

В этих обозначениях запишем тестовую статистику (3) в виде

$$\widehat{X}^2_{n_1 n_2} = n \sum_{i,j=1}^{2} \frac{(\nu_{ij} - \nu_{i\bullet}\nu_{\bullet j}/n)^2}{\nu_{i\bullet}\nu_{\bullet j}} \equiv n \sum_{i,j=1}^{2} \frac{\Delta_{ij}^2}{\nu_{i\bullet}\nu_{\bullet j}}.$$

Рассмотрим разности $\Delta_{ij} = \nu_{ij} - \nu_{i\bullet}\nu_{\bullet j}/n$, $i, j = 1, 2$. Легко видеть, что все суммы

$$\sum_i \Delta_{ij} = \sum_j \Delta_{ij} = 0.$$

Например,

$$\Delta_{11} + \Delta_{12} = \nu_{11} + \nu_{12} - \frac{\nu_{1\bullet}\nu_{\bullet 1}}{n} - \frac{\nu_{1\bullet}\nu_{\bullet 2}}{n} = \nu_{1\bullet} - \frac{\nu_{1\bullet}}{n}(\nu_{\bullet 1} + \nu_{\bullet 2}) = \nu_{1\bullet} - \nu_{1\bullet} = 0.$$

Таким образом, все четыре величины Δ_{ij} равны по модулю и потому

$$\widehat{X}^2_{n_1 n_2} = n\Delta_{11}^2 \sum_{i,j=1}^{2} \frac{1}{\nu_{i\bullet}\nu_{\bullet j}} = \frac{n^3 \Delta_{11}^2}{\nu_{1\bullet}\nu_{2\bullet}\nu_{\bullet 1}\nu_{\bullet 2}}.$$

Далее,

$$n\Delta_{11} = \nu_{1\bullet}\left(\frac{\nu_{11}}{\nu_{1\bullet}}(\nu_{1\bullet} + \nu_{2\bullet}) - (\nu_{11} + \nu_{21})\right) = \nu_{1\bullet}\nu_{2\bullet}\left(\frac{\nu_{11}}{\nu_{1\bullet}} - \frac{\nu_{21}}{\nu_{2\bullet}}\right),$$

поэтому, обозначая

$$Z_{n_1 n_2} = \left(\frac{\nu_{11}}{n_1} - \frac{\nu_{21}}{n_2}\right)\sqrt{\frac{n n_1 n_2}{\nu_{\bullet 1}\nu_{\bullet 2}}},$$

получим

$$\widehat{X}^2_{n_1 n_2} = \frac{n\nu_{1\bullet}\nu_{2\bullet}}{\nu_{\bullet 1}\nu_{\bullet 2}}\left(\frac{\nu_{11}}{\nu_{1\bullet}} - \frac{\nu_{21}}{\nu_{2\bullet}}\right)^2 = Z^2_{n_1 n_2}. \tag{7}$$

Представление (7) подсказывает идею использовать $Z_{n_1 n_2}$ в качестве другой тестовой статистики в рассматриваемой задаче. Эта статистика является, очевидно, более информативной, чем статистика $\widehat{X}^2_{n_1 n_2}$, так как знак $Z_{n_1 n_2}$ может нести дополнительную информацию о характере альтернативы в тех случаях, когда гипотеза H_0 неверна (знак тестовой статистики указывает на направление отклонения альтернативы от H_0). Действительно, если обозначить через p_i вероятность реализации A в испытаниях i-й серии ($i = 1, 2$), то $\mathcal{L}(\nu_{i1})$ $\mathrm{Bi}(n_i, p_i)$ и $\mathrm{E}(\nu_{11}/n_1 - \nu_{21}/n_2) = p_1 - p_2$. Следовательно, при альтернативе вида H_1^+ $p_1 > p_2$ статистика $Z_{n_1 n_2}$ будет иметь тенденцию смещаться вправо от нуля, а при альтернативе H_1^- $p_1 < p_2$ — влево. Таким образом, статистику $Z_{n_1 n_2}$ естественно использовать в задачах проверки гипотезы H_0 против «односторонних» альтернатив H_1^+ или H_1^- т. е. в задачах (H_0, H_1^+) и (H_0, H_1^-), в то время как статистика $\widehat{X}^2_{n_1 n_2}$ «работает» лишь в задаче $(H_0, H_1 = H_1^+ \cup H_1^-)$, т. е. при общей (двусторонней) альтернативе H_1 $p_1 \neq p_2$.

Для построения соответствующего критерия надо исследовать распределение статистики $Z_{n_1 n_2}$ при нулевой гипотезе. Мы будем рассматривать, как обычно, схему больших выборок, т. е. предполагать, что $n_1, n_2 \to \infty$. Тогда на основании теоремы Муавра—Лапласа $\mathcal{L}(\nu_{i1}/n_i | H_0) \sim \mathcal{N}(p, pq/n_i)$, $i = 1, 2$, а так как случайные величины ν_{11} и ν_{21} независимы по условию, то при гипотезе H_0

$$\mathcal{L}\left(\frac{\nu_{11}}{n_1} - \frac{\nu_{21}}{n_2}\right) \sim \mathcal{N}\left(0, \frac{npq}{n_1 n_2}\right).$$

Следовательно, статистика

$$\zeta_{n_1 n_2} = \left(\frac{\nu_{11}}{n_1} - \frac{\nu_{21}}{n_2}\right)\sqrt{\frac{n_1 n_2}{npq}} = Z_{n_1 n_2}\sqrt{\frac{\nu_{\bullet 1}\nu_{\bullet 2}}{n^2 pq}}$$

при гипотезе H_0 асимптотически нормальна с параметрами 0 и 1:

$$\mathcal{L}(\zeta_{n_1 n_2} | H_0) \sim \mathcal{N}(0, 1).$$

Но при гипотезе H_0 $\nu_{\bullet 1}/n \xrightarrow{\mathrm{P}} p$, $\nu_{\bullet 2}/n \xrightarrow{\mathrm{P}} q$ (ведь относительная частота исхода в n независимых испытаниях является состоятельной оценкой его вероятности), следовательно,

$$\sqrt{\frac{\nu_{\bullet 1}\nu_{\bullet 2}}{n^2 pq}} \xrightarrow{\mathrm{P}} 1.$$

Отсюда следует (см. справочник предельных теорем в п. 4 § 3.5, утверждение 2°(г)), что при гипотезе H_0 предельные распределения статистик $Z_{n_1 n_2}$ и $\zeta_{n_1 n_2}$ совпадают. Итак, при $n_1, n_2 \to \infty$

$$\mathcal{L}(Z_{n_1 n_2} | H_0) \to \mathcal{N}(0, 1).$$

Пусть теперь требуется проверить гипотезу однородности $H_0 \quad p_1 = p_2$ против правосторонней альтернативы $H_1^+ \quad p_1 > p_2$. Из предыдущих рассуждений следует, что критическую область разумно выбрать в виде $\{Z_{n_1 n_2} > t_\alpha\}$, так как большие **положительные** значения статистики $Z_{n_1 n_2}$ естественно рассматривать как свидетельство в пользу альтернативы H_1^+. Далее, поскольку при больших n_1 и n_2 $\mathbf{P}\{Z_{n_1 n_2} > t_\alpha | H_0\} \approx 1 - \Phi(t_\alpha) = \Phi(-t_\alpha)$, то при уровне значимости α критическая граница приближенно определяется условием $\Phi(-t_\alpha) = \alpha$, т. е. $t_\alpha = -\Phi^{-1}(\alpha) = \Phi^{-1}(1-\alpha)$ есть квантиль уровня $1 - \alpha$ стандартного нормального распределения.

Замечание. Поскольку для любой гипотезы, задаваемой вероятностями p_1, p_2 при $n_1, n_2 \to \infty$

$$\mathcal{L}\left(\frac{\nu_{11}}{n_1} - \frac{\nu_{21}}{n_2}\right) \sim \mathcal{N}\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right),$$

то аналогичными рассуждениями можно показать (это мы оставляем читателю в качестве упражиения), что для «близкой» альтернативы

$$H_{1n}^+ \quad \frac{p_1 - p_2}{\sqrt{p_1 q_1}} = \frac{a}{\sqrt{n}}, \quad a > 0,$$

при $n_1, n_2 \to \infty$ так, что $n_1/n \to \gamma$, выполняется предельное соотношение

$$\mathcal{L}(Z_{n_1 n_2} | H_{1n}^+) \longrightarrow \mathcal{N}\left(a\sqrt{\gamma(1-\gamma)}, 1\right)$$

Этот результат позволяет вычислять предельную мощность построенного критерия при таких альтернативах:

$$\mathbf{P}\{Z_{n_1 n_2} > t_\alpha | H_{1n}^+\} \longrightarrow \Phi\left(a\sqrt{\gamma(1-\gamma)} - t_\alpha\right).$$

Пример 1. Поступающие в вуз абитуриенты разбиты на два потока по 300 человек в каждом. Итоги экзамена по одному и тому же предмету на каждом потоке оказались следующими: на 1-м потоке баллы 2, 3, 4 и 5 получили соответственно 33, 43, 80 и 144 человека; соответствующие же данные для второго потока — 39, 35, 72 и 154. Можно ли при уровне значимости $\alpha = 0{,}05$ считать оба потока однородными? Здесь $k = 2$, $N = 4$ и статистика (5) принимает значение 2,18. Критическая же граница $\chi^2_{0{,}95;\,3} = 7{,}82$. Таким образом, $\widehat{X}^2_{n_1 n_2} < \chi^2_{0{,}95;\,3}$, т. е. согласие гипотезы однородности с данными хорошее. •

Пример 2 [15, с. 484]. В таблице на с. 347 приведено число детей, родившихся в Швеции в течение $k = 12$ месяцев 1935 г.

Оценкой для вероятности рождения мальчика является

$$\widehat{p} = \frac{45\,682}{88\,273} = 0{,}5175.$$

По формуле (6) получаем $\widehat{X}^2_{n_1 \ldots n_{12}} = 14{,}986$ с 11 степенями свободы. Поскольку $\chi^2_{0{,}95;\,11} = 19{,}7$, то можно считать (при уровне значимости $\alpha = 0{,}05$), что данные совместимы с гипотезой о постоянной вероятности рождения мальчиков по месяцам. •

Месяцы	Пол ребенка		$\Sigma = n_{i\cdot}$
	Мальчики	Девочки	
1	3743	3537	7280
2	3550	3407	6957
3	4017	3866	7883
4	4173	3711	7884
5	4117	3775	7892
6	3944	3665	7609
7	3964	3621	7585
8	3797	3596	7393
9	3712	3491	7203
10	3512	3391	6903
11	3392	3160	6552
12	3761	3371	7132
$\Sigma = \nu_{\cdot j}$	45 682	42 591	88 273 $= n$

3. Другие критерии в задаче о двух выборках

Кроме двух описанных критериев однородности разработаны и другие методы, основанные на различных принципах. Далее приводится краткий обзор некоторых из них, применяемых в случае двух выборок из непрерывных распределений.

а) *Критерий знаков.* Простой для проверки однородности критерий, не требующий сложных вычислений, представляет собой описанный в п. 4 § 4.2 критерий знаков. Основным недостатком этого критерия является «неэкономное» использование информации, содержащейся в результатах наблюдений, поэтому его обычно рекомендуют применять только на стадии предварительного анализа. Кроме того, этот критерий можно использовать лишь для выборок одинакового объема.

б) *Критерий пустых блоков.* Целый класс критериев однородности можно построить исходя из следующих соображений. Рассмотрим вариационный ряд выборки X: $X_{(1)} < X_{(2)} <$ $< X_{(n)}$ (поскольку выборка из непрерывного распределения, то с вероятностью 1 все порядковые статистики будут различны, см. п. 2 § 2.1). Он порождает естественное разбиение оси Ox на интервалы $(X_{(i-1)}, X_{(i)}]$, $i = 1, \ldots, n+1$ ($X_{(0)} = -\infty$, $X_{(n+1)} = \infty$), которые называются *выборочными блоками*. Пусть $\zeta_r = \zeta_r(n, m)$ — число этих блоков, каждый из которых содержит ровно r элементов второй выборки $Y = (Y_1, \quad, Y_m)$, $r = 0, 1, \ldots, m$. Тогда в качестве тестовой статистики можно взять, вообще говоря, произвольную линейную комбинацию $\sum_r c_r \zeta_r$. В частности, простейшая такая статистика ζ_0 — число блоков, не содержащих ни одного элемента второй выборки, — порождает *критерий пустых блоков*. Основой для построе-

ния и расчета этого критерия в случае больших выборок является следующий результат [24, с. 453]: если n, $m \to \infty$ так, что $m/n \to \rho > 0$, то

$$\mathcal{L}\big(\zeta_0(n,m)|H_0\big) \sim \mathcal{N}\left(\frac{n}{1+\rho}, \frac{n\rho^2}{(1+\rho)^3}\right).$$

Критерий пустых блоков

Отсюда получаем асимптотический вариант критерия пустых блоков: *при уровне значимости α гипотеза однородности H_0 отвергается тогда и только тогда, когда*

$$\boxed{\zeta_0(n,m) > \frac{n}{1+\rho} + \sqrt{n}\,\frac{\rho}{(1+\rho)^{3/2}}\,t_\alpha, \quad \Phi(-t_\alpha) = \alpha.} \tag{8}$$

Этот критерий является состоятельным против альтернатив (F_1, F_2), $F_1 \not\equiv F_2$, удовлетворяющих следующему условию: функция $F_2(F_1^{-1}(u))$, $u \in [0,1]$, имеет производную $g(u)$, отличную от 1 на множестве положительной лебеговой меры.

в) *Критерий серий.* В ряде случаев особый интерес представляют такие отклонения от нулевой гипотезы H_0 $F_1(x) \equiv F_2(x)$, когда одно распределение сдвинуто относительно другого, например, когда $F_1(x) > F_2(x)$, — в этом случае говорят, что случайная величина η с распределением F_2 «стохастически больше», чем ξ, распределение которой есть F_1 (при каждом x величина η с бо́льшей вероятностью превосходит x, чем ξ). Простой критерий, хорошо «улавливающий» такие отклонения, можно построить следующим образом. Объединим обе выборки $\boldsymbol{X}$ и $\boldsymbol{Y}$ в одну $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ и построим вариационный ряд этой объединенной выборки. После этого заменим все элементы выборки $\boldsymbol{X}$ буквой C, а все элементы $\boldsymbol{Y}$ — буквой $\overline{C}$. В результате получим некоторую последовательность n букв C и m букв $\overline{C}$. Число всех таких последовательностей равно C_{n+m}^n и интуитивно ясно, что при гипотезе H_0 (элементы обеих выборок $\boldsymbol{X}$ и $\boldsymbol{Y}$ должны «вести себя одинаково») все возможные последовательности равновероятны. Описанные выше отклонения от H_0 приводят к тому, что с повышенной вероятностью будут наблюдаться последовательности, в которых элементы одного вида имеют тенденцию смещаться к какому-нибудь краю последовательности (в приведенном выше примере буквы $\overline{C}$ смещаются к правому краю). Одной из статистик, с помощью которой можно количественно охарактеризовать степень перемешивания в получаемой последовательности букв C и $\overline{C}$, является число серий $W(n,m)$. *Серией* называется участок последовательности, состоящий из подряд идущих одинаковых букв и ограниченный с обеих сторон (или с одной стороны, если речь идет о концах последовательности) буквами другого вида. Число серий будет мало, если одинаковые буквы группируются в одном месте, поэтому в данном случае следует задать критическую область в виде $\{W(n,m) \leqslant t_\alpha(n,m)\}$. Такой критерий, предложенный в 1940 г. Вальдом и Вольфовитцем, и называется *критерием серий*.

Асимптотический (для больших выборок) вариант этого критерия можно рассчитать на основании следующего утверждения [24, с. 459]: если n, $m \to \infty$

так, что $m/n \to \rho > 0$, то

$$\mathcal{L}(W(n, m)|H_0) \sim \mathcal{N}\left(\frac{2n\rho}{1+\rho}, \frac{4n\rho^2}{(1+\rho)^3}\right).$$

Критерий серий

При этих условиях критерий серий формулируется следующим образом: *при уровне значимости α гипотезу однородности H_0 отвергают тогда и только тогда, когда*

$$\boxed{W(n, m) \leqslant \frac{2n\rho}{1+\rho} - \sqrt{n}\,\frac{2\rho}{(1+\rho)^{3/2}}\,t_\alpha, \quad \Phi(-t_\alpha) = \alpha.} \tag{9}$$

Этот критерий является состоятельным против тех же альтернатив, что и рассмотренный в п. б) критерий пустых блоков.

г) *Ранговые критерии.* Иногда исходная статистическая информация может быть задана не числовыми значениями наблюдений, а отношением порядка между ними (типа «больше—меньше»). Особенно часто это имеет место в психологических исследованиях. В таких случаях наблюдения *ранжируют*, т. е. упорядочивают по степени их предпочтения. Номер места, которое занимает конкретное наблюдение в таком упорядоченном ряду, называют *рангом* этого наблюдения. Таким образом, с самого начала информация может быть задана рангами наблюдений;

Ранги

статистические методы, которые применяют в таких ситуациях, называют *ранговыми методами*, статистики, являющиеся функциями только рангов, — *ранговыми статистиками*, а критерии, основанные на таких статистиках, — *ранговыми критериями*.

Ранговые методы можно применять и в тех случаях, когда заданы числовые значения наблюдений, т. е. выборка $\boldsymbol{X} = (X_1, \ldots, X_n)$, так как при этом всегда можно упорядочить элементы выборки, построив ее вариационный ряд. В этом случае рангом i-го наблюдения X_i является номер места R_i, которое занимает X_i в вариационном ряду $X_{(1)} \leqslant X_{(2)} \leqslant \quad \leqslant X_{(n)}$. Теория ранговых критериев является важной и хорошо разработанной областью современной математической статистики, она весьма полно изложена в [7]. В этом пункте мы рассмотрим в качестве иллюстрации лишь один пример рангового критерия однородности — *критерий Вилкоксона*, предложенный впервые Вилкоксоном (1945) для выборок одинакового объема и распространенный на случай выборок разных объемов Манном и Уитни (1947).

Критерий Вилкоксона

Рассмотрим, как и в предыдущем случае, объединенную выборку $(X_1, \ldots, X_n, Y_1, \quad, Y_m)$ и построим ее вариационный ряд. Пусть R_i — ранг величины X_i в этом ряду ($i = 1, \quad, n$) и $T = R_1 + \quad + R_n$ — сумма номеров мест, которые занимают в общем вариационном ряду элементы выборки X. Ранговая статистика T и есть тестовая статистика критерия Вилкоксона.

Введем случайные величины $Z_{rs} = I(X_r < Y_s)$ и положим

$$U = U(n, m) = \sum_{r=1}^{n} \sum_{s=1}^{m} Z_{rs}, \tag{10}$$

так что U есть общее число тех случаев, когда элементы выборки X предшествуют в общем вариационном ряду элементам выборки Y. Можно показать, что статистики T и U связаны линейным соотношением

$$T + U = nm + \frac{n(n+1)}{2},$$

поэтому критерий Вилкоксона эквивалентен критерию, основанному на статистике U, называемому иногда *критерием Манна–Уитни*.

Для любой пары (F_1, F_2) абсолютно непрерывных распределений среднее значение статистики U есть

$$\mathbf{E}U = nm\,\mathbf{E}Z_{11} = nm\,\mathbf{P}\{X_1 < Y_1\} = nm\int_{-\infty}^{\infty} F_1(x)f_2(x)\,dx \equiv anm, \quad (11)$$

следовательно, при нулевой гипотезе, т.е. при $F_1 \equiv F_2$, $\mathbf{E}U = nm/2$. Таким образом, с позиций статистики U альтернативы можно характеризовать величиной параметра a в (11): $a < 1/2$, либо $a > 1/2$, либо $a \neq 1/2$. Для расчета критерия можно использовать следующий результат для случая больших выборок [24, с. 467]: если $n, m \to \infty$, то

$$\mathcal{L}\big(U(n,m)|H_0\big) \sim \mathcal{N}\left(\frac{nm}{2}, \frac{nm(n+m+1)}{12}\right)$$

(это предельное распределение можно с хорошим приближением использовать уже при $n, m \geqslant 4$, $n + m \geqslant 20$).

Основываясь на этом результате уже можно рассчитать асимптотический вариант критерия согласия для гипотезы H_0. Критическая область имеет различный вид в зависимости от целей, для которых строится критерий. Если хотят получить критерий, состоятельный против альтернатив (F_1, F_2) с $a < 1/2$, то критическую область следует задать в виде $\{U(n,m) \leqslant t_\alpha(n,m)\}$, где при при больших n и m и заданном уровне значимости α критическая граница имеет вид

$$t_\alpha(n,m) = \frac{nm}{2} - \sqrt{\frac{nm(n+m+1)}{12}}\,t_\alpha, \quad \Phi(-t_\alpha) = \alpha.$$

Случай $a > 1/2$ сводится к рассмотренному перестановкой F_1 и F_2. Наконец, критерий, состоятельный против двусторонних альтернатив: $a \neq 1/2$, задается двусторонней критической областью

$$\left\{\left|U(n,m) - \frac{nm}{2}\right| > \sqrt{\frac{nm(n+m+1)}{12}}\,t_{\alpha/2}\right\}$$

§ 4.4. Гипотеза независимости

В данном параграфе будут рассмотрены несколько вариантов проверки гипотезы независимости, описанной в примере 6 Введения. В этом случае данные представляют собой независимые наблюдения (X_i, Y_i), $i = 1, \ldots, n$,

над некоторой двумерной случайной величиной $\xi = (\xi_1, \xi_2)$ с неизвестной функцией распределения $F_\xi(x, y)$, о которой требуется проверить гипотезу H_0 $F_\xi(x, y) = F_{\xi_1}(x)F_{\xi_2}(y)$, где $F_{\xi_i}(x)$, $i = 1, 2$, — некоторые одномерные функции распределения.

1. Критерий независимости хи-квадрат

Простой критерий согласия для гипотезы независимости H_0 можно построить, основываясь на методике хи-квадрат. Эту методику применяют для дискретных моделей с конечным числом исходов, поэтому условимся считать, что случайная величина ξ_1 принимает конечное число s некоторых значений, которые будем обозначать $a_1, \ldots, a_s$, а вторая компонента ξ_2 — k некоторых значений $b_1, \ldots, b_k$. Если данные имеют другую структуру, то их предварительно группируют отдельно по первой и второй компонентам. В этом случае множество возможных значений ξ_1 разбивается на s непересекающихся интервалов $\mathcal{E}_1^{(1)}, \ldots, \mathcal{E}_s^{(1)}$, множество значений ξ_2 — на k интервалов $\mathcal{E}_1^{(2)}, \ldots, \mathcal{E}_k^{(2)}$ и тем самым множество значений двумерной величины $\xi = (\xi_1, \xi_2)$ разбивается на $N = sk$ прямоугольников $\mathcal{E}_i^{(1)} \times \mathcal{E}_j^{(2)}$

Пусть ν_{ij} есть число наблюдений пары (a_i, b_j) (число элементов выборки, принадлежащих прямоугольнику $\mathcal{E}_i^{(1)} \times \mathcal{E}_j^{(2)}$, если данные группируются), так что

$$\sum_{i=1}^{s} \sum_{j=1}^{k} \nu_{ij} = n.$$

Таблица сопряженности двух признаков

Результаты наблюдений удобно для наглядности расположить в виде *таблицы сопряженности двух признаков* (в приложениях ξ_1 и ξ_2 обычно означают два признака, по которым производится классификация результатов наблюдений, и тогда H_0 есть гипотеза о независимости этих признаков):

ξ_1	ξ_2				$\sum$
	b_1	b_2	..	b_k	
a_1	ν_{11}	ν_{12}		ν_{1k}	$\nu_{1\bullet}$
a_2	ν_{21}	ν_{22}		ν_{2k}	$\nu_{2\bullet}$
a_s	ν_{s1}	ν_{s2}	.....	ν_{sk}	$\nu_{s\bullet}$
$\sum$	$\nu_{\bullet 1}$	$\nu_{\bullet 2}$		$\nu_{\bullet k}$	n

Пусть $p_{ij} = \mathrm{P}\{\xi_1 = a_i, \xi_2 = b_j\}$, $i = 1, \ldots, s$, $j = 1, \ldots, k$. Тогда гипотеза независимости означает, что существует $s + k$ постоянных $p_{i\bullet}$, $p_{\bullet j}$ таких, что

$$\sum_{i=1}^{s} p_{i\bullet} = \sum_{j=1}^{k} p_{\bullet j} = 1 \quad \text{и} \quad p_{ij} = p_{i\bullet}p_{\bullet j}.$$

Поскольку наблюдения независимы по условию, то частоты $(\nu_{ij}, i = 1, \ldots, s, j = 1, \ldots, k)$ распределены по полиномиальному закону и при гипотезе H_0 вероятности исходов p_{ij} имеют указанную специфическую структуру (они определяются значениями $r = s + k - 2$ неизвестных параметров $p_{i\bullet}$, $i = 1, \ldots, s-1$, $p_{\bullet j}$, $j = 1, \ldots, k-1$). Следовательно, гипотеза независимости в данном случае является специфической формой сложной гипотезы для полиномиальной модели, рассмотренной в п. 3 § 4.2. Чтобы применить критерий хи-квадрат (12) § 4.2, надо сначала найти оценки максимального правдоподобия для определяющих рассматриваемую модель неизвестных параметров при справедливости гипотезы H_0, т. е. максимизируя в данном случае функцию правдоподобия

$$\prod_{i,j}(p_{i\bullet}p_{\bullet j})^{\nu_{ij}} = \prod_i p_{i\bullet}^{\nu_{i\bullet}} \prod_j p_{\bullet j}^{\nu_{\bullet j}}$$

по параметрам $p_{i\bullet}$, $p_{\bullet j}$. Отсюда, как и в п. 2 § 4.3, имеем

$$\widehat{p}_{i\bullet} = \frac{\nu_{i\bullet}}{n}, \quad \widehat{p}_{\bullet j} = \frac{\nu_{\bullet j}}{n},$$

т. е. это обычные оценки параметров полиномиальных распределений

$$\mathcal{L}(\nu_{1\bullet}, \ldots, \nu_{s\bullet}) = M(n; p_{1\bullet}, \quad , p_{s\bullet})$$

и

$$\mathcal{L}(\nu_{\bullet 1}, \ldots, \nu_{\bullet k}) = M(n; p_{\bullet 1}, \quad , p_{\bullet k}).$$

Тестовая статистика для гипотезы независимости

С учетом этого тестовая статистика (11) § 4.2 принимает вид

$$\widehat{X}_n^2 = \sum_{i,j} \frac{(\nu_{ij} - n\widehat{p}_{i\bullet}\widehat{p}_{\bullet j})^2}{n\widehat{p}_{i\bullet}\widehat{p}_{\bullet j}} = n\sum_{i=1}^{s}\sum_{j=1}^{k} \frac{1}{\nu_{i\bullet}\nu_{\bullet j}}\left(\nu_{ij} - \frac{\nu_{i\bullet}\nu_{\bullet j}}{n}\right)^2 \qquad (1)$$

с числом степеней свободы

$$N - 1 - r = sk - 1 - (s + k - 2) = (s-1)(k-1).$$

Критерий [redacted] независимости χ^2

Итак, асимптотический (при больших n) вариант критерия независимости χ^2 при заданном уровне значимости α имеет вид

$$\boxed{H_0 \quad \text{отвергается} \quad \Longleftrightarrow \quad \widehat{X}_n^2 > \chi^2_{1-\alpha,(s-1)(k-1)}} \qquad (2)$$

где $\widehat{X}_n^2$ вычисляется по формуле (1).

Пример 1 [12, с. 781]. Каждый из $n = 6800$ индивидуумов классифицировался по двум признакам: цвету глаз (ξ_1) и цвету волос (ξ_2), при этом для ξ_1 было определено 3 категории (a_1, a_2, a_3), а для ξ_2 — 4 категории (b_1, b_2, b_3, b_4). Данные эксперимента приведены в таблице

ξ_1	ξ_2				$\sum$
	b_1	b_2	b_3	b_4	
a_1	1768	807	189	47	2811
a_2	946	1387	746	53	3132
a_3	115	438	288	16	857
$\sum$	2829	2632	1223	116	6800

Здесь значение статистики (1) равно 1075,2, а, например, $\chi^2_{0,999;\,6} = 22{,}5$, поэтому гипотезу H_0 о независимости этих признаков следует отклонить; вероятность ошибки при этом значительно меньше $\alpha = 0{,}001$ •

Альтернативные признаки

В приложениях часто каждый из признаков ξ_1 и ξ_2 имеет альтернативную структуру, т. е. $s = k = 2$. Типичным примером является классификация сдающих вступительные экзамены в вуз абитуриентов по признакам «принят (A) — не принят ($\bar{A}$)» и «мужчина (B) — женщина ($\bar{B}$)». Для таких ситуаций можно провести более глубокий анализ соответствующих статистических данных, которые удобно представлять в виде следующей таблицы «2 × 2» сопряженности признаков.

$\xi_1\backslash\xi_2$	$1(B)$	$0(\bar{B})$	$\sum$
$1(A)$	ν_{11}	ν_{12}	$\nu_{1\bullet}$
$0(\bar{A})$	ν_{21}	ν_{22}	$\nu_{2\bullet}$
$\sum$	$\nu_{\bullet 1}$	$\nu_{\bullet 2}$	n

Структура таблицы здесь такая же, как и таблицы «2 × 2» в случае проверки однородности (см. п. 2 § 4.3), поэтому ее анализ во многом аналогичен (хотя ее содержание имеет совсем другой смысл!). Так, совершенно аналогично можно установить, что статистика (1) имеет в данном случае представление $\widehat{X}^2_n = Z^2_n$, где

$$Z_n = \left(\frac{\nu_{11}}{\nu_{1\bullet}} - \frac{\nu_{21}}{\nu_{2\bullet}}\right)\sqrt{\frac{n\nu_{1\bullet}\nu_{2\bullet}}{\nu_{\bullet 1}\nu_{\bullet 2}}}.$$

Поскольку статистика Z_n является знакопеременной, ее знак может нести дополнительную информацию, указывая «направление» отклонения от нулевой гипотезы, и, следовательно, основываясь на ней, можно построить «более тонкий» критерий.

Установим предварительно следующую связь статистики Z_n с выборочным коэффициентом корреляции $\widehat{\rho}_n$ (см. (5) § 2.3) для наших данных: $Z_n = \sqrt{n}\widehat{\rho}_n$.

Доказательство. Действительно, пусть (X_i, Y_i), $i = 1, \ldots, n$, — независимые наблюдения над (ξ_1, ξ_2) и $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)$. Тогда,

очевидно, выборочные средние есть $\overline{X} \quad \nu_{1\bullet}/n$, $\overline{Y} = \nu_{\bullet 1}/n$, выборочные дисперсии равны

$$S_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \overline{X}(1-\overline{X}) = \frac{\nu_{1\bullet}\nu_{2\bullet}}{n^2},$$

$$S_2^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2 = \overline{Y}(1-\overline{Y}) = \frac{\nu_{\bullet 1}\nu_{\bullet 2}}{n^2},$$

наконец, выборочная ковариация (см. (4) § 2.4)

$$S_{12} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \overline{X}\,\overline{Y} = \frac{\nu_{11}}{n} - \frac{\nu_{1\bullet}\nu_{\bullet 1}}{n^2} =$$

$$= \frac{1}{n^2}\left[\nu_{11}(\nu_{1\bullet}+\nu_{2\bullet}) - \nu_{1\bullet}(\nu_{11}+\nu_{21})\right] = \frac{\nu_{1\bullet}\nu_{2\bullet}}{n^2}\left(\frac{\nu_{11}}{\nu_{1\bullet}} - \frac{\nu_{21}}{\nu_{2\bullet}}\right).$$

Отсюда имеем

$$\widehat{\rho}_n = \frac{S_{12}}{S_1 S_2} = \left(\frac{\nu_{11}}{\nu_{1\bullet}} - \frac{\nu_{21}}{\nu_{2\bullet}}\right)\sqrt{\frac{\nu_{1\bullet}\nu_{2\bullet}}{\nu_{\bullet 1}\nu_{\bullet 2}}} = \frac{Z_n}{\sqrt{n}}. \quad \blacksquare$$

В частности, отсюда следует (см. п. 3 § 2.3), что при $n \to \infty$

$$\frac{Z_n}{\sqrt{n}} \xrightarrow{\mathbf{P}} \rho = \operatorname{corr}(\xi_1, \xi_2). \tag{3}$$

Найдем, наконец, предельное (при $n \to \infty$) распределение статистики Z_n при гипотезе H_0. Для этого введем случайную величину

$$\zeta_n = \frac{1}{\sqrt{n\mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B})}}\sum_{i=1}^{n}(X_i - \mathbf{P}(A))(Y_i - \mathbf{P}(B)). \tag{4}$$

Если гипотеза H_0 справедлива, то X_i и Y_i независимы и потому

$$\mathbf{E}(X_i - \mathbf{P}(A))(Y_i - \mathbf{P}(B)) = \mathbf{E}(X_i - \mathbf{P}(A))\mathbf{E}(Y_i - \mathbf{P}(B)) = 0,$$

$$\mathbf{D}(X_i - \mathbf{P}(A))(Y_i - \mathbf{P}(B)) = \mathbf{E}(X_i - \mathbf{P}(A))^2(Y_i - \mathbf{P}(B))^2 =$$

$$= \mathbf{E}(X_i - \mathbf{P}(A))^2\mathbf{E}(Y_i - \mathbf{P}(B))^2 =$$

$$= \mathbf{D}X_i\mathbf{D}Y_i =$$

$$= \mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B}),$$

при этом слагаемые в (4) независимы и одинаково распределены по условию. Следовательно, случайная величина ζ_n представляет собой нормированную сумму независимых одинаково распределенных слагаемых и к ней можно применить центральную предельную теорему (см. (1) § 1.2), согласно которой $n \to \infty \quad \mathcal{L}(\zeta_n | H_0) \to \mathcal{N}(0,1)$.

Далее, так как

$$\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})=\sum_{i=1}^{n}(X_i-\mathbf{P}(A))(Y_i-\mathbf{P}(B))-n(\overline{X}-\mathbf{P}(A))(\overline{Y}-\mathbf{P}(B)),$$

то

$$Z_n=\sqrt{n}\,\frac{S_{12}}{S_1S_2}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{n\overline{X}(1-\overline{X})\overline{Y}(1-\overline{Y})}}=$$

$$=\left[\zeta_n-\frac{\sqrt{n}(\overline{X}-\mathbf{P}(A))}{\sqrt{\mathbf{P}(A)\mathbf{P}(\overline{A})}}\,\frac{\overline{Y}-\mathbf{P}(B)}{\sqrt{\mathbf{P}(B)\mathbf{P}(\overline{B})}}\right]\left[\frac{\mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B})}{\overline{X}(1-\overline{X})\overline{Y}(1-\overline{Y})}\right]^{1/2}$$

По теореме 3 § 2.2 об асимптотической нормальности выборочных моментов при $n\to\infty$

$$\mathcal{L}\left(\frac{\sqrt{n}(\overline{X}-\mathbf{P}(A))}{\sqrt{\mathbf{P}(A)\mathbf{P}(\overline{A})}}\right)\to\mathcal{N}(0,1),$$

а по закону больших чисел $\overline{X}\xrightarrow{\mathbf{P}}\mathbf{P}(A)$, $\overline{Y}\xrightarrow{\mathbf{P}}\mathbf{P}(B)$. Следовательно, при $n\to\infty$

$$\frac{\sqrt{n}(\overline{X}-\mathbf{P}(A))}{\sqrt{\mathbf{P}(A)\mathbf{P}(\overline{A})}}\,\frac{Y-\mathbf{P}(B)}{\sqrt{\mathbf{P}(B)\mathbf{P}(\overline{B})}}\xrightarrow{\mathbf{P}}0,$$

$$\left[\frac{\mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B})}{\overline{X}(1-\overline{X})\overline{Y}(1-\overline{Y})}\right]^{1/2}\xrightarrow{\mathbf{P}}1,$$

и потому предельные распределения Z_n и ζ_n совпадают (см. справочник предельных теорем в п. 4 § 3.5). Итак $\mathcal{L}(Z_n|H_0)\to\mathcal{N}(0,1)$ при $n\to\infty$.

Эти результаты уже дают возможность построить и рассчитать критерий проверки гипотезы независимости H_0, основываясь на статистике Z_n. Альтернативы здесь можно характеризовать величиной коэффициента корреляции $\rho=\operatorname{corr}(\xi_1,\xi_2)$, для которого нетрудно получить представление (это мы оставляем читателю в качестве простого упражнения)

$$\rho=\frac{\mathbf{P}(AB)-\mathbf{P}(A)\mathbf{P}(B)}{\sqrt{\mathbf{P}(A)\mathbf{P}(\overline{A})\mathbf{P}(B)\mathbf{P}(\overline{B})}}=\left[\mathbf{P}(A|B)-\mathbf{P}(A|\overline{B})\right]\left[\frac{\mathbf{P}(B)\mathbf{P}(\overline{B})}{\mathbf{P}(A)\mathbf{P}(\overline{A})}\right]^{1/2}\qquad(5)$$

Рассмотрим, например, правостороннюю альтернативу H_1^+ $\rho>0$ (или, что ввиду (5) эквивалентно, $\mathbf{P}(A|B)>\mathbf{P}(A|\overline{B})$), означающую положительную коррелированность признаков ξ_1 и ξ_2 (или положительную сопряженность событий A и B A в паре с B встречается с бóльшей вероятностью, чем в паре с $\overline{B}$). Тогда при гипотезе H_0 (см. (3)) $Z_n\xrightarrow{\mathbf{P}}0$, если $n\to\infty$, а при

альтернативе H_1^+ $Z_n \xrightarrow{\mathrm{P}} \infty$, поэтому естественно рассматривать слишком большие значения Z_n как свидетельство в пользу альтернативы H_1^+ Следовательно, если рассматривается задача (H_0, H_1^+), то критерий следует задать критической областью вида $\{Z_n > t_\alpha\}$ При этом $\mathbf{P}\{Z_n > t_\alpha | H_0\} \approx \Phi(-t_\alpha)$, т.е. $t_\alpha = \Phi^{-1}(1-\alpha)$ есть $(1-\alpha)$-квантиль стандартного нормального распределения $\mathcal{N}(0,1)$. Итак, в задаче $(H_0, H_1^+ \quad \rho > 0)$ для больших выборок критерий при уровне значимости α имеет вид: *если* $Z_n \leqslant t_\alpha = \Phi^{-1}(1-\alpha)$, *то гипотеза* H_0 *принимается, в противном случае она отклоняется (принимается альтернатива* H_1^+*).*

Аналогично рассматривается симметричная задача $(H_0,\ H_1^- : \rho < 0)$: здесь критическая область имеет вид $\{Z_n < -t_\alpha\}$, $\Phi(-t_\alpha) = \alpha$. Общий же критерий независимости χ^2 проверяет H_0 лишь против двусторонней альтернативы $H_1 = H_1^+ \cup H_1^- : \rho \neq 0$, и он задается в данном случае двусторонней критической областью $\{|Z_n| > t_{\alpha/2}\} = \{\widehat{X}_n^2 > \chi^2_{1-\alpha,1}\}$ (см. (2)), поскольку $t^2_{\alpha/2} = \chi^2_{1-\alpha,1}$.

Замечание. Можно показать, что если в задаче (H_0, H_1^+) «близкая» альтернатива задается условием

$$H_{1n}^+ \quad \rho = \rho(n) = \frac{a}{\sqrt{n}},$$

$a > 0$ — фиксировано, то $\mathcal{L}(Z_n | H_{1n}^+) \to \mathcal{N}(a, 1)$ и, следовательно, мощность критерия $\{Z_n > t_\alpha\}$ удовлетворяет предельному соотношению

$$\mathbf{P}\{Z_n > t_\alpha | H_{1n}^+\} \to \Phi(a - t_\alpha).$$

Пример 2. Имеются следующие данные о приеме в вуз (A — принят, B — мужчина):

	B	$\overline{B}$	$\sum$
A	97	40	137
$\overline{A}$	263	42	305
$\sum$	360	82	442

Проверим гипотезу о независимости признаков «результат» и «пол» против альтернативы $H_1^+ \quad \mathbf{P}(A|B) > \mathbf{P}(A|\overline{B})$. Здесь

$$Z_n = \left(\frac{97}{360} - \frac{40}{82}\right)\sqrt{\frac{360 \cdot 82 \cdot 442}{137 \quad 305}} = -3{,}86,$$

а $Z_n^2 = 14{,}89$. В то же время, например, $\chi^2_{0{,}999;1} = 10{,}8$. Это означает, что гипотеза о независимости признаков должна быть отвергнута (вероятность совершить при этом ошибку меньше 0,001) в пользу альтернативы

$$H_1^- \quad \mathbf{P}(A|B) < \mathbf{P}(A|\overline{B})$$

(этот факт можно интерпретировать как отсутствие дискриминации в отношении женщин при приеме в вуз). •

Пример 3. В следующей таблице приведены 818 случаев, классифицированных по двум признакам: наличию прививки против холеры (A) и отсутствие заболевания (B):

	B	$\bar{B}$	$\sum$
A	276	3	279
$\bar{A}$	473	66	539
$\sum$	749	69	818

Проверим гипотезу H_0 о независимости признаков A и B против альтернативы H_1^+ об их положительной сопряженности (т. е. об эффективности вакцинации). Здесь

$$Z_n = \left(\frac{276}{749} - \frac{3}{69}\right)\sqrt{\frac{749 \quad 69 \cdot 818}{279 \quad 539}} = 5{,}45,$$

а $Z_n^2 = 29{,}70$. Так как $\chi^2_{0{,}999;\,1} = 10{,}8$, то гипотеза H_0 должна быть отвергнута, а поскольку $\Phi^{-1}(0{,}999) = 3{,}09$, то принимается альтернатива H_1^+ Вероятность ошибочно отклонить H_0 в пользу H_1^+ при этом меньше 10^{-3} •

2. Критерий Спирмена

На практике для проверки гипотезы независимости часто используют ранговые критерии (см. п. 3 (г) § 4.3). Наиболее известным из них является *критерий Спирмена*, называемый по имени знаменитого психолога, который ввел его в 1904 г. Он основывается на коэффициенте ранговой корреляции ρ, определяемом следующим образом.

Обозначим через R_i ранг X_i среди наблюдений $X_1, \ldots, X_n$ (т. е. номер места, занимаемого величиной X_i в вариационном ряду $X_{(1)} \leqslant \quad \leqslant X_{(n)}$); аналогично, пусть S_i — ранг Y_i среди наблюдений $Y_1, \ldots, Y_n$. Тем самым исходные данные (X_i, Y_i), $i = 1, \quad ., n$ порождают n пар рангов (R_i, S_i), $i = 1, \ldots, n$. Переставив эти пары в порядке возрастания первой компоненты, обозначим полученное множество пар через $(1, T_1), \quad , (n, T_n)$. Статистика Спирмена ρ определяется теперь как коэффициент корреляции двух множеств рангов $(R_1, \quad , R_n)$ и $(S_1, \quad , S_n)$:

$$\rho = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\left[\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(S_i - \bar{S})^2\right]^{1/2}}$$

(здесь $\bar{R}$ и $\bar{S}$ — соответствующие арифметические средние). Но $(R_1, \ldots, R_n)$ и $(S_1, \quad , S_n)$ — некоторые перестановки множества $(1, 2, \quad , n)$, поэтому

$$\bar{R} = \bar{S} = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2},$$

$$\sum_{i=1}^{n}(R_i-\bar{R})^2=\sum_{i=1}^{n}(S_i-\bar{S})^2=\sum_{i=1}^{n}i^2-n\left(\frac{n+1}{2}\right)^2=\frac{n(n^2-1)}{12},$$

так как

$$\sum_{i=1}^{n}i^2=\frac{n(n+1)(2n+1)}{6}.$$

Статистика Спирмена

Отсюда

$$\rho=\frac{12}{n(n^2-1)}\sum_{i=1}^{n}\left(R_i-\frac{n+1}{2}\right)\left(S_i-\frac{n+1}{2}\right)=$$

$$=\frac{12}{n(n^2-1)}\sum_{i=1}^{n}\left(i-\frac{n+1}{2}\right)\left(T_i-\frac{n+1}{2}\right). \qquad (6)$$

Таким образом, ρ — линейная функция рангов T_i. Часто используют также формулы

$$\rho=1-\frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i-S_i)^2=1-\frac{6}{n(n^2-1)}\sum_{i=1}^{n}(T_i-i)^2 \qquad (7)$$

совпадение которых с (6) проверяется непосредственно.

При полном соответствии рангов ($R_i=S_i$, $i=1,\ \ldots,n$) статистика $\rho=1$, а при противоположных рангах ($T_i=n-i+1$, $i\ \ 1,\ \ldots,n$) $\rho=-1$, вообще же $-1\leqslant\rho\leqslant 1$. Далее, если гипотеза H_0 верна, то $\mathbf{E}\rho=0$, $\mathbf{D}\rho=1/(n-1)$, поэтому значения ρ, близкие к крайним, рассматривают как свидетельствующие против H_0, и, следовательно, критическую область критерия Спирмена задают в виде $\mathcal{T}_{1\alpha}=\{|\rho|>t_\alpha(n)\}$

Критерий Спирмена

Для определения числового значения критической границы $t_\alpha(n)$ при заданных объеме выборки n и уровне значимости α используют таблицы табулированного распределения статистики ρ при нулевой гипотезе, рассчитанные для $n=2,\ \ldots,30$ (в [3] приведена таблица для $n=4(1)10$). Для больших n можно воспользоваться асимптотическим результатом $\mathcal{L}(\sqrt{n}\rho|H_0)\to\mathcal{N}(0,1)$. Отсюда следует, что если выбрать $t_\alpha(n)=t_{\alpha/2}/\sqrt{n}$, $\Phi(-t_{\alpha/2})=\alpha/2$, то при больших n

$$\mathbf{P}\{\rho\in\mathcal{T}_{1\alpha}|H_0\}=\mathbf{P}\{\sqrt{n}|\rho|>t_{\alpha/2}|H_0\}\approx 2\Phi(-t_{\alpha/2})=\alpha,$$

т. е. уровень значимости критерия приблизительно равен α.

3. Критерий Кендалла

Другой известный ранговый критерий предложен М. Кендаллом (1938) и основан на статистике

$$\tau=\frac{2}{n(n-1)}S_\tau,\quad S_\tau=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\operatorname{sign}(T_j-T_i), \qquad (8)$$

где $\operatorname{sign} a = 1$, если $a > 0$, и -1, если $a < 0$. Если гипотеза H_0 верна, то

$$\mathbf{E}\tau = 0, \quad \mathbf{D}\tau = \frac{2(2n+5)}{9n(n-1)} \quad \text{и} \quad \mathcal{L}(\tau|H_0) \sim \mathcal{N}\left(0, \frac{4}{9n}\right)$$

при $n \to \infty$. Отсюда следует, что при больших n критическую область следует выбрать в виде

Критерий Кендалла

$$T_{1\alpha} = \left\{|\tau| > \frac{2t_{\alpha/2}}{3\sqrt{n}}\right\}.$$

Для $n = 4(1)10$ критическая граница критерия Кендалла рассчитывается с помощью таблицы, приведенной в [3]. Например, при $n = 9$ критической области $\{|\tau| \geqslant 0{,}5\}$ отвечает вероятность ошибки первого рода 0,076.

Практически статистику Кендалла τ рекомендуется вычислять по формуле

$$\tau = \frac{4N}{n(n-1)} - 1,$$

где N — число тех пар индексов (i, j), $i < j$, для которых $T_i < T_j$. Действительно, обозначим дополнительно через N_1 число тех пар индексов (i, j), $i < j$, для которых $T_i > T_j$. Тогда, очевидно, $S_\tau = N - N_1$, а $N + N_1 = n(n-1)/2$ — общее число пар (i, j), $i < j$. Складывая эти два равенства, получаем $S_\tau = 2N - n(n-1)/2$. Отсюда и из (8) получаем указанное представление для τ.

4. Случай m признаков

В этом пункте мы опишем типичную практическую ситуацию, в которой для анализа данных при проверке гипотезы независимости эффективно применяются ранговые методы. Рассмотрим совокупность индивидуумов, обладающих таким признаком, который, может быть, и не поддается точной количественной оценке, однако позволяет сравнивать индивидуумы друг с другом. Таким образом, в результате подобного сравнения всю совокупность можно «ранжировать», приписав каждому индивидууму порядковый номер, соответствующий итогам сравнения с остальными индивидуумами. Если индивидуумы могут обладать не одним, а двумя и более признаками, то для исследования их влияния друг на друга обычно рассматривают выборку из n независимых индивидуумов и ранжируют их по каждому признаку отдельно. Пусть количество исследуемых признаков есть $m \geqslant 2$, тогда результаты ранжировок по этим признакам можно записать в виде $(m \times n)$-матрицы, i-я строка которой

$$R = \begin{Vmatrix} r_{11} & r_{12} & & r_{1n} \\ r_{21} & r_{22} & & r_{2n} \\ & & & \\ r_{m1} & r_{m2} & & r_{mn} \end{Vmatrix}$$

содержит результаты ранжировки по i-му признаку ($i = 1, \ldots, m$), а столбцы соответствуют индивидуумам, принадлежащим исследуемой выборке (таким

образом, r_{ij} — это порядковый номер j-го индивидуума при сравнении его с другими индивидуумами по i-му признаку).

Коэффициент конкордации

В качестве единой выборочной меры связи m признаков Кендалл и Бэбингтон Смит (1939) предложили *коэффициент согласованности* W, называемый также *коэффициентом конкорданции* (concórdance — согласие, соответствие, конкорданция):

$$W = \frac{12}{m^2(n^3 - n)} S_W, \quad S_W = \sum_{j=1}^{n} \left[\sum_{i=1}^{m} r_{ij} - \frac{m(n+1)}{2} \right]^2 \tag{9}$$

Если по матрице R вычислить коэффициент ρ Спирмена (см. (6)–(7)) для каждой из $m(m-1)/2$ пар признаков, то арифметическое среднее таких ρ будет равняться $(mW - 1)/(m - 1)$. В частности, если $m = 2$, то $\rho = 2W - 1$.

Коэффициент W принимает значения из отрезка $[0, 1]$ и используется в качестве тестовой статистики для проверки гипотезы H_0 о независимости признаков (признаки называются независимыми, если для любого столбца матрицы R ранги $r_{1j}, r_{2j}, \ldots, r_{mj}$ являются независимыми случайными величинами). Если гипотеза H_0 верна, то

$$\mathbf{E}W = \frac{1}{m}, \quad \mathbf{D}W = \frac{2(m-1)}{m^3(n-1)}$$

и распределение W удовлетворительно аппроксимируется бета-распределением [3, с. 97–98]; в [3] имеются также таблицы для расчета критерия при $n = 3$, $m = 3(1)10$; $n = 4$, $m = 3(1)6$ и $n = 5$, $m = 3$. Например, при $n = 5$ и $m = 3$ критическая область критерия независимости трех признаков задается неравенством $\{W \geqslant 0{,}733\}$, если уровень значимости есть $\alpha = 0{,}038$.

Приведем несколько иллюстративных примеров расчета критерия, опирающегося на коэффициент ранговой корреляции ρ Спирмена.

Пример 4. В следующей таблице указана относительная оценка знаний по математике и литературе 8 учеников, занумерованных числами от 1 до 8, причем лучшим знаниям соответствует меньший ранг:

Ученик (i)	1	2	3	4	5	6	7	8
Ранг знаний по математике (R_i)	5	2	1	7	8	4	6	3
Ранг знаний по литературе (S_i)	4	3	2	1	7	6	5	8

Проверим по критерию Спирмена гипотезу H_0 о независимости между успехами по математике и литературе против альтернативы H_1^+ успехи по математике и литературе положительно коррелированы.

В данном случае в пользу альтернативы говорят «большие» значения статистики ρ, поэтому в задаче (H_0, H_1^+) критическую область разумно взять в виде одностороннего неравенства $\{\rho > t_\alpha(n)\}$. При $n = 8$ и, например,

$\alpha = 0{,}1$ критическая граница $t_\alpha(n) = 0{,}4762$, а величина ρ, вычисленная по формуле (7), равна

$$\rho = 1 - \frac{6}{8^3 - 8}\left(1^2 + 1^2 + 1^2 + 6^2 + 1^2 + 2^2 + 1^2 + 5^2\right) = 1 - \frac{6}{8^3 - 8}\cdot\frac{70}{} = \frac{1}{6} = 0{,}1667.$$

Так как $\rho < t_{0,1}(8)$, у нас нет оснований отвергать гипотезу независимости. •

Пример 5. Относительная оценка знаний математики и физики в группе из 7 учеников дала следующие результаты (R_i — ранг по математике, S_i — ранг по физике i-го ученика):

i	1	2	3	4	5	6	7
R_i	5	7	1	4	2	3	6
S_i	4	6	3	5	1	2	7

Проверим с уровнем значимости $\alpha = 0{,}05$ гипотезу H_0 о независимости между успехами по математике и физике. В данном случае критическая область критерия Спирмена имеет вид $\{|\rho| > t_{0,05}(7)\}$ и вычисления дают

$$t_{0,05}(7) = 0{,}7450, \quad \rho = 1 - \frac{6}{7^3 - 7}\cdot 10 = 1 - \frac{5}{28} = 0{,}8022.$$

Следовательно, гипотеза H_0 отвергается. •

Важное замечание

В заключение этой темы подчеркнем, что выше рассмотрена гипотеза независимости H_0 в ее самой общей форме, без каких-либо дополнительных априорных предположений (за исключением предположений типа непрерывности распределения наблюдений). В то же время такая гипотеза может возникать и в рамках параметрических моделей, т. е. когда априори предполагается, что все допустимые распределения наблюдений принадлежат заданному параметрическому семейству распределений. Если, например, мы работаем в рамках двумерной нормальной модели, то гипотеза независимости эквивалентна предположению о равенстве нулю коэффициента корреляции. Критерий проверки такой гипотезы, тестовой статистикой которого является выборочный коэффициент корреляции, рассмотрен ранее в п. 4 § 2.3.

§ 4.5. Гипотеза случайности

Хаос — ужас богов.

Древнее высказывание

В различных статистических задачах исходные данные $\boldsymbol{X} = (X_1, \quad , X_n)$ часто рассматривают как случайную выборку из некоторого распределения $\mathcal{L}(\xi)$, т. е. считают компоненты X_i вектора данных $\boldsymbol{X}$ независимыми и одинаково

распределенными случайными величинами. Но иногда это исходное предположение само нуждается в проверке как специальная статистическая гипотеза (см. пример 7 Введения). Такая гипотеза, называемая *гипотезой случайности*, записывается в виде

$$H_0 \quad F_X(x_1, \quad , x_n) = F(x_1) \quad F(x_n),$$

где $F(x)$ — некоторая одномерная функция распределения. Мы рассмотрим некоторые подходы к построению критериев согласия для проверки такой гипотезы.

1. Критерий инверсий

Предположим, что вектор X имеет непрерывное распределение. Если гипотеза H_0 верна, то компоненты вектора X «равноправны» и поэтому данные не должны быть ни в каком смысле упорядочены. Другими словами, ситуацию, соответствующую гипотезе H_0, можно охарактеризовать как «полный хаос», или «полный беспорядок». При отклонениях от H_0 исходные данные имеют тот или иной порядок, проявляются связи. Следовательно, при проверке H_0 надо основываться на тестовых статистиках, измеряющих степень «беспорядка» данных. Одной из таких естественных статистик является число инверсий в выборке, которое определяется следующим образом.

Инверсии [redacted] Построим вариационный ряд $X_{(1)} < \quad < X_{(n)}$ выборки X. Говорят, что компоненты X_i и X_j образуют *инверсию*, если $i < j$, но X_i стоит правее X_j в вариационном ряду, т. е. наблюдению с меньшим номером соответствует большее значение. Пусть η_i — число инверсий, образованных компонентой X_i (в вариационном ряду левее X_i стоит η_i элементов выборки с бóльшими номерами), $i = 1, \quad , n-1$. Тогда $T_n = T_n(X) = \eta_1 + \ldots + \eta_{n-1}$ — общее число инверсий для выборки X. Статистика T_n является естественной мерой «беспорядка» среди наблюдений и ее можно использовать для проверки гипотезы H_0. Крайние случаи, когда вариационный ряд имеет вид

$$X_1 < X_2 < \quad < X_n \quad \text{или} \quad X_n < X_{n-1} < \quad < X_1,$$

естественно рассматривать как свидетельства «полного отсутствия беспорядка», т. е. противоречащие гипотезе H_0. В первом из этих случаев статистика T_n принимает минимальное значение, равное 0, а во втором — она максимальна и равна

$$(n-1) + (n-2) + \quad + 1 = \frac{n(n-1)}{2}.$$

Таким образом, малые значения T_n и значения, близкие к $n(n-1)/2$, естественно рассматривать как критические для гипотезы H_0. Чтобы высказать более определенные суждения о виде критерия инверсий, надо исследовать распределение статистики T_n при справедливости гипотезы H_0.

Из соображений симметрии ясно, что при гипотезе H_0 любое из $n!$ относительных расположений элементов выборки в соответствующем вариационном ряду имеет одинаковую вероятность $1/n!$ Случайная величина η_i

определяется расположением компоненты X_i по отношению к $X_{i+1}, \ldots, X_n$ в вариационном ряду и не зависит от относительного расположения последних между собой, т. е. η_i не зависит от $\eta_{i+1}, \ldots, \eta_{n-1}$, $i = 1, \ldots, n-2$. Но это означает, что случайные величины $\eta_1, \ldots, \eta_{n-1}$ взаимно независимы.

Далее, η_i может с одной и той же вероятностью $1/(n-i+1)$ принимать значения $0, 1, \ldots, n-i$, поэтому ее производящая функция имеет вид (см. замечание в упр. 21 к гл. 1)

$$\varphi_i(z) = \sum_{r=0}^{n-i} \mathbf{P}\{\eta_i = r\} z^r = \frac{1}{n-i+1}(1 + z + \ldots + z^{n-i}),$$

а производящая функция статистики T_n — вид

$$\mathbf{E}z^{T_n} = \sum_r \mathbf{P}\{T_n = r\} z^r = \prod_{i=1}^{n-1} \varphi_i(z) = \frac{1}{n!} \prod_{r=1}^{n-1} (1 + z + \ldots + z^r).$$

Отсюда имеем (см. указание к упр. 24 гл. 1)

$$\begin{aligned}
\mathbf{E}\eta_i &= \varphi_i'(1) = \frac{n-i}{2}, \\
\mathbf{D}\eta_i &= \varphi_i''(1) + \mathbf{E}\eta_i - (\mathbf{E}\eta_i)^2 = \frac{(n-i)(n-i+2)}{12}, \\
\mathbf{E}T_n &= \sum_{i=1}^{n-1} \mathbf{E}\eta_i = \frac{n(n-1)}{4}, \qquad (1) \\
\mathbf{D}T_n &= \sum_{i=1}^{n-1} \mathbf{D}\eta_i = \frac{n(n-1)(2n+5)}{72}.
\end{aligned}$$

Итак, среднее значение статистики T_n при нулевой гипотезе совпадает с серединой промежутка $[0, n(n-1)/2]$, и в критическую область $\mathcal{T}_{1\alpha}$ следует включать все целые точки этого промежутка, достаточно удаленные от середины, т. е. можно положить

Критерий инверсий

$$\mathcal{T}_{1\alpha} = \left\{ \left| T_n - \frac{n(n-1)}{4} \right| > t_\alpha(n) \right\}. \qquad (2)$$

Критическую границу $t_\alpha(n)$ при заданном уровне значимости α выбирают из условия $\mathbf{P}\{T_n \in \mathcal{T}_{1\alpha} | H_0\} \leqslant \alpha$ или, что эквивалентно,

$$\mathbf{P}\left\{ \frac{n(n-1)}{4} - t_\alpha(n) \leqslant T_n \leqslant \frac{n(n-1)}{4} + t_\alpha(n) \right\} \geqslant 1 - \alpha \qquad (3)$$

($t_\alpha(n)$ — это минимальное число, удовлетворяющее данному соотношению). Соотношениями (2)–(3) и определяется *критерий инверсий*. Распределение статистики T_n протабулировано для значений $n = 2(1)12$ (соответствующая

таблица приведена в [3]) и может быть использовано для расчета критерия (2). При $n > 10$ можно использовать нормальное приближение

$$\mathcal{L}(T_n|H_0) \approx \mathcal{N}(\mathbf{E}T_n, \mathbf{D}T_n),$$

где $\mathbf{E}T_n$ и $\mathbf{D}T_n$ указаны в (1). В таких случаях можно приближенно полагать

$$t_\alpha(n) = \frac{t_{\alpha/2}n^{3/2}}{6}, \quad \Phi(-t_{\alpha/2}) = \frac{\alpha}{2}.$$

2. Проблема датчиков и обобщенный критерий хи-квадрат

В примере 7 Введения была сформулирована проблема контроля «качества» последовательностей, вырабатываемых датчиками (генераторами) случайных или псевдослучайных чисел, которая кратко называется *проблемой датчиков*. Научная и практическая значимость этой проблемы очень высока, так как спрос на такие датчики, а также на сами случайные последовательности в настоящее время непрерывно возрастает (появляются даже научно-производственные фирмы, занимающиеся производством и продажей больших массивов случайных чисел; например, с 1996 г. в мире распространяется компакт-диск «The Marsaglia random number CDROM», содержащий 4,8 млрд «истинно случайных» бит).



Эта проблема имеет большую историю. Еще в начале XX в. случайные последовательности целых чисел имитировались с помощью простейших случайных экспериментов: бросание монеты или игральной кости, извлечение шаров из урны, раскладывание карт, рулетка и т. д. В 1927 г. Л. Типпетом впервые были опубликованы таблицы, содержащие свыше 40 000 случайных цифр, «произвольно извлеченных из отчетов о переписи населения». В 1939 г. М. Дж. Кендалл и Б. Бэбингтон Смит с помощью специально сконструированного механического устройства — генератора случайных чисел создали таблицу, включающую 10^5 случайных цифр. В 1946 г. американский математик Джон фон Нейман впервые предложил компьютерный алгоритм генерации случайных чисел. В 1955 г. компания RAND Corporation опубликовала получившие широкую популярность таблицы, содержащие 10^6 случайных цифр, сгенерированных на ЭВМ.

Таблицы случайных чисел можно найти в каждом сборнике статистических таблиц. Так в [3] приведена таблица, содержащая 12 500 цифр от 0 до 9. Эти данные можно рассматривать как реализации независимых и одинаково распределенных случайных величин, принимающих значения 0, 1, 2, , 9 с одной и той же вероятностью, равной 0,1. Каждая пара таких цифр представляет собой реализацию случайной величины, которая может принимать любые целочисленные значения от 00 до 99 с одинаковыми вероятностями, равными 0,01. Аналогичный результат получится, если цифры сгруппировать по три, четыре и т. д.

Если каждую группу из k цифр, рассматриваемую как целое число, умножить на 10^{-k} то получим реализации случайной величины ξ, принимающей k-разрядные значения от 0 до $(1 - 10^{-k})$ с одинаковыми вероятностями,

равными 10^{-k}. Такое распределение вероятностей близко к равномерному распределению $U(0, 1)$, причем разность соответствующих функций распределения не превосходит 10^{-k}. Следовательно, реализации ξ можно рассматривать как реализации случайных величин, равномерно распределенных на отрезке $[0, 1]$. Эти реализации и называют обычно равномерно распределенными случайными числами. Из этих чисел с помощью алгоритмов, описанных в § 1.3, можно построить (сгенерировать) случайные числа с любым другим распределением, отличным от равномерного распределения $U(0, 1)$.

Итак, в математическом плане, основное требование к датчику случайных чисел состоит в том, что вырабатываемая им последовательность $\xi_1, \xi_2, \ldots$ должна быть в статистическом смысле неотличима от последовательности независимых и одинаково распределенных случайных величин, т. е. должна соответствовать гипотезе случайности H_0 (согласовываться с ней). Теперь мы уточним постановку задачи, которую будем обсуждать далее.

Пусть задана последовательность из n случайных величин (выборка)

$$\xi_1, \xi_2, \quad , \xi_n, \tag{4}$$

принимающих значения из конечного множества (алфавита) $\mathcal{E} = \{1, 2, \ldots, N\}$, отосительно которой требуется проверить гипотезу H_0 о том, что эти случайные величины в совокупности независимы и равномерно распределены на множестве $\mathcal{E}$, т. е. $\mathrm{P}\{\xi_i = j\} = 1/N$, $j = 1, \quad , N$, $i = 1, 2, \quad$ Гипотеза H_0 — это есть гипотеза об идеальном датчике: если гипотеза H_0 верна, то датчик, вырабатывающей последовательность (4), реализует «идеальную» схему независимых испытаний с равновероятными исходами. Но идеальных датчиков не существует; более того, как отмечено в примере 7 Введения (см. также п. 1 § 1.3), часто датчик вырабатывает последовательность (4) программным способом (т. е. эти числа могут быть даже детерминированными), но тогда эти числа должны быть «практически неотличимы» (или «очень похожи») от независимых и равновероятных чисел. В математическом плане это понимается так: выборка (4) не должна опровергать гипотезу H_0 никаким тестом. Практически же идут на следующий компромис: определяется некоторая совокупность статистических критериев (тестов) и, основываясь на выборке (4), подвергают гипотезу H_0 проверке по каждому из этих контрольных тестов. Если гипотеза H_0 успешно проходит такое тестирование, то она принимается как истинная, и числа (4) рекомендуются для использования в соответствующих целях как независимые и равномерно распределенные на множестве $\mathcal{E}$ случайные числа. Общепринятый в настоящее время стандартный набор таких тестов сформулировал Д. Кнут (1969) и он имеется в книге [13], однако работа по совершенствованию методов тестирования случайных и псевдослучайных последовательностей продолжается. В качестве примера одного из подходящих для проверки гипотезы H_0 тестов мы подробно опишем критерий типа хи-квадрат, основанный на подсчете частот s-цепочек в последовательности (4).

Идею построения этого теста можно объяснить следующим образом. Поскольку гипотеза H_0 — простая (она однозначно фиксирует распределение

наблюдений (4)), то, в принципе, для ее проверки можно было бы использовать классический критерий χ^2, описанный в п. 2 § 4.2. Но этот критерий «работает» лишь при условии независимости наблюдений (4): он состоятелен против альтернатив, которые учитывают изменение лишь вероятностей появления исходов, но сохраняют предположение о независимости наблюдений. Но ведь именно это предположение о независимости и требуется подвергнуть проверке! Поэтому критерий χ^2 Пирсона не решает эту проблему. Сохраняя идеологию метода хи-квадрат (сравнение эмпирических частот некоторых событий с их гипотетическими вероятностями с помощью меры хи-квадрат), можно попытаться сконструировать тест, «реагирующий» на альтернативы того или иного типа зависимости между данными (4), рассматривая вместо частот отдельных исходов в последовательности (4) частоты цепочек исходов некоторой фиксированной длины $s \geqslant 2$, причем цепочки должны быть перекрывающимися. Точное определение таково: *s-цепочкой* (или *s-граммой*) называется любая подпоследовательность последовательности (4), состоящая из s подряд идущих ее членов.

s-цепочки

Таким образом, из (4) можно образовать следующие s-цепочки: $(\xi_1, \ldots, \xi_s)$, $(\xi_2, \ldots, \xi_{s+1})$, $\ldots$, $(\xi_{n-s+1}, \ldots, \xi_n)$, всего $n - s + 1$ цепочек. Пусть теперь $\bar{i} = (i_1 \ldots i_s)$ — любая s-мерная комбинация элементов множества $\mathcal{E}$, всех различных таких комбинаций N^s. В методе хи-квадрат используются частоты s-цепочек и $(s+1)$-цепочек, подсчитываемые по следующим формулам:

$$\begin{aligned} \nu_{\bar{i}} &= \sum_{t=1}^{n-s+1} I(\xi_t = i_1,\ \xi_{t+1} = i_2, \ldots \xi_{t+s-1} = i_s), \\ \nu_{\bar{i}j} &= \sum_{t=1}^{n-s} I(\xi_t = i_1, \ldots \xi_{t+s-1} = i_s,\ \xi_{t+s} = j), \quad j \in \mathcal{E} \end{aligned} \tag{5}$$

(напомним, что $I(A)$ есть индикатор события A). При этом достаточно вычислить лишь частоты $\{\nu_{\bar{i}j}\}$ (их число равно N^{s+1}), так как можно положить

$$\nu_{\bar{i}} = \sum_{j=1}^{N} \nu_{\bar{i}j}$$

(на самом деле последняя сумма может на 1 отличаться от $\nu_{\bar{i}}$, но на излагаемых далее асимптотических результатах это никак не сказывается).

Далее, при гипотезе H_0 (все ξ_j независимы и равномерно распределены на $\mathcal{E}$)

$$\mathbf{E}\nu_{\bar{i}j} = \frac{n-s}{N^{s+1}} \sim \frac{n}{N^{s+1}}$$

(асимптотическое равенство справедливо при $n \to \infty$ и фиксированном s) и аналогично $\mathbf{E}\nu_{\bar{i}} \sim n/N^s$, поэтому для проверки H_0 можно сравнивать эмпирические частоты $\nu_{\bar{i}j}$ с, например, $\nu_{\bar{i}}/N$ при помощи «стандартной»

статистики хи-квадрат:

$$X^2_{n,s+1} \sum_{(\bar{i}j)\in\mathcal{E}^{s+1}} \frac{(\nu_{\bar{i}j} - \nu_{\bar{i}}/N)^2}{\nu_{\bar{i}}/N}. \tag{6}$$

Обоснованием именно такой конструкции тестовой статистики является следующий асимтотический результат [29].

Теорема 1. Пусть гипотеза H_0, означающая, что последовательность (4) получена по схеме независимых испытаний с N равновероятными исходами, справедлива. Тогда статистика (6) имеет предельное (при $n \to \infty$) распределение $\chi^2(N^s(N-1))$. Соответствующий критерий, задаваемый при уровне значимости α критической областью

$$\{X^2_{n,s+1} > \chi^2_{1-\alpha, N^s(N-1)}\}, \tag{7}$$

состоятелен против альтернатив, при которых последовательность (4) представляет собой однородную эргодическую цепь Маркова порядка s или меньше.

Здесь нам встретились некоторые новые понятия, которые мы должны пояснить, прежде чем комментировать этот результат. Понятие цепи Маркова нам уже встречалось ранее. Именно, в п. 10 § 1.3 приведено определение простой (т. е. порядка 1) однородной цепи Маркова. Обобщение на произвольный порядок s выглядит следующим образом (сравни с (6) § 1.3): при любых $k > s$ и $i_1, \quad , i_k \in \mathcal{E}$ условная вероятность $\mathbf{P}\{\xi_k = i_k \mid \xi_m = i_m,\ m < k\}$ не зависит от значений i_m для $m < k - s$, т. е. вероятности, с которыми ξ_k принимает возможные значения, зависят лишь от значений s предыдущих случайных величин последовательности (4). В этом случае говорят, что цепь Маркова управляется переходными вероятностями

$$p_{i_1 \ldots i_s; j} = \mathbf{P}\{\xi_k = j \mid \xi_{k-s} = i_1, \qquad \xi_{k-1} = i_s\}. \tag{8}$$

Если эти переходные вероятности не зависят от «времени» k, то цепь называется *однородной*. Чтобы полностью определить вероятностный закон, управляющий цепью Маркова порядка s, нужно еще задать начальное распределение цепи, т. е. вероятности $\mathbf{P}\{\xi_k = i_k,\ k \leqslant s\}$. Наконец, цепь Маркова является *эргодической*, если для любых $i_1, \quad , i_s \in \mathcal{E}$ существуют пределы

$$\lim_{t\to\infty} \mathbf{P}\{\xi_t = i_1, \qquad \xi_{t+s-1} = i_s\} = p_{i_1 \ldots} \tag{9}$$

не зависящие от начального распределения цепи; эти пределы $p_{i_1,\ldots,i_s}$ называются *стационарными вероятностями* комбинаций $\bar{i} = (i_1, \ldots, i_s)$. При этом для эргодичности цепи достаточно, чтобы все переходные вероятности (8) были положительны (в цепи возможны все переходы).

Цепь Маркова порядка s — это специфическая, но в то же время весьма общая модель зависимых («на глубину s») испытаний, подходящая для описания многих реальных процессов, поэтому в обсуждаемой проблеме датчиков

подобные, как говорят, *марковские* альтернативы представляют особый интерес, и качество тех или иных критериев для проверки гипотезы случайности H_0 прежде всего проверяют именно на таких альтернативах: критерий должен быть состоятельным против марковских альтернатив. Теорема 1 и утверждает, что обобщенный критерий χ^2, определяемый соотношениями (6)–(7), с этих позиций является вполне подходящим тестом для гипотезы случайности: если мы хотим контролировать отклонения от нулевой гипотезы типа марковской зависимости до порядка s включительно, то мы должны использовать частоты $(s+1)$-цепочек в исходной последовательности (4) для конструирования тестовой статистики (6). Сама же тестовая статистика и основанный на ней критерий (7) вполне приемлемы с позиций трудоемкости практических вычислений.

Упражнения

Если действовать не будешь,
ни к чему ума палата.

Шота Руставели

1 Среди n = 10 000 чисел $0, 1, \ldots, 9$ числа, не превосходящие 4, встретились 5089 раз. Проверить по критерию χ^2 Пирсона гипотезу о равновероятности чисел при уровне значимости $\alpha = 0{,}05$. При каком уровне значимости эта гипотеза отвергается?

2 Смоделировать выборку объема $n = 1000$ из распределения Бернулли Bi(1, 3/5) и проверить по критерию χ^2 соответствие данных теоретической модели.

◀ *Указание.* Использовать, например, таблицу случайных чисел и алгоритм п. 2 § 1.3. ▶

3 При $n = 4000$ независимых испытаниях события A_1, A_2, A_3, составляющие полную группу, осуществились соответственно 1905, 1015 и 1080 раз. Проверить, согласуются ли данные при уровне значимости 0,05 с гипотезой H_0 $p_1 = 1/2$, $p_2 = p_3 = 1/4$, где $p_i = \mathrm{P}\{A_i\}$

◀ *Указание.* Применить критерий χ^2 с критической границей $\chi^2_{0.95;2} = 5{,}99$. ▶

4 В десятичной записи числа π среди первых 10 002 знаков после запятой цифры $0, 1, \ldots, 9$ встречаются соответственно 968, 1026, 1021, 974, 1014, 1046, 1021, 970, 948, 1014 раз. Проверить по критерию χ^2 гипотезу о равновероятности «случайных» чисел $0, 1, \ldots, 9$. При каком уровне значимости эта гипотеза отвергается?

5 Используя таблицу значений какой-либо функции ($\cos x$, e^x, $\ln x$ и т. д.) выписать 100 цифр, выбирая из каждого значения функции второй знак справа, и проверить для такой выборки по критерию χ^2 гипотезу о равновероятности цифр $0, 1, \ldots, 9$ при уровне значимости 0,01.

6 Согласуются ли данные, приведенные в примере 2 в § 2.2 с гипотезой о симметричности костей? Ответить на такой же вопрос в отношении данных, приведенных в упр. 9 к гл. 2.

7 Смоделировать выборку объема 100 из распределения Bi(3, 1/2) и проверить соответствие полученных данных теоретической модели.

◀ *Указание.* Использовать алгоритм п. 2 § 1.3. ▶

8 Смоделировать выборку объема 100 из равномерного распределения $\mathcal{U}(0, 1)$ и проверить по полученным данным гипотезу о равномерности распределения по критериям Колмогорова и χ^2

9 Смоделировать выборку объема $n = 100$ из показательного распределения $\Gamma(1, 1)$ и, сгруппировав полученные данные по $N = 4$ равновероятным (при гипотезе H_0 $F(x) = 1 - e^{-x}$ $x \geqslant 0$) интервалом, проверить по критерию χ^2 гипотезу H_0 при уровне значимости $\alpha = 0{,}1$.

◀ *Указание.* Использовать алгоритм п. 7 § 1.3. ▶

10 Проверить по критерию Колмогорова соответствие данных, полученных в упр. 6 к гл. 2, гипотезе о нормальной распределенности чисел.

11 В генетической модели Фишера [2, с. 79] считается, что вероятности появления потомства, классифицируемого по четырем типам, имеют вид $p_1(\theta) = (2 + \theta)/4$, $p_2(\theta) = p_3(\theta) = (1 - \theta)/4$, $p_4(\theta) = \theta/4$, $\theta \in (0, 1)$ — неизвестный параметр. Построить критерий χ^2 соответствия этой модели реальным данным.

12 При 8000 независимых испытаниях события A_1, A_2 и A_3, составляющие полную группу, осуществились соответственно 2014, 5012 и 974 раза. Верна ли при уровне значимости 0,05 гипотеза: $\mathrm{P}\{A_1\} = 0{,}5 - 2\theta$, $\mathrm{P}\{A_2\} = 0{,}5 + \theta$, $\mathrm{P}\{A_3\} = \theta$, где $\theta \in (0, 1/4)$?

13 Проверить гипотезу H_0 $\mathcal{L}(\xi) \in \Pi(\theta)$, $\theta > 0$ — неизвестный параметр, для данных, приведенных в упр. 11 к гл. 2.

◀ *Указание.* В качестве оценки параметра принять выборочное среднее. ▶

14 Через равные промежутки времени в тонком слое раствора золота регестрировалось число ξ частиц золота, попавших в поле зрения микроскопа. По данным наблюдений, приведенных в следующей таблице (ν_i — число опытов, для которых $\xi = i$, $i = 0, 1, 2, \ldots$), проверить гипотезу пуассоновости H_0 $\mathcal{L}(\xi) \in \Pi(\theta)$, $\theta > 0$ — неизвестный параметр, оцениваемый выборочным средним.

i	0	1	2	3	4	5	6	7	Σ
ν_i	112	168	130	68	32	5	1	1	$n = 518$

15 В течение эпидемии гриппа среди 2000 контролируемых индивидуумов одно заболевание наблюдалось у 181 человека, а дважды болели гриппом лишь 9 человек. Согласуются ли при уровне значимости 0,05 эти данные с гипотезой, согласно которой число заболеваний отдельного индивидуума — биномиальная $\mathrm{Bi}(2, \theta)$ случайная величина?

16 Исходя из формул (8)–(9) § 4.2, убедиться в том, что среднее и дисперсия статистики $\mathring{X}_n^2$ при «близких» альтернативах вида $\underline{p}(n) = \underline{p}^\circ + \underline{\beta}/\sqrt{n}$, $\underline{\beta} \neq \underline{0}$, имеют при $n \to \infty$ следующие асимптотические представления:

$$\mathrm{E}\left(\mathring{X}_n^2 | \underline{p}(n)\right) = N - 1 + \sum_{j=1}^{N} \frac{\beta_j^2}{p_j^\circ} + O\left(\frac{1}{\sqrt{n}}\right),$$

$$\mathrm{D}\left(\mathring{X}_n^2 | \underline{p}(n)\right) = 2(N - 1) + 4\sum_{j=1}^{N} \frac{\beta_j^2}{p_j^\circ} + O\left(\frac{1}{\sqrt{n}}\right).$$

17 Исходя из формул (15) § 4.2, получить, что при

$$p_j = p_j(n) = \frac{1}{N}\left(1 + \frac{b_j}{n^{1/4}}\right), \quad j = 1, \ldots, N \quad \text{и}$$

$$n, N \to \infty, \quad \frac{n}{N} \to \rho > 0, \quad \max_{1 \leqslant j \leqslant N} |b_j| \leqslant c < \infty$$

$$\mathrm{E}\mu_0(n, N) = Ne^{-\rho} + \sqrt{N}\rho^{3/2}\frac{e^{-\rho}b^2(N)}{2}(1 + o(1)), \quad b^2(N) = \frac{1}{N}\sum_{j=1}^{N} b_j^2,$$

$$\mathrm{D}\mu_0(n, N) = Ne^{-\rho}\left(1 - (1 + \rho)e^{-\rho}\right)(1 + o(1)).$$

18 Следующая таблица содержит данные о смертности среди матерей, родивших первого ребенка, в четыре различные периода времени [2, с. 102] (n_j — число матерей, ν_j — число смертных исходов):

n_j	1072	1133	2455	1995
ν_j	22	23	49	33

По критерию однородности χ^2 с двумя исходами проверить гипотезу H_0 о том, что в уровнях смертности между этими периодами не существует различия.

19 Смоделировать выборку объема $n = 200$ равномерно распределенных чисел (X_i). С помощью критерия Смирнова проверить, что две подвыборки $(X_{2i},\ i = 1,\quad 100)$ и $(X_{2i+1},\ i = 0, 1, \ldots, 99)$ являются выборками из одного и того же распределения.

20 Смоделировать последовательность $(X_i,\ i = 1, 2,\quad 100)$ случайных величин, с равными вероятностями принимающих значения 1, 2, 3, 4, 5. Образовать из этой последовательности две выборки $(X_{2i},\ i = 1, 2,\quad, 50)$ и $(X_{2i+1},\ i = 0, 1,\quad, 49)$ и проверить для них гипотезу однородности.

◀ *Указание.* Использовать алгоритм п. 4 § 1.3. ▶

21 Пусть случайные величины $\eta_1, \ldots, \eta_N$ независимы и $\mathcal{L}(\eta_j) = \Pi(\theta_j)$, $j = 1, \ldots, N$, где параметры θ_j неизвестны; дополнительно известно, что $\eta_1 + \quad + \eta_N = n$. Построить при этом условии критерий χ^2 для проверки гипотезы однородности

$$H_0 \quad \theta_1 = \quad = \theta_N.$$

◀ *Указание.* Воспользоваться упр. 22 к гл. 1. ▶

22 Из 300 абитуриентов, поступивших в вуз, 97 человек имели 5 в школе и 48 получили 5 на вступительных экзаменах по тому же предмету, причем только 18 человек имели 5 и в школе, и на экзамене. С уровнем значимости 0,1 проверить гипотезу о независимости оценок 5 в школе и на экзамене.

23 Используя данные, полученные в упр. 20, образовать последовательность пар (X_{2i-1}, X_{2i}), $i = 1, 2,\quad, 50$, и проверить по этим данным гипотезу независимости.

24 Можно ли с уровнем значимости 0,001 считать, что последовательность чисел 1,05; 1,12; 1,37; 1,50; 1,51; 1,73; 1,85; 1,98 является реализацией случайного вектора, все 8 компонент которого — независимые одинаково распределенные непрерывные случайные величины?

◀ *Указание.* Применить критерий инверсий. ▶

25 Смоделировать выборку из 100 равномерных псевдослучайных чисел. С помощью статистики T_n — числа инверсий — проверить гипотезу случайности для этих данных.

26 Доказать, что число инверсий T_n в случайной выборке объема n из непрерывного распределения распределено асимптотически нормально $\mathcal{N}(n(n-1)/4, n^3/36)$.

◀ *Указание.* Используя производящую функцию $\mathrm{E}z^{T_n}$ (см. п. 1 § 4.5), перейти к характеристической функции и убедиться в том, что при $n \to \infty$

$$\mathrm{E}\exp\left\{it\left(T_n - \frac{n(n-1)}{4}\right)\frac{6}{n^{3/2}}\right\} \to \exp\left\{-\frac{t^2}{2}\right\} \quad \blacktriangleright$$

Глава 5

Параметрические гипотезы

Статистика — это искусство уточнять то, что является неизвестным.

Параметрические гипотезы, рассматриваемые в настоящей главе, — это гипотезы об истинном значении неизвестного параметра, определяющего априори заданное параметрическое семейство распределений. Здесь изложены и продемонстрированы на конкретных примерах основные принципы построения оптимальных или асимптотически оптимальных (для больших выборок) критериев проверки таких гипотез, в основе которых лежит метод отношения правдоподобия Ю. Неймана и Э. Пирсона. Отдельные параграфы посвящены последовательному анализу А. Вальда и статистическому анализу конечных цепей Маркова.

§ 5.1. Общие положения

Важный класс статистических гипотез составляют гипотезы о параметрических моделях. В этом случае априори постулируется, что класс $\mathcal{F}$ допустимых распределений наблюдаемой случайной величины ξ является некоторым параметрическим семейством $\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$, т. е. все допустимые распределения идентифицируются соответствующим параметром θ $(\theta_1, \quad, \theta_r)$ со значениями из параметрического множества $\Theta \subseteq R^r$ поэтому гипотезы о распределении $\mathcal{L}(\xi)$, по существу, относятся к неизвестному параметру θ и называются *параметрическими*. Примерами параметрических гипотез являются утверждения:

Параметрические гипотезы

1) H_0 $\theta = \theta_0$, где $\theta_0 \in \Theta$ — некоторое фиксированное значение параметра;
2) H_0 $\theta_1 = \theta_2 = \quad = \theta_r$;
3) H_0 $g(\theta) = g_0$, где $g(\theta)$ — некоторая (в общем случае векторная) функция, g_0 — фиксированное значение.

В общем случае основная (нулевая) параметрическая гипотеза задается указанием некоторого подмножества $\Theta_0 \subset \Theta$, элементом которого является, по предположению, неизвестное истинное значение параметра θ, что записывается в виде H_0 $\theta \in \Theta_0$. Точки $\theta \in \Theta_1 = \Theta \backslash \Theta_0$ называются *альтернативами*, а утверждение H_1 $\theta \in \Theta_1$ — альтернативной гипотезой. Таким образом, в параметрическом случае альтернативная гипотеза конкретизируется и имеет такой же вид, как и основная гипотеза H_0. Другими словами, в данном случае

отклонения от основной гипотезы имеют вполне определенный вид, ограниченный рамками рассматриваемого параметрического семейства. Следовательно, и соответствующий статистический критерий должен быть направлен на выявление именно таких отклонений от H_0, учитывать их специфику. Следует ожидать, что наличие такой значительной априорной информации о допустимых распределениях наблюдений (данных) должно привести к более сильным выводам, найти свое отражение в свойствах соответствующих *параметрических* критериев.

Общие принципы построения критериев, изложенные в § 4.1, относятся и к рассматриваемому случаю. Напомним кратко, что если в задаче ($H_0 : \theta \in \Theta_0$, H_1 $\theta \in \Theta_1$) применяется некоторый «критерий $\mathfrak{X}_{1\alpha}$», т. е.

$$\boxed{H_0 \text{ отвергается} \iff X \in \mathfrak{X}_{1\alpha},}$$

где X — исходные данные и при заданном уровне значимости α критическая область $\mathfrak{X}_{1\alpha}$ удовлетворяет условию

$$W(\theta) = W(\theta; \mathfrak{X}_{1\alpha}) = \mathbf{P}_\theta\{X \in \mathfrak{X}_{1\alpha}\} \leqslant \alpha, \quad \forall\, \theta \in \Theta_0 \tag{1}$$

(функция $W(\theta; \mathfrak{X}_{1\alpha})$, $\theta \in \Theta$, называется *функцией мощности* критерия $\mathfrak{X}_{1\alpha}$), то **проблема максимум** состоит в построении такого критерия $\mathfrak{X}^*_{1\alpha}$, для которого при произвольном $\mathfrak{X}_{1\alpha}$, удовлетворяющем (1),

$$\begin{aligned} &W(\theta; \mathfrak{X}^*_{1\alpha}) \leqslant W(\theta; \mathfrak{X}_{1\alpha}) \quad \text{при} \quad \theta \in \Theta_0 \quad \text{и} \\ &W(\theta; \mathfrak{X}^*_{1\alpha}) \geqslant W(\theta; \mathfrak{X}_{1\alpha}) \quad \text{при} \quad \theta \in \Theta_1. \end{aligned} \tag{2}$$

Такой критерий $\mathfrak{X}^*_{1\alpha}$ (когда он существует) называется *равномерно наиболее мощным критерием* (р. н. м. к.) в задаче (H_0, H_1).



Р. н. м. к.

Равномерно наиболее мощный критерий не всегда существует, поскольку, как правило, критерий, максимизирующий мощность при определенной альтернативе $\theta \in \Theta_1$, зависит от этой альтернативы, и экстремальная задача (2) при ограничении (1) имеет решение только в некоторых специальных случаях. С примерами таких ситуаций мы встретимся в дальнейшем.

Часто ограничиваются рассмотрением подкласса *несмещенных* критериев, для которых одновременно с (1) выполняется условие

$$W(\theta) \geqslant \alpha, \quad \forall\, \theta \in \Theta_1. \tag{3}$$

В ряде задач, для которых р. н. м. к. не существуют, могут иметь место р. н. м. несмещенные критерии.

Наконец, напомним, что в случае рандомизированного критерия, задаваемого критической функцией $\varphi(x)$, функция мощности «критерия φ» определяется равенством



Рандомизированный критерий

$$W(\theta) = W(\theta; \varphi) = \mathbf{E}_\theta \varphi(X) = \int \varphi(x)\, dF_X(x; \theta). \tag{4}$$

Излагаемая ниже теория параметрических критериев базируется на методе отношения правдоподобия, предложенном Ю. Нейманом и Э. Пирсоном (1933)

в качестве основы общего метода построения критериев. В основе этой теории лежит знаменитая (*фундаментальная*) *лемма Неймана—Пирсона*, дающая вид наиболее мощного критерия в задаче различения двух простых гипотез. С этого результата мы и начнем (в случае, когда множество Θ_1 состоит из одной точки, т. е. альтернатива H_1 простая, вместо р. н. м. к. используется термин *наиболее мощный критерий*).

§ 5.2. Выбор из двух простых гипотез. Критерий Неймана—Пирсона

1. Постановка задачи и предварительные соображения

Рассмотрим простейший частный (но играющий принципиальную роль в последующей теории) случай, когда обе гипотезы H_0 и H_1 — простые. Формализуется это так: постулируется (априори предполагается), что допустимыми распределениями наблюдаемой случайной величины ξ являются лишь два заданных распределения (две функции распределения) $F_0(x)$ и $F_1(x)$ и требуется по выборке $X = (X_1, \quad , X_n)$ из $\mathcal{L}(\xi)$ проверить гипотезу $H_0 \quad F_\xi(x) = F_0(x)$ против альтернативы $H_1 \quad F_\xi(x) = F_1(x)$. Чтобы подчеркнуть, что это частный случай параметрических гипотез, можно рассмотреть семейство распределений вида $F(x;\theta) = (1-\theta)F_0(x) + \theta F_1(x)$, где параметрическое множество Θ состоит из двух точек: $\Theta = \{0, 1\}$ и проверяемая (основная) гипотеза есть $H_0 \quad \theta = \theta_0 = 0$, а альтернатива $H_1 \quad \theta = \theta_1 = 1$. Тогда $F_i(x) = F(x;\theta_i)$, $i = 0, 1$, и речь идет о выборе одного из этих распределений в качестве истинного.

Согласно изложенному выше общему принципу, задача построения наилучшего (наиболее мощного) критерия $\mathfrak{X}^*_{1\alpha}$ сводится в данном случае к решению экстремальной задачи максимизации по $\mathfrak{X}_{1\alpha}$ мощности $W(\theta_1, \mathfrak{X}_{1\alpha})$:

$$W(\theta_1; \mathfrak{X}_{1\alpha}) = \mathbf{P}_{\theta_1}\{X \in \mathfrak{X}_{1\alpha}\} = \max \tag{1}$$

при ограничении

$$W(\theta_0; \mathfrak{X}_{1\alpha}) = \mathbf{P}_{\theta_0}\{X \in \mathfrak{X}_{1\alpha}\} \leqslant \alpha.$$

Нетрудно понять, как надо строить оптимальную критическую область $\mathfrak{X}^*_{1\alpha}$. Рассмотрим правдоподобие данных $L(\underline{x};\theta) = f(x_1;\theta) \quad f(x_n;\theta)$ при каждой из гипотез H_0 и H_1 (см. п. 1 § 3.2). Тогда, например, для непрерывных данных проблема (1) запишется в виде

$$\begin{aligned} W(\theta_0; \mathfrak{X}_{1\alpha}) &= \int\limits_{\mathfrak{X}_{1\alpha}} L(\underline{x};\theta_0)\, d\underline{x} = \alpha, \\ W(\theta_1; \mathfrak{X}_{1\alpha}) &= \int\limits_{\mathfrak{X}_{1\alpha}} L(\underline{x};\theta_1)\, d\underline{x} = \max, \end{aligned} \tag{2}$$

т. е. среди всех областей $\mathfrak{X}_{1\alpha}$ надо выбрать ту, для точек которой правдоподобие $L(\underline{x}; \theta_1)$ будет наибольшим в сравнении с $L(\underline{x}; \theta_0)$. Следовательно все выборочные точки $\underline{x} = (x_1, \ldots, x_n)$ надо «упорядочить» соответственно величине отношения

Отношение правдоподобия

$$l(\underline{x}) = \frac{L(\underline{x}; \theta_1)}{L(\underline{x}; \theta_0)} = \frac{\prod_{i=1}^{n} f(x_i; \theta_1)}{\prod_{i=1}^{n} f(x_i; \theta_0)} \tag{3}$$

и затем включить в $\mathfrak{X}^*_{1\alpha}$ все те из них, для которых $l(\underline{x}) \geqslant c$, а граница $c = c_\alpha$ должна быть выбрана так, чтобы обеспечить первое условие в (2). Функция $l(\underline{x})$ в (3) называется *отношением правдоподобия*, она и порождает тестовую статистику $T(X) = l(X)$ критерия Неймана—Пирсона в рассматриваемой задаче (H_0, H_1). Таким образом, критерий Неймана—Пирсона основывается на *статистике отношения правдоподобия*

$$l(X) = \frac{L(X; \theta_1)}{L(X; \theta_0)}$$

и задается критической областью

Критерий Неймана—Пирсона

$$\mathfrak{X}^*_{1\alpha} = \{\underline{x} = (x_1, \ldots, x_n) \quad l(\underline{x}) \geqslant c_\alpha\}.$$

То, что такая конструкция приводит к оптимальному критерию, и утверждает фундаментальная лемма Неймана—Пирсона.

2. Случай абсолютно непрерывных распределений

Прежде чем давать ее точную формулировку, условимся о некоторых соглашениях. Рассмотрим сначала случай абсолютно непрерывных распределений F_0 и F_1 и будем считать, что соответствующие плотности $f_0(x)$ и $f_1(x)$ удовлетворяют условию $f_j(x) > 0$, $j = 0, 1$ (это исключает тривиальные ситуации, когда при одной из гипотез некоторые значения x «запрещены», и следовательно, наблюдая в опыте такие значения, мы однозначно определяем, какая из гипотез верна). Определим теперь функцию

$$\psi(c) = \mathbf{P}_0\{l(X) \geqslant c\} = \int\limits_{\underline{x}: l(\underline{x}) \geqslant c} L(\underline{x}; \theta_0)\, d\underline{x}$$

(для сокращения записи пишем $\mathbf{P}_i$ вместо $\mathbf{P}_{\theta_i}$, $i = 0, 1$, и $\mathbf{P}_i\{S\}$ вместо $\mathbf{P}_{\theta_i}\{X \in S\}$). С ростом c эта функция может только убывать, кроме того, $\psi(0) = 1$. Далее,

$$1 \geqslant \mathbf{P}_1\{l(X) \geqslant c\} = \int\limits_{\underline{x}: l(\underline{x}) \geqslant c} L(\underline{x}; \theta_1)\, d\underline{x} \geqslant c \int\limits_{\underline{x}: l(\underline{x}) \geqslant c} L(\underline{x}; \theta_0)\, d\underline{x} = c\psi(c),$$

поэтому $\psi(c) \leqslant 1/c$, т. е. $\psi(c) \to 0$ при $c \to \infty$. Будем предполагать, что существует такое значение $c = c_\alpha$, что $\psi(c_\alpha) = \alpha$ (в частности, это всегда имеет место, если функция $\psi(c)$ непрерывна). Тогда для соответствующего множества $\mathfrak{X}^*_{1\alpha} = \{\underline{x} \quad l(\underline{x}) \geqslant c_\alpha\}$ имеем

$$W(\theta_0; \mathfrak{X}^*_{1\alpha}) = \mathbf{P}_0\{\mathfrak{X}^*_{1\alpha}\} = \int\limits_{\mathfrak{X}^*_{1\alpha}} L(\underline{x}; \theta_0)\, d\underline{x} = \psi(c_\alpha) = \alpha. \tag{4}$$

Справедливо следующее утверждение.

Теорема 1 (Лемма Неймана—Пирсона). При сделанных предположениях критическая область $\mathfrak{X}^*_{1\alpha}$ задает наиболее мощный критерий для гипотезы $H_0 \quad F_\xi(x) = F_0(x)$ относительно альтернативы $H_1 \quad F_\xi(x) = F_1(x)$ среди всех критериев уровня значимости α.

Доказательство. Рассмотрим любой другой критерий $\mathfrak{X}_{1\alpha}$, так что

$$W(\theta_0; \mathfrak{X}_{1\alpha}) = \mathbf{P}_0\{\mathfrak{X}_{1\alpha}\} = \int\limits_{\mathfrak{X}_{1\alpha}} L(\underline{x}; \theta_0)\, d\underline{x} = \alpha. \tag{5}$$

Из формул (4) и (5) получаем

$$\mathbf{P}_0\{\mathfrak{X}^*_{1\alpha} - \mathfrak{X}_{1\alpha}\mathfrak{X}^*_{1\alpha}\} = \alpha - \mathbf{P}_0\{\mathfrak{X}_{1\alpha}\mathfrak{X}^*_{1\alpha}\} = \mathbf{P}_0\{\mathfrak{X}_{1\alpha} - \mathfrak{X}_{1\alpha}\mathfrak{X}^*_{1\alpha}\}.$$

Согласно определению множества $\mathfrak{X}^*_{1\alpha}$, вне этого множества $l(\underline{x}) < c_\alpha$, т. е.

$$c_\alpha L(\underline{x}; \theta_0) > L(\underline{x}; \theta_1),$$

поэтому

$$\begin{aligned}\mathbf{P}_1\{\mathfrak{X}^*_{1\alpha} - \mathfrak{X}_{1\alpha}\mathfrak{X}^*_{1\alpha}\} &\geqslant c_\alpha \mathbf{P}_0\{\mathfrak{X}^*_{1\alpha} - \mathfrak{X}_{1\alpha}\mathfrak{X}^*_{1\alpha}\} = \\ &= c_\alpha \mathbf{P}_0\{\mathfrak{X}_{1\alpha} - \mathfrak{X}_{1\alpha}\mathfrak{X}^*_{1\alpha}\} > \mathbf{P}_1\{\mathfrak{X}_{1\alpha} - \mathfrak{X}_{1\alpha}\mathfrak{X}^*_{1\alpha}\}.\end{aligned}$$

Прибавляя к обеим частям этого неравенства $\mathbf{P}_1\{\mathfrak{X}_{1\alpha}\mathfrak{X}^*_{1\alpha}\}$, получим

$$\mathbf{P}_1\{\mathfrak{X}^*_{1\alpha}\} = W(\theta_1; \mathfrak{X}^*_{1\alpha}) > W(\theta_1; \mathfrak{X}_{1\alpha}) = \mathbf{P}_1\{\mathfrak{X}_{1\alpha}\}. \ \blacksquare$$

Замечание. Конечно, оптимальный критерий должен удовлетворять условию несмещенности (3) § 5.1. Покажем, что критерий Неймана—Пирсона $\mathfrak{X}^*_{1\alpha}$ обладает этим свойством. Пусть в (4) $c_\alpha > 1$. Тогда

$$W(\theta_1; \mathfrak{X}^*_{1\alpha}) = \int\limits_{\mathfrak{X}^*_{1\alpha}} L(\underline{x}; \theta_1)\, d\underline{x} \geqslant c_\alpha \int\limits_{\mathfrak{X}^*_{1\alpha}} L(\underline{x}; \theta_0)\, d\underline{x} = \alpha c_\alpha > \alpha.$$

Если же $c_\alpha \leqslant 1$, то

$$W(\theta_1; \mathfrak{X}^*_{1\alpha}) = 1 - \mathbf{P}_1\{\overline{\mathfrak{X}}^*_{1\alpha}\} > 1 - c_\alpha \mathbf{P}_0\{\overline{\mathfrak{X}}^*_{1\alpha}\} \geqslant 1 - \mathbf{P}_0\{\overline{\mathfrak{X}}^*_{1\alpha}\} = \mathbf{P}_0\{\mathfrak{X}^*_{1\alpha}\} = \alpha.$$

Следовательно, в любом случае $W(\theta_1; \mathfrak{X}^*_{1\alpha}) > \alpha$.

Рассмотрим теперь вопрос о расчете критерия $\mathfrak{X}^*_{1\alpha}$, т. е. о вычислении критической границы c_α (см. (4)) и вероятности ошибки 2-го рода

$$\beta = \beta(\alpha, n) = 1 - W(\theta_1; \mathfrak{X}^*_{1\alpha}).$$

Точное вычисление этих основных параметров критерия возможен в том случае, когда распределение тестовой статистики $l(X)$ известно как при гипотезе H_0, так как при альтернативе H_1 (см. примеры ниже). Но это не всегда осуществимо, поэтому часто для решения этой задачи применяется асимптотический подход (при объеме выборки $n \to \infty$), основанный на использовании центральной предельной теоремы (ЦПТ). Использование ЦПТ

Обозначим

$$Z_i = \ln \frac{f_1(X_i)}{f_0(X_i)}, \quad i = 1, \quad , n,$$

тогда $S_n = Z_1 + \ldots + Z_n$ есть сумма независимых одинаково распределенных при каждой гипотезе случайных величин и событие $\{l(X) \geqslant c_\alpha\}$ можно записать в виде $\{S_n \geqslant \ln c_\alpha\}$. Обозначим, далее,

$$a_j = \mathbf{E}_{\theta_j} Z_1 = \int f_j(x) \ln \frac{f_1(x)}{f_0(x)}\, dx, \quad \sigma_j^2 = \mathbf{D}_{\theta_j} Z_1, \quad j = 0, 1$$

(предполагается, что эти моменты существуют). Из неравенства $\ln t \leqslant t - 1$, $t > 0$, и представления

$$a_0 = \int f_0(x) \left[\ln \frac{f_1(x)}{f_0(x)} - \left(\frac{f_1(x)}{f_0(x)} - 1 \right) \right] dx$$

следует, что $a_0 \leqslant 0$, и аналогично, $a_1 \geqslant 0$. Будем далее считать, что $a_0 < 0$, $a_1 > 0$ и $\sigma_j^2 > 0$, $j = 0, 1$. Тогда при $n \to \infty$ на основании ЦПТ (см. (1) § 1.2)

$$\mathcal{L}_{\theta_j}(S_n) \sim \mathcal{N}(na_j, n\sigma_j^2), \quad j = 0, 1.$$

Отсюда следует, что для больших выборок для расчета критерия Неймана—Пирсона можно воспользоваться нормальной аппроксимацией для суммы S_n. Имеем

$$\alpha = \mathbf{P}_0\{S_n \geqslant \ln c_\alpha\} = \mathbf{P}_0\left\{ \frac{S_n - na_0}{\sqrt{n}\sigma_0} \geqslant \frac{\ln c_\alpha - na_0}{\sqrt{n}\sigma_0} \right\} \approx \Phi(-t_\alpha),$$

если c_α выбрано так, что Асимметрический вариант критерия Неймана—Пирсона

$$\ln c_\alpha = na_0 + t_\alpha \sigma_0 \sqrt{n}, \quad \Phi(-t_\alpha) = \alpha. \quad (6)$$

Соотношение (6) решает вопрос о приближенном определении критической границы в критерии Неймана—Пирсона в случае больших выборок.

Аналогично, с учетом (6) для вероятности ошибки 2-го рода $\beta(\alpha, n)$ при больших n можем записать

$$\beta(\alpha, n) = \mathbf{P}_1\{S_n < \ln c_\alpha\} = \mathbf{P}_1\left\{ \frac{S_n - na_1}{\sqrt{n}\sigma_1} < \frac{a_0 - a_1}{\sigma_1}\sqrt{n} + t_\alpha \frac{\sigma_0}{\sigma_1} \right\}. \quad (7)$$

Здесь $(a_0 - a_1)\sqrt{n} \to -\infty$ при $n \to \infty$ и прямое применение ЦПТ для вычисления этой вероятности при больших n становится уже некорректным (из ЦПТ можно лишь заключить, что $\beta(\alpha, n) \to 0$ при $n \to \infty$); говорят,

Большие уклонения

что в данном случае значение суммы S_n находится в *зоне больших уклонений* (от среднего значения na_1 при рассматриваемой гипотезе H_1). Таким образом, для определения значения $\beta(\alpha, n)$ при больших n необходимо использовать более глубокие (нежели ЦПТ) результаты из теории суммирования независимых случайных величин — так называемую *теорию больших уклонений*. Применение этой теории к нашей задаче дает в типичных случаях следующий результат [5, с. 289]: при $n \to \infty$

$$\beta(\alpha, n) \sim \frac{1}{\sigma_0\sqrt{2\pi n}} \exp\left\{na_0 + \sqrt{n}\sigma_0 t_\alpha - \frac{t_\alpha^2}{2}\right\}, \tag{8}$$

т. е. ошибка β убывает экспоненциально с возрастанием числа наблюдений n, если при этом гипотетические распределения F_0 и F_1 остаются фиксированными (т. е. фиксированы параметры a_j, σ_j^2, $j = 0, 1$).

«Близкие» гипотезы

Однако, можно представить себе более сложную ситуацию, когда вместе с ростом числа наблюдений n изменяются и распределения F_j, $j = 0, 1$, т. е. их параметры a_j, σ_j^2 являются некоторыми функциями от n : $a_j = a_j(n)$, $\sigma_j^2 = \sigma_j^2(n)$. Тогда соотношение (7) позволяет ответить на вопрос о том, для каких «близких» гипотез критерий Неймана—Пирсона $\mathfrak{X}_{1\alpha}^*$ имеет невырождающуюся в пределе мощность, т. е.

$$W(\theta_1; \mathfrak{X}_{1\alpha}^*) = 1 - \beta(\alpha, n) \to \gamma, \quad \alpha < \gamma < 1. \tag{9}$$

Это будет так, если, например,

$$\frac{\sqrt{n}\big(a_0(n) - a_1(n)\big)}{\sigma_1(n)} \to -\delta < 0, \quad \sigma_0(n) \sim \sigma_1(n):$$

в этом случае из (7) имеем $\gamma = \Phi(\delta - t_\alpha)$.

Таким образом, мы можем сделать следующий принципиальный вывод: при больших выборках наиболее мощный критерий Неймана—Пирсона способен различать гипотезы, сближающиеся со скоростью порядка $1/\sqrt{n}$ (расстояние между гипотезами измеряется взвешенной разностью средних значений случайной величины $Z = \ln\left(f_1(\xi)/f_0(\xi)\right)$ при этих гипотезах).

Пример 1 (Критерий Неймана—Пирсона для среднего нормальной модели). Пусть о нормально распределенной случайной величине ξ с известной дисперсией σ^2 и неизвестным средним θ имеются две гипотезы: H_0 $\theta = \theta_0$ и $H_1^+ : \theta = \theta_1$ (для определенности будем считать, что $\theta_0 < \theta_1$). Если $X = (X_1, \ldots, X_n)$ — выборка из $\mathcal{L}(\xi)$ и $\underline{x}$ — ее реализация, то отношение правдоподобия имеет вид

$$l(\underline{x}) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(x_i - \theta_1)^2 - (x_i - \theta_0)^2\right]\right\} =$$

$$= \exp\left\{\frac{n}{\sigma^2}(\theta_1 - \theta_0)\bar{x} - \frac{n}{2\sigma^2}(\theta_1^2 - \theta_0^2)\right\}$$

и неравенство $l(\underline{x}) \geqslant c$ эквивалентно неравенству

$$\bar{x} \geqslant \sigma^2 \frac{\ln c}{n(\theta_1 - \theta_0)} + \frac{\theta_1 + \theta_0}{2},$$

которое удобно переписать в виде

$$\frac{\sqrt{n}}{\sigma}(\bar{x} - \theta_0) \geqslant \frac{\sigma}{\sqrt{n}(\theta_1 - \theta_0)} \ln c + \frac{\sqrt{n}}{2\sigma}(\theta_1 - \theta_0) \equiv t(c).$$

Так как $\mathcal{L}_{\theta_0}(\overline{X}) = \mathcal{N}(\theta_0, \sigma^2/n)$, то отсюда имеем

$$\psi(c) = \mathbf{P}_{\theta_0}\{l(X) \geqslant c\} = \mathbf{P}_{\theta_0}\left\{\frac{\sqrt{n}}{\sigma}(\overline{X} - \theta_0) \geqslant t(c)\right\} = \Phi(-t(c)).$$

При $c > 0$ функция $t(c)$ непрерывна, поэтому непрерывна и функция $\psi(c)$, и для любого $\alpha \in (0, 1)$ однозначно определена величина c_α, где $t(c_\alpha) = t_\alpha$, а $\Phi(-t_\alpha) = \alpha$.

Критерий Неймана—Пирсона в задаче (H_0, H_1^+)

Таким образом, в данном случае критерий Неймана—Пирсона задается критической областью

$$\mathfrak{X}_{1\alpha}^{*+} = \left\{\underline{x} \;\; \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} \geqslant t_\alpha\right\}, \quad \Phi(-t_\alpha) = \alpha. \tag{10}$$

Подчеркнем следующие важные обстоятельства: в данном случае критическая область наиболее мощного критерия $\{l(X) \geqslant c_\alpha\}$ оказалась выраженной через выборочное среднее $\overline{X}$ — достаточную статистику для модели $\mathcal{N}(\theta, \sigma^2)$ (см. п. 1 § 3.3), а сама критическая область (10) не зависит от конкретного значения альтернативы $\theta_1 (> \theta_0)$.

Вычислим мощность критерия (10). Так как $\mathcal{L}_{\theta_1}(\overline{X}) = \mathcal{N}(\theta_1, \sigma^2/n)$, то

$$\begin{aligned} W(\theta_1; \mathfrak{X}_{1\alpha}^{*+}) &= \mathbf{P}_{\theta_1}\left\{\overline{X} \geqslant \theta_0 + \frac{\sigma}{\sqrt{n}} t_\alpha\right\} = \\ &= \mathbf{P}_{\theta_1}\left\{\frac{\sqrt{n}}{\sigma}(\overline{X} - \theta_1) \geqslant -\frac{\sqrt{n}}{\sigma}(\theta_1 - \theta_0) + t_\alpha\right\} = \\ &= \Phi\left(\frac{\sqrt{n}}{\sigma}(\theta_1 - \theta_0) - t_\alpha\right). \end{aligned} \tag{11}$$

В частности, отсюда видно, что $W(\theta_1; \mathfrak{X}_{1\alpha}^{*+}) > \Phi(-t_\alpha) = \alpha$ (несмещенность).

Критерий Неймана—Пирсона в задаче (H_0, H_1^-)

Аналогично, при $\theta_1 < \theta_0$, т. е. в задаче (H_0, H_1^-), критическая область наиболее мощного критерия имеет вид

$$\mathfrak{X}_{1\alpha}^{*-} = \left\{\underline{x} \;\; \sqrt{n}\,\frac{\bar{x} - \theta_0}{\sigma} \leqslant -t_\alpha\right\}, \quad \Phi(-t_\alpha) = \alpha, \tag{12}$$

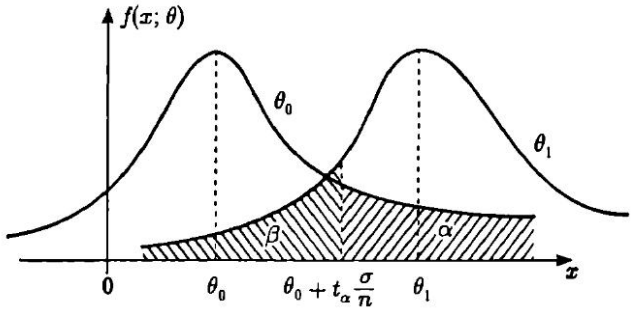


Рис. 1

а его мощность

$$W(\theta_1; \mathfrak{X}^{*-}_{1\alpha}) = \Phi\left(\frac{\sqrt{n}}{\sigma}(\theta_0 - \theta_1) - t_\alpha\right).$$

Из (11) следует, что вероятность ошибки 2-го рода критерия (10) равна

$$\beta = \beta(\alpha, n) = 1 - W(\theta_1; \mathfrak{X}^{*+}_{1\alpha}) = \Phi\left(t_\alpha - \sqrt{n}\,\frac{\theta_1 - \theta_0}{\sigma}\right). \qquad (13)$$

Наглядной иллюстрацией этих результатов служит рис. 1.

Рассмотрим следующую задачу. Пусть заранее заданы вероятности ошибок α и β. Определим, каким должно быть минимальное число $n^* = n^*(\alpha, \beta)$ наблюдений, необходимых для того, чтобы ошибочные заключения могли быть сделаны с вероятностями, не превосходящими α и β

Из (13) следует, что $\beta(\alpha, n)$ с ростом n убывает к нулю, поэтому искомое число n^* есть наименьшее из тех n, для которых $\beta(\alpha, n) \leqslant \beta$. Для определения n имеем два уравнения:

$$\Phi(-t_\alpha) = \alpha \quad \text{и} \quad \Phi\left(t_\alpha - \sqrt{n}\,\frac{\theta_1 - \theta_0}{\sigma}\right) = \beta.$$

Пусть ζ_p обозначает p-квантиль стандартного нормального распределения $\mathcal{N}(0, 1)$, т. е. $\Phi(\zeta_p) = p$. Тогда

$$-t_\alpha = \zeta_\alpha, \quad t_\alpha - \sqrt{n}\,\frac{\theta_1 - \theta_0}{\sigma} = \zeta_\beta, \quad \text{т. е.} \quad n = \sigma^2 \frac{(\zeta_\alpha + \zeta_\beta)^2}{(\theta_1 - \theta_0)^2}.$$

Число наблюдений в критерии Неймана—Пирсона

Но n должно быть целым числом, поэтому надо положить

$$n^* = \left[\sigma^2 \frac{(\zeta_\alpha + \zeta_\beta)^2}{(\theta_1 - \theta_0)^2}\right] + 1, \qquad (14)$$

где $[a]$ — целая часть a. Отсюда следует, что при фиксированных ошибках α и β число необходимых наблюдений пропорционально дисперсии σ^2 и обратно пропорционально квадрату разности между гипотетическими средними.

Если, например, $\alpha = \beta = 0{,}001$, то $\zeta_\alpha = \zeta_\beta = -3{,}09$ и из формулы (14) имеем

$$n^* = \frac{38{,}2\sigma^2}{(\theta_1 - \theta_0)^2} + 1.$$

При $\sigma^2 = 1$, $\theta_1 - \theta_0 = 1$ значение $n^* = 39$. Таким образом, если требуется различить гипотезы с указанными параметрами, то только при числе испытаний, не меньшем 39, можно быть уверенным в том, что, следуя критерию (10), будет принято ошибочное решение с вероятностью, не большей 0,001.

«Близкие» альтернативы

Наконец, здесь легко решается и вопрос о поведении критерия при «близких» альтернативах (см. конец п. 2). Пусть, например, решается задача (H_0, H_1^+) и рассматриваются «близкие» альтернативы вида

$$H_{1n}^+ \quad \theta = \theta_{1n} = \theta_0 + \varepsilon_n,$$

где $\varepsilon_n \downarrow 0$ при $n \to \infty$. Тогда из (11) имеем

$$W(\theta_{1n}; \mathfrak{X}_{1\alpha}^{*+}) = \Phi\left(\frac{\sqrt{n}}{\sigma}\varepsilon_n - t_\alpha\right) \longrightarrow$$

$$\longrightarrow \begin{cases} 1, & \text{если} \quad \sqrt{n}\varepsilon_n \to \infty, \\ \gamma = \Phi\left(\dfrac{a}{\sigma} - t_\alpha\right), & \text{если} \quad \varepsilon_n = \dfrac{a}{\sqrt{n}}, \quad a > 0, \\ \alpha & \text{если} \quad \varepsilon_n = o\left(\dfrac{1}{\sqrt{n}}\right). \end{cases}$$

Отсюда следует, что в данном случае «пороговые» альтернативы имеют вид

$$\theta_{1n} = \theta_0 + \frac{a}{\sqrt{n}},$$

$a > 0$ — константа; более близкие альтернативы $\left(\theta_{1n} - \theta_0 = o(1/\sqrt{n})\right)$ от H_0 асимптотически не отличаются, а более далекие $\left(\sqrt{n}(\theta_{1n} - \theta_0) \to \infty\right)$ улавливаются критерием с вероятностью, стремящейся к 1 при $n \to \infty$, когда они верны (т. е. против таких «близких» альтернатив критерий сохраняет свойство состоятельности). Итак, для больших выборок критерий Неймана—Пирсона $\mathfrak{X}_{1\alpha}^{*+}$ (как и критерий $\mathfrak{X}_{1\alpha}^{*-}$ (12)) способен различать гипотезы, сближающиеся со скоростью $1/\sqrt{n}$.

Сравнение двух средних

В качестве важного приложения этих результатов построим критерий для проверки гипотезы о равенстве средних двух независимых нормальных выборок. Более конкретно, речь идет об известной нам *задаче сравнения двух средних* (см. пример 2 § 2.5) в следующем варианте: пусть $\overline{X}$ и $\overline{Y}$ — выборочные средние двух независимых выборок объемов n и m из распределений $\mathcal{N}(\theta_1, \sigma_1^2)$ и $\mathcal{N}(\theta_2, \sigma_2^2)$ соответственно (дисперсии σ_i^2 известны) и требуется проверить гипотезу о равенстве средних $H_0 : \Delta = \theta_1 - \theta_2 = 0$ против альтернативы $H_1^+ : \Delta = \Delta_0 > 0$.

В качестве тестовой статистики здесь естественно взять $T = (\overline{X} - \overline{Y})/\sigma$, где $\sigma^2 = \sigma_1^2/n + \sigma_2^2/m$ (см. таблицу в п. 1 § 3.8). Легко находим, что распределение этой статистики при обеих гипотезах имеет вид: $\mathcal{L}(T|H_0) = \mathcal{N}(0, 1)$, $\mathcal{L}(T|H_1^+) = \mathcal{N}(\Delta_0/\sigma, 1)$, таким образом, в данном случае речь идет о различении двух простых гипотез о среднем нормального распределения с единичной дисперсией по одному наблюдению над случайной величиной T. Следовательно, можно воспользоваться вариантом критерия Неймана—Пирсона (10) при $n = 1$. В итоге имеем, что искомый критерий имеет вид $\mathfrak{X}_{1\alpha}^{*+} = \{T \geqslant t_\alpha\}$, $\Phi(-t_\alpha) = \alpha$, а вероятность ошибки 2-го рода (см. (13)) $\beta = \Phi(t_\alpha - \Delta_0/\sigma)$. •

Пример 2 (Проверка гипотез для дисперсии нормального распределения). Пусть $X = (X_1, \quad, X_n)$ — выборка из распределения $\mathcal{N}(\mu, \theta^2)$ (μ известно), о котором выдвинуты две простые гипотезы $H_0 \quad \theta = \theta_0$ и $H_1^- \quad \theta = \theta_1 < \theta_0$. Чтобы построить критерий Неймана—Пирсона для различения этих гипотез, найдем сначала статистику отношения правдоподобия

$$l(X) = \frac{\prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi}\theta_1} e^{-(X_i-\mu)^2/2\theta_1^2}}{\prod\limits_{i=1}^{n} \frac{1}{\sqrt{2\pi}\theta_0} e^{-(X_i-\mu)^2/2\theta_0^2}} = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left\{-\frac{\theta_0^2 - \theta_1^2}{2\theta_0^2\theta_1^2} T\right\},$$

$$T \equiv \sum_{i=1}^{n} (X_i - \mu)^2$$

Здесь $\mathcal{L}_\theta(T/\theta^2) = \chi^2(n)$ (см. пример 3 § 3.8) и неравенство $l(X) \geqslant c$ эквивалентно неравенству $T \leqslant t$, поэтому при уровне значимости α критическая граница $t = t_\alpha$ определяется из условия

$$\alpha = \mathbf{P}\{T \leqslant t_\alpha | H_0\} = F_n\left(\frac{t_\alpha}{\theta_0^2}\right),$$

где $F_n(t)$ — функция распределения закона $\chi^2(n)$. Отсюда $t_\alpha/\theta_0^2 = \chi_{\alpha,n}^2$ — α-квантиль распределения $\chi^2(n)$ и искомый критерий имеет вид

$$\mathfrak{X}_{1\alpha}^{*-} = \{T \leqslant \theta_0^2 \chi_{\alpha,n}^2\}.$$

Его мощность

$$W(\theta_1; \mathfrak{X}_{1\alpha}^{*-}) = \mathbf{P}\{T \leqslant \theta_0^2 \chi_{\alpha,n}^2 | H_1\} = F_n\left(\chi_{\alpha,n}^2 \frac{\theta_0^2}{\theta_1^2}\right).$$

Аналогично находим, что при $\theta_1 > \theta_0$ критерий имеет вид

$$\mathfrak{X}_{1\alpha}^{*+} = \{T \geqslant \theta_0^2 \chi_{1-\alpha,n}^2\}$$

и имеет мощность

$$W(\theta_1; \mathfrak{X}_{1\alpha}^{*+}) = 1 - F_n\left(\chi_{1-\alpha,n}^2 \frac{\theta_0^2}{\theta_1^2}\right). \quad \bullet$$

Пример 3 (Многомерное нормальное распределение, критерий Неймана—Пирсона для вектора средних). Пусть $\underline{\xi} = (\xi_1, \ldots, \xi_k)$, $k \geqslant 2$, — нормальный случайный вектор, имеющий при гипотезе H_i распределение $\mathcal{N}(\underline{\mu}^{(i)}, \Sigma)$, $i = 0, 1$ (общая дисперсионная матрица известна и предполагается невырожденной, см. п. 2 § 1.2). Построим критерий Неймана—Пирсона для различения гипотез H_0 и H_1 по одному наблюдению над $\underline{\xi}$.

В данном случае отношение правдоподобия в соответствии с (9) § 1.2 есть

$$l(\underline{x}) = \frac{f(\underline{x}|\underline{\mu}^{(1)}, \Sigma)}{f(\underline{x}|\underline{\mu}^{(0)}, \Sigma)} = \exp\left\{\frac{1}{2}\left[(\underline{x}-\underline{\mu}^{(0)})'\Sigma^{-1}(\underline{x}-\underline{\mu}^{(0)}) - (\underline{x}-\underline{\mu}^{(1)})'\Sigma^{-1}(\underline{x}-\underline{\mu}^{(1)})\right]\right\}.$$

Непосредственно проверяется, что выражение в квадратных скобках равно

$$(\underline{\mu}^{(0)} - \underline{\mu}^{(1)})'\Sigma^{-1}(\underline{\mu}^{(0)} + \underline{\mu}^{(1)}) - 2(\underline{\mu}^{(0)} - \underline{\mu}^{(1)})'\Sigma^{-1}\underline{x},$$

поэтому обозначив

$$\underline{a} = \Sigma^{-1}(\underline{\mu}^{(0)} - \underline{\mu}^{(1)}),$$

критическую область критерия $\mathfrak{X}_1 = \{l(\underline{x}) \geqslant c\}$ можно записать в виде

$$\mathfrak{X}_1(c) = \left\{\underline{x} \quad \underline{a}'\underline{x} - \frac{1}{2}\underline{a}'(\underline{\mu}^{(0)} + \underline{\mu}^{(1)}) \leqslant c_1 = -\ln c\right\} \tag{15}$$

Рассмотрим случайную величину

$$Y = \underline{a}'\underline{\xi} - \frac{1}{2}\underline{a}'(\underline{\mu}^{(0)} + \underline{\mu}^{(1)}).$$

Как линейная функция от $\underline{\xi}$ эта величина распределена по нормальному закону при обеих гипотезах, следовательно, чтобы найти ее распределения $\mathcal{L}(Y|H_i)$, $i = 0, 1$, достаточно определить только ее первые и вторые моменты. Имеем

$$\mathbf{E}(Y|H_i) = \underline{a}'\underline{\mu}^{(i)} - \frac{1}{2}\underline{a}'(\underline{\mu}^{(0)} + \underline{\mu}^{(1)}) = \begin{cases} \rho/2 & \text{при} \quad i = 0, \\ -\rho/2 & \text{при} \quad i = 1, \end{cases}$$

где

$$\rho = \underline{a}'(\underline{\mu}^{(0)} - \underline{\mu}^{(1)}) = (\underline{\mu}^{(0)} - \underline{\mu}^{(1)})'\Sigma^{-1}(\underline{\mu}^{(0)} - \underline{\mu}^{(1)}), \tag{16}$$

а

$$\mathbf{D}(Y|H_i) = \mathbf{D}(\underline{a}'\underline{\xi}|H_i) = \underline{a}'\Sigma\underline{a} = \rho, \quad i = 0, 1.$$

Таким образом, $\mathcal{L}(Y|H_i) = \mathcal{N}((-1)^i\rho/2, \rho)$, $i = 0, 1$, величина ρ, определенная в (16), называется *расстоянием Махаланобиса* [1] между распределениями $\mathcal{N}(\underline{\mu}^{(i)}, \Sigma)$, $i = 0, 1$

Расстояние Махаланобиса

[1] Махаланобис Прасанта Чандра (1893–1972) — индийский экономист, статистик. Иностр. член АН СССР (1958).

(или между гипотезами H_0 и H_1). Отсюда находим вероятности ошибок 1-го и 2-го рода критерия $\mathfrak{X}_1(c)$:

$$\alpha(c) = \mathbf{P}\{Y \leqslant c_1 | H_0\} = \Phi\left(\frac{c_1 - \rho/2}{\sqrt{\rho}}\right),$$
$$\beta(c) = \mathbf{P}\{Y > c_1 | H_1\} = \Phi\left(-\frac{c_1 + \rho/2}{\sqrt{\rho}}\right). \tag{17}$$

Если задан уровень значимости (вероятность ошибки 1-го рода) α, то критическая граница c_1 в (15) должна быть взята в виде

$$c_1 = \frac{\rho}{2} + \sqrt{\rho}\,\zeta_\alpha, \quad \Phi(\zeta_\alpha) = \alpha.$$

В итоге получаем, что в рассматриваемой задаче (H_0, H_1) критерий Неймана—Пирсона при уровне значимости α задается критической областью

$$\mathfrak{X}^*_{1\alpha} = \left\{\underline{x} \;\; \underline{a}'(\underline{x} - \underline{\mu}^{(0)}) \leqslant \sqrt{\rho}\,\zeta_\alpha\right\} \tag{18}$$

и его мощность равна $\Phi(\sqrt{\rho} + \zeta_\alpha)$. •

Замечание. Излагаемый подход к построению оптимального критерия в задаче различения двух простых гипотез, связанный с именами Неймана и Пирсона, является общепринятым (классическим) вариантом, но он не есть единственно возможный. Другой вариант, также основанный на сравнении правдоподобий данных при рассматриваемых гипотезах, представляет следующее утверждение.

Теорема 2. Пусть по наблюдению X требуется различить два распределения с плотностями $f_0(x)$ (гипотеза H_0) и $f_1(x)$ (гипотеза H_1). Рассмотрим класс критериев вида

$$\mathfrak{X}_1(c) = \{x \;\; f_1(x) \geqslant c f_0(x)\}, \quad c > 0, \tag{19}$$

и пусть $\alpha(c)$ и $\beta(c)$ — соответствующие вероятности ошибок 1-го и 2-го родов. Тогда

1) $\dfrac{\beta(c)}{1 - \alpha(c)} \leqslant c \leqslant \dfrac{1 - \beta(c)}{\alpha(c)}$,
2) $\alpha(c) + \beta(c) \leqslant 1$,
3) $\min\limits_c (\alpha(c) + \beta(c)) = \alpha(1) + \beta(1)$,
4) если X есть случайная выборка объема n и

$$\int f_0(x) \ln \frac{f_1(x)}{f_0(x)}\, dx = \delta < 0,$$

то вероятности α_n и β_n ошибок оптимального критерия $\mathfrak{X}_1(1)$ удовлетворяют условию $\alpha_n, \beta_n \to 0$ при $n \to \infty$.

Доказательство.

1) Утверждение следует из очевидных соотношений

$$\alpha(c) = \int\limits_{\mathfrak{X}_1(c)} f_0(x)\,dx \leqslant \frac{1}{c}\int\limits_{\mathfrak{X}_1(c)} f_1(x)\,dx = \frac{1}{c}(1-\beta(c)),$$

$$\beta(c) = \int\limits_{\mathfrak{X}_0(c)} f_1(x)\,dx \leqslant c\int\limits_{\mathfrak{X}_0(c)} f_0(x)\,dx = c(1-\alpha(c)), \quad \mathfrak{X}_0(c) = \overline{\mathfrak{X}}_1(c).$$

2) Пусть $c > 1$, тогда, используя первое из предыдущих соотношений, имеем

$$\alpha(c) + \beta(c) \leqslant \frac{1}{c}(1-\beta(c)) + \beta(c) < 1.$$

При $c \leqslant 1$ используем второе неравенство:

$$\alpha(c) + \beta(c) \leqslant \alpha(c) + c(1-\alpha(c)) \leqslant 1.$$

3) Пусть $c > 1$. Тогда исходим из соотношения

$$\alpha(c) + \beta(c) = \int\limits_{\mathfrak{X}_1(c)} f_0(x)\,dx + \int\limits_{\mathfrak{X}_0(c)} f_1(x)\,dx = 1 - \int\limits_{\mathfrak{X}_1(c)} (f_1(x) - f_0(x))\,dx.$$

Очевидно, $\mathfrak{X}_1(c) \subseteq \mathfrak{X}_1(1)$ и на множестве $\mathfrak{X}_1(1)$ разность $f_1(x) - f_0(x) \geqslant 0$. Следовательно,

$$\int\limits_{\mathfrak{X}_1(c)} (f_1(x) - f_0(x))\,dx \leqslant \int\limits_{\mathfrak{X}_1(1)} (f_1(x) - f_0(x))\,dx,$$

откуда $\alpha(c) + \beta(c) \geqslant \alpha(1) + \beta(1)$.

При $c < 1$ рассуждаем аналогично, исходя из соотношения

$$\alpha(c) + \beta(c) = 1 - \int\limits_{\mathfrak{X}_0(c)} (f_0(x) - f_1(x))\,dx.$$

4) Пусть

$$\boldsymbol{X} = (X_1, \ldots, X_n), \quad f_j(\underline{x}) = \prod_{i=1}^{n} f_j(x_i), \quad j = 0, 1.$$

Рассмотрим статистику

$$T_n(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} \ln \frac{f_1(X_i)}{f_0(X_i)}.$$

Если верна гипотеза H_0, то на основании закона больших чисел при $n \to \infty$

$$T_n(\boldsymbol{X}) \xrightarrow{\mathrm{P}} \mathrm{E}\left(\ln \frac{f_1(X_1)}{f_0(X_1)}\middle| H_0\right) = \int f_0(x) \ln \frac{f_1(x)}{f_0(x)}\,dx = \delta,$$

при этом, согласно неравенству Йенсена (о нем см. в п. 5 § 4.2), всегда $\delta \leqslant 0$. Далее, из (19) следует, что в данном случае множество $\mathfrak{X}_1(1)$ можно записать в виде

$$\mathfrak{X}_1(1) = \{\underline{x} = (x_1, \ldots, x_n) : T_n(\underline{x}) \geqslant 0\}, \tag{20}$$

поэтому при $\delta < 0$.

$$\alpha_n = \mathbf{P}\{T_n(\boldsymbol{X}) \geqslant 0 | H_0\} \leqslant \mathbf{P}\{|T_n(\boldsymbol{X}) - \delta| \geqslant |\delta| | H_0\} \longrightarrow 0,$$

если $n \to \infty$. В силу симметрии (если поменять ролями H_0 и H_1) также и $\beta_n \to 0$. ■

Таким образом, если в концепции Неймана—Пирсона, вместо минимизации вероятности ошибки 2-го рода при фиксированной вероятности ошибки 1-го рода, исходить из минимизации суммы вероятностей ошибок, то мы придем к оптимальному критерию $\mathfrak{X}_1(1)$ вида (19) (или (20)). При этом при бесконечно большой выборке критерий $\mathfrak{X}_1(1)$ различает гипотезы H_0 и H_1 (в случае $\delta < 0$) безошибочно.

Пример 3 (Продолжение). В этой задаче критерием, минимизирующим сумму вероятностей ошибок, является (см. (15))

$$\mathfrak{X}_1(1) = \left\{\underline{x} \quad \underline{a}'\underline{x} \leqslant \frac{1}{2}\,\underline{a}'(\underline{\mu}^{(0)} + \underline{\mu}^{(1)})\right\}, \tag{21}$$

и эта сумма равна в силу (17)

$$\alpha(1) + \beta(1) = 2\Phi\left(-\frac{\sqrt{\rho}}{2}\right). \qquad \bullet$$

3. Критерий Неймана—Пирсона в случае дискретных распределений

Аналогичные рассуждения можно провести и для дискретных распределений, для которых вероятности $f_j(z_k) > 0$, $j = 0, 1$, для всех z_k — возможных значений наблюдаемой случайной величины ξ. Общий принцип здесь остается прежним: «упорядочивают» выборочные точки $\underline{x} = (x_1, \ldots, x_n)$ в соответствии с величиной отношения правдоподобия

$$l(\underline{x}) = \frac{L(\underline{x}; \theta_1)}{L(\underline{x}; \theta_0)} = \frac{\prod\limits_{i=1}^{n} f_1(x_i)}{\prod\limits_{i=1}^{n} f_0(x_i)}$$

и включают в критическое множество $\mathfrak{X}_1$ максимальное число их, согласующееся с требованием

$$\sum_{\underline{x} \in \mathfrak{X}_1} L(\underline{x}; \theta_0) \leqslant \alpha.$$

Однако в отличие от непрерывного случая здесь в силу дискретности, вообще говоря, уже нельзя получить точное значение α за счет выбора границы c в неравенстве $l(\underline{x}) \leqslant c$, определяющем множество $\mathfrak{X}_1$: может быть так, что, включив в $\mathfrak{X}_1$ очередную точку мы еще не достигнем уровня α, а включив следующую — превзойдем его. Детальнее это можно объяснить следующим образом. Обозначим $< l_k < l_{k+1} <$ возможные значения статистики $l(X)$. Тогда при заданном уровне α можно определить такое $k = k(\alpha)$, что

$$\sum_{\underline{x}:\, l(\underline{x}) \geqslant l_{k+1}} L(\underline{x}; \theta_0) < \alpha \leqslant \sum_{\underline{x}:\, l(\underline{x}) \geqslant l_k} L(\underline{x}; \theta_0). \tag{22}$$

Если здесь в правой части имеет место знак равенства, то полагая

$$\mathfrak{X}^*_{1\alpha} = \{\underline{x} : l(\underline{x}) \geqslant l_{k(\alpha)}\}$$

и повторяя рассуждения, проведенные в п. 2, можно показать, что критическая область $\mathfrak{X}^*_{1\alpha}$ является оптимальной, т. е. основанный на ней критерий

$$\boxed{H_0 \quad \text{отвергается} \iff l(X) \in \mathfrak{X}^*_{1\alpha}} \tag{23}$$

имеет наибольшую мощность при альтернативе H_1 среди всех критериев уровня значимости α.

Пусть теперь в (22) имеют место строгие неравенства (так чаще всего и бывает). Обозначим

$$\begin{aligned} \alpha_j &= \sum_{\underline{x}:\, l(\underline{x}) \geqslant l_{k+1}} L(\underline{x}; \theta_j), \quad j = 0, 1, \\ p_j &= \mathbf{P}_j\{l(X) = l_k\} = \sum_{\underline{x}:\, l(\underline{x}) = l_k} L(\underline{x}; \theta_j), \quad j = 0, 1. \end{aligned} \tag{24}$$

Тогда соотношение (22) можно записать в виде $\alpha_0 < \alpha < \alpha_0 + p_0$. Чтобы в этом случае построить критерий с уровнем значимости α, надо использовать прием *рандомизации*, т. е. поступить следующим образом. Когда выполняется равенство $l(\underline{x}) = l_{k(\alpha)}$, произвести дополнительный случайный эксперимент с двумя исходами $\overline{R}$ и R, вероятности которых равны соответственно $(\alpha - \alpha_0)/p_0$ и $1 - (\alpha - \alpha_0)/p_0$ (для этого можно воспользоваться алгоритмом п. 2 § 1.3), и отвергнуть гипотезу H_0, если наблюдается $\overline{R}$, и принять H_0 в противном случае. Если же $l(\underline{x}) > l_{k(\alpha)}$, то гипотезу H_0 следует отвергнуть, а при $l(\underline{x}) < l_{k(\alpha)}$ — принять. Другими словами, мы строим рандомизированный критерий с критической функцией

$$\varphi^*(\underline{x}) = \begin{cases} 1 & \text{при} \quad l(\underline{x}) > l_{k(\alpha)}, \\ \dfrac{\alpha - \alpha_0}{p_0} & \text{при} \quad l(\underline{x}) = l_{k(\alpha)}, \\ 0 & \text{при} \quad l(\underline{x}) < l_{k(\alpha)}. \end{cases} \tag{25}$$

Вероятность ошибки 1-го рода такого критерия в соответствии с (24) равна

$$\mathbf{P}\{H_1|H_0\} = \mathbf{E}_0\varphi^*(X) = \mathbf{P}_0\{l(X) > l_{k(\alpha)}\} + \frac{\alpha - \alpha_0}{p_0}\,\mathbf{P}_0\{l(X) = l_{k(\alpha)}\} =$$

$$= \alpha_0 + \frac{\alpha - \alpha_0}{p_0}\,p_0 = \alpha,$$

т. е. критерий φ^* имеет уровень значимости в точности равный α. Его мощность вычисляется аналогично:

$$W(\theta_1; \varphi^*) = \mathbf{E}_1\varphi^*(X) = \alpha_1 + \frac{(\alpha - \alpha_0)p_1}{p_0}.$$

Покажем, что это наиболее мощный критерий в задаче (H_0, H_1) Рассмотрим произвольный критерий φ с уровнем значимости α $\mathbf{E}_0\varphi(X) = \alpha$, и обозначим $\mathfrak{X}^{\pm} = \{\underline{x} : \varphi^*(\underline{x}) - \varphi(\underline{x}) \gtrless 0\}$ (знак $+(-)$ соответствует неравенству $>$ $(<)$). Если $\underline{x} \in \mathfrak{X}^+$, то $\varphi^*(\underline{x}) > 0$ и поэтому $L(\underline{x}; \theta_1) \geqslant l_k L(\underline{x}; \theta_0)$; если $\underline{x} \in \mathfrak{X}^-$, то $\varphi(\underline{x}) < 1$ и, следовательно, $L(\underline{x}; \theta_1) \leqslant l_k L(\underline{x}; \theta_0)$. Таким образом,

$$\sum_{\underline{x}} \big(\varphi^*(\underline{x}) - \varphi(\underline{x})\big)\big(L(\underline{x}; \theta_1) - l_k L(\underline{x}; \theta_0)\big) = \sum_{\underline{x}\in\mathfrak{X}^+} + \sum_{\underline{x}\in\mathfrak{X}^-} \geqslant 0.$$

Отсюда для разности мощностей получаем

$$\mathbf{E}_1\varphi^*(X) - \mathbf{E}_1\varphi(X) = \sum_{\underline{x}} \big(\varphi^*(\underline{x}) - \varphi(\underline{x})\big)L(\underline{x}; \theta_1) \geqslant$$

$$\geqslant l_k \sum_{\underline{x}} \big(\varphi^*(\underline{x}) - \varphi(\underline{x})\big)L(\underline{x}; \theta_0) =$$

$$= l_k\big(\mathbf{E}_0\varphi^*(X) - \mathbf{E}_0\varphi^*(X)\big) = l_k(\alpha - \alpha) = 0,$$

т. е. $W(\theta_1; \varphi^*) \geqslant W(\theta_1; \varphi)$.

Итак, в дискретном случае также всегда можно построить наиболее мощный критерий с заданным уровнем значимости α, но такой критерий, вообще говоря, является уже рандомизированным. Отметим, что рандомизация бывает необходима лишь на пограничном множестве $\{\underline{x} \;\; l(\underline{x}) = l_{k(\alpha)}\}$ и только в том случае, когда желательно иметь вероятность ошибки 1-го рода равной точно α. На практике предпочитают в таких случаях (т. е. когда в (22) имеют место строгие неравенства) несколько изменить уровень значимости так, чтобы отпала необходимость в рандомизации, т. е. заменить α на меньший уровень α_0 (см. (24)) и использовать нерандомизированный критерий $\mathfrak{X}^*_{1\alpha_0} = \{l(\underline{x}) \geqslant l_{k(\alpha)+1}\}$ вида (23), который, по предыдущему, является наиболее мощным критерием уровня значимости α_0 $(<\alpha)$. Можно также использовать еще один нерандомизированный критерий $\mathfrak{X}^*_{1\alpha'} = \{l(\underline{x}) \geqslant l_{k(\alpha)}\}$ с бо́льшим уровнем значимости $\alpha' = \alpha_0 + p_0$. Далее мы проиллюстрируем эти общие положения несколькими примерами.

Пример 4 (Бернуллиевская модель, проверка простых гипотез). Пусть о неизвестной вероятности «успеха» θ в бернуллиевской модели $\mathrm{Bi}(1, \theta)$ имеются две простые гипотезы: $H_0 \quad \theta = \theta_0$ и $H_1^+ \quad \theta = \theta_1 > \theta_0$. Здесь для любой точки $\underline{x} = (x_1, \ldots, x_n)$, $x_i \in \{0,1\}$, $i = 1, \ldots, n$, правдоподобие

$$L(\underline{x}; \theta_j) = \theta_j^r (1 - \theta_j)^{n-r}, \quad j = 0, 1,$$

где $r = r(\underline{x}) = \sum_{i=1}^{n} x_i$ — наблюдавшееся число «успехов», а их отношение

$$l(\underline{x}) = \left[\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right]^r \left(\frac{1-\theta_1}{1-\theta_0}\right)^n$$

Функция $\varphi(\theta) = \theta/(1-\theta)$ возрастает на интервале $(0, 1)$, поэтому при $\theta_1 > \theta_0$ $\varphi(\theta_1)/\varphi(\theta_0) > 1$ и неравенство $l(\underline{x}) \geqslant c$ эквивалентно неравенству

$$r(\underline{x}) \geqslant \frac{\ln c - n\rho_1}{\rho},$$

где

$$\rho_1 = \ln\left(\frac{1-\theta_1}{1-\theta_0}\right), \quad \rho = \ln\left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right).$$

Следовательно, критическая область критерия Неймана—Пирсона выражается в данном случае через статистику $r(X)$ и имеет вид

$$\mathfrak{X}_{1\alpha}^{*+} = \{\underline{x} \quad r(\underline{x}) \geqslant r_\alpha\}. \tag{26}$$

Далее, поскольку $\mathcal{L}(r(X)|H_0) = \mathrm{Bi}(n, \theta_0)$, то для определения при заданном уровне α критической границы r_α имеем в соответствии с (22) условие

$$\alpha_0 \equiv \sum_{m=r_\alpha+1}^{n} C_n^m \theta_0^m (1-\theta_0)^{n-m} < \alpha \leqslant \sum_{m=r_\alpha}^{n} C_n^m \theta_0^m (1-\theta_0)^{n-m} \equiv \alpha' \tag{27}$$

Если здесь в правой части имеет место знак равенства ($\alpha = \alpha'$), то тем самым критерий (26) полностью определен; подчеркнем, что он не зависит от альтернативы θ_1, от нее зависит лишь мощность критерия

$$W(\theta_1; \mathfrak{X}_{1\alpha}^{*+}) = \mathbf{P}\{r(X) \geqslant r_\alpha | H_1\} = \sum_{m=r_\alpha}^{n} C_n^m \theta_1^m (1-\theta_1)^{n-m}$$

Чаще всего, однако, при заранее заданном α в (27) имеет место случай строгого неравенства ($\alpha < \alpha'$); следовательно, критерий, определяемый критической областью (26) будет иметь уровень значимости несколько бóльший, чем α, а именно, равный α'. В данном случае можно использовать также наиболее мощный критерий $\{r(X) \geqslant r_\alpha + 1\}$ с уровнем значимости $\alpha_0 < \alpha$.

Если бы в данной ситуации требовалось построить критерий с уровнем значимости, точно равным α, то следовало бы прибегнуть к рандомизации и поступить следующим образом. Положив

$$p_0 = \mathbf{P}\{r(X) = r_\alpha | H_0\} = C_n^{r_\alpha} \theta_0^{r_\alpha} (1-\theta_0)^{n-r_\alpha} = \alpha' - \alpha_0,$$

задать критическую функцию (см. (25))

$$\varphi_\alpha^*(\underline{x}) = \begin{cases} 1 & \text{при} \quad r(\underline{x}) > r_\alpha, \\ \dfrac{\alpha - \alpha_0}{p_0} & \text{при} \quad r(\underline{x}) = r_\alpha, \\ 0 & \text{при} \quad r(\underline{x}) < r_\alpha, \end{cases} \tag{28}$$

т. е. условиться отвергать гипотезу H_0, если $r(\underline{x}) > r_\alpha$, и принимать ее, если $r(\underline{x}) < r_\alpha$; если же $r(\underline{x}) = r_\alpha$, то H_0 отвергается с вероятностью $(\alpha - \alpha_0)/p_0$ и принимается с дополнительной вероятностью $(\alpha' - \alpha)/p_0$. Тогда вероятность ошибки 1-го рода такого рандомизированного критерия равна

$$\mathbf{P}\{H_1|H_0\} = \mathbf{E}_{\theta_0}\varphi_\alpha^*(X) =$$

$$= \mathbf{P}_{\theta_0}\{r(X) > r_\alpha\} + \frac{\alpha - \alpha_0}{p_0}\,\mathbf{P}_{\theta_0}\{r(X) = r_\alpha\} = \alpha_0 + \frac{\alpha - \alpha_0}{p_0}\,p_0 = \alpha.$$

Мощность этого критерия вычисляется по формуле

$$W(\theta_1; \varphi_\alpha^*) = \mathbf{E}_{\theta_1}\varphi_\alpha^*(X) =$$

$$\sum_{m=r_\alpha+1}^{n} C_n^m \theta_1^m (1-\theta_1)^{n-m} + (\alpha - \alpha_0)\left(\frac{\theta_1}{\theta_0}\right)^{r_\alpha}\left(\frac{1-\theta_1}{1-\theta_0}\right)^{n-r_\alpha}$$

Случай $\theta_1 < \theta_0$ рассматривается аналогично, при этом критическая область имеет вид $\{r(\underline{x}) \leqslant r_\alpha\}$.

Большие выборки

Для случая больших выборок ($n \to \infty$) по теореме Муавра—Лапласа

$$\mathcal{L}_\theta(r(X)) \sim \mathcal{N}(n\theta, n\theta(1-\theta)),$$

поэтому условие (27) можно заменить приближенным условием

$$\mathbf{P}_{\theta_0}\{r(X) \geqslant r_\alpha\} \approx \Phi\left(\frac{n\theta_0 - r_\alpha}{\sqrt{n\theta_0(1-\theta_0)}}\right) = \alpha,$$

т. е.

$$\frac{n\theta_0 - r_\alpha}{\sqrt{n\theta_0(1-\theta_0)}} = \zeta_\alpha.$$

Отсюда получаем асимптотический вариант критерия Неймана—Пирсона в нашей задаче

$$\mathfrak{X}_{1\alpha}^{*+} = \left\{r(X) \geqslant n\theta_0 - \zeta_\alpha\sqrt{n\theta_0(1-\theta_0)}\right\}, \quad \Phi(\zeta_\alpha) = \alpha. \tag{29}$$

Вычислим, наконец, мощность критерия (29) при «близкой» альтернативе

$$H_{1n}^+ \quad \theta = \theta_{1n} = \theta_0 + \frac{\delta}{\sqrt{n}}, \quad \delta > 0.$$

Имеем

$$W(\theta_{1n}) = \mathbf{P}_{\theta_{1n}}\left\{r(X) \geqslant n\theta_0 - \zeta_\alpha\sqrt{n\theta_0(1-\theta_0)}\right\} =$$

$$= \mathbf{P}_{\theta_{1n}}\left\{\frac{r(X) - n\theta_{1n}}{\sqrt{n\theta_{1n}(1-\theta_{1n})}} \geqslant -\frac{\delta}{\sqrt{\theta_{1n}(1-\theta_{1n})}} - \zeta_\alpha\sqrt{\frac{\theta_0(1-\theta_0)}{\theta_{1n}(1-\theta_{1n})}}\right\} \to$$

$$\to \Phi\left(\frac{\delta}{\sqrt{\theta_0(1-\theta_0)}} + \zeta_\alpha\right) \quad \bullet$$

Пример 5 (Пуассоновская модель, проверка простых гипотез). Рассмотрим задачу различения двух простых гипотез H_i: $\theta = \theta_i$, $i = 0, 1$, $0 < \theta_0 < \theta_1$, для пуассоновской модели $\Pi(\theta)$. Схема рассуждений здесь такая же, как в предыдущем примере. Статистика отношения правдоподобия есть

$$l(X) = \left(\frac{\theta_1}{\theta_0}\right)^{T_n} e^{n(\theta_0 - \theta_1)}, \quad T_n = \sum_{i=1}^{n} X_i,$$

т. е. выражается через достаточную статистику модели $\Pi(\theta)$ (см. п. 1 § 3.3), при этом $\mathcal{L}_\theta(T_n) = \Pi(n\theta)$ (см. п. 3 § 1.1). В терминах статистики T_n критерий задается критической областью вида $\{T_n \geqslant t_\alpha\}$, где критическая граница t_α при уровне значимости α определяется условием

$$\alpha_0 \equiv \sum_{m=t_\alpha+1}^{\infty} e^{-n\theta_0}\frac{(n\theta_0)^m}{m!} < \alpha \leqslant \sum_{m=t_\alpha}^{\infty} e^{-n\theta_0}\frac{(n\theta_0)^m}{m!} \equiv \alpha' \qquad (30)$$

При $\alpha = \alpha'$ оптимальный критерий имеет вид $\mathfrak{X}^*_{1\alpha} = \{T_n \geqslant t_\alpha\}$; если же $\alpha < \alpha'$, то можно использовать нерандомизированные наиболее мощные критерии $\mathfrak{X}^*_{1\alpha'} = \{T_n \geqslant t_\alpha\}$ и $\mathfrak{X}^*_{1\alpha_0} = \{T_n \geqslant t_\alpha + 1\}$ с уровнями значимости соответственно $\alpha' > \alpha$ и $\alpha_0 < \alpha$. В случае $\alpha < \alpha'$ можно построить также рандомизированный критерий с уровнем значимости в точности равным α: его критическая функция $\varphi^*_\alpha(T)$ имеет такой же вид, как (28) в предыдущем примере (с учетом обозначений, принятых в (30), в частности, здесь $p_0 = \alpha' - \alpha_0 = e^{-n\theta_0}\dfrac{(n\theta_0)^{t_\alpha}}{t_\alpha!}$).

В любом случае мощность вычисляется по формуле

$$W(\theta_1) = \sum_{m=t_\alpha+1}^{\infty} e^{-n\theta_1}\frac{(n\theta_1)^m}{m!} + (\alpha - \alpha_0)\left(\frac{\theta_1}{\theta_0}\right)^{t_\alpha} e^{-n(\theta_1-\theta_0)}$$



Большие выборки

Если $n \to \infty$, то по центральной предельной теореме $\mathcal{L}_\theta(T_n) \sim \mathcal{N}(n\theta, n\theta)$, и рассуждая как в предыдущем примере, получим, что асимптотическая форма критерия Неймана—Пуассона имеет вид $\{T_n \geqslant n\theta_0 - \zeta_\alpha\sqrt{n\theta_0}\}$, а его мощность при близкой альтернативе $\theta_1 = \theta_{1n} = \theta_0 + \delta/\sqrt{n}$, $\delta > 0$, удовлетворяет предельному соотношению

$$\lim_{n\to\infty} W(\theta_{1n}) = \Phi\left(\frac{\delta}{\sqrt{\theta_0}} + \zeta_\alpha\right).$$

Подчеркнем, что в любом случае вид оптимальной критической области здесь также не зависит от альтернативы θ_1. •

Пример 6 (Полиномиальная модель, проверка простых гипотез). Пусть по наблюдению вектора частот исходов $\underline{\nu} = (\nu_1, \ldots, \nu_N)$ в n испытаниях требуется различить две простые гипотезы о векторе вероятностей исходов в полиномиальной модели $M(n; \theta)$ (см. п. 6 § 1.1): $H_0 \quad \theta = \underline{p} = (p_1, \ldots, p_N)$ и $H_1 \quad \theta = \underline{q} = (q_1, \ldots, q_N)$ (все $0 < p_i, q_i < 1$). Здесь статистика отношения правдоподобия

$$l(\underline{\nu}) = \prod_{i=1}^{N} \left(\frac{q_i}{p_i}\right)^{\nu_i}$$

поэтому, переходя к логарифмам, запишем критическую область наиболее мощного критерия в виде

$$\mathfrak{X}_1^* = \left\{ T_n \equiv \sum_{i=1}^{N} \nu_i \ln\left(\frac{q_i}{p_i}\right) \geqslant l \right\}. \tag{31}$$

Для расчета критической границы и мощности такого критерия можно воспользоваться центральной предельной теоремой для полиномиального распределения (см. упр. 33 к гл. 1), справедливой при $n \to \infty$. Тестовая статистика T_n в (31) как линейная функция асимптотически нормального вектора $\underline{\nu}$ будет иметь асимптотически нормальное распределение при обеих гипотезах (см. п. 2 § 1.2), поэтому нам нужно вычислить лишь средние и дисперсии статистики T_n при гипотезах H_0 и H_1. Из формул (16) § 1.1 имеем

$$\mathbf{E}(T_n|H_0) = n \sum_{i=1}^{N} p_i \ln\left(\frac{q_i}{p_i}\right) \equiv na_0,$$

$$\mathbf{E}(T_n|H_1) = n \sum_{i=1}^{N} q_i \ln\left(\frac{q_i}{p_i}\right) \equiv na_1,$$

$$\mathbf{D}(T_n|H_0) = n \sum_{i=1}^{N} p_i(1-p_i) \ln^2\left(\frac{q_i}{p_i}\right) - n \sum_{i \neq j} p_i p_j \ln\left(\frac{q_i}{p_i}\right) \ln\left(\frac{q_j}{p_j}\right) =$$

$$= n\left[\sum_{i=1}^{N} p_i \ln^2\left(\frac{q_i}{p_i}\right) - \left(\sum_{i=1}^{N} p_i \ln\left(\frac{q_i}{p_i}\right)\right)^2\right] \equiv nb_0^2, \tag{32}$$

и аналогично

$$\mathbf{D}(T_n|H_1) = n\left[\sum_{i=1}^{N} q_i \ln^2\left(\frac{q_i}{p_i}\right) - \left(\sum_{i=1}^{N} q_i \ln\left(\frac{q_i}{p_i}\right)\right)^2\right] \equiv nb_1^2.$$

Большие выборки

Таким образом при больших n можно воспользоваться нормальной аппроксимацией

$$\mathcal{L}(T_n|H_i) \sim \mathcal{N}(na_i, nb_i^2), \quad i = 0, 1,$$

с указанными в (32) параметрами. Отсюда уже стандартными рассуждениями получаем, что асимптотический (при $n \to \infty$) вариант критерия (31) при уровне значимости α есть

$$\mathfrak{X}^*_{1\alpha} = \left\{ \sum_{i=1}^{N} \nu_i \ln \frac{q_i}{p_i} \geqslant n a_0 + \sqrt{n}\, b_0\, t_\alpha \right\}, \quad \Phi(-t_\alpha) = \alpha. \tag{33}$$

Для вероятности ошибки 2-го рода получаем соотношение

$$\beta_n = \mathbf{P}\{H_0 | H_1\} = \mathbf{P}\left\{ \frac{T_n - n a_1}{\sqrt{n} b_1} \leqslant \sqrt{n}\, \frac{a_0 - a_1}{b_1} + \frac{b_0}{b_1} t_\alpha \,\middle|\, H_1 \right\}. \tag{34}$$

По неравенству Йенсена $a_0 < 0$, $a_1 > 0$, поэтому, если гипотезы фиксированы, а $n \to \infty$, то отсюда лишь следует, что $\beta_n \to 0$. Более определенные заключения можно сделать, рассматривая «близкие» альтернативы, которые в данном случае удобно задать в виде

«Близкие» альтернативы

$$q_i = q_i(n) = p_i\left(1 + \frac{\delta_i}{\sqrt{n}}\right), \quad i = 1, \ldots, N, \quad \max_i |\delta_i| \leqslant c < \infty, \quad \sum_{i=1}^{N} p_i \delta_i = 0.$$

В этом случае из формул (32) имеем (поскольку $\ln(1+\varepsilon) = \varepsilon - \varepsilon^2/2 + O(\varepsilon^3)$)

$$a_1 - a_0 = \sum_{i=1}^{N} (q_i - p_i) \ln \frac{q_i}{p_i} =$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} p_i \delta_i \left(\frac{\delta_i}{\sqrt{n}} + O\left(\frac{1}{n}\right) \right) = \frac{1}{n} \sum_{i=1}^{N} p_i \delta_i^2 + O\left(\frac{1}{n^{3/2}}\right),$$

$$b_0^2 \sim b_1^2 \sim \frac{1}{2n} \sum_{i=1}^{N} p_i \delta_i^2.$$

Поэтому, обозначив

$$\delta^2 = \sum_{i=1}^{N} p_i \delta_i^2,$$

из (34) получим, что при $n \to \infty$

$$\beta_n \longrightarrow \Phi(-\sqrt{2}\,\delta + t_\alpha),$$

сам же критерий (33) может быть записан в виде

$$\mathfrak{X}^*_{1\alpha} = \left\{ T_n \geqslant -\frac{\delta^2}{2} + \frac{\delta}{\sqrt{2}} t_\alpha \right\}. \quad \bullet$$

§ 5.3. Сложные гипотезы

Случаи, когда основная и альтернативная гипотезы являются простыми, встречаются в приложениях сравнительно редко. На практике чаще имеют место ситуации, когда обе гипотезы (или по крайней мере одна из них) сложные. В этих случаях равномерно наиболее мощные (р. н. м.) критерии существуют только если модель $\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$ обладает определенными свойствами. Рассмотрим некоторые наиболее типичные случаи.

1. Р. н. м. критерии против сложных односторонних альтернатив. Модели с монотонным отношением правдоподобия

Пусть проверяется простая гипотеза $H_0 \quad \theta = \theta_0$ против сложной альтернативы $H_1 : \theta \in \Theta_1 = \Theta \backslash \{\theta_0\}$. Построим критерий Неймана—Пирсона $\mathfrak{X}^*_{1\alpha} = \mathfrak{X}^*_{1\alpha}(\theta_0, \theta_1)$ для фиксированной альтернативы $\theta_1 \in \Theta_1$, как это описано в § 5.2. Если этот критерий (т. е. вид соответствующей критической области) зависит от θ_1, то это значит, что не существует критической области, которая была бы наилучшей для всех $\theta_1 \in \Theta_1$. Другими словами, в этом случае не существует р. н. м. к. Однако может оказаться, что критерий $\mathfrak{X}^*_{1\alpha}(\theta_0, \theta_1)$ не зависит от конкретной альтернативы $\theta_1 \quad \mathfrak{X}^*_{1\alpha}(\theta_0, \theta_1) = \mathfrak{X}^*_{1\alpha}(\theta_0)$. Тогда соответствующая критическая область максимизирует мощность при любой допустимой альтернативе и поэтому $\mathfrak{X}^*_{1\alpha}(\theta_0)$ является р. н. м. к. в задаче $(H_0 \quad \theta = \theta_0,\ H_1 \quad \theta \in \Theta_1)$.

Такие случаи уже встречались нам в § 5.2. Так, в примере 1 (соотношение (10)) построен критерий Неймана—Пирсона в задаче $(H_0 \quad \theta = \theta_1,\ H_1 \quad \theta = \theta_1 > \theta_0)$ для нормальной модели $\mathcal{N}(\theta, \sigma^2)$, который оказался не зависящим от альтернативы θ_1, следовательно, этот критерий $\mathfrak{X}^{*+}_{1\alpha}$ является р. н. м. критерием при сложной *правосторонней* альтернативе $H_1^+ \quad \theta > \theta_0$. Аналогично, р. н. м. критерием при сложной *левосторонней* альтернативе $H_1^- \quad \theta < \theta_0$ в этой модели является критерий (12). Такие же заключения можно сделать о критериях, построенных в примерах 2, 4, 5. Таким образом, здесь мы имеем примеры существования р. н. м. критериев проверки простой гипотезы H_0 при сложных *односторонних* альтернативах в случае скалярного параметра. Если же в этих задачах рассматривать полный класс альтернатив $H_1 = H_1^+ \cup H_1^- \quad \theta \neq \theta_0$, то р. н. м. к. не существует, так как при альтернативах $\theta > \theta_0$ и $\theta < \theta_0$ критерии различаются.

Рассмотренные примеры характерны следующими двумя обстоятельствами:

1) в каждом из них статистика отношения правдоподобия $l(X)$ зависит от данных X через достаточную статистику соответствующей модели (например, в примере 1 — это выборочное среднее $\overline{X}$, в примере 2 — выборочная дисперсия (с точностью до коэффициента) и т. д.);
2) $l(X)$ является *монотонной* функцией от достаточной статистики, в силу чего, неравенство $l(X) \geqslant c$, определяющее оптимальную критическую область, эквивалентно аналогичному (одностороннему) неравенству для соответствующей достаточной статистики. Эти факторы имеют общую природу, и они приводят к следующему общему результату.

Предположим, что θ — скаляр и альтернатива H_1 — односторонняя (т. е. либо $H_1 = H_1^- \quad \theta < \theta_0$, либо $H_1 = H_1^+ \quad \theta > \theta_0$), основная же гипотеза $H_0 \quad \theta = \theta_0$ — простая. Пусть, далее, семейство $\mathcal{F} = \{F(x;\theta), \theta \in \Theta\}$ обладает достаточной статистикой $T = T(X)$ (см. теорему 1 § 3.3) и статистика отношения правдоподобия

$$l(X) = \frac{g(T(X);\theta_1)}{g(T(X);\theta_0)}$$

является *монотонной* функцией от T (тогда говорят, что модель $\mathcal{F}$ обладает *монотонным отношением правдоподобия*, — многие стандартные модели таковы, например, $\mathcal{F} = \mathcal{N}(\theta, \sigma^2)$, $\mathcal{N}(\mu, \theta^2)$, $\mathrm{Bi}(k, \theta)$, $\Pi(\theta)$, $\overline{\mathrm{Bi}}(r, \theta)$, $\Gamma(\theta, \lambda)$).

Монотонное отношение правдоподобия

Тогда в *задаче* (H_0, H_1) *существует р. н. м. критерий, совпадающий с критерием Неймана—Пирсона при произвольной фиксированной альтернативе из* H_1.

Доказательство. Пусть, например, $H_1 = H_1^+$ и при фиксированном $\theta_1 > \theta_0$ отношение $g(T;\theta_1)/g(T;\theta_0) \uparrow$ по T. Построим критерий Неймана—Пирсона в задаче (θ_0, θ_1). В данном случае неравенство $l(X) \geqslant c \iff T(X) \geqslant c_\alpha^+$, причем граница c_α^+ определяется лишь распределением

$$F(x;\theta_0) \quad \mathbf{P}_{\theta_0}\{T(X) \geqslant c_\alpha^+\} = \alpha.$$

Таким образом, критерий $\mathfrak{X}_{1\alpha}^{*+} = \{T(X) \geqslant c_\alpha^+\}$ не зависит от конкретной альтернативы $\theta_1 > \theta_0$, следовательно, он есть р. н. м. к. для задачи (H_0, H_1^+). Если $g(T;\theta_1)/g(T;\theta_0) \downarrow$ по T, то неравенство для T заменяется на противоположное. Аналогично рассматривается и случай левосторонней альтернативы H_1^- ■

Замечание. Можно доказать [16, с. 101], что для моделей с монотонным отношением правдоподобия р. н. м. критерий в задаче $(H_0 : \theta = \theta_0, H_1^+ \quad \theta > \theta_0)$ является одновременно р. н. м. критерием проверки сложной гипотезы $H_0 \quad \theta \leqslant \theta_0$ против H_1^+ того же уровня значимости. Аналогичное утверждение справедливо и для двойственной проблемы проверки сложной гипотезы $H_0 : \theta \geqslant \theta_0$ против левосторонней альтернативы $H_1^- \quad \theta < \theta_0$.

В частности, для экспоненциальной модели, введенной в п. 3 § 3.2, плотность которой имеет вид

$$f(x;\theta) = \exp\{A(\theta)B(x) + C(\theta) + D(x)\}, \tag{1}$$

статистика

$$T(X) = \sum_{i=1}^{n} B(X_i)$$

является достаточной и

$$l(X) = \exp\{(A(\theta_1) - A(\theta_0))T(X) + n(C(\theta_1) - C(\theta_0))\}.$$

Следовательно, если функция $A(\theta)$ строго монотонна, то $l(X)$ — монотонная функция T, поэтому для такой модели всегда существует р. н. м. к. против

односторонних альтернатив. Вид критических областей указан в следующей таблице

	$H_1^+ \quad \theta > \theta_0$	$H_1^- \quad \theta < \theta_0$
$A(\theta) \uparrow$	$T(X) \geqslant c_\alpha^+$	$T(X) \leqslant c_\alpha^-$
$A(\theta) \downarrow$	$T(X) \leqslant c_\alpha^+$	$T(X) \geqslant c_\alpha^-$

Пример 1 (Выборочный контроль: биномиальный выбор). В процессе производства изделия обычно подвергают выборочному статистическому контролю. Предположим, что каждое изделие независимо от других может оказаться дефектным с некоторой одной и той же, но неизвестной вероятностью θ $(0 < \theta < 1)$ и исправным с дополнительной вероятностью $1-\theta$ Пусть для контроля взято n изделий и результат описывается вектором $X = (X_1, \quad , X_n)$, где $X_i = 1$, если i-е изделие дефектно, и $X_i = 0$ в противном случае. Часто бывает нужно проверить гипотезу $H_0 \quad \theta \geqslant \theta_0$, где θ_0 — некоторая критическая доля брака (если гипотеза H_0 истинна, то процесс производства необходимо приостановить и усовершенствовать для уменьшения доли брака). При сделанных предположениях число $r(X) = X_1 + \quad + X_n$ дефектных изделий в выборке имеет биномиальное распределение $\mathrm{Bi}(n, \theta)$, к которому применима изложенная выше теория. Следовательно, существует р. н. м. критерий проверки гипотезы H_0 против альтернативы $H_1^- \quad \theta < \theta_0$, который задается критической областью вида $\{r(X) \leqslant r_\alpha\}$. Критическая граница r_α вычисляется здесь так же, как в примере 4 § 5.2. •

Пример 2 (Обратный биномиальный выбор). В условиях предыдущего примера иногда применяют другой план контроля, который называется *обратным биномиальным выбором*. В этом случае испытания продолжают до получения заданного числа r «успехов» (здесь под «успехом» понимается обнаружение дефектного изделия). Пусть Y_i — число испытаний между $(i-1)$-м и i-м «успехами». Тогда $Y_1, \ \ldots, Y_r$ — независимые геометрические случайные величины:

$$\mathbf{P}_\theta\{Y_1 = y\} = \theta(1-\theta)^y \qquad y = 0, 1, 2, \ldots,$$

т. е. $Y = (Y_1, \quad , Y_r)$ — выборка из распределения $\overline{\mathrm{Bi}}(1, 1-\theta)$ (см. п. 2 § 1.1), которое является распределением экспоненциального типа (1) с

$$T_r(Y) = \sum_{i=1}^{r} Y_i \quad \text{и} \quad A(\theta) = \ln(1-\theta).$$

Так как $A(\theta)$ — убывающая функция θ, то в задаче $(H_0 \quad \theta \geqslant \theta_0, H_1^- \quad \theta < \theta_0)$ р. н. м. к. существует и имеет вид (см. таблицу) $\{T_r(Y) \geqslant c_\alpha^-\}$. Этот вывод соответствует и интуитивному представлению, так как при больших значениях θ естественно ожидать, что r-й «успех» наступит достаточно быстро, поэтому большие значения T естественно трактовать как свидетельствующие против гипотезы H_0 (в пользу альтернативы H_1^-). Для расчета критической границы

используют тот факт, что $\mathcal{L}_\theta(T_r(Y)) = \overline{\mathrm{Bi}}(r, 1-\theta)$, следовательно, при заданном уровне значимости α граница c_α^- определяется условием (см. (6) § 1.1)

$$\alpha_0 \equiv \sum_{x=c_\alpha^-+1}^{\infty} C_{r+x-1}^x \theta_0^r (1-\theta_0)^x < \alpha \leqslant \sum_{x=c_\alpha^-}^{\infty} C_{r+x-1}^x \theta_0^r (1-\theta_0)^x \equiv \alpha' \qquad (2)$$

Дальнейшие рассуждения такие же, как в примерах 4 и 5 § 5.2. Если $\alpha = \alpha'$, то критерий нерандомизированный и есть $\mathfrak{X}_{1\alpha}^{*-} = \{T_r(Y) \geqslant c_\alpha^-\}$ Если же $\alpha < \alpha'$, то можно использовать любой из нерандомизированных критериев $\mathfrak{X}_{1\alpha'}^{*-} = \{T_r(Y) \geqslant c_\alpha^-\}$ или $\mathfrak{X}_{1\alpha_0}^{*-} = \{T_r(Y) \geqslant c_\alpha^- + 1\}$ с уровнями значимости соответственно α' и α_0, либо строить рандомизированный критерий с помощью критической функции φ_α^* в (28) § 5.2 (с учетом обозначений, принятых в (2), в частности, здесь $p_0 = \alpha' - \alpha_0 = C_{r+c_\alpha^- -1}^{c_\alpha^-} \theta_0^r (1-\theta_0)^{c_\alpha^-}$).

Особенно прост будет критерий при $r = 1$, т. е. когда тестовой статистикой является Y_1 — число испытаний, предшествующих появлению первого дефектного изделия. В этом случае условие (2) записывается в виде

$$\alpha_0 \equiv (1-\theta_0)^{c_\alpha^-+1} < \alpha \leqslant (1-\theta_0)^{c_\alpha^-} \equiv \alpha',$$

поэтому при уровне значимости $\alpha = (1-\theta_0)^s$ при заданном целом $s \geqslant 1$ критерий задается критической областью $\mathfrak{X}_{1\alpha}^{*-} = \{Y_i \geqslant s\}$ и его мощность при произвольной альтернативе $\theta_1 < \theta_0$ есть

$$W(\theta_1) = \mathbf{P}_{\theta_1}\{Y_i \geqslant s\} = \sum_{y=s}^{\infty} \theta_1 (1-\theta_1)^y = (1-\theta_1)^s$$

Наконец, при больших r можно воспользоваться центральной предельной теоремой и получить нормальную аппроксимацию для распределения статистики $T_r(Y)$, именно (см. (7) § 1.1), при $r \to \infty$

$$\mathcal{L}_\theta(T_r(Y)) \sim \mathcal{N}\left(r\frac{1-\theta}{\theta},\ r\frac{1-\theta}{\theta^2}\right).$$

Эта аппроксимация позволяет просто рассчитать асимптотический вариант критерия. Имеем

$$\mathbf{P}_{\theta_0}\{T_r(Y) \geqslant c_\alpha^-\} \approx \Phi\left(\frac{r(1-\theta_0) - \theta_0 c_\alpha^-}{\sqrt{r(1-\theta_0)}}\right) = \alpha,$$

откуда

$$c_\alpha^- = r\frac{1-\theta_0}{\theta_0} - \zeta_\alpha \frac{\sqrt{r(1-\theta_0)}}{\theta_0}, \quad \Phi(\zeta_\alpha) = \alpha.$$

В итоге получаем, что при больших r и заданном уровне значимости α критерий имеет вид

$$\mathfrak{X}_{1\alpha}^{*-} = \left\{T_r(Y) \geqslant r\frac{1-\theta_0}{\theta_0} - \zeta_\alpha \frac{\sqrt{r(1-\theta_0)}}{\theta_0}\right\},$$

и его мощность $W_r(\theta_1)$ при «близких» альтернативах вида $\theta_1 = \theta_{1r} = \theta_0 - \delta/\sqrt{r}$, $\delta > 0$, удовлетворяет предельному соотношению

$$\lim_{r\to\infty} W_r(\theta_{1r}) = \Phi\left(\frac{\delta}{\theta_0\sqrt{1-\theta_0}} + \zeta_\alpha\right). \tag{3}$$

•

Пример 3 (Показательная модель, р. н. м. к. для односторонних гипотез). Предположим, что время «жизни» некоторого технического устройства есть случайная величина ξ, распределенная по показательному закону $\Gamma(\theta, 1)$ (см. п. 3 § 1.2) с неизвестным параметром $\theta > 0$. Пусть $X_1, \ldots, X_n$ — времена «жизни» n контрольных устройств, и по этой информации требуется проверить гипотезу $H_0 \quad \theta \geqslant \theta_0$ при альтернативе $H_1^- \quad \theta < \theta_0$. Так как модель $\Gamma(\theta, 1)$ входит в экспоненциальное семейство (1) (для нее $A(\theta) = -1/\theta$ и $T(X) = X_1 + \quad + X_n$), то в нашей задаче р. н. м. критерий существует и задается критической областью вида $\{T(X) \leqslant c\}$. Это также ожидаемый результат, так как параметр θ есть среднее значение ξ и потому малые значения $T(X)$ естественно интерпретировать в пользу альтернативы.

Заметим далее, что из свойств гамма-распределения вытекает соотношение

$$\mathcal{L}_\theta\left(\frac{2T(X)}{\theta}\right) = \chi^2(2n)$$

и, следовательно, границу критической области можно определить по таблицам квантилей распределения $\chi^2(2n)$. Именно, так как

$$\mathbf{P}_{\theta_0}\left\{\frac{2T(X)}{\theta_0} \leqslant \chi^2_{\alpha,\,2n}\right\} = \alpha,$$

то при уровне значимости α р. н. м. к. в нашей задаче имеет вид

$$\mathfrak{X}^{*-}_{1\alpha} = \left\{T(X) \leqslant \frac{\theta_0}{2}\chi^2_{\alpha,\,2n}\right\}.$$

Мощность этого критерия при произвольной альтернативе $\theta_1 < \theta_0$ вычисляется по формуле

$$W(\theta_1) = \mathbf{P}_{\theta_1}\left\{\frac{2T(X)}{\theta_1} \leqslant \frac{\theta_0}{\theta_1}\chi^2_{\alpha,\,2n}\right\} = F_{2n}\left(\frac{\theta_0}{\theta_1}\chi^2_{\alpha,\,2n}\right),$$

где $F_{2n}(t)$ — функция распределения закона $\chi^2(2n)$. •

2. Двусторонние альтернативы, р. н. м. несмещенные критерии

Как уже отмечалось выше, даже для моделей экспоненциального типа (1) с монотонным отношением правдоподобия при двусторонней альтернативе $H_1 = H_1^- \cup H_1^+ \quad \theta \neq \theta_0$ (θ — скалярный параметр) р. н. м. критерия, вообще говоря, не существует. В задачах с такими альтернативами, т. е. в задаче ($H_0 \quad \theta = \theta_0$, $H_1 \quad \theta \neq \theta_0$), обычно поступают следующим образом: используют статистику $T(X)$, с помощью которой строятся р. н. м. критерии против

односторонних альтернатив H^- и H^+, и задают критическую область в виде $\mathfrak{X}_{1\alpha} = \{T(\underline{x}) \leqslant c^-_{\alpha_1}\} \cup \{T(\underline{x}) \geqslant c^+_{\alpha_2}\}$, т. е. объединяют две соответствующие односторонние критические области с уровнями значимости α_1 и α_2 такими, что $\alpha_1 + \alpha_2 = \alpha$. Таким образом, получается целое семейство критериев с уровнем значимости α (условие $\alpha = \alpha_1 + \alpha_2$ оставляет одну степень свободы). В этом семействе обычно и находится оптимальный критерий. Прежде чем формулировать соответствующие общие утверждения рассмотрим характерный иллюстративный пример.

Пример 4 (Нормальная модель с неизвестным средним, двусторонняя альтернатива). Рассмотрим применение описанной методики к задаче $(H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0)$ в модели $\mathcal{N}(\theta, \sigma^2)$. Из примера 1 § 5.2, как отмечено в начале п. 1, следует, что р. н. м.к. уровня значимости α_1 при левосторонней альтернативе $H_1^- \quad \theta < \theta_0$ имеет вид (12) § 5.2 (с заменой α на α_1), а р. н. м.к. уровня значимости α_2 при правосторонней альтернативе $H_1^+ \quad \theta > \theta_0$ — вид (10) § 5.2 (с заменой $\alpha \to \alpha_2$). Следовательно, при произвольных α_1, α_2 таких, что $\alpha_1 + \alpha_2 = \alpha$, объединенная критическая область

$$\mathfrak{X}_{1\alpha} = \left\{ \underline{x} \quad \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} \leqslant -t_{\alpha_1} \quad \text{либо} \quad \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma} \leqslant t_{\alpha_2} \right\} \tag{4}$$

задает критерий уровня значимости α в задаче (H_0, H_1). Интуитивно предпочтителен выбор «симметричного» критерия, т. е. когда $\alpha_1 = \alpha_2 = \alpha/2$; в этом случае получаем критерий

$$\tilde{\mathfrak{X}}_{1\alpha} = \left\{ \underline{x} : \frac{\sqrt{n}|\bar{x} - \theta_0|}{\sigma} \geqslant t_{\alpha/2} \right\}. \tag{5}$$

Покажем, что этот критерий один обладает свойством несмещенности среди всех критериев вида (4). Мощности односторонних критериев указаны в примере 1 § 5.2, откуда получаем для мощности произвольного критерия (4) выражение

$$W(\theta; \mathfrak{X}_{1\alpha}) = \mathbf{P}_\theta\{X \in \mathfrak{X}^{*-}_{1\alpha_1}\} + \mathbf{P}_\theta\{X \in \mathfrak{X}^{*+}_{1\alpha_2}\} =$$
$$= \Phi\left(-\frac{\sqrt{n}}{\sigma}\Delta - t_{\alpha_1}\right) + \Phi\left(\frac{\sqrt{n}}{\sigma}\Delta - t_{\alpha_2}\right) \equiv \psi(\Delta),$$
$$\Delta = \theta - \theta_0.$$

Исследуем поведение функции $\psi(\Delta)$. Ее первые две производные равны

$$\psi'(\Delta) = \frac{1}{\sigma}\sqrt{\frac{n}{2\pi}}\left[\exp\left\{-\frac{1}{2}\left(\frac{\sqrt{n}}{\sigma}\Delta - t_{\alpha_2}\right)^2\right\} - \exp\left\{-\frac{1}{2}\left(\frac{\sqrt{n}}{\sigma}\Delta + t_{\alpha_1}\right)^2\right\}\right],$$

$$\psi''(\Delta) = \frac{n}{\sigma^2\sqrt{2\pi}}\left[\left(t_{\alpha_2} - \frac{\sqrt{n}}{\sigma}\Delta\right)\exp\left\{-\frac{1}{2}\left(\frac{\sqrt{n}}{\sigma}\Delta - t_{\alpha_2}\right)^2\right\} + \right.$$
$$\left. + \left(t_{\alpha_1} + \frac{\sqrt{n}}{\sigma}\Delta\right)\exp\left\{-\frac{1}{2}\left(\frac{\sqrt{n}}{\sigma}\Delta + t_{\alpha_1}\right)^2\right\}\right]$$

Отсюда

$$\psi'(\Delta) = 0 \iff t_{\alpha_2} - \frac{\sqrt{n}}{\sigma}\Delta = t_{\alpha_1} + \frac{\sqrt{n}}{\sigma}\Delta \iff \Delta = \Delta_0 = \sigma\frac{t_{\alpha_2} - t_{\alpha_1}}{2\sqrt{n}},$$

а

$$\psi''(\Delta_0) = \frac{n}{\sigma^2\sqrt{2\pi}}(t_{\alpha_1} + t_{\alpha_2}) \exp\left\{-\frac{1}{8}(t_{\alpha_1} + t_{\alpha_2})^2\right\}.$$

При малых α_1 и α_2 величины t_{α_1} и t_{α_2} положительны, следовательно, $\psi''(\Delta_0) > 0$, т. е. Δ_0 — точка минимума функции $\psi(\Delta)$. Поскольку $\Delta_0 = 0 \iff \alpha_1 = \alpha_2 = \alpha/2$, а $\psi(0) = \Phi(-t_{\alpha_1}) + \Phi(-t_{\alpha_2}) = \alpha_1 + \alpha_2 = \alpha$ (все критерии (4) имеют уровень значимости α), то лишь для симметричного критерия (5) выполняется условие несмещенности $W(\theta; \widetilde{\mathfrak{X}}_{1\alpha}) > \alpha$, $\forall\, \theta \neq \theta_0$. В остальных случаях, т. е. при $\alpha_1 \neq \alpha_2$, критерий $\mathfrak{X}_{1\alpha}$ при некоторых альтернативах будет иметь мощность меньшую α, т. е. будет смещенным (см. рис. 1). Таким образом, среди всех критериев вида (4) лишь симметричный критерий (5) — несмещенный и, следовательно, ему надо отдать предпочтение.

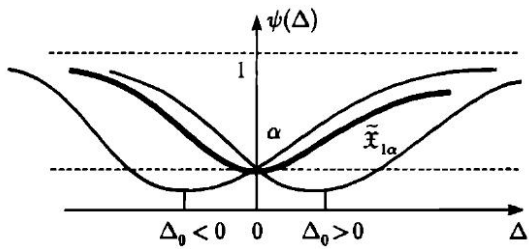


Рис. 1

Критерий (5) на самом деле обладает наибольшей мощностью среди всех *несмещенных* критериев уровня значимости α, т. е. является р. н. м. несмещенным критерием. Это утверждение является следствием приводимой ниже теоремы 1 об общем виде р. н. м. несмещенного критерия. •

Пусть Θ — интервал действительной оси и θ_0 — внутренняя точка Θ. Пусть, далее, допустимые распределения абсолютно непрерывны и функция правдоподобия $L(\underline{x}; \theta)$ имеет во всех внутренних точках Θ производную $\partial L(\underline{x}; \theta)/\partial\theta \equiv L_1(\underline{x}; \theta)$ такую, что $|L_1(\underline{x}; \theta)| \leqslant M(\underline{x})$, где $M(\underline{x})$ — интегрируемая функция на выборочном пространстве $\mathfrak{X}$ (тогда

$$\frac{\partial}{\partial\theta}\int\limits_S L(\underline{x}; \theta)\, d\underline{x} = \int\limits_S L_1(\underline{x}; \theta)\, d\underline{x}$$

для любого подмножества $S \subseteq \mathfrak{X}$). Определим, наконец, множество

$$\widetilde{\mathfrak{X}}_{1\alpha} = \{\underline{x} : L(\underline{x}; \theta_1) \geqslant cL(\underline{x}; \theta_0) + c_1 L_1(\underline{x}; \theta_0)\}, \qquad (6)$$

где $\theta_1 \neq \theta_0$ и постоянные $c \geqslant 0$ и c_1 определяются из условий

$$\mathbf{P}_{\theta_0}\{\widetilde{\mathfrak{X}}_{1\alpha}\} = \int\limits_{\widetilde{\mathfrak{X}}_{1\alpha}} L(\underline{x}; \theta_0)\, d\underline{x} = \alpha, \quad \mathbf{P}'_{\theta_0}\{\widetilde{\mathfrak{X}}_{1\alpha}\} = \int\limits_{\widetilde{\mathfrak{X}}_{1\alpha}} L_1(\underline{x}; \theta_0)\, d\underline{x} = 0. \tag{7}$$

Теорема 1. Если определенное в (6) множество не зависит от θ_1, то $\widetilde{\mathfrak{X}}_{1\alpha}$ есть р. н. м. несмещенный критерий уровня значимости α для проверки простой гипотезы $H_0 \quad \theta = \theta_0$ против двусторонней альтернативы $H_1 \quad \theta \neq \theta_0$.

Доказательство. Рассмотрим любой несмещенный критерий $\mathfrak{X}_{1\alpha}$ проверки гипотезы $H_0 \quad \theta = \theta_0$. Так как мощность $W(\theta_0; \mathfrak{X}_{1\alpha}) = \alpha$ и производная функции мощности по θ существует, то $W'(\theta_0; \mathfrak{X}_{1\alpha}) = 0$. Как и в доказательстве теоремы 1 § 5.2, запишем равенство

$$\mathbf{P}_{\theta_0}\{\widetilde{\mathfrak{X}}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} = \alpha - \mathbf{P}_{\theta_0}\{\mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} = \mathbf{P}_{\theta_0}\{\mathfrak{X}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\}; \tag{8}$$

кроме того из (7) имеем

$$\mathbf{P}'_{\theta_0}\{\widetilde{\mathfrak{X}}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} + \mathbf{P}'_{\theta_0}\{\mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} = 0$$

и аналогично

$$\mathbf{P}'_{\theta_0}\{\mathfrak{X}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} + \mathbf{P}'_{\theta_0}\{\mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} = 0,$$

т. е.

$$\mathbf{P}'_{\theta_0}\{\widetilde{\mathfrak{X}}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} = -\mathbf{P}'_{\theta_0}\{\mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} = \mathbf{P}'_{\theta_0}\{\mathfrak{X}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\}. \tag{9}$$

Из определения (6) множества $\widetilde{\mathfrak{X}}_{1\alpha}$ и равенств (8) и (9) получаем следующую цепочку соотношений:

$$\mathbf{P}_{\theta_1}\{\widetilde{\mathfrak{X}}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} \geqslant c\mathbf{P}_{\theta_0}\{\widetilde{\mathfrak{X}}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} + c_1\mathbf{P}'_{\theta_0}\{\widetilde{\mathfrak{X}}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} =$$
$$= c\mathbf{P}_{\theta_0}\{\mathfrak{X}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} + c_1\mathbf{P}'_{\theta_0}\{\mathfrak{X}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\} \geqslant \mathbf{P}_{\theta_1}\{\mathfrak{X}_{1\alpha} - \mathfrak{X}_{1\alpha}\widetilde{\mathfrak{X}}_{1\alpha}\}.$$

Отсюда следует, что $\mathbf{P}_{\theta_1}\{\widetilde{\mathfrak{X}}_{1\alpha}\} \geqslant \mathbf{P}_{\theta_1}\{\mathfrak{X}_{1\alpha}\}$, т. е. критерий $\widetilde{\mathfrak{X}}_{1\alpha}$ имеет наибольшую мощность среди всех несмещенных критериев того же уровня значимости α при альтернативе θ_1. Но по условию, множество $\widetilde{\mathfrak{X}}_{1\alpha}$ не зависит от θ_1, поэтому этот критерий является р. н. м. несмещенным критерием против любой против любой допустимой альтернативы $\theta_1 \neq \theta_0$. ■

Замечание. Аналогичное утверждение справедливо также и для дискретных моделей с теми же оговорками, что и при построении критерия Неймана—Пирсона в § 5.2.

Пример 4 (продолжение). Применим доказанный результат к критерию (5). Здесь

$$L_1(\underline{x}; \theta) = \frac{n}{\sigma^2}(\overline{x} - \theta)L(\underline{x}; \theta), \tag{10}$$

следовательно, неравенство в (6) эквивалентно неравенству

$$\frac{L(\underline{x}; \theta_1)}{L(\underline{x}; \theta_0)} \geqslant c + c_1\frac{n(\overline{x} - \theta_0)}{\sigma^2},$$

которое легко привести к виду

$$\exp\left\{MT(\underline{x}) - \frac{M^2}{2}\right\} \geqslant c + c_1' T(\underline{x}), \tag{11}$$

где

$$T(\underline{x}) = \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sigma}, \quad M = \frac{\sqrt{n}(\theta_1 - \theta_0)}{\sigma}, \quad c_1' = \frac{c_1\sqrt{n}}{\sigma}.$$

Неравенство, противоположное (11), эквивалентно неравенствам

$$t^{(1)} \leqslant T(\underline{x}) \leqslant t^{(2)} \tag{12}$$

Статистика $T(\boldsymbol{X})$ при нулевой гипотезе имеет стандартное нормальное распределение $\mathcal{N}(0, 1)$, симметричное относительно нуля. Отсюда и из (10) следует, что для обеспечения второго из соотношений (7) необходимо, чтобы интервал (12) был симметричен относительно нуля. Это означает, что $t^{(2)} = -t^{(1)}$ Для одновременного выполнения и первого из соотношений (7) надо положить $t^{(2)} = t_{\alpha/2}$, где $\Phi(-t_{\alpha/2}) = \alpha/2$. Таким образом, условия (7) выполняются тогда и только тогда, когда область (6) имеет вид $\widetilde{\mathfrak{X}}_{1\alpha} = \{\underline{x} \quad |T(\underline{x})| \geqslant t_{\alpha/2}\}$, т. е. мы пришли к (5). Следовательно, по теореме 1 критерий (5) является р. н. м. несмещенным критерием в задаче ($H_0 \quad \theta = \theta_0$, $H_1 \quad \theta \neq \theta_0$). Функция мощности этого критерия

$$W(\theta; \widetilde{\mathfrak{X}}_{1\alpha}) = \Phi\left(\frac{\sqrt{n}(\theta - \theta_0)}{\sigma} - t_{\alpha/2}\right) + \Phi\left(\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} - t_{\alpha/2}\right)$$

симметрична относительно точки $\theta = \theta_0$, в которой она имеет минимум, равный α. При $\theta \neq \theta_0$ мощность строго больше α и стремится к 1 при $\theta \to \pm\infty$. В соответствии с теоремой 1 график этой функции лежит выше соответствующего графика любого другого несмещенного критерия в данной задаче (все графики проходят через точку (θ_0, α)).

Сравним функцию мощности $W(\theta; \widetilde{\mathfrak{X}}_{1\alpha})$ с функциями мощности соответствующих односторонних р. н. м.к $W(\theta; \mathfrak{X}^{*-}_{1\alpha})$ и $W(\theta; \mathfrak{X}^{*+}_{1\alpha})$ (рис. 2). Функция $W(\theta; \widetilde{\mathfrak{X}}_{1\alpha})$ всегда заключена между этими двумя функциями (они изображены

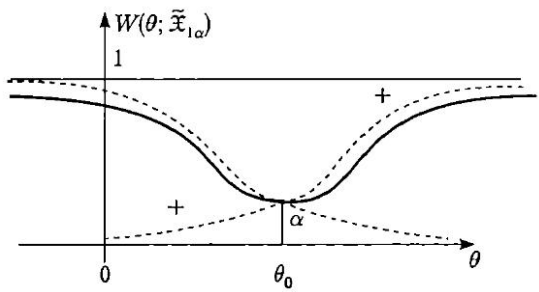


Рис. 2

на рисунке пунктирными линиями с соответствующим знаком), за исключением точки $\theta = \theta_0$, в которой все три графика совпадают. Это означает, что критерий $\widetilde{\mathfrak{X}}_{1\alpha}$ всегда менее мощный, чем один из односторонних критериев $\mathfrak{X}^{*+}_{1\alpha}$ или $\mathfrak{X}^{*-}_{1\alpha}$, но всегда более мощный, чем другой односторонний критерий.•

Более полно теория р. н. м. несмещенных критериев изложена в [16, гл. 4].

3. Локальные наиболее мощные критерии

На практике особый интерес представляют малые отклонения от нулевой гипотезы $H_0 \quad \theta = \theta_0$. В этом случае при исследовании свойств критерия можно ограничиться только анализом локального поведения его функции мощности $W(\theta)$ в окрестности точки θ_0. При таком подходе часто удается построить *локальный наиболее мощный* (л. н. м.) критерий, даже тогда, когда р. н. м. критерия не существует.

Пусть θ — скалярный параметр и выполнены условия регулярности, сформулированные перед теоремой 1. Кроме того, будем считать, что для любой критической области $\mathfrak{X}_{1\alpha}$ функция мощности $W(\theta) \quad W(\theta; \mathfrak{X}_{1\alpha})$ $(W(\theta_0) = \alpha)$ допускает разложение в ряд Тейлора в окрестности точки θ_0:

$$W(\theta) = \alpha + (\theta - \theta_0)W'(\theta_0) + \frac{(\theta - \theta_0)^2}{2}W''(\theta_0) + \tag{13}$$

Пусть теперь гипотеза $H_0 \quad \theta = \theta_0$ проверяется против правосторонней альтернативы $H_1^+ \quad \theta > \theta_0$. Тогда для получения л. н. м. одностороннего критерия необходимо максимизировать величину

$$W'(\theta_0) = \int\limits_{\mathfrak{X}_{1\alpha}} L_1(\underline{x}; \theta_0)\, d\underline{x}$$

при ограничении

$$W(\theta_0) = \int\limits_{\mathfrak{X}_{1\alpha}} L(\underline{x}; \theta_0)\, d\underline{x} = \alpha.$$

Повторяя рассуждения, использованные при доказательстве теоремы 1 § 5.2, можно получить, что оптимальная критическая область имеет вид

$$\mathfrak{X}^+_{1\alpha} = \left\{\underline{x} \quad \frac{L_1(\underline{x}; \theta_0)}{L(\underline{x}; \theta_0)} \geqslant c^+_\alpha\right\}, \tag{14}$$

где граница c^+_α определяется условием $W(\theta_0; \mathfrak{X}^+_{1\alpha}) = \alpha$.

Аналогично, критическая область

$$\mathfrak{X}^-_{1\alpha} = \left\{\underline{x} \quad \frac{L_1(\underline{x}; \theta_0)}{L(\underline{x}; \theta_0)} \leqslant c^-_\alpha\right\}, \tag{15}$$

задает л. н. м. критерий в случае левосторонней альтернативы $H_1^- \quad \theta < \theta_0$.

При двусторонней альтернативе $H_1 \quad \theta \neq \theta_0$ необходимо ввести дополнительное условие несмещенности критерия: $W'(\theta_0) = 0$; тогда л. н. м. критерием является критерий, максимизирующий $W''(\theta_0)$ — коэффициент при

$(\theta-\theta_0)^2$ в разложении (13). Как и при доказательстве теоремы 1 выше, можно показать, что в этом случае оптимальная критическая область имеет вид

$$\widetilde{\mathfrak{X}}_{1\alpha} = \{\underline{x} : L_2(\underline{x};\theta_0) \geqslant cL(\underline{x};\theta_0) + c_1 L_1(\underline{x};\theta_0)\}, \tag{16}$$

где

$$L_2(\underline{x};\theta) = \frac{\partial L_1(\underline{x};\theta)}{\partial\theta}$$

и константы c и c_1 определяются условиями $W(\theta_0)=\alpha$, $W'(\theta_0)=0$.

Пример 5 (Нормальная модель, л. н. м. критерий для среднего). Рассмотрим ту же задачу, что в примере 4, и построим в ней л. н. м. критерий. Здесь (см. (10))

$$L_2(\underline{x};\theta) = \left[\frac{n^2}{\sigma^4}(\overline{x}-\theta)^2 - \frac{n}{\sigma^2}\right] L(\underline{x};\theta),$$

поэтому

$$\frac{L_1(\underline{x};\theta_0)}{L(\underline{x};\theta_0)} = \frac{n}{\sigma^2}(\overline{x}-\theta_0), \quad \frac{L_2(\underline{x};\theta_0)}{L(\underline{x};\theta_0)} = \frac{n^2}{\sigma^4}(\overline{x}-\theta_0)^2 - \frac{n}{\sigma^2}.$$

Следовательно, область (16) задается условием $T^2(\underline{x}) \geqslant k_1 T(\underline{x}) + k_2$ (T определено в (11)). Чтобы было выполнено условие несмещенности, эта область должна быть симметричной (поскольку распределение статистики $T(X)$ при гипотезе H_0 симметрично относительно нуля), что имеет место только при $k_1 = 0$. Таким образом, условие можно переписать в виде $|T(\underline{x})| \geqslant k$. Наконец, из условия $W(\theta_0)=\alpha$ находим $k = t_{\alpha/2}$, где $\Phi(-t_{\alpha/2}) = \alpha/2$. Окончательно получаем, что область (16) имеет вид $\{|T| \geqslant t_{\alpha/2}\}$, т. е. результат совпадает с полученным в примере 4. Это ожидаемый результат, так как л. н. м. несмещенный критерий должен совпадать с р. н. м. несмещенным критерием, когда последний существует. •

Вернемся к случаю односторонних альтернатив и проанализируем более подробно условия в (14)–(15). Вспомнив определение вклада выборки $V(X;\theta)$ (см. (1) § 3.2), можем записать

$$\frac{L_1(\underline{x};\theta_0)}{L(\underline{x};\theta_0)} = \frac{\partial}{\partial\theta}\ln L(\underline{x};\theta_0) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta_0) = V(\underline{x};\theta_0).$$

Как показано в п. 1 § 3.2, для регулярных моделей

$$\mathbf{E}_{\theta_0} V(X;\theta_0) = 0, \quad \mathbf{D}_{\theta_0} V(X;\theta_0) = ni(\theta_0),$$

где $i(\theta)$ — информация Фишера (см. (5)–(5′) § 3.2). Если n велико, то, по центральной предельной теореме,

$$\mathcal{L}_{\theta_0}\left(\frac{V(X;\theta_0)}{\sqrt{ni(\theta_0)}}\right) \sim \mathcal{N}(0,1), \tag{17}$$

следовательно, для больших выборок при вычислении границ в (14)–(15) можно использовать нормальную аппроксимацию (17): в (14) полагать

$c_\alpha^+ \approx t_\alpha\sqrt{ni(\theta_0)}$, а в (15) полагать $c_\alpha^- \approx -t_\alpha\sqrt{ni(\theta_0)}$, где $\Phi(-t_\alpha) = \alpha$. Тем самым получено общее асимптотическое решение задачи построения л. н. м. односторонних критериев для больших выборок.

Применяя прием объединения двух односторонних критических областей, можно построить и асимптотический двусторонний критерий

$$\left\{|V(X;\theta_0)| \geqslant t_{\alpha/2}\sqrt{ni(\theta_0)}\right\} \tag{18}$$

проверки гипотезы $H_0 \quad \theta = \theta_0$ против альтернативы $H_1 \quad \theta \neq \theta_0$.

Из свойства симметрии предельного распределения статистики $\dfrac{V(X;\theta_0)}{\sqrt{n}}$ вытекает, что критерий (18) — несмещенный и не существует другого несмещенного критерия, который имел бы асимптотически бóльшую мощность. Другими словами, *критерий* (18) *является асимптотически л. н. м. несмещенным критерием в задаче* ($H_0 \quad \theta = \theta_0$, $H_1 \quad \theta \neq \theta_0$).

Построенный приближенный критерий (18) совпадает с полученным в примере 5 точным критерием для нормальной модели $\mathcal{N}(\theta, \sigma^2)$, основанным на статистике

$$T(X) = \frac{\sqrt{n}(\overline{X} - \theta_0)}{\sigma},$$

поскольку в этом случае

$$V(X;\theta_0) = \frac{\sqrt{n}T(X)}{\sigma},$$

а $i(\theta) = \sigma^{-2}$ (см. табл. 1 § 3.2).

Пример 6 (Модель Коши, л. н. м. критерий для нее). Пусть требуется проверить гипотезу $H_0 \quad \theta = \theta_0$ против альтернативы $H_1 \quad \theta \neq \theta_0$ для модели Коши $\mathcal{K}(\theta)$ (см. указание к табл. 1 § 3.2). Используя результаты примера 8 § 3.5, получим, что при больших n л. н. м. несмещенный критерий имеет вид

$$\widetilde{\mathfrak{X}}_{1\alpha} = \left\{\underline{x} \quad \frac{2\sqrt{2}}{\sqrt{n}}\left|\sum_{i=1}^{n}\frac{x_i - \theta_0}{1 + (x_i - \theta_0)^2}\right| \geqslant t_{\alpha/2}\right\}. \qquad \bullet$$

4. Проверка гипотез и доверительное оценивание

Между задачами проверки гипотез о параметре θ и его доверительным оцениваением имеется тесная связь, позволяющая просто получать решение одной из этих задач, когда решение другой известно. Это делается с помощью следующих рассуждений.

Рассмотрим для произвольного $\theta_0 \in \Theta$ какой-либо критерий $\mathfrak{X}_{1\alpha} = \mathfrak{X}_{1\alpha}(\theta_0)$ проверки простой гипотезы $H_0 \quad \theta = \theta_0$. Пусть $\mathfrak{X}_{0\alpha}(\theta_0) = \overline{\mathfrak{X}}_{1\alpha}(\theta_0)$ — область принятия гипотезы H_0. Тем самым в выборочном пространстве $\mathfrak{X}$ задано семейство подмножеств $\{\mathfrak{X}_{0\alpha}(\theta),\ \theta \in \Theta\}$. Определим теперь при каждом $\underline{x} \in \mathfrak{X}$ подмножество $\mathcal{G}_\gamma(\underline{x}) \subset \Theta$, $\gamma = 1 - \alpha$, положив $\mathcal{G}_\gamma(\underline{x}) = \{\theta : \mathfrak{X}_{0\alpha}(\theta) \ni \underline{x}\}$.

Таким образом, в параметрическом множестве Θ получаем семейство подмножеств $\{\mathcal{G}_\gamma(\underline{x}),\ \underline{x} \in \mathfrak{X}\}$

Рассмотрим случайное множество $\mathcal{G}_\gamma(X)$. События

$$\{\theta \in \mathcal{G}_\gamma(X)\} \quad \text{и} \quad \{X \in \mathfrak{X}_{0\alpha}(\theta)\}$$

эквивалентны, так как по построению каждое из них влечет за собой другое, поэтому их вероятности при каждом θ совпадают. Но, по построению, $\mathbf{P}_\theta\{X \in \mathfrak{X}_{0\alpha}(\theta)\} = 1 - \alpha$, следовательно, $\mathbf{P}_\theta\{\theta \in \mathcal{G}_\gamma(X)\}$ $1 - \alpha$, т. е. $\mathcal{G}_\gamma(X)$ является γ-доверительной областью для θ (см. напр., соотношение (33) § 3.5). Эти рассуждения можно и обратить, именно, если имеется семейство γ-доверительных областей $\{\mathcal{G}_\gamma(\underline{x}),\ \underline{x} \in \mathfrak{X}\}$ для θ, то множество $\mathfrak{X}_{0,\alpha} = \{\underline{x}\ \ \theta_0 \in \mathcal{G}_\gamma(\underline{x})\}$, $\alpha = 1 - \gamma$, определяет область принятия гипотезы H_0 $\theta = \theta_0$ с уровнем значимости α.

Таким образом, задача построения доверительной области для параметра θ (доверительного интервала, если θ — скалярный параметр) и задача проверки простой гипотезы о θ взаимно обратны. При этом р. н. м. критерий (когда он существует) соответствует наикратчайшей доверительной области и наоборот.

Проиллюстрируем этот принцип несколькими полезными примерами.

Пример 7 (Нормальная модель, оптимальный доверительный интервал для среднего). Рассмотрим модель $\mathcal{N}(\theta, \sigma^2)$ (σ^2 известно). В примере 4 построен оптимальный (р. н. м. несмещенный) критерий (5) в задаче (H_0 $\theta = \theta_0$, $H_1 : \theta \neq \theta_0$). Здесь

$$\mathfrak{X}_{0\alpha}(\theta) = \left\{\underline{x}\ \ \frac{\sqrt{n}|\bar{x} - \theta|}{\sigma} < t_{\alpha/2}\right\}, \quad \Phi(-t_{\alpha/2}) = \alpha/2,$$

и множество $\mathcal{G}_\gamma(\underline{x})$, $\gamma = 1 - \alpha$, имеет вид

$$\mathcal{G}_\gamma(\underline{x}) = \left\{\theta\ \ \frac{\sqrt{n}|\bar{x} - \theta|}{\sigma} < t_{\alpha/2}\right\} = \left\{\theta\ \ \bar{x} - \frac{t_{\alpha/2}\sigma}{n} < \theta < \bar{x} + \frac{t_{\alpha/2}\sigma}{n}\right\}.$$

Следовательно, интервал $(\overline{X} \mp t_{\alpha/2}\,\sigma/\sqrt{n})$ является наикратчайшим среди всех $(1 - \alpha)$-доверительных интервалов для неизвестного среднего закона $\mathcal{N}(\theta, \sigma^2)$ — это значительное усиление результата примера 2 § 3.8 (подчеркнем, что $t_{\alpha/2} = c_{1-\alpha}$).

Далее, так как р. н. м. критерий против правосторонней альтернативы H_1^+ $\theta > \theta_0$ задается критической областью (10) § 5.2, то

$$\mathfrak{X}_{0\alpha}(\theta) = \left\{\underline{x}\ \ \frac{\sqrt{n}(\bar{x} - \theta)}{\sigma} < t_\alpha\right\}, \quad \Phi(-t_\alpha) = \alpha,$$

и потому множество $\mathcal{G}_\gamma(\underline{x})$ $\{\theta\ \ \theta > \bar{x} - t_\alpha\sigma/\sqrt{n}\}$ Отсюда имеем наилучший односторонний (нижний) $(1 - \alpha)$-доверительный интервал для θ: $\left(\overline{X} - t_\alpha\sigma/\sqrt{n}, \infty\right)$. Аналогично, исходя из (12) § 5.2, получаем наилучший верхний $(1 - \alpha)$-доверительный интервал для θ: $(-\infty, \overline{X} + t_\alpha\sigma/\sqrt{n})$. •

Пример 8 (Оптимальный доверительный интервал для дисперсии нормальной модели). В примере 3 § 3.8 был построен центральный γ-доверительный интервал для неизвестной дисперсии модели $\mathcal{N}(\mu, \theta^2)$. В то же время в упр. 12 к данной главе указан вид р. н. м. несмещенного критерия для гипотезы $H_0 \quad \theta = \theta_0$ против двусторонней альтернативы $H_1 \quad \theta \neq \theta_0$, откуда имеем

$$\mathfrak{X}_{0\alpha}(\theta) = \left\{\underline{x} \quad \theta^2 \chi^2_{\alpha_1, n} < T(\underline{x}) < \theta^2 \chi^2_{1-\alpha_2, n}\right\}, \quad T(\underline{x}) = \sum_{i=1}^{n} (x_i - \mu)^2$$

Следовательно, множество $\mathcal{G}_\gamma(\underline{x})$, $\gamma = 1 - \alpha$, имеет здесь вид

$$\mathcal{G}_\gamma(\underline{x}) = \left\{\theta^2 \quad \frac{T(\underline{x})}{\chi^2_{1-\alpha_2, n}} < \theta^2 < \frac{T(\underline{x})}{\chi^2_{\alpha_1, n}}\right\}.$$

Это означает, что наикратчайший $(1-\alpha)$-доверительный интервал для θ^2 есть

$$\left(\frac{T(X)}{\chi^2_{1-\alpha_2,}}, \frac{T(X)}{\chi^2_{\alpha_1, n}}\right),$$

где $(\chi^2_{\alpha_1, n}, \chi^2_{1-\alpha_2, n})$ указаны в упр. 12. Следовательно, центральный интервал в данном случае не является наилучшим. Приведем краткую таблицу значений пар $(\chi^2_{\alpha_1, n}; \chi^2_{1-\alpha_2, n})$ и $(\chi^2_{(1-\gamma)/2, n}; \chi^2_{(1+\gamma)/2, n})$ для $\gamma = 1 - \alpha = 0{,}95$:

n	$(\chi^2_{\alpha_1, n}; \chi^2_{1-\alpha_2, n})$	$(\chi^2_{(1-\gamma)/2, n}; \chi^2_{(1+\gamma)/2, n})$
2	(0,08; 9,53)	(0,05; 7,38)
5	(0,99; 14,37)	(0,83; 12,83)
10	(3,52; 21,73)	(3,25; 20,48)
20	(9,96; 35,23)	(9,59; 34,17)

•



«Мешающий» параметр

Методику, устанавливающую связь между задачами доверительного оценивания и проверки гипотез, можно применять и в более общих ситуациях, когда требуется оценить лишь подвектор $\theta^{(1)}$ параметрического вектора $\theta \quad (\theta^{(1)}, \theta^{(2)})$ (о $\theta^{(2)}$ в таком случае говорят как о «мешающем» параметре), а также когда интерес представляет заданная параметрическая функция $\tau(\theta)$.

Пусть, например, имеется семейство γ-доверительных областей $\mathcal{G}_\gamma(\underline{x})$ для $\theta^{(1)}$, т. е. семейство подмножеств параметрического множества

$$\Theta^{(1)} = \{\theta^{(1)} \quad (\theta^{(1)}, \theta^{(2)}) \in \Theta\},$$

удовлетворяющих условию

$$\mathbf{P}_\theta\{\theta^{(1)} \in \mathcal{G}_\gamma(X)\} \geqslant \gamma, \quad \forall\, \theta.$$

Определим для фиксированного $\theta_0^{(1)}$ подмножество

$$\mathfrak{X}_{0,\,1-\gamma} = \{\underline{x} \quad \mathcal{G}_\gamma(\underline{x}) \ni \theta_0^{(1)}\}$$

в выборочном пространстве $\mathfrak{X}$. Тогда по построению

$$\mathbf{P}_{(\theta_0^{(1)},\,\theta^{(2)})}\{X \in \mathfrak{X}_{0,\,1-\gamma}\} = \mathbf{P}_{(\theta_0^{(1)},\,\theta^{(2)})}\{\theta_0^{(1)} \in \mathcal{G}_\gamma(X)\} \geqslant \gamma,$$

или

$$\mathbf{P}_{(\theta_0^{(1)},\,\theta^{(2)})}\{X \in \overline{\mathfrak{X}}_{0,\,1-\gamma}\} \leqslant 1-\gamma.$$

Следовательно, область $\mathfrak{X}_{1\alpha} = \overline{\mathfrak{X}}_{0,\,1-\gamma}$ $\alpha = 1-\gamma$, задает критерий уровня значимости α для проверки гипотезы H_0 $\theta^{(1)} = \theta_0^{(1)}$ (это гипотеза сложная, так как оставляет неопределенным значение $\theta^{(2)}$). Как и выше, можно провести обратные рассуждения: от задачи проверки гипотезы перейти к задаче доверительного оценивания.

Пример 9 (проверка гипотез для параметров общей нормальной модели). В примере 1 § 2.5 построены доверительные интервалы для неизвестных среднего θ_1 и дисперсии θ_2^2 в модели $\mathcal{N}(\theta_1, \theta_2^2)$. Используя результат (12) из этого примера и описанную методику, сразу получаем критерий проверки гипотезы о среднем H_0 $\theta_1 = \theta_{10}$ (θ_2 произвольно) против альтернативы H_1 $\theta_1 \neq \theta_{10}$:

$$\mathfrak{X}_{1\alpha} = \left\{\underline{x} \quad \sqrt{n-1}\,\frac{|\overline{x} - \theta_{10}|}{S(\underline{x})} \geqslant t_{1-\alpha/2,\,n-1}\right\}. \tag{19}$$

Известно [16, с. 225], что критерий (19) является оптимальным (р. н. м. несмещенным), поэтому и доверительный интервал (12) § 2.5 является наикратчайшим. Ü

Аналогично, используя результат (9) § 2.5, получаем следующий критерий для проверки гипотезы о дисперсии H_0 $\theta_2 = \theta_{20}$ (θ_1 произвольно) против альтернативы H_1 $\theta_2 \neq \theta_{20}$:

$$\mathfrak{X}_{1\alpha} = \{\underline{x} \quad nS^2(\underline{x}) \leqslant \theta_{20}^2 \chi^2_{\alpha_1,\,n-1} \quad \text{либо} \quad nS^2(\underline{x}) \geqslant \theta_{20}^2 \chi^2_{1-\quad -1}\}, \tag{20}$$

где $\chi^2_{\alpha_1,\,n-1}$ и $\chi^2_{1-\alpha_2,\,n-1}$ — те же, что в примере 8 с заменой n на $n-1$.

Критерий (20) также р. н. м. несмещенный [16, с. 224], поэтому и соответствующий доверительный интервал (9) § 2.5 — наикратчайший (он не совпадает с центральным интервалом (10)). •

Пример 10 (гипотеза о равенстве средних двух нормальных моделей). Рассмотрим ситуацию, описанную в примере (2) § 2.5. Используя построенный там доверительный интервал для разности средних $\Delta = \theta^{(1)} - \theta^{(2)}$ получаем критерий проверки гипотезы о равенстве средних H_0 $\Delta = 0$, который определяется критической областью

$$\mathfrak{X}_{1\alpha} = \left\{(\underline{x}, \underline{y}) \quad |\overline{x} - \overline{y}| \geqslant t_{1-\alpha/2,\,n+m-2}\sqrt{\frac{n+m}{nm(n+m-2)}\left(nS^2(\underline{x}) + mS^2(\underline{y})\right)}\right\} \tag{21}$$

Эта область задает р. н. м. несмещенный критерий [16, с. 234], тем самым и соответствующий доверительный интервал для Δ является наиболее точным.

Если же теоретические дисперсии известны, то, используя результат, указанный в таблице § 3.8, получаем, что критерий имеет вид

$$\mathfrak{X}_{1\alpha} = \left\{(\underline{x}, \underline{y}) \quad |\overline{x} - \overline{y}| \geqslant \zeta_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right\}, \quad \Phi(\zeta_p) = p. \tag{22}$$

•

§ 5.4. Критерий отношения правдоподобия

1. Метод отношения правдоподобия для общих гипотез

Одним из наиболее универсальных методов построения критериев проверки сложных параметрических гипотез является *метод отношения правдоподобия*, суть которого состоит в следующем. Для проверки гипотезы $H_0 \quad \theta \in \Theta_0 \subset \Theta$ против альтернативы $H_1 \quad \theta \in \Theta_1 = \Theta \backslash \Theta_0$ вводится *статистика отношения правдоподобия*

$$\Lambda_n = \Lambda_n(X, \Theta_0) = \frac{\sup\limits_{\theta \in \Theta_1} L_n(X; \theta)}{\sup\limits_{\theta \in \Theta_0} L_n(X; \theta)}, \quad X = (X_1, \quad , X_n),$$

представляющая собой естественное обобщение на случай сложных гипотез статистики $l(X)$ Неймана—Пирсона (см. (3) § 5.2); как и в критерии Неймана—Пирсона в случае простых гипотез, здесь в критическую область включаются большие значения этой статистики.

В литературе, однако, традиционно в качестве тестовой статистики в задаче (H_0, H_1) используют несколько отличную статистику КОП

$$\lambda_n = \lambda_n(X; \Theta_0) = \frac{\sup\limits_{\theta \in \Theta_0} L_n(X; \theta)}{\sup\limits_{\theta \in \Theta} L_n(X; \theta)}, \tag{1}$$

которая предпочтительнее Λ_n с позиций практических вычислений, и критическую область задают в виде

$$\mathfrak{X}_1 = \{\underline{x} : \lambda_n(\underline{x}; \Theta_0) \leqslant c\}. \tag{2}$$

Критическую границу c в (2) выбирают так, чтобы критерий имел заданный уровень значимости α:

$$\mathbf{P}_\theta\{X \in \mathfrak{X}_1\} = \mathbf{P}_\theta\{\lambda_n(X; \Theta_0) \leqslant c\} \leqslant \alpha, \quad \forall\, \theta \in \Theta_0. \tag{3}$$

Соотношениями (1)–(3) определяется *критерий отношения правдоподобия* (КОП) в задаче $(H_0 \quad \theta \in \Theta_0, \; H_1 \quad \theta \in \Theta_1)$.

Как было показано в § 5.2, метод отношения правдоподобия в случае простых гипотез H_0 и H_1 приводит к оптимальному критерию. В общем случае

сложных гипотез это свойство оптимальности не выполняется. Тем не менее этот метод широко применяют, получая удовлетворительные решения во многих практических задачах. Кроме того, как будет показано ниже, при некоторых естественных условиях регулярности на исходную статистическую модель $\mathcal{F} = \{F(x;\theta),\ \theta \in \Theta\}$ КОП обладает свойством асимптотической оптимальности для больших выборок.

Пример 1 (КОП для среднего нормальной модели). Продемонстрируем изложенный метод на уже рассмотренной в примере 9 § 5.3 задаче о среднем ($H_0 \quad \theta_1 = \theta_{10}$, $H_1 \quad \theta_1 \neq \theta_{10}$) для нормальной модели $\mathcal{N}(\theta_1, \theta_2^2)$. Здесь $\Theta = \{\theta = (\theta_1, \theta_2) \quad \theta_1 \in R^1,\ \theta_2 > 0\}$ — полуплоскость, $\Theta_0 = \{\theta \quad \theta_1 = \theta_{10},\ \theta_2 > 0\}$ полупрямая. В соответствии с (1) мы должны вычислить абсолютный супремум $\sup\limits_{\theta\in\Theta} L(X;\theta)$ и условный (при гипотезе H_0) супремум $\sup\limits_{\theta\in\Theta_0} L(X;\theta)$ функции правдоподобия выборки X. Но такие задачи мы уже решали при нахождении оценок максимального правдоподобия (о. м. п.) для параметров модели в § 3.5. Так, в примере 1 § 3.5 мы нашли, что

$$\sup_{\theta\in\Theta} L(X;\theta) = L\big(X;\widehat{\theta} = (\overline{X}, S)\big) = (2\pi e S^2)^{-n/2}, \qquad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Условный же супремум есть (см. табл. 2 § 3.2)

$$\sup_{\theta\in\Theta_0} L(X;\theta) = \sup_{\theta_2} L\big(X, (\theta_{10}, \theta_2)\big) = L\big(X; (\theta_{10}, S_0)\big) = (2\pi e S_0^2)^{-n/2},$$

где

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \theta_{10})^2$$

— о. м. п. для дисперсии θ_2^2 при гипотезе H_0; при этом легко видеть, что

$$S_0^2 = S^2 + (\overline{X} - \theta_{10})^2$$

Отсюда имеем

$$\lambda_n = \left(\frac{S_0^2}{S^2}\right)^{-n/2} = \left(1 + \frac{t_{n-1}^2}{n-1}\right)^{-n/2} \qquad t_{n-1} = \sqrt{n-1}\,\frac{\overline{X} - \theta_{10}}{S},$$

и неравенство $\{\lambda_n \leqslant c\} \Longleftrightarrow \{|t_{n-1}| \geqslant t\}$ Но статистика t_{n-1} есть стьюдентово отношение (см. (29) § 1.2), соответствующее гипотезе H_0, т. е. $\mathcal{L}(t_{n-1}|H_0) = S(n-1)$. Таким образом, статистика t_{n-1} при гипотезе H_0 имеет распределение, не зависящее от «мешающего» параметра θ_2, а именно распределение Стьюдента с $n-1$ степенями свободы. Тем самым мы можем рассчитать критическую границу t при заданном уровне значимости α: $t = t_{1-\alpha/2,\,n-1}$, и в итоге прийти к критерию (19) в примере 9 § 5.3. Таким образом, в данном случае метод отношения правдоподобия приводит к оптимальному критерию. •

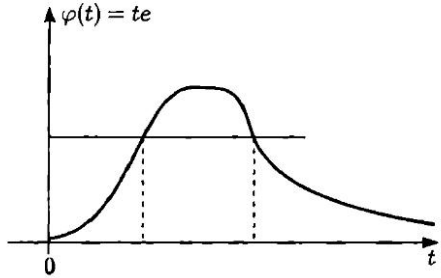


Рис. 1

Пример 2 (КОП для дисперсии нормальной модели). Рассмотрим, к какому результату приводит аналогичный подход в задаче о дисперсии ($H_0 : \theta_2 = \theta_{20}$, $H_1 : \theta_2 \neq \theta_{20}$) нормальной модели $\mathcal{N}(\theta_1, \theta_2^2)$. Абсолютный супремум $\sup\limits_{\theta \in \Theta} L(\boldsymbol{X}; \theta)$ здесь тот же, что и в предыдущем примере, поэтому достаточно вычислить условный супремум $\sup\limits_{\theta \in \Theta_0} L(\boldsymbol{X}; \theta)$. В данном случае

$$\Theta_0 = \{\theta = (\theta_1, \theta_2) \quad \theta_2 = \theta_{20},\ \theta_1 \in R^1\},$$

т. е. надо найти о. м. п. $\widehat{\theta}_1$ для модели $\mathcal{N}(\theta_1, \theta_{20}^2)$ с известной дисперсией и затем вычислить $L\big(\boldsymbol{X}; (\widehat{\theta}_1, \theta_{20})\big)$. Но $\widehat{\theta}_1 = \overline{X}$ (см. табл. 2 § 3.2), а из представления (2) § 3.3 для функции правдоподобия $L(\boldsymbol{X}; \theta)$ следует, что

$$\sup_{\theta \in \Theta_0} L(\boldsymbol{X}; \theta) = L\big(\boldsymbol{X}; (\overline{X}, \theta_{20})\big) = (2\pi\theta_{20}^2)^{-n/2} \exp\left\{-\frac{n}{2\theta_{20}^2} S^2\right\}.$$

Отсюда имеем (см. (1))

$$\lambda_n = \left(\frac{eS^2}{\theta_{20}^2}\right)^{n/2} \exp\left\{-\frac{n}{2}\frac{S^2}{\theta_{20}^2}\right\} = (T_n e^{-T_n + 1})^{n/2}, \qquad T_n = T_n(\boldsymbol{X}) = \frac{S^2}{\theta_{20}^2},$$

и (см. рис. 1) неравенство $\{\lambda_n \leqslant c\} \Longleftrightarrow \{T_n \leqslant t_1 \text{ либо } T_n \geqslant t_2\}$. Таким образом, КОП в данном случае имеет вид

$$\mathfrak{X}_{1\alpha} = \big\{\underline{x} \quad S^2(\underline{x}) \leqslant \theta_{20}^2 t_1 \quad \text{либо} \quad S^2(\underline{x}) \geqslant \theta_{20}^2 t_2\big\}.$$

Для вычисления критических границ t_1 и t_2 воспользуемся теоремой Фишера (теорема 2 § 2.5), согласно которой

$$\mathcal{L}_\theta\left(\frac{n}{\theta_2^2} S^2\right) = \chi^2(n-1).$$

Следовательно, мы имеем уравнение

$$\alpha = \mathbf{P}\{\boldsymbol{X} \in \mathfrak{X}_{1\alpha} | H_0\} = F_{n-1}(nt_1) + 1 - F_{n-1}(nt_2),$$

где $F_{n-1}(t)$ — функция распределения закона $\chi^2(n-1)$. Полагая $\alpha_1 = F_{n-1}(nt_1)$, $\alpha_2 = 1 - F_{n-1}(nt_2)$, будем иметь $nt_1 = \chi^2_{\alpha_1, n-1}$, $nt_2 = \chi^2_{1-\alpha_2, n-1}$, и тем самым

$$\mathfrak{X}_{1\alpha} = \left\{\underline{x} \quad nS^2(\underline{x}) \leqslant \theta_{20}^2 \chi^2_{\alpha_1, n-1} \quad \text{либо} \quad nS^2(\underline{x}) \geqslant \theta_{20}^2 \chi^2_{1-\alpha_2, n-1}\right\},$$
$$\alpha_1 + \alpha_2 = \alpha.$$

Здесь α_1 и α_2 могут быть любыми положительными числами, в сумме дающими α, поэтому мы имеем на самом деле целый класс критериев уровня значимости α. Эту неопределенность можно устранить, исследуя полученные критерии на несмещенность. Для этого вычислим функцию мощности $W(\theta_2) = W(\theta_2; \mathfrak{X}_{1\alpha})$. Снова воспользовавшись теоремой Фишера, получим

$$W(\theta_2) = \mathbf{P}_\theta\left\{\frac{n}{\theta_2^2} S^2 \leqslant \frac{\theta_{20}^2}{\theta_2^2}\chi^2_{\alpha_1, n-1}\right\} + \mathbf{P}_\theta\left\{\frac{n}{\theta_2^2} S^2 \geqslant \frac{\theta_{20}^2}{\theta_2^2}\chi^2_{1-\alpha_2, n-1}\right\} =$$
$$= F_{n-1}\left(\frac{\theta_{20}^2}{\theta_2^2}\chi^2_{\alpha_1, n-1}\right) + 1 - F_{n-1}\left(\frac{\theta_{20}^2}{\theta_2^2}\chi^2_{1-\alpha_2, n-1}\right)$$

Чтобы обеспечить несмещенность: $W(\theta_2) \geqslant \alpha$ при $\theta_2 \neq \theta_{20}$ (напомним, что $W(\theta_{20}) = \alpha$), надо решить уравнение $W'(\theta_{20}) = 0$, т. е. (далее $F'_{n-1}(t) = k_{n-1}(t)$)

$$\chi^2_{\alpha_1, n-1} k_{n-1}(\chi^2_{\alpha_1, n-1}) = \chi^2_{1-\alpha_2, n-1} k_{n-1}(\chi^2_{1-\alpha_2, n-1}).$$

Вместе с условием $\alpha_1 + \alpha_2 = \alpha$ последнее уравнение однозначно определяет выбор пары (α_1, α_2), а тем самым и $(\chi^2_{\alpha_1, n-1}, \chi^2_{1-\alpha_2, n-1})$, обеспечивающих несмещенность критерия $\mathfrak{X}_{1\alpha}$. В итоге мы приходим к критерию (20), указанному в примере 9 § 5.3, т. е. и в данном случае метод приводит к оптимальному критерию.

Если рассмотренные примеры носили сугубо иллюстративный характер (мы получили в них уже известные решения, следуя новому методу), то два следующих примера по существу демонстрируют возможности метода отношения правдоподобия. В них будут рассмотрены типичные, так называемые, *задачи для многих выборок*, для решения которых этот метод особенно эффективен. •

Задачи для многих выборок

Пример 3 (гипотеза однородности для нескольких нормальных выборок). Пусть имеется $k \geqslant 2$ независимых выборок $X_j = (X_{j1}, \quad, X_{jn_j})$, при этом X_j — выборка из нормального распределения $\mathcal{N}(\theta_{1j}, \theta_2^2)$, $j = 1, \quad, k$ (все параметры $\theta = (\theta_{11}, \quad, \theta_{1k}, \theta_2)$ неизвестны). Требуется проверить *гипотезу однородности* H_0 $\theta_{11} = \theta_{12} = \quad = \theta_{1k}$ (при гипотезе H_0 все выборки — из одного и того же распределения $\mathcal{N}(\theta_1, \theta_2^2)$ с некоторыми (неизвестными) параметрами). Построим КОП для проверки этой гипотезы. Обозначим $\overline{X}_j$, S_j^2 выборочные среднее и дисперсию j-й выборки, $j = 1, \quad, k$, и положим $n = n_1 + \ldots + n_k$ — общее число наблюдений,

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{k} n_j \overline{X}_j$$

— среднее всех данных, S^2 — выборочная дисперсия всех данных:

$$S^2 = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\overline{X}_{ji} - \overline{X})^2$$

наконец,

$$S_0^2 = \frac{1}{n}\sum_{j=1}^{k} n_j S_j^2.$$

Пусть $L_j = L_j(\theta_{1j}, \theta_2)$ обозначает функцию правдоподобия для j-й выборки, т. е. (см. (2) § 3.3)

$$L_j = (2\pi\theta_2^2)^{-n_j/2}\exp\left\{-\frac{n_j}{2\theta_2^2}S_j^2 - \frac{n_j}{2\theta_2^2}(\overline{X}_j - \theta_{1j})^2\right\}, \quad j = 1, \ldots, k. \tag{4}$$

Поскольку выборки независимы, то функция правдоподобия всех данных есть

$$L(\theta) = \prod_{j=1}^{k} L_j(\theta_{1j}, \theta_2) = (2\pi\theta_2^2)^{-n/2}\exp\left\{-\frac{n}{2\theta_2^2}S_0^2 - \frac{1}{2\theta_2^2}\sum_{j=1}^{k} n_j(\overline{X}_j - \theta_{1j})^2\right\} =$$

$$= (2\pi e S_0^2)^{-n/2}\exp\left\{-\frac{1}{2\theta_2^2}\sum_{j=1}^{k} n_j(\overline{X}_j - \theta_{1j})^2 - n\left[\frac{1}{2}\left(\frac{S_0^2}{\theta_2^2} - 1\right) - \ln\frac{S_0}{\theta_2}\right]\right\}. \tag{5}$$

Рассуждая далее как в примере 1 § 3.5, убеждаемся в том, что показатель экспоненты в (5) всегда $\leqslant 0$ и знак «=» $\Longleftrightarrow \theta_{1j} = \overline{X}_j$, $j = 1, \quad , k$, $\theta_2 = S_0$. Таким образом,

$$\sup_{\theta} L(\theta) = L(\overline{X}_1, \quad , \overline{X}_k, S_0) = (2\pi e S_0^2)^{-n/2} \tag{6}$$

При гипотезе же H_0 все данные — это одна выборка объема n из нормального распределения $\mathcal{N}(\theta_1, \theta_2^2)$, поэтому условный супремум $L(\theta)$ совпадает в данном случае с безусловным супремумом примеров 1 и 2, т. е. равен $(2\pi e S^2)^{-n/2}$ Отсюда находим статистику отношения правдоподобия (1) (λ-статистику)

$$\lambda_{n_1 \ldots n_k} = \left(\frac{S^2}{S_0^2}\right)^{-n/2} \quad \left(\frac{nS^2}{\sum_{j=1}^{k} n_j S_j^2}\right)^{-n/2} \tag{7}$$

В общем случае произвольного числа выборок k «работать» с этой статистикой трудно и общее решение (вид КОП) будет получено нами позже в рамках соответствующей асимптотической теории для больших выборок. Но в частном (важном) случае двух выборок ($k = 2$) можно продвинуться дальше и получить явный вид КОП.

Две выборки

Итак, пусть далее $k = 2$. Запишем разложение

$$nS^2 = \sum_{j=1}^{2}\sum_{i=1}^{n_j}(X_{ji} - \overline{X})^2 =$$

$$= \sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1 + \frac{n_2}{n}(\overline{X}_1 - \overline{X}_2)\right)^2 + \sum_{i=1}^{n_2}\left(X_{2i} - \overline{X}_2 + \frac{n_1}{n}(\overline{X}_2 - \overline{X}_1)\right)^2$$

$$= n_1S_1^2 + n_2S_2^2 + \frac{n_1n_2}{n}(\overline{X}_1 - \overline{X}_2)^2$$

Тогда из (7) имеем

$$\lambda_{n_1n_2} = \left(1 + \frac{n_1n_2}{n}\frac{(\overline{X}_1 - \overline{X}_2)^2}{n_1S_1^2 + n_2S_2^2}\right) = \left(1 + \frac{t_{n-2}^2}{n-2}\right)^{-n/2} \tag{8}$$

где

$$t_{n-2} = (\overline{X}_1 - \overline{X}_2)\sqrt{\frac{n_1n_2(n-2)}{n(n_1S_1^2 + n^2S_2^2)}}$$

имеет при гипотезе H_0 распределение Стьюдента $S(n-2)$ (см. пример 2 § 2.5, соотношение (14) при $\Delta = 0$ с учетом принятых здесь обозначений). Из (8) следует, что неравенство $\{\lambda_{n_1n_2} \leqslant c\}$ эквивалентно неравенству $\{|t_{n-2}| \geqslant t\}$,

КОП для двух выборок

а так как $\mathcal{L}(t_{n-2}|H_0) = S(n-2)$, то при уровне значимости α критическая граница $t = t_{1-\alpha/2,\,n-2}$, и мы получаем вид КОП в данном случае:

$$\mathfrak{X}_{1\alpha} = \{|t_{n-2}| \geqslant t_{1-\alpha/2,\,n-2}\}. \tag{9}$$

Но такой критерий (с точностью до обозначений) мы уже получили ранее в примере 10 (соотношение (21)). Таким образом, в случае проверки гипотезы о равенстве средних двух нормальных выборок КОП совпадает с оптимальным критерием. •

Пример 4 (гипотеза о равенстве дисперсий нормальных выборок). Пусть $S_1^2, \quad , S_k^2$ — выборочные дисперсии, построенные по независимым выборкам объемов $n_1, \quad , n_k$ из соответствующих нормальных распределений $\mathcal{N}(\theta_{1j}, \theta_{2j}^2)$, $j = 1, \quad , k$. Построим КОП для проверки гипотезы о равенстве дисперсий $H_0 \quad \theta_{21} = \quad = \theta_{2k}$.

Пусть $L_j(\theta_{1j}, \theta_{2j})$ — функция правдоподобия для j-й выборки, $j = 1, \ldots, k$, а $L = \prod_{j=1}^{k} L_j(\theta_{1j}, \theta_{2j})$ — для всех данных. Как и в предыдущем примере 3 находим, что условный (при гипотезе H_0) супремум L дается формулой (6). Безусловный же (абсолютный) супремум L равен

$$\prod_{j=1}^{k} \sup L_j(\theta_{1j}, \theta_{2j}) = \prod_{j=1}^{k}(2\pi eS_j^2)^{-n_j/2}.$$

Таким образом, в данном случае λ-статистика (1) имеет вид

$$\lambda_{n_1 \ldots n_k} = \frac{1}{(S_0^2)^{n/2}} \prod_{j=1}^{k} (S_j^2)^{n_j/2} = \prod_{j=1}^{k} \left(\frac{S_j^2}{S_0^2} \right)^{n_j/2} \tag{10}$$

Анализ общего случая мы отложим до следующего параграфа, где будут доказаны общие предельные теоремы для больших выборок, а сейчас ограничимся частным случаем двух ($k=2$) выборок. В этом случае (10) можно преобразовать следующим образом

Две выборки

$$\begin{aligned}
\lambda_{n_1 n_2} &= \left(\frac{nS_1^2}{n_1 S_1^2 + n_2 S_2^2} \right)^{n_1/2} \left(\frac{nS_2^2}{n_1 S_1^2 + n_2 S_2^2} \right)^{n_2/2} \\
&= \frac{n^{n/2}}{n_1^{n_1/2} n_2^{n_2/2}} \frac{S_2^n}{\left(S_1^2 + \dfrac{n_2}{n_1} S_2^2 \right)^{n_1/2} \left(\dfrac{n_1}{n_2} S_1^2 + S_2^2 \right)^{n_2/2}} = \\
&= \frac{n^{n/2}}{n_1^{n_1/2} n_2^{n_2/2}} \left(\frac{n_1 S_1^2}{n_2 S_2^2} \right)^{n_1/2} \left(1 + \frac{n_1 S_1^2}{n_2 S_2^2} \right)^{-n/2} \\
&= C\, F_{n_1-1,\, n_2-1}^{n_1/2} \left(1 + \frac{n_1 - 1}{n_2 - 1} F_{n_1-1,\, n_2-1} \right)^{-n/2}
\end{aligned}$$

где

$$F_{n_1-1,\, n_2-1} = \frac{n_1(n_2-1)S_1^2}{n_2(n_1-1)S_2^2}$$

имеет при гипотезе H_0 распределение Снедекора $S(n_1-1, n_2-1)$ (см. п. 7 § 1.2).

При этом (см. рис. 2) неравенство $\{\lambda_{n_1 n_2} \leqslant c\} \Longleftrightarrow \{F_{n_1-1,\, n_2-1} \leqslant c_1$ либо $F_{n_1-1,\, n_2-1} \geqslant c_2\}$, $c_1 < c_2$. Обозначая $F_{n,m}(t)$ функцию распределения закона $S(n, m)$, для определения критических границ c_1 и c_2 при заданном уровне значимости α имеем уравнение

$$\begin{aligned}
\alpha &= \mathbf{P}\{F_{n_1-1,\, n_2-1} \leqslant c_1 \text{ либо } F_{n_1-1,\, n_2-1} \geqslant c_2 | H_0\} = \\
&= F_{n_1-1,\, n_2-1}(c_1) + 1 - F_{n_1-1,\, n_2-1}(c_2).
\end{aligned}$$

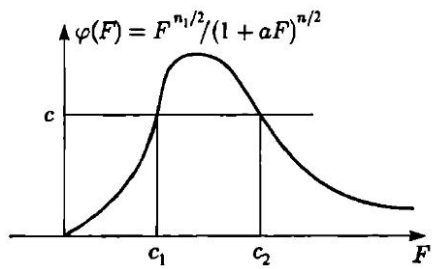


Рис. 2

КОП для двух выборок

Следовательно, если $c_1 = F_{\alpha_1, n_1-1, n_2-1}$ — α_1-квантиль распределения $S(n_1-1, n_2-1)$, а $c_2 = F_{1-\alpha_2, n_1-1, n_2-1}$, то при любых α_1 и α_2 таких, что $\alpha_1 + \alpha_2 = \alpha$, КОП имеет вид

$$\mathfrak{X}_{1\alpha} = \{F_{n_1-1, n_2-1} \leqslant F_{\alpha_1, n_1-1, n_2-1}, \text{ либо } F_{n_1-1, n_2-1} \geqslant F_{1-\alpha_2, n_1-1, n_2-1}\} \quad (11)$$

(сравни с упр. 16, где построен критерий при $\alpha_1 = \alpha_2 = \alpha/2$ методом обращения центрального доверительного интервала для отношения теоретических дисперсий). Чтобы критерий вида (11) был несмещенным, надо вычислить функцию мощности $W(\theta_{21}, \theta_{22})$ и потребовать выполнения условия $W(\theta_{21}, \theta_{22}) \geqslant \alpha$ при $\theta_{21} \neq \theta_{22}$ (при нулевой гипотезе $W(\theta_2, \theta_2) = \alpha$). Для этого заметим, что по теореме Фишера $\mathcal{L}(n_i S_i^2/\theta_{2i}^2) = \chi^2(n_i - 1)$, $i = 1, 2$, и потому отношение Снедекора в общем случае произвольных дисперсий θ_{21} и θ_{22} имеет вид (см. п. 7 § 1.2)

$$F = \frac{n_1 S_1^2}{\theta_{21}^2(n_1-1)} \Big/ \frac{n_2 S_2^2}{\theta_{22}^2(n_2-1)} = \frac{n_1(n_2-1)S_1^2}{n_2(n_1-1)S_2^2}\frac{\theta_{22}^2}{\theta_{21}^2},$$

при этом $\mathcal{L}(F) = S(n_1-1, n_2-1)$. С учетом этого для функции мощности критерия (11) при произвольной альтернативе $\theta = (\theta_{21}, \theta_{22})$: $\theta_{21} \neq \theta_{22}$ имеем представление

$$W(\theta_{21}, \theta_{22}) = \mathbf{P}_\theta\{F_{n_1-1, n_2-1} \leqslant c_1 \text{ либо } F_{n_1-1, n_2-1} \geqslant c_2\} =$$
$$= F_{n_1-1, n_2-1}\left(c_1\frac{\theta_{22}^2}{\theta_{21}^2}\right) + 1 - F_{n_1-1, n_2-1}\left(c_2\frac{\theta_{22}^2}{\theta_{21}^2}\right)$$

Условие несмещенности можно записать как равенство нулю производной правой части по $\tau = \theta_{22}^2/\theta_{21}^2$ при $\tau = 1$ (ниже $f_{n_1-1, n_2-1}(t) = F'_{n_1-1, n_2-1}(t)$):

$$c_1 f_{n_1-1, n_2-1}(c_1) = c_2 f_{n_1-1, n_2-1}(c_2). \quad (12)$$

Этим равенством вместе с условием $\alpha_1 + \alpha_2 = \alpha$ границы c_1 и c_2 (следовательно и α_1, α_2) определяются уже однозначно, тем самым мы получаем несмещенный критерий вида (11), который является оптимальным [16, с. 230]. •

В рассмотренных примерах удалось найти точное распределение статистики (1) (точнее, некоторых взаимно однозначных функций от нее) при нулевой гипотезе, что и позволило рассчитать соответствующие критерии. Такое обстоятельство характерно для нормальной модели. Для других моделей точное распределение λ_n удается найти не всегда, поэтому для малых выборок метод отношения правдоподобия не всегда позволяет получить решение задачи (рассчитать КОП). В то же время для больших выборок при достаточно общих условиях можно рассчитать асимптотический вариант КОП и установить его асимптотическую оптимальность. Далее излагаются основные элементы асимптотической теории критерия отношения правдоподобия для больших выборок.

2. КОП для больших выборок

Будем далее предполагать, что выполняются условия регулярности, обеспечивающие существование, единственность и асимптотическую нормальность о. м. п. $\widehat{\theta}_n = (\widehat{\theta}_{1n}, \quad , \widehat{\theta}_{rn})$ параметра $\theta = (\theta_1, \ldots, \theta_r)$ (см. п. 4 § 3.5). Рассмотрим сначала случай простой нулевой гипотезы (далее пишем L_n вместо L).

Теорема 1. Пусть требуется проверить простую гипотезу $H_0 \quad \theta = \theta_0$, где $\theta_0 = (\theta_{10}, \quad , \theta_{r0})$ — заданная внутренняя точка параметрического множества Θ. Тогда при $n \to \infty$ и при выполнении указанных условий регулярности КОП задается асимптотически критической областью

$$\mathfrak{X}_{1\alpha} = \left\{\underline{x} = (x_1, \quad , x_n) \quad -2 \ln \lambda_n(\underline{x}; \theta_0) \geqslant \chi^2_{1-\alpha, r}\right\}, \tag{13}$$

т. е. при $n \to \infty$

$$\mathbf{P}_{\theta_0}\{X \in \mathfrak{X}_{1\alpha}\} = \mathbf{P}_{\theta_0}\left\{-2 \ln \lambda_n(X; \theta_0) \geqslant \chi^2_{1-\alpha, r}\right\} \longrightarrow \alpha.$$

Доказательство. Достаточно установить, что в условиях теоремы

$$\mathcal{L}_{\theta_0}\left(-2 \ln \lambda_n(X; \theta_0)\right) \to \chi^2(r). \tag{14}$$

Если справедлива гипотеза H_0, то в силу состоятельности о. м. п. $\widehat{\theta}_n$ ее значение при больших n находится вблизи точки θ_0, поэтому для $\ln L_n(\theta_0) = \ln L_n(X; \theta_0)$ можно записать разложение Тейлора относительно точки θ_0:

$$\ln L_n(\theta_0) = \ln L_n(\widehat{\theta}_n) + \frac{1}{2} \sum_{i,j=1}^{r} \frac{\partial^2 \ln L_n(\theta_n^*)}{\partial \theta_i \partial \theta_j} (\widehat{\theta}_{in} - \theta_{i0})(\widehat{\theta}_{jn} - \theta_{j0}),$$

где θ^* — некоторая промежуточная точка между θ_0 и $\widehat{\theta}_n$. Отсюда следует, что

$$-2 \ln \lambda_n = 2\left[\ln L_n(\widehat{\theta}_n) - \ln L_n(\theta_0)\right] = \sum_{i,j=1}^{r} \frac{1}{n} \frac{\partial^2 L_n(\theta_n^*)}{\partial \theta_i \partial \theta_j} \sqrt{n}(\widehat{\theta}_{in} - \theta_{i0}) \sqrt{n}(\widehat{\theta}_{jn} - \theta_{j0}). \tag{15}$$

Так как $\widehat{\theta}_n \xrightarrow{\mathbf{P}} \theta_0$, то и $\theta_n^* \xrightarrow{\mathbf{P}} \theta_0$, а поскольку вторые производные функции правдоподобия, по предположению, непрерывны по θ, то по теореме 1 § 2.2 при больших n величина

$$\frac{1}{n} \frac{\partial^2 \ln L_n(\theta_n^*)}{\partial \theta_i \partial \theta_j}$$

будет мало отличаться от величины

$$\frac{1}{n} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta_i \partial \theta_j} = \frac{1}{n} \sum_{k=1}^{n} \frac{\partial^2 \ln f(X_k, \theta_0)}{\partial \theta_i \partial \theta_j},$$

которая на основании закона больших чисел, в свою очередь, будет мало отличаться от среднего

$$\mathbf{E}_{\theta_0}\left(\frac{\partial^2 \ln f(X_1;\theta_0)}{\partial\theta_i\partial\theta_j}\right).$$

Таким образом, матрица предельных (при $n \to \infty$) значений коэффициентов квадратичной формы в (15) совпадает с информационной матрицей $I(\theta_0)$ (см. (21) § 3.2). Далее, по теореме 2 § 3.5 (соотношение (19)) случайный вектор $\underline{\eta}_n = \sqrt{n}(\widehat{\theta}_n - \theta_0)$ имеет при гипотезе H_0 такое же распределение, как нормальный $\mathcal{N}(\underline{0}, I^{-1}(\theta_0))$ случайный вектор $\underline{\eta}$. Таким образом, правая часть (15) имеет в пределе такое же распределение, как и квадратичная форма $Q = \underline{\eta}' I(\theta_0)\underline{\eta}$. Но, по теореме 1 § 2.5 (см. замечание к ней), $\mathcal{L}(Q) = \chi^2(r)$, что и доказывает (14). ■

Замечание. Предельные распределения статистик $-2\ln\lambda_n$ и $Q_n^{(1)} = \underline{\eta}_n' I(\widehat{\theta}_n)\underline{\eta}_n$ при нулевой гипотезе совпадают (следует из хода предыдущих рассуждений), поэтому КОП (13) асимптотически эквивалентен критерию вида

$$\mathfrak{X}_{1\alpha}^{(1)} = \{Q_n^{(1)} \geqslant \chi^2_{1-\alpha,r}\}. \tag{16}$$

Если здесь матрицу $I(\widehat{\theta}_0)$ заменить на близкую к ней при больших n матрицу $I(\theta_0)$, то получим еще один критерий, асимптотически эквивалентный (13):

$$\mathfrak{X}_{1\alpha}^{(0)} = \{n(\widehat{\theta}_n - \theta_0)' I(\theta_0)(\widehat{\theta}_n - \theta_0) \geqslant \chi^2_{1-\alpha,r}\}. \tag{17}$$

Замечание. Из доказательства асимптотической нормальности о. м. п. следует (см. соотношение (26) § 3.5), что в случае одномерного параметра предельные распределения $\eta_n = \sqrt{n}(\widehat{\theta}_n - \theta_0)$ и $\dfrac{V_n(\theta_0)}{\sqrt{n}i(\theta_0)}$ при гипотезе H_0 совпадают. В случае векторного параметра этот вывод остается в силе, если вторую статистику заменить на $I^{-1}(\theta_0)\dfrac{\underline{V}_n(\theta_0)}{\sqrt{n}}$, где (см. п. 5 § 3.2) $\underline{V}_n(\theta) = (V_{1n}(\theta), \quad , V_{rn}(\theta))$ и $V_{in}(\theta) = \dfrac{\partial \ln L_n(\theta)}{\partial\theta_i}$, $i = 1, \quad , r$. Отсюда и из предыдущего замечания следует, что при гипотезе H_0 статистики $Q_n^{(1)}$ и $Q_n^{(2)} = \dfrac{1}{n}\underline{V}_n'(\theta_0)I^{-1}(\theta_0)\underline{V}_n(\theta_0)$ имеют одно и то же предельное распределение $\chi^2(r)$, поэтому КОП асимптотически эквивалентен также критерию вида

$$\mathfrak{X}_{1\alpha}^{(2)} = \{Q_n^{(2)} \geqslant \chi^2_{1-\alpha,r}\}. \tag{18}$$

Достоинством этого варианта критерия является то, что при его использовании не требуется вычислять о. м. п. $\widehat{\theta}_n$ и его, следовательно, целесообразно применять в тех случаях, когда явный вид $\widehat{\theta}_n$ получить невозможно.

Итак, в задаче проверки простой гипотезы H_0 $\theta = \theta_0$ против сложной альтернативы H_1 $\theta \in \Theta\backslash\{\theta_0\}$ все четыре критерия (13), (16), (17) и (18) асимптотически эквивалентны, т. е. уровень значимости каждого из них стремится к α при $n \to \infty$.

Рассмотрим важный для приложений пример применения изложенных результатов к полиномиальному распределению $M(n; \underline{p} = (p_1, \quad , p_N))$.

Пример 5 (Метод отношения правдоподобия для полиномиального распределения). Все необходимые вычисления для данного распределения проведены ранее в примерах 10 § 3.2 и 12 § 3.5. Используя их, находим, что λ-статистика (1) для проверки простой гипотезы $H_0 \quad \underline{p} = \underline{p}^\circ$ имеет вид

$$\lambda_n(\underline{p}^\circ) = \frac{\prod_{j=1}^{N} p_j^{0\nu_j}}{\prod_{j=1}^{N}\left(\frac{\nu_j}{n}\right)^{\nu_j}} = \prod_{j=1}^{N}\left(\frac{np_j^\circ}{\nu_j}\right)^{\nu_j}$$

Отсюда

$$-2\ln\lambda_n(\underline{p}^\circ) = 2\sum_{j=1}^{N}\nu_j \ln\frac{\nu_j}{np_j^\circ}.$$

Заметим, что в данном случае вектор

$$\underline{\eta}_n = \sqrt{n}(\widehat{\theta}_n - \theta_0) = \sqrt{n}\left(\frac{\nu_1}{n} - p_1^\circ, \qquad \frac{\nu_{N-1}}{n} - p_{N-1}^\circ\right) =$$

$$= \underline{\nu}^* = \left(\frac{\nu_j - np_j^\circ}{\sqrt{n}}, \ j = 1, \quad , N-1\right),$$

а $I(\theta_0) = \mathring{\Sigma}_{N-1}^{-1}$ (эти обозначения использовались в доказательстве теоремы 1 § 4.2), поэтому квадратичная форма

$$Q_n = \underline{\eta}_n' I(\theta_0)\underline{\eta}_n = \underline{\nu}^{*\prime}\mathring{\Sigma}_{N-1}^{-1}\underline{\nu}^* = \mathring{X}_n^2$$

(см. (4) § 4.2). Далее для квадратичной формы $Q_n^{(1)}$ в (16) имеем представление (см. (29) § 3.2)

$$Q_n^{(1)} = n\left[\sum_{j=1}^{N-1}\left(\frac{n}{\nu_j} + \frac{n}{\nu_N}\right)\left(\frac{\nu_j}{n} - p_j^\circ\right)^2 + \sum_{\substack{i,j=1\\ i\neq j}}^{N-1}\frac{n}{\nu_N}\left(\frac{\nu_i}{n} - p_i^\circ\right)\left(\frac{\nu_j}{n} - p_j^\circ\right)\right] =$$

$$= \sum_{j=1}^{N}\frac{(\nu_j - np_j^\circ)^2}{\nu_j}.$$

Наконец (см. (28) § 3.2)

$$V_{jn}(\theta) = \sum_{i=1}^{n}\frac{\partial \ln f(X_i;\theta)}{\partial p_j} = \sum_{i=1}^{n}\left(\frac{I(X_i = j)}{p_j} - \frac{I(X_i = N)}{p_N}\right) = \frac{\nu_j}{p_j} - \frac{\nu_N}{p_N},$$

$$j = 1, \ldots, N-1,$$

и потому

$$Q_n^{(2)} = \frac{1}{n}\underline{V}_n'(\theta_0)\mathring{\Sigma}_{N-1}\underline{V}_n(\theta_0) =$$

$$= \frac{1}{n}\left[\sum_{j=1}^{N-1}\left(\frac{\nu_j}{p_j^\circ} - \frac{\nu_N}{p_N^\circ}\right)^2 p_j^\circ(1-p_j^\circ) - \sum_{\substack{i,j=1\\ i\neq j}}^{N-1}\left(\frac{\nu_i}{p_i^\circ} - \frac{\nu_N}{p_N^\circ}\right)\left(\frac{\nu_j}{p_j^\circ} - \frac{\nu_N}{p_N^\circ}\right)p_i^\circ p_j^\circ\right] =$$

$$= \frac{1}{n}\left[\sum_{j=1}^{N-1}\left(\frac{\nu_j - np_j^\circ}{p_j^\circ} - \frac{\nu_N - np_N^\circ}{p_N^\circ}\right)^2 p_j^\circ - \left(\sum_{j=1}^{N-1}\left(\frac{\nu_j}{p_j^\circ} - \frac{\nu_N}{p_N^\circ}\right)p_j^\circ\right)^2\right]$$

Далее простыми преобразованиями с учетом равенств

$$\sum_{j=1}^{N-1}\nu_j = n - \nu_N, \quad \sum_{j=1}^{N-1}p_j^\circ = 1 - p_N^\circ$$

приходим к представлению $Q_n^{(2)} = \mathring{X}_n^2$.

Таким образом, в рассматриваемой задаче критерий можно строить, используя любую из тестовых статистик

$$-2\ln\lambda_n(\underline{p}^\circ) = 2\sum_{j=1}^{N}\nu_j \ln\frac{\nu_j}{np_j^\circ},$$

$$Q_n = Q_n^{(2)} = \mathring{X}_n^2 = \sum_{j=1}^{N}\frac{(\nu_j - np_j^\circ)^2}{np_j^\circ}, \tag{19}$$

$$Q_n^{(1)} = \sum_{j=1}^{N}\frac{(\nu_j - np_j^\circ)^2}{\nu_j}.$$

Если справедлива гипотеза H_0 $\underline{p} = \underline{p}^\circ$, то в пределе при $n \to \infty$ все эти статистики имеют одно и то же распределения $\chi^2(N-1)$, поэтому при заданном уровне значимости α критическую границу выбирают равной $\chi^2_{1-\alpha,\,N-1}$. Статистика $\mathring{X}_n^2$ (статистика хи-квадрат К. Пирсона) нам уже встречалась ранее в п. 2 § 4.2, где она была введена из других соображений. В § 4.2 (теорема 1) было получено то же самое предельное распределение статистики $\mathring{X}_n^2$ при нулевой гипотезе и построен асимптотический вариант критерия согласия χ^2 (см. (6) § 4.2). Здесь эти же результаты получены как простое следствие метода отношения правдоподобия применительно к полиномиальному распределению. Таким образом, для полиномиального распределения оба метода: метод хи-квадрат и метод отношения правдоподобия, — асимптотически эквивалентны. С позиций практических применений классическая статистика $\mathring{X}_n^2$ предпочтительнее (проще вычисляется) статистики $-2\ln\lambda_n(\underline{p}^\circ)$, но по сравнению со статистикой $Q_n^{(1)}$ у нее преимуществ нет. •

3. Дальнейшие асимптотические свойства КОП

В этом пункте мы рассмотрим поведение КОП при альтернативах в случае больших выборок. Прежде всего мы покажем, что этот критерий состоятелен, т. е. для любой фиксированной альтернативы $\theta \neq \theta_0$ его мощность

$$W_n(\theta) = \mathbf{P}_\theta\{-2\ln\lambda_n(X;\theta_0) \geqslant \chi^2_{1-\alpha,r}\} \longrightarrow 1 \quad \text{при} \quad n \to \infty. \tag{20}$$

Доказательство. Для простоты рассмотрим случай скалярного параметра $(r = 1)$.

Пусть $\theta_1 \neq \theta_0$ — произвольная фиксированная внутренняя точка Θ. Запишем статистику $\lambda_n(X;\theta_0) = \lambda_n(\theta_0)$ в виде

$$\lambda_n(\theta_0) = \frac{L_n(\theta_1)}{L_n(\widehat{\theta}_n)}\,\frac{L_n(\theta_0)}{L_n(\theta_1)} \equiv \lambda_n(\theta_1)\nu_n(\theta_0,\theta_1).$$

Отсюда

$$-2\ln\lambda_n(\theta_0) = -2\ln\lambda_n(\theta_1) + n v_n(\theta_0,\theta_1), \tag{21}$$

где

$$v_n(\theta_0,\theta_1) = -\frac{2}{n}\ln\nu_n(\theta_0,\theta_1) = \frac{2}{n}\Big[\ln L_n(\theta_1) - \ln L_n(\theta_0)\Big] =$$

$$= 2\Big[\frac{1}{n}\sum_{i=1}^{n}\ln f(X_i;\theta_1) - \frac{1}{n}\sum_{i=1}^{n}\ln f(X_i;\theta_0)\Big].$$

Введем функцию $H(\theta) = \mathbf{E}_{\theta_1}\ln f(X_1;\theta)$. На основании закона больших чисел, при гипотезе $\theta = \theta_1$ и $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n}\ln f(X_i;\theta_k) \xrightarrow{\mathbf{P}_{\theta_1}} H(\theta_k), \quad k = 0, 1.$$

Далее имеем

$$H^{(j)}(\theta) = \mathbf{E}_{\theta_1}\left(\frac{\partial^j \ln f(X_1;\theta)}{\partial\theta^j}\right), \quad j = 1, 2.$$

Отсюда (см. (3) и (5′) § 3.2) $H'(\theta_1) = 0$, $H''(\theta_1) = -i(\theta_1) < 0$. Это означает, что в точке θ_1 функция $H(\theta)$ имеет максимум, поэтому случайная величина $v_n(\theta_0,\theta_1)$ сходится по вероятности к положительному числу $2[H(\theta_1) - H(\theta_0)]$. По теореме 1, $\mathcal{L}_{\theta_1}\big(-2\ln\lambda_n(\theta_1)\big) \longrightarrow \chi^2(1)$, следовательно, из (21) имеем, что при альтернативе θ_1 статистика $-2\ln\lambda_n(\theta_0)$ по вероятности неограниченно возрастает при $n \to \infty$. Отсюда следует соотношение (20) при $\theta = \theta_1$. ■

Таким образом, для КОП (13) функция мощности $W_n(\theta)$ сходится при $n \to \infty$ к функции

$$\zeta(\theta) = \begin{cases} \alpha & \text{при} \quad \theta = \theta_0, \\ 1 & \text{при} \quad \theta \neq \theta_0. \end{cases}$$

«Близкие» альтернативы

Можно также исследовать поведение мощности КОП и при «близких» альтернативах, т. е. когда альтернатива $\theta_1 = \theta_1(n) \to \theta_0$ при $n \to \infty$. Приведем (без доказательства) результат, дающий вид предельного распределения тестовой статистики в (13) при «пороговых» альтернативах.

Теорема 2. Пусть близкая альтернатива имеет вид

$$\theta_1(n) = \theta_0 + \frac{\underline{\delta}}{\sqrt{n}},$$

где $\underline{\delta} = (\delta_1, \ldots, \delta_r)$ — фиксированный ненулевой вектор, задающий отклонение альтернативы от нулевой гипотезы. Тогда в условиях теоремы 1 при $n \to \infty$

$$\mathcal{L}_{\theta_1(n)}\left(-2\ln \lambda_n(\boldsymbol{X}; \theta_0)\right) \longrightarrow \chi^2(r; \lambda^2), \quad \lambda^2 = \underline{\delta}' I(\theta_0)\underline{\delta} \tag{22}$$

(определение нецентрального распределения хи-квадрат см. в п. 2 § 2.5).

Из этой теоремы следует, что для близких альтернатив вида

$$\theta_1(n) = \theta_0 + \frac{\underline{\delta}}{\sqrt{n}}$$

мощность КОП удовлетворяет при $n \to \infty$ соотношению

$$W_n\left(\theta_1(n)\right) = \mathrm{P}_{\theta_1(n)}\left\{-2\ln \lambda_n(X; \theta_0) \geqslant \chi^2_{1-\alpha, r}\right\} \longrightarrow$$
$$\longrightarrow 1 - F_r\left(\chi^2_{1-\alpha, r}; \lambda^2\right), \tag{23}$$

где $F_r\left(t; \lambda^2\right)$ — функция распределения закона $\chi^2(r; \lambda^2)$.

Для более близких альтернатив (случай $\lambda^2 \to 0$ или $\sqrt{n}(\theta_1(n) - \theta_0) \to \underline{0}$) мощность стремится при $n \to \infty$ к уровню значимости α, т. е. такие альтернативы по КОП асимптотически от нулевой гипотезы не отличаются. Более же далекие альтернативы (случай $\lambda^2 \to \infty$ или $\sqrt{n}(\theta_1(n) - \theta_0) \to \infty$) критерий улавливает с вероятностью, стремящейся к 1 при $n \to \infty$, когда они верны, поскольку для таких альтернатив мощность стремится к 1. Эти выводы справедливы и для критериев (16)–(18).

Пример 5 (продолжение). Для близких альтернатив

$$\underline{p}(n) = \underline{p}^{\circ} + \frac{\underline{\beta}}{\sqrt{n}}, \quad \sum_{j=1}^{N} \beta_j = 0,$$

мощность любого из критериев, основанных на тестовых статистиках, указанных в (19), равна в пределе (при $n \to \infty$) указанной в соотношении (10) § 4.2 величине. Это утверждение является специализацией общего результата (23) для полиномиальной модели. •

4. Сложная нулевая гипотеза

Все важные асимптотические свойства КОП сохраняются и для случая сложной нулевой гипотезы. Мы ограничимся лишь формулировками соответствующих утверждений, отсылая интересующихся деталями доказательств к оригинальной работе на эту тему[2)]

Пусть требуется проверить сложную гипотезу $H_0 \quad \theta \in \Theta_0 \subset \Theta$, т. е. подмножество Θ_0 не сводится к одной точке. Альтернативная гипотеза $H_1 \quad \theta \in \Theta \backslash \Theta_0$ также сложная. Рассмотрим типичную ситуацию, когда Θ_0 — евклидово подпространство размерности меньшей, чем размерность всего параметрического множества Θ (напомним, что Θ — либо все r-мерное евклидово пространство R^r при некотором $r \geqslant 2$, либо его подмножество той же размерности r). Пусть $s = \dim \Theta_0$ (от dimension (англ.) — размерность), так что $s < r = \dim \Theta$; тогда значение параметра $\theta \in \Theta_0$ можно выразить в виде функции некоторого нового параметра $\delta = (\delta_1, \ldots, \delta_s)$ с соответствующим множеством возможных значений Δ, т. е. Θ_0 имеет параметрическое представление в терминах параметра δ:

$$\Theta_0 = \{\theta = (\theta_1, \quad, \theta_r) \quad \theta = h(\delta),\ \delta = (\delta_1, \ \ldots, \delta_s) \in \Delta\}. \tag{24}$$

В общем случае будем предполагать, что функции $h(\delta) = \big(h_1(\delta), \quad, h_r(\delta)\big)$ удовлетворяют следующим стандартным условиям регулярности: они трижды непрерывно дифференцируемы и ранг матрицы их производных $\|\partial h_i(\delta)/\partial \delta_j\|$ равен s всюду в Δ.

Опишем также другой, эквивалентный, способ задания нулевой гипотезы. Обозначим через $\theta^{(2)}$ совокупность тех координат вектора θ, которые являются взаимно однозначными функциями от δ. Пусть для простоты $\theta^{(2)} = (\theta_{r-s+1}, \quad, \theta_r)$. Тогда δ можно выразить через $\theta^{(2)} \quad \delta = \delta(\theta^{(2)})$. Если теперь ввести функции $H_j(\theta) = \theta_j - h_j\big(\delta(\theta^{(2)})\big)$, $j = 1, \ldots, r-s$, то гипотеза $H_0 \quad \theta \in \Theta_0$ примет вид

$$H_0 \quad H_j(\theta) = 0, \quad j = 1, \quad, r - s. \tag{25}$$

Тем самым гипотезу H_0 можно задать указанием некоторого числа связей (ограничений) на координаты параметрического вектора $\theta = (\theta_1, \ldots, \theta_r)$.

В приложениях часто встречается случай, когда Θ_0 задается как *цилиндрическое подмножество* Θ, например,

$$\Theta_0 = \{\theta \quad \theta = (\theta_{10}, \ . \ , \theta_{r-s,0}, \theta_{r-} \qquad, \theta_r)\}, \tag{26}$$

где θ_{i0}, $i = 1, \quad, r - s$, — некоторые фиксированные значения координат $\theta_1, \quad, \theta_{r-s}$, а остальные координаты $\theta_{r-s+1}, \ldots, \theta_r$ (их число есть s) являются свободными параметрами (их принято называть «мешающими» параметрами). Если Θ_0 имеет вид (26), то гипотезу $H_0 \quad \theta \in \Theta_0$ называют *линейной*.

[2)] *Davidson R., Lever W.* The limiting distribution of the likelihood ratio statistic under a class of local alternatives // Sankhya, A. 1970. V. 32. № 2. P. 209–224.

Пусть проверяемая гипотеза задана в форме (24). Если функцию правдоподобия данных (выборки X) рассматривать при этой гипотезе, то она является функцией нового параметра δ: $L_n(X;\theta) = L_n(X;h(\delta))$ и при выполнении всех сформулированных условий регулярности о. м. п. $\widehat{\delta}_n$ параметра δ будет обладать стандартными асимптотическими свойствами (см. теорему 2 § 3.5), а λ-статистика (1) может быть записана в виде

$$\lambda_n(X;\Theta_0) = \frac{L_n\big(X;h(\widehat{\delta}_n)\big)}{L_n(X;\widehat{\theta}_n)}.$$

Для больших выборок (при $n \to \infty$) справедливы следующие результаты:

1) $$\mathcal{L}\big(-2\ln\lambda_n(X;\Theta_0)|H_0\big) \longrightarrow \chi^2(r-s); \tag{27}$$

2) КОП, основанный на критической области

$$\mathfrak{X}_{1\alpha} = \big\{\underline{x} \quad -2\ln\lambda_n(\underline{x};\Theta_0) \geqslant \chi^2_{1-\alpha,r-s}\big\}, \tag{28}$$

состоятелен.

Применим эти результаты к анализу λ-статистик (7) и (10), выведенных нами ранее в задачах для многих нормальных выборок.

Пример 6 (продолжение примера 3). В примере 3, очевидно, $\dim\Theta = k+1$, $\dim\Theta_0 = 2$, а $-2\ln\lambda_{n_1\ldots n_k} = n(\ln S^2 \quad \ln S_0^2)$, поэтому асимптотический вариант КОП (28) имеет вид

$$\mathfrak{X}_{1\alpha} = \big\{n(\ln S^2 - \ln S_0^2) \geqslant \chi^2_{1-\alpha,k-1}\big\}. \tag{29}$$

•

Пример 7 (продолжение примера 4). Здесь общее число неизвестных параметров равно $2k = \dim\Theta$, а число параметров, определяющих гипотезу H_0, равно $k+1 = \dim\Theta_0$, т. е. в (28) $r-s = 2k-(k+1) = k-1$; тестовая же статистика в силу (10) есть

$$-2\ln\lambda_{n_1\ldots n_k} \quad \sum_{j=1}^{k} n_j(\ln S_0^2 - \ln S_j^2).$$

В итоге критерий (28) принимает в данной задаче вид

$$\mathfrak{X}_{1\alpha} = \left\{\sum_{j=1}^{k} n_j(\ln S_0^2 - \ln S_j^2) \geqslant \chi^2_{1-\alpha,k-1}\right\}. \tag{30}$$

•

Пример 8 (Гипотеза однородности для схемы Бернулли). Пусть $\overline{X}_1, \quad, \overline{X}_k$ — выборочные средние для независимых выборок объемов $n_1, \quad, n_k$ из совокупностей $\mathrm{Bi}(1,\theta_1), \quad, \mathrm{Bi}(1,\theta_k)$ соответственно. Построим асимптотический (при $n_1, \ldots, n_k \to \infty$) вариант КОП для гипотезы однородности $H_0 : \theta_1 = \ldots = \theta_k$.

Обозначим через $L_j(\theta_j) = \theta_j^{n_j\overline{X}_j}(1-\theta_j)^{n_j(1-\overline{X}_j)}$ функцию правдоподобия для j-й выборки, $j = 1, \ldots, k$. Тогда, в силу независимости выборок, функция правдоподобия для всех данных есть $L(\theta_1, \ldots, \theta_k) \quad L_1(\theta_1) \quad L_k(\theta_k)$. Далее, так как *о. м. п.* параметра бернуллиевской модели совпадает со средним арифметическим выборки (см. п. 1 § 3.5), то отсюда имеем

$$\max_{\theta_1,\ldots,\theta_k} L(\theta_1, \ldots, \theta_k) = \prod_{j=1}^{k} \max_{\theta_j} L_j(\theta_j) = \prod_{j=1}^{k} L_j(\overline{X}_j) = \prod_{j=1}^{k} \overline{X}_j^{n_j\overline{X}_j}(1-\overline{X}_j)^{n_j(1-\overline{X}_j)},$$

а при гипотезе H_0

$$\max_{\theta} L(\theta, \ldots, \theta) = \max_{\theta} \theta^{n\overline{X}}(1-\theta)^{n(1-\overline{X})} = \overline{X}^{n\overline{X}}(1-\overline{X})^{n(1-\overline{X})},$$

где $n = n_1 + \quad + n_k$, $\overline{X} = (n_1\overline{X}_1 + \quad + n_k\overline{X}_k)/n$ — выборочное среднее всех данных. Отсюда имеем, что в данном случае статистика отношения правдоподобия имеет вид

$$\lambda_{n_1,\ldots,n_k} = \frac{\overline{X}^{n\overline{X}}(1-\overline{X})^{n(1-\overline{X})}}{\prod_{j=1}^{k} \overline{X}_j^{n_j\overline{X}_j}(1-\overline{X}_j)^{n_j(1-\overline{X}_j)}} = \prod_{j=1}^{k}\left(\frac{\overline{X}}{\overline{X}_j}\right)^{n_j\overline{X}_j}\left(\frac{1-\overline{X}}{1-\overline{X}_j}\right)^{n_j(1-\overline{X}_j)}$$

Наконец, здесь $\dim\Theta = k$, $\dim\Theta_0 = 1$, следовательно, критерий (28) принимает вид

$$\mathfrak{X}_{1\alpha} = \left\{ 2\sum_{j=1}^{k} n_j\left[\overline{X}_j(\ln\overline{X}_j - \ln\overline{X}) + \right.\right.$$

$$\left.\left. + (1-\overline{X}_j)\left(\ln(1-\overline{X}_j) - \ln(1-\overline{X})\right)\right] \geqslant \chi^2_{1-\alpha,k-1} \right\}. \qquad (31)$$

Напомним, что эту задачу мы уже решали в п. 2 § 4.3, и полученный там критерий однородности χ^2, основанный на статистике (6), записывается в наших обозначениях в виде

$$\mathfrak{X}_{1\alpha} = \left\{ \widehat{X}^2_{n_1,\ldots,n_k} = \frac{1}{\overline{X}(1-\overline{X})}\sum_{j=1}^{k} n_j(\overline{X}_j - \overline{X})^2 \geqslant \chi^2_{1-\alpha,k-1} \right\}. \quad \bullet$$

Пример 9 (Гипотеза однородности для пуассоновских выборок). Пусть $X_j = (X_{j1}, \quad, X_{jn_j})$, $j = 1, \ldots, k$, — независимые выборки из совокупностей $\Pi(\theta_1), \quad, \Pi(\theta_k)$ соответственно. Требуется по этим данным проверить гипотезу однородности $H_0 \quad \theta_1 = \quad = \theta_k$. Схема решения здесь такая же, как в предыдущем примере. Используя те же обозначения и учитывая, что в данном случае

$$L_j(\theta_j) = \frac{e^{-n_j\theta_j}\theta_j^{n_j\overline{X}_j}}{\prod_{i=1}^{n_j} X_{ji}!}, \quad j = 1, \quad, k,$$

найдем, что

$$\lambda_{n_1 \ldots n_k} = \frac{\overline{X}^{n\overline{X}}}{\prod\limits_{j=1}^{k} \overline{X}_j^{n_j \overline{X}_j}} = \prod_{j=1}^{k} \left(\frac{\overline{X}}{\overline{X}_j} \right)^{n_j \overline{X}_j}$$

(напомним (см. п. 1 § 3.5), что здесь о. м. п. $\widehat{\theta}_j = \overline{X}_j$, а при гипотезе H_0 о. м. п. $\widehat{\theta} = \overline{X}$). Если $n_1, \quad , n_k \to \infty$, то асимптотический вариант КОП (28) принимает в данном случае вид

$$\mathfrak{X}_{1\alpha} = \left\{ 2 \sum_{j=1}^{k} n_j \overline{X}_j (\ln \overline{X}_j - \ln \overline{X}) \geqslant \chi^2_{1-\alpha, k-1} \right\}. \tag{32}$$

•

5. Доверительные области максимального правдоподобия

Изложенные результаты дают возможность получить общее решение задачи доверительного оценивания параметров распределений для общих параметрических моделей. При этом используют общую методику, изложенную в п. 4 § 5.3. Далее будем предполагать, что выполняются стандартные условия регулярности, формулируемые в асимптотической теории КОП.

Рассмотрим КОП уровня значимости $\alpha = 1-\gamma$ для произвольной простой гипотезы, фиксирующей значение параметра $\theta = (\theta_1, \quad , \theta_r)$. Согласно (13), в случае больших выборок область принятия этой гипотезы имеет вид

$$\mathfrak{X}_{0\gamma} = \left\{ \underline{x} \quad -2 \ln \lambda_n(\underline{x}; \theta) < \chi^2_{\gamma, r} \right\}$$

Отсюда имеем, что

$$\mathcal{G}_\gamma(X) = \left\{ \theta \quad -2 \ln \lambda_n(X; \theta) < \chi^2_{\gamma, r} \right\} \tag{33}$$

есть асимптотическая γ-доверительная область для параметра θ. Определенное таким образом случайное подмножество параметрического множества Θ называется (асимптотической) *доверительной областью максимального правдоподобия*. Таким образом, множество (33) строится на основании статистики отношения правдоподобия

$$\lambda_n(X; \theta) = \frac{L_n(X; \theta)}{L_n(X, \widehat{\theta}_n)}$$

и включает в себя все значения параметра, при которых функция правдоподобия $L_n(X; \theta)$ близка к своему максимальному значению $L_n(X; \widehat{\theta}_n)$, или, как говорят, все *достаточно правдоподобные* значения параметра. В частности, наиболее правдоподобное значение параметра — о. м. п. $\widehat{\theta}_n$ — всегда принадлежит $\mathcal{G}_\gamma(X)$ (см. рис. 3).

Приближенную доверительную область максимального правдоподобия можно построить и для произвольной совокупности координат параметрического вектора θ при наличии мешающих параметров. Пусть, например, требуется оценить координаты $\theta^{(1)} = (\theta_1, \ldots, \theta_{r-s})$ (остальные координаты

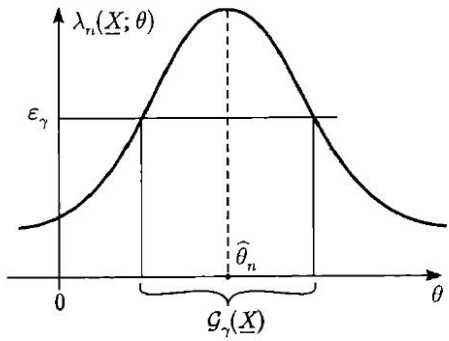


Рис. 3

$\theta^{(2)} = (\theta_{r-s+1}, \ldots, \theta_r)$) являются мешающими параметрами). В этом случае следует рассмотреть линейную гипотезу относительно $\theta^{(1)}$ (см. предыдущий пункт), вычислить статистику

$$\lambda_n(X;\theta^{(1)}) = \frac{\sup\limits_{\theta^{(2)}} L_n\big(X;(\theta^{(1)},\theta^{(2)})\big)}{L_n(X;\widehat{\theta}_n)}$$

и на основании (28) положить

$$\mathcal{G}_\gamma(X) = \big\{\theta^{(1)} \quad -2\ln\lambda_n(X;\theta^{(1)}) < \chi^2_{\gamma,r-s}\big\}.$$

Это и есть искомая асимптотическая γ-доверительная область для $\theta^{(1)}$ так как из (27)–(28) следует, что при $n \to \infty$

$$\mathbf{P}_\theta\big\{\theta^{(1)} \in \mathcal{G}_\gamma(X)\big\} = \mathbf{P}_\theta\big\{-2\ln\lambda_n(X;\theta^{(1)}) < \chi^2_{\gamma,r-s}\big\} \longrightarrow \gamma, \quad \forall\,\theta \in \Theta.$$

§ 5.5. Проверка гипотез для конечных цепей Маркова

Изложенные выше подходы к проверке гипотез для ситуаций, когда исходные статистические данные (выборка) представляют собой результаты последовательных независимых наблюдений над некоторой случайной величиной, можно применять и для моделей с зависимыми наблюдениями. Одной такой весьма общей и важной для приложений моделью является цепь Маркова некоторого порядка $s \geqslant 1$, определение которой дано в п. 2 § 4.5. В данном параграфе на примере этой модели мы обсудим специфические проблемы проверки гипотез для зависимых испытаний и приведем некоторые общие результаты для соответствующих критериев типа хи-квадрат и отношения правдоподобия.

Полное освещение этой темы имеется в [9].

1. Гипотезы для конечных цепей Маркова

Пусть задана последовательность из n случайных величин (выборка объема n)

$$\xi_1, \xi_2, \ldots, \xi_n, \tag{1}$$

принимающих значения из конечного множества (алфавита) $\mathcal{E} = \{1, 2, \ldots, N\}$, и пусть априори предполагается, что эти случайные величины связаны в однородную цепь Маркова порядка $s \geqslant 1$ (это предположение будем далее называть общей гипотезой $\mathcal{H}_s$). Точные определения всех этих понятий дано в п. 2 § 4.5 после теоремы 1. Вероятностный закон, управляющий последовательностью (1), задается переходными вероятностями (см. (8) § 4.5)

$$p_{\bar{i}:j} = \mathbf{P}\{\xi_k = j \mid \xi_{k-s} = i_1, \quad \xi_{k-1} = i_s\}, \quad \bar{i} = (i_1, \quad, i_s), \tag{2}$$

и начальным распределением $\mathbf{P}\{\xi_k = i_k,\ k \leqslant s\}$ (которое, однако, в излагаемой ниже асимптотической (при $n \to \infty$) теории не играет никакой роли, и потому мы на нем не будем акцентировать внимание, понимая под гипотезами утверждения лишь о переходных вероятностях (2)). Тем самым, можно сказать, что модель (гипотеза) $\mathcal{H}_s$ определяется значениями N^{s+1} величин $\{p_{\bar{i}:j}\}$ Различные утверждения, фиксирующие вид или значения переходных вероятностей (2), называются гипотезами о цепи Маркова.

Пусть, например, $\{p^\circ_{\bar{i}:j}\}$ — некоторые заданные значения величин $p_{\bar{i}:j}$. Тогда утверждение $H^0 \quad p_{\bar{i}:j} = p^\circ_{\bar{i}:j}$, $i_1, \quad, i_s$, $j \in \mathcal{E}$, есть простая гипотеза о цепи Маркова, однозначно фиксирующая переходные вероятности цепи (2). В свою очередь, в зависимости от вида этих вероятностей $p^\circ_{\bar{i}:j}$ гипотеза H^0 может означать и цепь порядка s (общий случай) и более частные случаи, состоящие в том, что цепь имеет меньший порядок $r < s$ (если $p^\circ_{i_1,\ldots,i_s:j} = p^\circ_{i_{s-r+1},\ldots,i_s:j}$, т. е. зависимость «на глубину» $s-(s-r)=r$) или даже образует последовательность независимых испытаний (если $p^\circ_{\bar{i}:j} = p^\circ_j$, т. е. нет зависимости от прошлого); в последнем случае при $p^\circ_j = 1/N$, $j = 1, \quad, N$, гипотеза H^0 означает схему независимых испытаний с равновероятными исходами (именно такую гипотезу мы рассматривали в п. 2 § 4.5 в связи с проблемой датчиков).

Часто проверяемая гипотеза H^0 определяется не полным заданием переходных вероятностей (2), а указанием лишь их функциональной зависимости от некоторого параметра $\theta = (\theta_1, \quad, \theta_k)$ с соответствующей областью возможных его значений Θ, являющейся открытым подмножеством k-мерного евклидова пространства R^k:

$$H^0 = H^0(\Theta) \quad p_{\bar{i}:j} = p_{\bar{i}:j}(\Theta), \quad \theta \in \Theta. \tag{3}$$

В этом случае мы имеем дело с проверкой сложной гипотезы $H^0(\Theta)$ внутри более общей сложной гипотезы $\mathcal{H}_s$. В частности, проверяемая гипотеза может заключаться в том, что последовательность (1) является цепью Маркова

порядка, меньшего s, но без указания точных значений переходных вероятностей, Например, пусть

$$p_{i_1,\dots,i_s;j}(\theta) = \theta_{i_{s-r+1},\dots} \tag{4}$$

где $\theta_\bullet > 0$ таковы, что

$$\sum_{j=1}^{N} \theta_{i_{s-r+1},\dots,i_s;j} = 1, \quad i_{s-r+1}, \dots, i_s \in \mathcal{E},$$

т. е. сами переходные вероятности являются неизвестными параметрами и при этом зависимость от прошлого имеет «глубину» r (здесь параметр $k = N^r(N-1)$). В этом случае проверяемая гипотеза состоит в том, что цепь имеет порядок $r < s$. Если $r = 0$, то имеем гипотезу о независимости испытаний при конкурирующей гипотезе типа марковской зависимости порядка s.

Сделаем теперь следующее общее замечание. В п. 2 § 4.5 (соотношение (5)) были введены важные эмпирические характеристики выборки (1) — частоты $\nu_{\bar{i}j}$ появления комбинаций $(\bar{i}j) = (i_1 \quad i_s j)$, называемых $(s+1)$-цепочками, в последовательности (1):

$$\nu_{\bar{i}j} = \sum_{t=1}^{n-s} I\big(\xi_t = i_1,\ \xi_{t+1} = i_2, \qquad \xi_{t+s-1} = i_s,\ \xi_{t+s} = j\big). \tag{5}$$

Число всех таких $(s+1)$-цепочек равно N^{s+1}, так как каждый символ i_1, i_s, j пробегает все множество $\mathcal{E} = \{1, 2, \quad, N\}$. Однако не все соответствующие частоты $\nu_{\bar{i}j}$ будут отличными от нуля. Так, если переходная вероятность $p_{\bar{i};j}$, определенная в (2), при некоторых $\bar{i}, j$ равна нулю, то комбинация $(\bar{i}j)$ не может появиться в выборке (1), и потому $\nu_{\bar{i}j} = 0$. Конечно, в силу случайности, $\nu_{\bar{i}j}$ может принимать нулевое значение и при $p_{\bar{i};j} > 0$, но в теории эргодических цепей Маркова доказывается, что в больших выборках (при $n \to \infty$) вероятность такого события близка к нулю, и потому можно считать, что для больших выборок с вероятностью, близкой к 1, $\nu_{\bar{i}j} > 0 \Longleftrightarrow p_{\bar{i};j} > 0$. С учетом этого замечания будем далее ограничиваться рассмотрением лишь тех комбинаций $(\bar{i}j)$, для которых $p_{\bar{i};j} > 0$. Множество всех таких комбинаций обозначим $\mathcal{E}_0^{s+1}$ а его объем — $d^0 = |\mathcal{E}_0^{s+1}|$. В частности, если все переходные вероятности $p_{\bar{i};j} > 0$, то

$$\mathcal{E}_0^{s+1} = \mathcal{E}^{s+1} = \underbrace{\mathcal{E} \times \quad \times \mathcal{E}}_{s+1 \text{ раз}} \quad \text{и} \quad d^0 = |\mathcal{E}^{s+1}| = N^{s+1}$$

Далее, если проверяемая гипотеза задана в параметрическом виде (3), то множество

$$\mathcal{E}_0^{s+1} = \{(\bar{i}j) \quad p_{\bar{i};j}(\theta) > 0\}$$

может, вообще говоря, зависеть от параметра θ и тем самым для каждого значения θ быть своим. Этот нерегулярный случай мы будем в дальнейшем

исключать из рассмотрения; более того, мы будем предполагать выполненным следующее общее условие регулярности.

Условие регулярности A

Условие A.

1) *Множество* $\mathcal{E}_0^{s+1} = \{(\bar{i}j) \quad p_{\bar{i}:j}(\theta) > 0\}$ *не зависит от* θ *и обладает следующим свойством: любая цепь Маркова с переходными вероятностями* $p_{\bar{i}:j}$ *такими, что* $p_{\bar{i}:j} > 0 \Longleftrightarrow (\bar{i}j) \in \mathcal{E}_0^{s+1}$, *является эргодической*;

2) *функции* $p_{\bar{i}:j}(\theta)$ *трижды непрерывно дифференцируемы по* θ;

3) *матрица* $\left\| \dfrac{\partial p_{\bar{i}:j}(\theta)}{\partial \theta_m} \right\|$ *размера* $(d^0 \times k)$ *имеет ранг* k *всюду в* Θ.

Обозначим теперь $\mathcal{H}_s'$ класс эргодических цепей Маркова порядка s, для переходных вероятностей которых выполняется условие $p_{\bar{i}:j} > 0 \Longleftrightarrow (\bar{i}j) \in \mathcal{E}_0^{s+1}$ В силу условия A (п. 1) — это цепи, в которых возможны такие же переходы $\bar{i} \to j$, как и при гипотезе $H^0(\Theta)$, но вид переходных вероятностей при этом никак не конкретизируется (за исключением необходимого *условия стохастичности* $\sum_{j=1}^{N} p_{\bar{i}:j} \quad 1, \forall\, \bar{i}$). Если справедлива некоторая альтернатива $H \in \mathcal{H}_s \backslash \mathcal{H}_s'$: для некоторой комбинации $(\bar{i}j) \notin \mathcal{E}_0^{s+1}$ переходная вероятность $p_{\bar{i}:j} > 0$, то, как отмечено выше, $\mathbf{P}\{\nu_{\bar{i}j} > 0 | H\} \to 1$ при $n \to \infty$. Но при гипотезе $H^0(\Theta)$ событие $\{\nu_{\bar{i}j} > 0\}$ невозможно, и потому появление в выборке (1) комбинации $(\bar{i}j)$ свидетельствует о том, что гипотеза $H^0(\Theta)$ ложна. Таким образом, альтернативы $H \in \mathcal{H}_s \backslash \mathcal{H}_s'$ по длинной реализации цепи (1) улавливаются с вероятностью, близкой к 1, и потому проблема, собственно, состоит в различении гипотезы $H^0(\Theta)$ и альтернатив из класса $\mathcal{H}_s'$.

Наконец, выделим еще один тип задач для цепей Маркова, возникающих в приложениях. Иногда требуется проверить нулевую гипотезу H^0 внутри более узкого класса гипотез, нежели общий класс $\mathcal{H}_s$. Формализация подобных ситуаций выглядит следующим образом. Пусть априори постулируется, что мы имеем дело с цепью Маркова вида (3), а проверяемая гипотеза есть утверждение $H^0(\Theta_0, \Theta) \quad \theta \in \Theta_0$, где Θ_0 — заданное подмножество Θ, при этом $\dim \Theta_0 = r < k$. Обычно подмножество Θ_0 задается с помощью новой параметризации в терминах некоторого параметра $\delta = (\delta_1, \quad , \delta_r)$ с множеством возможных его значений Δ, где Δ — заданное открытое подмножество r-мерного евклидова пространства R^r В этом случае Θ_0 есть образ Δ при некотором отображении $h = (h_1, \ldots, h_k)$:

$$\Theta_0 = \{\theta \quad \theta = h(\delta) = (h_1(\delta), \ldots, h_k(\delta)), \ \delta \in \Delta\} \tag{6}$$

В дальнейшем мы будем предполагать, что отображение h удовлетворяет стандартным условиям регулярности:

Условие B. *Функции* $h(\delta) = \big(h_1(\delta), \ldots, h_k(\delta)\big)$ *трижды непрерывно дифференцируемы в* Δ *и*

Условие регулярности B

$$\operatorname{rank} \left\| \frac{\partial h_i(\delta)}{\partial \delta_j} \right\| = r \quad \textit{всюду в } \Delta.$$

Из изложенного видно, насколько более разнообразны могут быть гипотезы, подлежащие проверке, в рамках марксовской модели зависимых испытаний в сравнении со схемой независимых испытаний. Тем не менее знакомые уже нам стандартные (можно даже сказать — классические) методы хи-квадрат и отношения правдоподобия позволяют построить удобные для применений критерии проверки всех таких гипотез.

2. Критерий хи-квадрат для простой гипотезы

Пусть требуется проверить простую гипотезу

Простая гипотеза для цепи Маркова

$$H^0 \quad p_{\bar{i}:j} = p^\circ_{\bar{i}:j} > 0 \quad \Longleftrightarrow \quad (\bar{i}j) \in \mathcal{E}_0^{s+1}$$

В соответствии со сказанным в п. 1, будем рассматривать в цепи (1) лишь те комбинации $(\bar{i}j)$, для которых $p^\circ_{\bar{i}:j} > 0$, т. е. при $(\bar{i}j) \in \mathcal{E}_0^{s+1}$. В методе хи-квадрат сравниваются наблюдаемые частоты $\{\nu_{\bar{i}j}, (\bar{i}j) \in \mathcal{E}_0^{s+1}\}$ с соответствующими «гипотетическими частотами» $\{\nu_{\bar{i}} p^\circ_{\bar{i}:j}\}$ и в качестве тестовой статистики используют статистику

$$\mathring{X}^2_{n,\,s+1} = \sum_{(\bar{i}j) \in \mathcal{E}_0^{s+1}} \frac{(\nu_{\bar{i}j} - \nu_{\bar{i}}\, p^\circ_{\bar{i}:j})^2}{\nu_{\bar{i}}\, p^\circ_{\bar{i}:j}} \tag{6'}$$

(здесь $\nu_{\bar{i}j}$ вычисляются по формуле (5), а $\nu_{\bar{i}} = \sum_{j=1}^{N} \nu_{\bar{i}j}$, см. замечание к формулам (5) § 4.5). Расчет соответствующего критерия производится на основании следующего утверждения [29].

Теорема 1. Если справедлива гипотеза H_0, то статистика (6′) имеет в пределе при $n \to \infty$ распределение $\chi^2(d^0 - N^s)$. Соответствующий критерий, задаваемый при уровне значимости α критической областью

$$\{\mathring{X}^2_{n,\,s+1} > \chi^2_{1-\alpha,\,d^0-N^s}\}, \tag{7}$$

состоятелен против любой фиксированной альтернативы $H \in \mathcal{H}_s$.

Выделим важный частный случай, когда все переходные вероятности $p^\circ_{\bar{i}:j} > 0$. Тогда $\mathcal{E}_0^{s+1} = \mathcal{E}^{s+1}$, $d^0 = N^{s+1}$ и критерий (7) принимает вид

Критерий χ^2 для цепи Маркова

$$\left\{ \mathring{X}^2_{n,s+1} = \sum_{(\bar{i}\bar{j})\in\mathcal{E}^{s+1}} \frac{(\nu_{\bar{i}\bar{j}} - \nu_{\bar{i}}\, p^{\circ}_{\bar{i}:\bar{j}})^2}{\nu_{\bar{i}}\, p^{\circ}_{\bar{i}:\bar{j}}} > \chi^2_{1-\alpha,\, N^s(N-1)} \right\} \qquad (7')$$

С частным случаем критерия (7'), когда все $p^{\circ}_{\bar{i}:\bar{j}} = 1/N$, мы уже встречались в п. 2 § 4.5 при обсуждении проблемы датчиков случайных чисел.

3. Критерий хи-квадрат для сложной гипотезы

Рассмотрим теперь проблему проверки сложной гипотезы $H^0(\Theta)$ вида (3) внутри общей гипотезы $\mathcal{H}_s$ и будем предполагать выполненным условие регулярности А. Критерий согласия для гипотезы $H^0(\Theta)$ строится по аналогии со случаем простой гипотезы, т. е. с использованием стандартной статистики типа хи-квадрат

$$X^2_{n,s+1}(\theta) = \sum_{(\bar{i}\bar{j})\in\mathcal{E}_0^{s+1}} \frac{(\nu_{\bar{i}\bar{j}} - \nu_{\bar{i}}\, p_{\bar{i}:\bar{j}}(\theta))^2}{\nu_{\bar{i}} p_{\bar{i}:\bar{j}}(\theta)}$$

Конечно, эту статистику еще нельзя использовать для построения критерия, так как она зависит от неизвестного параметра θ. Но если параметр θ предварительно оценить по данным (5), т. е. заменить его некоторой оценкой $\widehat{\theta}_n = \widehat{\theta}_n(\{\nu_{\bar{i}\bar{j}}\})$, и затем заменить функции $p_{\bar{i}:\bar{j}}(\theta)$ оценками $\widehat{p}_{\bar{i}:\bar{j}} = p_{\bar{i}:\bar{j}}(\widehat{\theta}_n)$, то мы придем уже к настоящей тестовой статистике

$$\widehat{X}^2_{n,s+1} = X^2_{n,s+1}(\widehat{\theta}_n) = \sum_{(\bar{i}\bar{j})\in\mathcal{E}_0^{s+1}} \frac{(\nu_{\bar{i}\bar{j}} - \nu_{\bar{i}}\, \widehat{p}_{\bar{i}:\bar{j}})^2}{\nu_{\bar{i}}\, \widehat{p}_{\bar{i}:\bar{j}}}. \qquad (8)$$

Это достаточно сложная для расчетов, а главное — для исследования ее распределения статистика, но тем не менее для таких статистик имеется хорошо разработанная асимптотическая (при $n \to \infty$) теория, которая утверждает, что если оценки $\widehat{\theta}_n$ получаются по методу максимального правдоподобия, то статистика (8) имеет при гипотезе $H^0(\Theta)$ предельное распределение χ^2, как и в случай простой гипотезы. Отличие проявляется лишь в числе степеней свободы предельного распределения, которое уменьшается на величину, равную числу оцениваемых параметров, определяющих гипотезу $H^0(\Theta)$. То есть в данном случае имеется полная аналогия с теорией метода хи-квадрат для схемы независимых испытаний.

Приведем ключевые элементы этой теории. Рассмотрим функцию правдоподобия выборки (1) при гипотезе $H^0(\Theta)$, т. е. вероятность наблюденной реализации цепи в предположении, что гипотеза $H^0(\Theta)$ истинна:

$$L_n(\theta) = p_{\xi_1\ldots\xi_s:\xi_{s+1}}(\theta)\, p_{\xi_2\ldots\xi_{s+1}:\xi_{s+2}}(\theta) \qquad p_{\xi_{n-s}\ldots\xi_{n-1}:\xi_n}(\theta) = \prod_{(\bar{i}\bar{j})\in\mathcal{E}_0^{s+1}} \left(p_{\bar{i}:\bar{j}}(\theta)\right)^{\nu_{\bar{i}\bar{j}}} \qquad (9)$$

По аналогии со случаем независимых испытаний назовем оценкой максимального правдоподобия (о. м. п.) $\widehat{\theta}_n$ параметра θ решение системы уравнений

$$\frac{\partial \ln L_n(\theta)}{\partial \theta_m} = \sum_{(i,j)\in\mathcal{E}_0^{s+1}} \frac{\nu_{i,j}}{p_{i:j}(\theta)} \frac{\partial p_{i:j}(\theta)}{\partial \theta_m} = 0, \quad m = 1, \quad , k. \tag{10}$$

Относительно существования, единственности и свойств решения системы (10) известно следующее утверждение [29].

Теорема 2. Пусть выполнено условие А, тогда с вероятностью, стремящейся к 1 при $n \to \infty$, система (10) имеет единственное решение $\widehat{\theta}_n$, являющееся состоятельной оценкой параметра θ. Более того, если θ^0 есть истинное значение θ, то нормированный случайный вектор $\sqrt{n}(\widehat{\theta}_n - \theta^0)$ распределен асимптотически нормально с центром в нуле и матрицей вторых моментов $\mathcal{I}^{-1}(\theta^0)$, где

$$\mathcal{I}(\theta) = \left\| \mathbf{E}_\theta \left(\frac{\partial \ln p_{\xi_1 \ldots \xi_s : \xi_{s+1}}(\theta)}{\partial \theta_m} \frac{\partial \ln p_{\xi_1 \ldots \xi_s : \xi_{s+1}}(\theta)}{\partial \theta_l} \right) \right\|$$

Эта теорема — аналог хорошо известных утверждений об асимптотических свойствах о. м. п. для схемы независимых испытаний; она указывает метод оценивания неизвестных параметров для цепей Маркова.

Наконец, основой для построения и расчета критерия проверки гипотезы $H^0(\Theta)$ служит следующее утверждение.

Теорема 3. Пусть выполнено условие А и $\widehat{p}_{i:j} = p_{i:j}(\widehat{\theta}_n)$, где $\widehat{\theta}_n$ — о. м. п. параметра θ. Тогда, если гипотеза $H^0(\Theta)$ справедлива, то определенная в (8) статистика $\widehat{X}^2_{n,s+1}$ распределена в пределе при $n \to \infty$ по закону

$$\chi^2(d^0 - N^s - k),$$

а критерий, основанный на критической области

$$\{\widehat{X}^2_{n,s+1} > \chi^2_{1-\alpha, d^0 - N^s - k}\}, \tag{11}$$

имеет асимптотически уровень значимости α и является состоятельным против любой фиксированной альтернативы $H \in \mathcal{H}_s$.

Пример 1 (гипотеза о порядке цепи Маркова). Пусть гипотеза H^0_r означает, что наблюдаемая последовательность (1) представляет собой цепь Маркова порядка r, в которой разрешены все переходы. Требуется построить критерий проверки этой гипотезы внутри общей гипотезы $\mathcal{H}_s$ (допускается марковость порядка $s > r$). В этом случае нулевая гипотеза задается так, как указано в (4), т. е. неизвестными параметрами, определяющими H^0_r, являются положитель-

ные числа $\theta_{i_{s-r+1}\dots}$ образующие параметрический вектор θ размерности $k = N^r(N-1)$.

Найдем о. м. п. для этих параметров. Поскольку координаты вектора θ удовлетворяют линейным связям (условию стохастичности)

$$\sum_{j=1}^{N} \theta_{i_{s-r}\dots i_s j} = 1, \quad \forall\, i_{s-r}\dots, i_s \in \mathcal{E},$$

то мы имеем задачу на условный экстремум. Введя множители Лагранжа

$$\underline{\lambda} = \left(\lambda_{i_{s-r+1}\dots} \quad (i_{s-r+1}, \dots, i_s) \in \mathcal{E}^r\right),$$

сводим задачу к отысканию абсолютного максимума функции

$$F(\theta, \underline{\lambda}) = \sum_{i_{s-r+1}, \dots, i_s, j} \nu_{*i_{s-r+1}\dots i_s j} \ln \theta_{i_{s-r+1}\dots i_s j} - \sum_{i_{s-r+1},} \lambda_{i_{s-r}\dots i_s} \left(\sum_{j=1}^{N} \theta_{i_{s-r+1}\dots i_s j} - 1\right),$$

где символ $*$ на месте недостающих индексов в комбинации $(i_1 \quad i_s j)$ означает, что по этим индексам произведено суммирование.

Для этого мы должны решить систему уравнений

$$\frac{\partial F(\theta, \underline{\lambda})}{\partial \theta_{i_{s-r+1}\dots i_s j}} = \frac{\nu_{*i_{s-r+1}\dots i_s j}}{\theta_{i_{s-r+1}\dots}} - \lambda_{i_{s-r+1}\dots} = 0, \quad i_{s-r+1}, \dots, i_s, j \in \mathcal{E},$$

$$\sum_{j=1}^{N} \theta_{i_{s-r+1}\dots i_s j} = 1, \quad i_{s-r+1}, \quad , i_s \in \mathcal{E}.$$

Отсюда легко находим решение $\widehat{\theta}_n = (\widehat{\theta}_{i_{s-r+1}\dots} \quad i_{s-r+1}, \quad , i_s, j \in \mathcal{E})$:

$$\widehat{\theta}_{i_{s-r+1}\dots i_s j} = \frac{\nu_{*i_{s-r+1}\dots i_s j}}{\nu_{*i_{s-r+1}\dots}}, \quad i_{s-r+1}, \quad , i_s, j \in \mathcal{E}, \tag{12}$$

которое говорит о том, что оценками переходных вероятностей при гипотезе H_r^0 являются наблюденные относительные частоты соответствующих переходов в последовательности (1). В итоге получаем, что статистика (8) в данном случае имеет вид

Критерий χ^2 для гипотезы о порядке цепи Маркова

$$\widehat{X}^2_{n,\, s+1,\, r} = \sum_{(ij) \in \mathcal{E}^{s+1}} \frac{\left(\nu_{ij} - \dfrac{\nu_{i*}\nu_{*i_{s-r+1}\dots i_s j}}{\nu_{*i_{s-r+1}\dots i_s *}}\right)^2}{\dfrac{\nu_{i*}\nu_{*i_{s-r+1}\dots i_s j}}{\nu_{*i_{s-r+1}\dots}}}, \tag{13}$$

и если гипотеза H_r^0 справедлива, то при $n \to \infty$ эта статистика имеет предельное распределение $\chi^2\left((N^s - N^r)(N-1)\right)$.

Выделим важный частный случай $r=0$, соответствующий гипотезе H_0^0 о независимости случайных величин в последовательности (1), когда в качестве альтернатив допускается марковость порядка s. В этом случае $p_{\tilde{i}:j}=\theta_j>0$ и $\widehat{\theta}_j=\nu_{*j}/n$, поскольку

$$\nu_*=\sum_{(\tilde{i}j)}\nu_{\tilde{i}j}=n.$$

Таким образом, для проверки гипотезы независимости H_0^0 внутри общей гипотезы $\mathcal{H}_s$ служит статистика

Критерий независимости χ^2

$$\widehat{X}^2_{n,\,s+1,\,0}=\sum_{(\tilde{i}j)\in\mathcal{E}^{s+1}}\frac{\left(\nu_{\tilde{i}j}-\dfrac{\nu_{\tilde{i}*}\nu_{*j}}{n}\right)^2}{\dfrac{\nu_{\tilde{i}*}\nu_{*j}}{n}} \tag{14}$$

Если гипотеза H_0^0 справедлива, то статистика (14) при $n\to\infty$ имеет распределение $\chi^2\big((N^s-1)(N-1)\big)$. •

Пример 2 (Гипотеза о циклическом случайном блуждании). Предположим, что в длинной реализации (1) простой ($s=1$) цепи Маркова наблюдались лишь переходы $i\to i\pm 1$ (при $i=1$ значение $1-1$ интерпретируется как N, при $i=N$ значение $N+1$ интерпретируется как 1). Таким образом, здесь множество $\mathcal{E}_0^2=\{(i,i\pm 1),\ i=1,2,\ \ ,N\}$ разрешенных переходов состоит из $d^0=2N$ элементов и не совпадает с множеством $\mathcal{E}^2$ всех пар (i,j). Пусть требуется проверить гипотезу

$$H^0\ \ p_{i:i+1}=\theta,\ p_{i:i-1}=1-\theta,$$

где $\theta\in(0,1)$ — неизвестный параметр (это и есть гипотеза о *циклическом случайном блуждании*). Проверим, что в данном случае выполнено условие А. Рассмотрим произвольную (простую) цепь Маркова с матрицей переходных вероятностей вида

$$\mathbf{P}=\begin{Vmatrix} 0 & p_{12} & 0 & & & 0 & p_{1N}\\ p_{21} & 0 & p_{23} & & & 0 & 0\\ 0 & p_{32} & 0 & & & & \\ & & & & & & \\ & & & & & & 0\\ 0 & & & p_{N-1:N-2} & 0 & p_{N-1:N} & \\ p_{N1} & 0 & & 0 & p_{N:N-1} & 0 & \end{Vmatrix},$$

где все $p_{i:i\pm1}>0$. Легко видеть, что $\mathbf{P}^N$ состоит только из положительных элементов. Это означает, что за N шагов с положительной вероятностью из любого состояния можно достичь любого другого состояния цепи, что является

достаточным условием эргодичности. Таким образом, п. 1 условия А здесь выполняется, а п. 2 и 3 очевидны. Следовательно, для построения критерия проверки гипотезы H^0 можно воспользоваться изложенной выше теорией. Для этого найдем прежде всего оценку $\widehat{\theta}_n$ для параметра θ.

Обозначим

$$u = \sum_{i=1}^{N} \nu_{i,i+1}$$

число переходов «вперед» $(i \to i+1)$ в выборке (1). Тогда функция правдоподобия (9) запишется в данном случае в виде

$$L_n(\theta) = \theta^u (1-\theta)^{n-u}$$

и уравнение

$$\frac{\partial \ln L_n(\theta)}{\partial \theta} = 0$$

имеет единственное решение $\widehat{\theta}_n = u/n$. Отсюда, в силу теоремы 3, имеем, что статистика

$$\widehat{X}^2_{n,2} = \sum_{i=1}^{N} \frac{\left(\nu_{i,i+1} - \nu_i \dfrac{u}{n}\right)^2}{\nu_i \dfrac{u}{n}} + \sum_{i=1}^{N} \frac{\left(\nu_{i,i-1} - \nu_i\left(1 - \dfrac{u}{n}\right)\right)^2}{\nu_i\left(1 - \dfrac{u}{n}\right)},$$

где $\nu_i = \nu_{i,i-1} + \nu_{i,i+1}$, распределена в пределе при $n \to \infty$ по закону $\chi^2(N-1)$, если гипотеза H^0 справедлива. Следовательно, при уровне значимости α критерий для гипотезы H^0 имеет вид

$$\boxed{H^0 \quad \text{отвергается} \iff \{\widehat{X}^2_{n,2} > \chi^2_{1-\alpha,N-1}\}.} \tag{15}$$

Если бы нужно было проверить соответствующую простую гипотезу о циклическом блуждании: $\theta = \theta^0$, где $\theta^0 \in (0,1)$ — заданное число, то в соответствии с п. 3 следовало бы воспользоваться статистикой (см. (6′))

$$\mathring{X}^2_{n,2} = \sum_{i=1}^{N} \frac{(\nu_{i,i+1} - \nu_i\theta^0)^2}{\nu_i\theta^0} + \sum_{i=1}^{N} \frac{(\nu_{i,i-1} - \nu_i(1-\theta^0))^2}{\nu_i(1-\theta^0)}, \tag{16}$$

которая при справедливости проверяемой гипотезы распределена при $n \to \infty$ по закону $\chi^2(N)$. •

4. Критерий отношения правдоподобия для общих параметрических гипотез

Рассмотрим, наконец, проблему проверки гипотезы $H^0(\Theta_0, \Theta)$, где Θ_0 задано в виде (6) и при этом выполнено условие D Тогда функция правдоподобия (9) при гипотезе $H^0(\Theta_0, \Theta)$ становится функцией от нового параметра δ:

$L_n(h(\delta))$, и система уравнений

$$\frac{\partial \ln L_n(h(\delta))}{\partial \delta_l} = \sum_{(\bar{i}\bar{j})\in\mathcal{E}_0^{s+1}} \frac{\nu_{\bar{i}\bar{j}}}{p_{\bar{i}:\bar{j}}(h(\delta))} \frac{\partial p_{\bar{i}:\bar{j}}(h(\delta))}{\partial \delta_l} = 0, \quad l = 1, \ldots, r, \tag{17}$$

будет иметь состоятельное решение $\widehat{\delta}_n$, если гипотеза $H^0(\Theta_0, \Theta)$ справедлива. Общий принцип построения критерия отношения правдоподобия (КОП) для проверки гипотезы $H^0(\Theta_0, \Theta)$ остается таким же, как и в случае независимых испытаний: строится статистика отношения правдоподобия (λ-статистика)

$$\lambda_n = \frac{L_n\big(h(\widehat{\delta}_n)\big)}{L_n(\widehat{\theta}_n)}, \tag{18}$$

где $\widehat{\theta}_n$ — решение системы (10), и критическая область задается в виде $\{\lambda_n < c_\alpha\}$, где при заданном уровне значимости α критическая граница c_α выбирается так, чтобы выполнялось условие

$$\mathbf{P}_\theta\{\lambda_n < c_\alpha\} \leqslant \alpha, \quad \forall\, \theta \in \Theta_0. \tag{19}$$

Основой для расчета такого критерия служит следующая теорема [29].

Теорема 4. Пусть выполнены условия A и B. Тогда, если гипотеза $H^0(\Theta_0, \Theta)$ справедлива, то при $n \to \infty$ статистика

$$-2\ln\lambda_n = 2\big[\ln L_n(\widehat{\theta}_n) - \ln L_n\big(h(\widehat{\delta}_n)\big)\big] = 2\sum_{(\bar{i}\bar{j})\in\mathcal{E}_0^{s+1}} \nu_{\bar{i}\bar{j}} \ln \frac{p_{\bar{i}:\bar{j}}(\widehat{\theta}_n)}{p_{\bar{i}:\bar{j}}\big(h(\widehat{\delta}_n)\big)} \tag{20}$$

имеет предельное распределение $\chi^2(k-r)$; если δ^0 — истинное значение параметра δ, то статистика

$$2\big[\ln L_n\big(h(\widehat{\delta}_n)\big) - \ln L_n\big(h(\delta^0)\big)\big] = 2\sum_{(\bar{i}\bar{j})\in\mathcal{E}_0^{s+1}} \nu_{\bar{i}\bar{j}} \ln \frac{p_{\bar{i}:\bar{j}}\big(h(\widehat{\delta}_n)\big)}{p_{\bar{i}:\bar{j}}\big(h(\delta^0)\big)} \tag{21}$$

распределена в пределе при $n \to \infty$ по закону $\chi^2(r)$; при этом статистики (20) и (21) асимптотически независимы.

Итак, при заданном уровне значимости α КОП проверки гипотезы $H^0(\Theta_0, \Theta)$ $\theta \in \Theta_0$ внутри общей гипотезы $H^0(\Theta) : \theta \in \Theta$ имеет для больших выборок вид

$$\boxed{H^0(\Theta_0, \Theta) \quad \text{отвергается} \iff \{-2\ln\lambda_n > \chi^2_{1-\alpha,\,k-r}\}.} \tag{22}$$

При этом, напомним, $r = \dim\Theta_0 < \dim\Theta = k$.

Аналогично с помощью статистики (21) строится критерий проверки простой гипотезы $\delta = \delta^0$ внутри гипотезы: $\delta \in \Delta$.

Пример 3 (продолжение примера 2). Предположим, что класс цепей, в которых разрешены лишь переходы в соседние состояния ($i \to i \pm 1$), выделен условием цикличности: $p_{i;i+1} = \delta$, $\delta \in (0, 1)$, и требуется при этом проверить простую гипотезу $H^0 : \delta = \delta^0$ В силу теоремы 4, в этой задаче надо воспользоваться статистикой (21), которая в обозначениях примера 2 имеет в данном случае вид

$$T_n = 2u \ln \frac{u}{n\delta^0} + 2(n-u) \ln \frac{1 - u/n}{1 - \delta^0}. \tag{23}$$

Если гипотеза H^0 истинна, то эта статистика асимптотически (при $n \to \infty$) распределена по закону $\chi^2(1)$, откуда и следует соответствующий критерий: $\{T_n > \chi^2_{1-\alpha,1}\}$. •

5. Критерий однородности

В этом пункте мы рассмотрим применение метода отношения правдоподобия к еще одной общей статистической проблеме — проверке гипотезы однородности для многих выборок в марковской модели. В общем виде эта проблема формулируется следующим образом.

Пусть $\xi^{(\tau)} = \{\xi_t^{(\tau)}\}$, $\tau = 1, \quad , v$, — независимые марковские цепи порядка s с одним и тем же пространством состояний $\mathcal{E} = \{1, \ldots, N\}$ и одним и тем же множеством возможных переходов

$$\mathcal{E}_0^{s+1} \quad p_{\bar{i};j}(\theta) > 0 \iff (\bar{i}j) \in \mathcal{E}_0^{s+1} \qquad \theta = (\theta_1, \quad , \theta_k) \in \Theta.$$

В дальнейшем предполагается выполненным стандартное условие регулярности А. Обозначим через $\theta^{(\tau)} \in \Theta$ значение параметра θ, порождающего процесс $\xi^{(\tau)}$ (так что $\xi^{(\tau)}$ управляется переходными вероятностями $p_{\bar{i};j}(\theta^{(\tau)})$, $\tau = 1, \ldots, v$). Значения $\theta^{(1)}, \ldots, \theta^{(v)}$ неизвестны и, вообще говоря, различны. Эту модель будем называть общей гипотезой $\mathcal{H}$, ее размерность (число определяющих ее параметров) $\dim \mathcal{H} = vk$. Предположим теперь, что мы располагаем выборкой из каждого процесса, так что наши исходные статистические данные есть

$$\xi_1^{(\tau)}, \quad , \xi_{n_\tau}^{(\tau)}, \quad \tau = 1, \quad , v. \tag{24}$$

По этой информации требуется построить критерий проверки *гипотезы однородности* $H^0 \quad \theta^{(1)} = \quad = \theta^{(v)}$ внутри гипотезы $\mathcal{H}$. Размерность этой гипотезы есть k, и она (гипотеза) означает, что наблюдаемые процессы $\xi^{(\tau)}$ являются независимыми реализациями одного и того же марковского процесса ξ из описанного выше класса процессов; другими словами, в этом отношении наши данные (24) однородны.

Схема применения λ-метода в данном случае выглядит следующим образом. Выпишем функцию правдоподобия данных (24) при гипотезах $\mathcal{H}$ и H^0.

Поскольку выборки независимы, то в соответствии с (9) имеем

$$L_{n_1\dots n_v}(\theta_1^{(1)}, \quad , \theta^{(v)}) = \prod_{\tau=1}^{v} L_{n_\tau}(\theta^{(\tau)}),$$
$$L_{n_\tau}(\theta^{(\tau)}) = \prod_{(\bar{i}\bar{j})\in\mathcal{E}_0^{s+1}} \left(p_{\bar{i}:\bar{j}}(\theta^{(\tau)})\right)^{\nu_{\bar{i}\bar{j}}^{(\tau)}} \tag{25}$$

$$L_n(\theta) = L_{n_1\dots n_v}(\theta, \dots, \theta) = \prod_{(\bar{i}\bar{j})\in\mathcal{E}_0^{s+1}} \left(p_{\bar{i}:\bar{j}}(\theta)\right)^{\nu_{\bar{i}\bar{j}}}, \tag{26}$$

где $\nu_{\bar{i}\bar{j}}^{(\tau)}$ — число наблюдаемых комбинаций $(\bar{i}\bar{j})$ в выборке с номером τ (см. (5)), $\nu_{\bar{i}\bar{j}} = \sum_{\tau=1}^{v} \nu_{\bar{i}\bar{j}}^{(\tau)}$ и $n = n_1 + \quad + n_\tau$. Далее, исходя из (25) и (26), мы оцениваем неизвестные параметры $\theta^{(\tau)}$ $\tau = 1, \quad , v$, при гипотезе $\mathcal{H}$ и θ — при гипотезе H^0 по методу максимального правдоподобия, решая соответствующие уравнения типа (10). Пусть соответствующие о. м. п. есть $\widehat{\theta}_{n_\tau}^{(\tau)}$, $\tau = 1, \quad , v$, и $\widehat{\theta}_n$. Тогда тестовая статистика (λ-статистика) имеет вид

$$-2\ln\lambda_{n_1\dots n_v} = 2\left[\ln L_{n_1\dots n_v}\left(\widehat{\theta}_{n_1}^{(1)}, \quad , \widehat{\theta}_{n_v}^{(v)}\right) - \ln L_n(\widehat{\theta}_n)\right] =$$
$$= 2\sum_{\tau=1}^{v}\sum_{(\bar{i}\bar{j})\in\mathcal{E}_0^{s+1}} \nu_{\bar{i}\bar{j}}^{(\tau)} \ln\left(\frac{p_{\bar{i}:\bar{j}}\left(\widehat{\theta}_{n_\tau}^{(\tau)}\right)}{p_{\bar{i}:\bar{j}}(\widehat{\theta}_n)}\right), \tag{27}$$

а сам критерий однородности формулируется следующим образом:

$$\boxed{H^0 \text{ отвергается} \iff \{-2\ln\lambda_{n_1\dots n_v} > t\}.} \tag{28}$$

Обоснованием этой методики является следующая теорема об асимптотическом распределении статистики (27), дающая одновременно и способ вычисления критической границы t в (28).

Теорема 5. Пусть выполнено условие A и $n_1 \quad n_v \to \infty$ так, что $n_\tau/n \to \lambda_\tau$, $0 < \lambda_\tau < 1$, $\tau = 1, \quad , v$. Тогда, если гипотеза однородности H^0 справедлива, то статистика (27) распределена в пределе по закону $\chi^2(k(v-1))$, где

$$k(v-1) = \dim\mathcal{H} - \dim H^0$$

Основываясь на этой теореме, при выполнении ее условий мы выбираем в (28) при заданном уровне значимости α величину t как квантиль уровня $1-\alpha$ распределения $\chi^2(k(v-1))$: $t = \chi^2_{1-\alpha,k(v-1)}$, и тогда

$$\mathbf{P}\left\{-2\ln\lambda_{n_1\dots n_v} > \chi^2_{1-\alpha,k(v-1)} | H^0\right\} \longrightarrow \alpha. \tag{29}$$

Соотношениями (27)–(29) полностью определен состоятельный в классе $\mathcal{H}$ критерий однородности в общем случае произвольной параметрической модели и произвольного числа выборок.

Выделим важный специальный случай описанной общей схемы, когда гипотеза $\mathcal{H}$ задана указанием лишь множества $\mathcal{E}_0^{s+1}$ возможных переходов без конкретизации вида переходных вероятностей $p_{\bar{i}:j}$, $(\bar{i}j) \in \mathcal{E}_0^{s+1}$. В этом случае сами эти вероятности являются неизвестными параметрами модели; их число (с учетом условия стохастичности $\sum_j p_{\bar{i}:j} = 1$, $\forall\, i \in \mathcal{E}^s$) для каждого процесса $\xi^{(\tau)}$ есть $d^0 - N^s$, т. е. $\dim \mathcal{H} = v(d^0 - N^s)$, а $\dim H^0 = d^0 - N^s$ а их о. м. п. являются наблюденные относительные частоты соответствующих переходов (см. пример 1):

$$\widehat{p}_{\bar{i}:j}^{(\tau)} = \frac{\nu_{\bar{i}j}^{(\tau)}}{\nu_{\bar{i}*}^{(\tau)}}, \quad \tau = 1, \ldots, v, \qquad \widehat{p}_{\bar{i}:j} = \frac{\nu_{\bar{i}j}}{\nu_{\bar{i}*}}, \quad (\bar{i}j) \in \mathcal{E}_0^{s+1}$$

Отсюда имеем, что в рассматриваемом случае статистика (27) принимает вид

$$-2 \ln \lambda_{n_1 \ldots n_v} = 2 \sum_{\tau=1}^{v} \sum_{(\bar{i}j) \in \mathcal{E}_0^{s+1}} \nu_{\bar{i}j}^{(\tau)} \ln \frac{\nu_{\bar{i}j}^{(\tau)} \nu_{\bar{i}*}}{\nu_{\bar{i}*}^{(\tau)} \nu_{\bar{i}j}}, \tag{30}$$

и при гипотезе однородности H^0 она имеет предельное (в условиях теоремы 5) распределение $\chi^2\big((d^0 - N^s)(v-1)\big)$.

Замечание. Можно показать, что в рассматриваемой задаче тестовая статистика (30) асимптотически эквивалентна статистике типа хи-квадрат

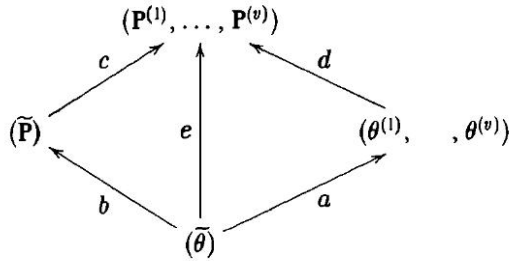
$$\sum_{\tau=1}^{v} \sum_{(\bar{i}j) \in \mathcal{E}_0^{s+1}} \frac{\left(\nu_{\bar{i}j}^{(\tau)} - \nu_{\bar{i}*}^{(\tau)} \dfrac{\nu_{\bar{i}j}}{\nu_{\bar{i}*}}\right)^2}{\nu_{\bar{i}*}^{(\tau)} \dfrac{\nu_{\bar{i}j}}{\nu_{\bar{i}*}}}. \tag{31}$$

Пример 4 (Модель случайного блуждания, проверка гипотез однородности для нее). Рассмотрим класс простых цепей Маркова, множество разрешенных переходов в которых имеет вид как в примерах 2 и 3:

$$\mathcal{E}_0^2 = \{(i, i \pm 1),\ i = 1, 2, \ldots, N\}.$$

Пусть мы имеем v независимых выборок в этой модели и $\{\nu_{i,\,i\pm1}^{(\tau)}\}$, $\tau = 1, \ldots, v$, — соответствующие наблюденные частоты переходов. Можно рассмотреть следующие четыре гипотезы. Прежде всего — это общая гипотеза $\mathcal{H}$, согласно которой каждая из наблюдаемых цепей управляется своей матрицей переходных вероятностей, вид которой указан в примере 2; если обозначить $\mathbf{P}^{(\tau)}$ матрицу процесса $\xi^{(\tau)}$, то эту гипотезу будем обозначать $(\mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(v)})$. Гипотезу, согласно которой $\mathbf{P}^{(1)} = \ldots = \mathbf{P}^{(v)} = \widetilde{\mathbf{P}}$, обозначим $(\widetilde{\mathbf{P}})$. Далее, выделим

общую гипотезу о циклическом случайном блуждании (см. пример 2) и обозначим $\theta^{(1)}, \ldots, \theta^{(v)}$ значения соответствующих параметров для наблюдаемых процессов — такую гипотезу обозначим $(\theta^{(1)}, \ldots, \theta^{(v)})$. Наконец, гипотезу, согласно которой $\theta^{(1)} = \ldots = \theta^{(v)} = \widetilde{\theta}$, обозначим $(\widetilde{\theta})$. Таким образом, $(\widetilde{\mathbf{P}})$ — это общая гипотеза однородности, а $(\widetilde{\theta})$ — гипотеза однородности для модели циклического случайного блуждания. Соотношения между этими гипотезами можно изобразить следующей диаграммой:



Здесь каждое ребро представляет собой статистическую проблему проверки гипотезы, являющейся началом ребра, внутри более общей гипотезы, являющейся его концом. Например, ребро a соответствует проверке гипотезы $(\widetilde{\theta})$ внутри гипотезы $(\theta^{(1)}, \ldots, \theta^{(v)})$. Из пяти проблем a, b, c, d и e проблемы a, c и e — это проблемы однородности.

Выпишем для каждой гипотезы оценки параметров, максимальное значение логарифма функции правдоподобия и ее размерность (т. е. число степеней свободы) (см. табл.).

Гипотеза	о. м. п.	$\max \ln L$	dim
$(\mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(v)})$	$\left\{\dfrac{\nu_{i,i+1}^{(\tau)}}{\nu_{i*}^{(\tau)}}\right\}$	$\displaystyle\sum_{\tau=1}^{v}\sum_{(ij)\in\mathcal{E}_0^2}\nu_{ij}^{(\tau)}\ln\frac{\nu_{ij}^{(\tau)}}{\nu_{i*}^{(\tau)}}$	vN
$(\theta^{(1)}, \ldots, \theta^{(v)})$	$\left\{\dfrac{u^{(\tau)}}{n_\tau}\right\}$	$\displaystyle\sum_{\tau=1}^{v}\left[u^{(\tau)}\ln\frac{u^{(\tau)}}{n_\tau}+(n_\tau-u^{\tau})\ln\left(1-\frac{u^{(\tau)}}{n_\tau}\right)\right]$	v
$(\widetilde{\mathbf{P}})$	$\left\{\dfrac{\nu_{i,i+1}}{\nu_{i*}}\right\}$	$\displaystyle\sum_{(ij)\in\mathcal{E}_0^2}\nu_{ij}\ln\frac{\nu_{ij}}{\nu_{i*}}$	N
$(\widetilde{\theta})$	$\dfrac{u}{n}$	$\displaystyle u\ln\frac{u}{n}+(n-u)\ln\left(1-\frac{u}{n}\right)$	

Здесь в дополнение к введенным ранее обозначениям положено

$$u^{(\tau)}=\sum_{i=1}^{N}\nu_{i,i+1}^{(\tau)} \quad \text{и} \quad u=u^{(1)}+\ldots+u^{(v)}.$$

Чтобы построить критерий для соответствующей проблемы (см. диаграмму), надо взять строки этой таблицы для двух соответствующих гипотез и вычислить разность между выражением в третьем столбце выше расположенной строки и соответствующим выражением нижней из этих двух строк; будучи умноженной на 2, эта разность дает соответствующую тестовую статистику. Число степеней свободы в предельном распределении χ^2 для этой статистики (если справедлива более узкая, т. е. ниже расположенная гипотеза) равно разности размерностей рассматриваемых гипотез. Например, асимптотический (в условиях теоремы 5) критерий однородности для проверки гипотезы $(\widetilde{\theta})$ внутри гипотезы $(\theta^{(1)}, \quad , \theta^{(v)})$ (проблема a) имеет (при уровне значимости α) вид

$$\left\{2\left(\sum_{\tau=1}^{v}\left[u^{(\tau)}\ln\frac{u^{(\tau)}}{n_\tau}+(n_\tau-u^{(\tau)})\ln\left(1-\frac{u^{(\tau)}}{n_\tau}\right)\right]-\right.\right.$$
$$\left.\left.-\left[u\ln\frac{u}{n}+(n-u)\ln\left(1-\frac{u}{n}\right)\right]\right)>\chi^2_{1-\alpha,\,v-1}\right\} \tag{32}$$

Если же мы хотим проверить ту же гипотезу $(\widetilde{\theta})$ внутри общей гипотезы $(\mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(v)})$ (проблема e), то надо использовать критерий

$$\left\{2\left(\sum_{\tau=1}^{v}\sum_{(ij)\in\mathcal{E}_0^2}\nu_{ij}^{(\tau)}\ln\frac{\nu_{ij}^{(\tau)}}{\nu_{i*}^{(\tau)}}-\right.\right.$$
$$\left.\left.-\left[u\ln\frac{u}{n}+(n-u)\ln\left(1-\frac{u}{n}\right)\right]\right)>\chi^2_{1-\alpha,\,vN-1}\right\} \tag{33}$$

Отметим также, что критерии для проблем b и d по существу построены в примере 2, а критерий для проблемы c есть частный случай критерия, порождаемого тестовой статистикой (30). •

6. Оценивание порядка цепи

Важнейшим параметром марковской модели является порядок цепи. В приложениях этот параметр, как правило, неизвестен и потому представляет большой теоретический и практический интерес вопрос о его оценивании по наблюдаемой реализации цепи (1). Один из возможных подходов к этой проблеме состоит в использовании построенного в примере 1 критерия проверки гипотезы о порядке цепи. Основываясь на тестовой статистике (13) можно предложить следующий алгоритм последовательного типа для определения порядка цепи в рамках модели $\mathcal{H}_s$.

С самого начала мы задаемся некоторым достаточно большим значением s и априори предполагаем, что наблюдаемая последовательность (1) является реализацией некоторой однородной эргодической цепи Маркова порядка не выше s. Подсчитав частоты $\nu_{\tilde{i}j}$, $\tilde{i}=(i_1, \ldots, i_s)$, $(s+1)$-цепочек в последовательности (1), с помощью статистики $\widehat{X}^2_{n,\,s+1,\,0}$ (14) мы проверяем гипотезу

независимости H_0^0 $r = 0$ (r — истинный порядок цепи, $r \leqslant s$). Если при уровне значимости α гипотеза H_0^0 принимается, т. е. наблюдается событие

$$A_0 = \left\{\widehat{X}^2_{n,\,s+1,\,0} \leqslant \chi^2_{1-\alpha,\,(N^s-1)(N-1)}\right\}, \tag{34}$$

то процедура заканчивается на 1-м шаге и принимается решение, что $r = 0$. В противном случае, т. е. если наблюдается противоположное к (34) событие, выполняется 2-й шаг процедуры, а именно, с помощью статистики $\widehat{X}^2_{n,\,s+1,\,1}$ (13) проверяется гипотеза H_1^0 $r = 1$ (цепь имеет порядок 1). Если гипотеза H_1^0 принимается, т. е. наблюдается событие

$$A_1 = \left\{\widehat{X}^2_{n,\,s+1,\ } \leqslant \chi^2_{1-\alpha,\,(N^s-N)(N-1)}\right\}, \tag{35}$$

то процедура заканчивается на 2-м шаге и принимается решение, что $r = 1$. Если же наблюдается дополнительное событие $\overline{A}_1$, то выполняется следующий, 3-й шаг процедуры, состоящий в проверке гипотезы H_2^0 $r = 2$: вычисляется статистика $\widehat{X}^2_{n,\,s+1,2}$ (13) и ее значение сравнивается с критическим уровнем $\chi^2_{1-\alpha,\,(N^s-N^2)(N-1)}$. Если наблюдается событие

$$A_2 = \left\{\widehat{X}^2_{n,\,s+1,2} \leqslant \chi^2_{1-\alpha,\,(N^s-N^2)(N-1)}\right\}, \tag{36}$$

то процесс заканчивается принятием решения $r = 2$, в противном случае процесс продолжается аналогичным образом и т. д. Действуя таким образом, мы можем остановиться на каком-то значении $r \leqslant s-1$, либо, проделав s шагов и отклонив на s-м шаге гипотезу H_{s-}^0 $r = s-1$, делаем заключение, что $r \geqslant s$.

Таким образом, обозначая через r^* оценку для истинного порядка цепи, можем записать итог в следующей форме:

$$\begin{aligned} r^* &= \min\left\{r \quad \text{гипотеза } H_r^0 \text{ принимается}\right\} \equiv \\ &\equiv \min\left\{r \quad \widehat{X}^2_{n,\,s+1,\,r} \leqslant \chi^2_{1-\alpha,\,(N^s-N^r)(N-1)}\right\}, \end{aligned} \tag{37}$$

если $r^* \leqslant s-1$, и $r^* \geqslant s$, если все гипотезы H_r^0, $r \leqslant s-1$, последовательно отвергаются. Отметим также следующее из (37) соотношение

$$\begin{aligned} \mathbf{P}\{r^* > r\} &= \mathbf{P}\{\overline{A}_0\overline{A}_1 \quad \overline{A}_r\} = \\ &= \mathbf{P}\left\{\widehat{X}^2_{n,\,s+1,\,t} > \chi^2_{1-\alpha,\,(N^s-N^t)(N-1)},\ t = 0, 1, \ldots, r\right\}, \\ & r \leqslant s-1. \end{aligned} \tag{38}$$

§ 5.6. Понятие о последовательном анализе. Критерий Вальда.

В примере 1 § 5.2 при рассмотрении задачи о выборе одной из двух гипотез о среднем нормальной модели была установлена связь между числом необходимых наблюдений и значениями вероятностей ошибок 1-го и 2-го родов α

и β (соотношение (14) § 5.2). Это число n^* можно рассчитать заранее (до проведения испытаний), и оно не зависит от исходов самих испытаний (от реализации выборки). Альтернативой к правилам проверки гипотез, основанным на выборках фиксированного объема, являются *последовательные* правила (критерии), когда вопрос о числе необходимых испытаний решают в процессе их проведения, и это число зависит от результатов испытаний, т. е. само является случайной величиной. Последовательные правила впервые были предложены А. Вальдом (1943), и их изучение составляет предмет важного раздела математической статистики — *последовательного анализа*. В данном параграфе на примере различения двух простых гипотез о распределении наблюдаемой случайной величины ξ мы кратко обсудим основные особенности последовательности критериев. С общей теорией последовательного анализа можно ознакомиться в [6] и [28].

1. Определение критерия Вальда

Пусть, как обычно, x_i — наблюдавшаяся реализация случайной величины ξ в i-м испытании, $i = 1, 2,$ и

$$L_{jn} = L(x_1 \quad , x_n; \theta_j) = \prod_{i=1}^{n} f_j(x_i)$$

— функция правдоподобия для первых n испытаний при гипотезе H_j $\theta = \theta_j$, $j = 0, 1$ (напомним, что $f_j(x)$ — плотность распределения ξ (или вероятность в дискретном случае) при гипотезе H_j, и испытания предполагаются независимыми). Согласно теории Неймана—Пирсона (см. § 5.2) наилучшая процедура проверки гипотезы H_0 против альтернативы H_1 состоит в принятии или отклонении гипотезы H_0 в зависимости от того, меньше или больше отношение правдоподобия L_{1n}/L_{0n} некоторой выбранной границы c, при этом объем выборки n фиксируется заранее и не зависит от наблюдений. Однако если объем выборки сделать зависящим от исходов испытаний, то можно добиться выигрыша в среднем числе испытаний до принятия окончательного решения. Эта идея реализуется в определяемом ниже *последовательном критерии отношения правдоподобия*, который принято называть по имени его автора *критерием Вальда*.

Зададим две положительные константы $A_0 < 1 < A_1$ и будем проверить испытания (наблюдения над ξ) до тех пор, пока не будет впервые нарушено какое-нибудь из неравенств

$$A_0 < \frac{L_{1n}}{L_{0n}} < A_1. \tag{1}$$

Если в момент прекращения испытаний (*момент остановки*) $L_{1n}/L_{0n} \leqslant A_0$, то принимается гипотеза H_0, если же $L_{1n}/L_{0n} \geqslant A_1$, то принимается гипотеза H_1. Такая последовательная процедура характеризуется, как обычно, вероятностями ошибок 1-го и 2-го родов $\alpha = \mathbf{P}\{H_1|H_0\}$ и $\beta = \mathbf{P}\{H_0|H_1\}$, а также средним числом $\mathbf{E}_j(\nu) = \mathbf{E}(\nu|H_j)$ испытаний ν до момента остановки

$(j = 0, 1)$. Если вероятности ошибок α и β заданы, то любой критерий с такими ошибками называют *критерием силы* (α, β). В классе критериев данной силы (α, β) предпочтительным является тот, который требует меньшего числа наблюдений. Критерий, минимизирующий одновременно $\mathbf{E}_0(\nu)$ и $\mathbf{E}_1(\nu)$ называют *оптимальным*. Таким свойством оптимальности и обладает критерий Вальда. В частности, этот критерий требует в среднем меньше испытаний, чем критерий Неймана—Пирсона с такими же вероятностями ошибок (α, β). Эти и другие свойства критерия Вальда рассмотрены ниже.

2. О числе испытаний до момента остановки в критерии Вальда

Пусть плотности $f_j(x) > 0$, $j = 0, 1$, и нетождественны. Это означает, что определена и невырождена случайная величина $Z = \ln\left(f_1(\xi)/f_0(\xi)\right)$; будем предполагать, что существуют $\mathbf{E}_\theta Z \neq 0$ и $\mathbf{D}_\theta Z = \sigma^2(\theta) > 0$ $(\theta = \theta_0, \theta_1)$. Обозначим также $Z_i = \ln\left(f_1(X_i)/f_0(X_i)\right)$, $i = 1, 2, \ldots$, где $X_1, X_2, \ldots$ последовательные независимые наблюдения над ξ. Тогда $Z_1, Z_2, \ldots$ являются независимыми наблюдениями над Z, и если $z_1, z_2, \ldots$ — наблюдавшиеся реализации этих величин, то в этих обозначениях процедура проверки гипотез заканчивается принятием H_0 или H_1 при первом n, при котором нарушается какое-нибудь из неравенств

$$a_0 = \ln A_0 < z_1 + \ldots + z_n < \ln A_1 = a_1, \quad a_0 < 0, \quad a_1 > 0. \tag{2}$$

Критерию можно дать следующую наглядную геометрическую интерпретацию. Рассмотрим блуждающую на плоскости частицу, находящуюся в начальный момент в начале координат, ордината которой в каждый целочисленный момент времени $t = 1, 2, \ldots$ получает приращение z_t. Тогда ордината частицы в момент $t = n$ будет равна $z_1 + \ldots + z_n$ и блуждание продолжается, пока траектория частицы впервые не выйдет за пределы полосы, ограниченной двумя горизонтальными прямыми на уровнях a_0 и a_1 (см. рис. 1).

Выход траектории на верхнюю (нижнюю) границу приводит к принятию гипотезы $H_1(H_0)$. Прежде всего возникает естественный вопрос, не может ли

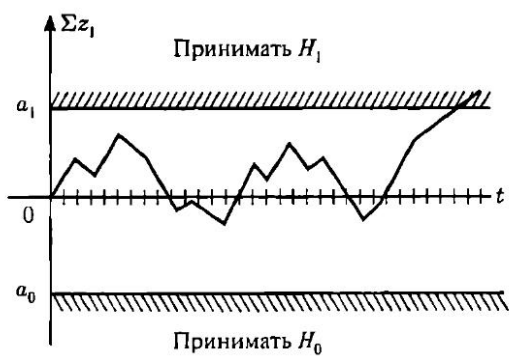


Рис. 1

блуждание продолжаться бесконечно долго, — в этом случае процесс проверки гипотез никогда бы не закончился. Ответ на этот вопрос дает следующее утверждение.

Теорема 1. Критерий Вальда с вероятностью 1 заканчивается за конечное число шагов, т. е. $\lim\limits_{n\to\infty} \mathbf{P}_\theta\{\nu > n\} = 0$ $(\theta = \theta_0, \theta_1)$.

Доказательство. Зафиксируем некоторое число r и введем случайные величины

$$\eta_1 = Z_1 + \quad + Z_r, \quad \eta_2 = Z_{r+1} + \quad + Z_{2r},$$

Тогда событие $\{\nu > rk\}$, эквивалентное событию $a_0 = Z_1 + \quad + Z_i < a_1$, $i \leqslant rk$, включается в событие $a_0 < \eta_1 + \quad + \eta_j < a_1$, $j \leqslant k$, которое, в свою очередь, включается в событие $|\eta_j| < b = a_1 - a_0$, $j \leqslant k$. Случайные величины η_j независимы и одинаково распределены, поэтому

$$\mathbf{P}_\theta\{\nu > rk\} \leqslant \mathbf{P}_\theta\{|\eta_j| < b,\ j \leqslant k\} = p^k(\theta), \tag{3}$$

где

$$p(\theta) = \mathbf{P}_\theta\{|\eta_1| < b\} = \mathbf{P}_\theta\{\eta_1^2 < b^2\}.$$

Но

$$\mathbf{E}_\theta\eta_1^2 \geqslant \mathbf{D}_\theta\eta_1 = \sum_{i=1}^{r} \mathbf{D}_\theta Z_i = r\sigma^2(\theta),$$

что может быть сделано больше, чем b^2, если выбрать

$$r > \max\left(\frac{b^2}{\sigma^2(\theta_0)},\ \frac{b^2}{\sigma^2(\theta_1)}\right).$$

Отсюда следует, что выбором r можно обеспечить неравенство $p(\theta) < 1$ (так как в противном случае распределение величины η^2 было бы сосредоточено на интервале $(0, b^2)$ и ее среднее было бы меньше, чем b^2). Переходя в неравенстве (3) к пределу при $k \to \infty$, получаем

$$\lim_{n\to\infty} \mathbf{P}_\theta\{\nu > n\} = \lim_{k\to\infty} \mathbf{P}_\theta\{\nu > rk\} = 0. \blacksquare$$

Итак, в критерии Вальда число испытаний ν до момента остановки с вероятностью 1 принимает конечное значение. Справедливо и более сильное утверждение, а именно: *все моменты случайной величины ν конечны*. В частности, конечны и средние значения $\mathbf{E}_0(\nu)$ и $\mathbf{E}_1(\nu)$; вычисление этих величин при заданных ошибках (α, β) рассмотрено ниже в п. 4.

3. О выборе границ в критерии Вальда

Установим теперь замечательную связь границ A_0 и A_1 в (1) с вероятностями ошибок α и β.

Теорема 2. Границы A_0 и A_1 критерия Вальда силы (α, β) удовлетворяют неравенством

$$A_0 \geqslant A_0' = \frac{\beta}{1-\alpha}, \quad A_1 \leqslant A_1' = \frac{1-\beta}{\alpha}; \tag{4}$$

при этом, если границы A_0 и A_1 в (1) заменить их оценками A_0' и A_1', то сила полученного критерия будет равна (α', β'), где

$$\alpha' \leqslant \frac{\alpha}{1-\beta}, \quad \beta' \leqslant \frac{\beta}{1-\alpha} \quad \text{и} \quad \alpha' + \beta' \leqslant \alpha + \beta. \tag{5}$$

Доказательство. Обозначим через $\mathfrak{X}_{0n}$ ($\mathfrak{X}_{1n}$) множество тех результатов наблюдений $(x_1, \quad , x_n)$, для которых процедура заканчивается на n-м шаге принятием H_0 (соответственно H_1), например,

$$\mathfrak{X}_{0n} = \left\{(x_1 \quad , x_n) \quad A_0 < \frac{L_{1k}}{L_{0k}} < A_1, \ k = 1, \quad , n-1, \ \frac{L_{1n}}{L_{0n}} \leqslant A_0\right\}.$$

В силу теоремы 1, для $\theta = \theta_0, \theta_1$

$$\sum_{n=1}^{\infty} \mathbf{P}_\theta\{\nu = n\} = \sum_{n=1}^{\infty} \mathbf{P}_\theta\{\mathfrak{X}_{0n}\} + \sum_{n=1}^{\infty} \mathbf{P}_\theta\{\mathfrak{X}_{1n}\} = 1.$$

Далее имеем

$$\alpha = \mathbf{P}\{H_1|H_0\} = \sum_{n=1}^{\infty} \mathbf{P}_{\theta_0}\{\mathfrak{X}_{1n}\} \leqslant \frac{1}{A_1} \sum_{n=1}^{\infty} \mathbf{P}_{\theta_1}\{\mathfrak{X}_{1n}\} =$$
$$= \frac{1}{A_1}(1 - \mathbf{P}\{H_0|H_1\}) = \frac{1-\beta}{A_1},$$

так как в точках множества $\mathfrak{X}_{1n}$ выполняется неравенство $L_{0n} \leqslant L_{1n}/A_1$. Аналогично получаем

$$\beta = \mathbf{P}\{H_0|H_1\} = \sum_{n=1}^{\infty} \mathbf{P}_{\theta_1}\{\mathfrak{X}_{0n}\} \leqslant A_0 \sum_{n=1}^{\infty} \mathbf{P}_{\theta_0}\{\mathfrak{X}_{0n}\} =$$
$$= A_0(1 - \mathbf{P}\{H_1|H_0\}) = A_0(1-\alpha),$$

поскольку в точках множества $\mathfrak{X}_{0n}$ выполняется неравенство $L_{1n} \leqslant A_0 L_{0n}$. Тем самым неравенства (4) доказаны.

Рассмотрим теперь критерий Вальда с границами A_0' и A_1', определенными в (4), и пусть α' и β' — его ошибки. Тогда на основании доказанного должны выполняться неравенства

$$\frac{\beta}{1-\alpha} \geqslant \frac{\beta'}{1-\alpha'}, \quad \frac{1-\beta}{\alpha} \leqslant \frac{1-\beta'}{\alpha'}.$$

Отсюда имеем

$$\beta' \leqslant \frac{(1-\alpha')\beta}{1-\alpha} \leqslant \frac{\beta}{1-\alpha}, \quad \alpha' \leqslant \frac{(1-\beta')\alpha}{1-\beta} \leqslant \frac{\alpha}{1-\beta}.$$

Складывая неравенства

$$\beta'(1-\alpha) \leqslant \beta(1-\alpha') \quad \text{и} \quad -\alpha(1-\beta') \leqslant -\alpha'(1-\beta),$$

получаем, что

$$\beta' - \alpha \leqslant \beta - \alpha' \quad \text{или} \quad \alpha' + \beta' \leqslant \alpha + \beta. \blacksquare$$

Комментарии

На практике теорему 2 используют следующим образом. Если требуется построить критерий Вальда силы (α, β), то границы в (1) полагают равными соответственно A_0' и A_1', как определено в (4). В этом случае последнее неравенство в (5) гарантирует, что сумма действительных ошибок α' и β' такого критерия не превосходит суммы $\alpha+\beta$ заданных ошибок. Далее, обычно α и β не превышают 0,1, поэтому из остальных неравенств (5) следует, что разность между действительными и заданными ошибками в этих случаях незначительна. Таким образом, можно утверждать, что такой специальный выбор границ в критерии Вальда приводит к более сильному критерию.

Отметим в этой связи интересную особенность последовательного критерия по сравнению с обычными критериями, основанными на выборках фиксированного объема. В обычном критерии при определении критической границы при заданном уровне значимости и вычислении мощности надо знать распределение тестовой статистики и при нулевой гипотезе, и при альтернативе. При расчете же критерия Вальда не возникает проблемы отыскания распределений. Действительно, границы A_0' и A_1' зависят только от заданных ошибок α и β, а отношение L_{1n}/L_{0n} можно вычислить на основании наблюдаемых данных. Необходимость отыскания распределений при использовании последовательного критерия возникает только при исследовании числа наблюдений ν до момента остановки.

4. О среднем числе наблюдений в критерии Вальда

Установим предварительно одно интересное равенство, называемое *тождеством Вальда*. Пусть $S_\nu = Z_1 + \quad + Z_\nu$ — ордината блуждающей частицы в момент остановки (см. п. 2), тогда

Тождество Вальда

$$\mathbf{E}_\theta(S_\nu) = \mathbf{E}_\theta(Z)\mathbf{E}_\theta(\nu). \tag{6}$$

Доказательство. Введем случайные величины $Y_1, Y_2, \quad$ где $Y_n = 1$, если решение не принято до n-го шага, и $Y_n = 0$ в противном случае. Тогда, очевидно, Y_n есть функция только $Z_1, \ldots, Z_{n-1}$ и, следовательно, не зависит от Z_n, и можно записать

$$S_\nu = \sum_{n=1}^{\infty} Y_n Z_n.$$

Отсюда по свойству математического ожидания ($\mathbf{E}(\xi\eta)) = \mathbf{E}(\xi)\mathbf{E}(\eta)$, если случайные величины ξ и η независимы) имеем

$$\mathbf{E}_\theta(S_\nu) = \sum_{n=1}^{\infty} \mathbf{E}_\theta(Y_n Z_n) = \mathbf{E}_\theta(Z) \sum_{n=1}^{\infty} \mathbf{E}_\theta(Y_n) = \mathbf{E}_\theta(Z) \sum_{n=1}^{\infty} \mathbf{P}_\theta\{Y_n = 1\}. \quad (7)$$

Событие $\{Y_n \quad 1\}$ эквивалентно событию $\{\nu \geqslant n\}$, а так как для любой целочисленной положительной случайной величины $\varkappa$ (каппа) справедлива формула

$$\mathbf{E}\varkappa = \sum_{n=1}^{\infty} \mathbf{P}\{\varkappa \geqslant n\} \quad \text{(показать!)},$$

то из (7) следует тождество Вальда (6). ■

Пусть теперь заданы значения ошибок α и β. Как показано в п. 3, практически при малых α и β границы в (2) можно заменить соответственно на

$$a_0' = \ln A_0' = \ln \frac{\beta}{1-\alpha} \quad \text{и} \quad a_1' = \ln A_1' = \ln \frac{1-\beta}{\alpha}. \quad (8)$$

При малых α и β ширина интервала (a_0', a_1') велика, а величина $\mathbf{E}_\theta(Z)$ от α и β не зависит и, по предположению (см. п. 2), конечна. Поэтому можно пренебречь эффектом превышения суммой S_ν границы в момент остановки, т. е. положить $S_\nu \approx a_j'$, если принимается гипотеза H_j, $j = 0, 1$. Другими словами, можно считать среднее значение суммы S_ν приближенно равным соответствующей границе a_0' или a_1'. Из этих рассуждений и тождества (6) имеем

$$\mathbf{E}_{\theta_0}(\nu)\mathbf{E}_{\theta_0}(Z) = \mathbf{E}_{\theta_0}(S_\nu) \approx a_0'\mathbf{P}\{H_0|H_0\} + a_1'\mathbf{P}\{H_1|H_0\} = (1-\alpha)a_0' + \alpha a_1'.$$

Аналогично

$$\mathbf{E}_{\theta_1}(\nu)\mathbf{E}_{\theta_1}(Z) = \mathbf{E}_{\theta_1}(S_\nu) \approx a_0'\mathbf{P}\{H_0|H_1\} + a_1'\mathbf{P}\{H_1|H_1\} = \beta a_0' + (1-\beta)a_1'.$$

Таким образом, $\mathbf{E}_j(\nu) = \mathbf{E}_{\theta_j}(\nu)$ можно вычислять по следующим приближенным формулам:

$$\mathbf{E}_0(\nu) \approx \frac{(1-\alpha)a_0' + \alpha a_1'}{\mathbf{E}_0(Z)}, \quad \mathbf{E}_1(\nu) \approx \frac{\beta a_0' + (1-\beta)a_1'}{\mathbf{E}_1(Z)}, \quad (9)$$

где a_0' или a_1' определены в (8).

Пример 1 (критерий Вальда для нормальной модели). Построим и рассчитаем критерий Вальда для задачи различения двух простых гипотез о среднем нормальной модели, которая рассмотрена в примере 1 § 5.2 с позиций теории Неймана—Пирсона. Здесь

$$z_i = \ln \frac{f_1(x_i)}{f_0(x_i)} = \frac{(x_i-\theta_0)^2 - (x_i-\theta_1)^2}{2\sigma^2} = \frac{\theta_1-\theta_0}{\sigma^2}\left(x_i - \frac{\theta_0+\theta_1}{2}\right) \quad (10)$$

и критерий Вальда силы (α, β) определяется в данном случае следующим образом. Наблюдения продолжаются до тех пор, пока впервые не нарушится какое-нибудь из неравенств

$$\frac{\sigma^2}{\theta_1-\theta_0}\ln\frac{\beta}{1-\alpha} < \sum_{i=1}^{n} x_i - \frac{n}{2}(\theta_0+\theta_1) < \frac{\sigma^2}{\theta_1-\theta_0}\ln\frac{1-\beta}{\alpha};$$

если нарушается левое (правое) неравенство, то принимается гипотеза H_0 (H_1). Вычислим среднее число наблюдений до момента остановки. Из (10) имеем

$$\mathbf{E}_j(Z) = \frac{\theta_1-\theta_0}{\sigma^2}\left(\theta_j - \frac{\theta_0+\theta_1}{2}\right) = \begin{cases} -\dfrac{(\theta_1-\theta_0)^2}{2\sigma^2} & \text{при} \quad j=0, \\ \dfrac{(\theta_1-\theta_0)^2}{2\sigma^2} & \text{при} \quad j=1. \end{cases}$$

Отсюда и из (9) находим

$$\mathbf{E}_0(\nu) \approx -\frac{2\sigma^2}{(\theta_1-\theta_0)^2}\left[(1-\alpha)\ln\frac{\beta}{1-\alpha} + \alpha\ln\frac{1-\beta}{\alpha}\right],$$
$$\mathbf{E}_1(\nu) \approx \frac{2\sigma^2}{(\theta_1-\theta_0)^2}\left[\beta\ln\frac{\beta}{1-\alpha} + (1-\beta)\ln\frac{1-\beta}{\alpha}\right],$$

Для критерия Неймана—Пирсона с теми же ошибками α и β необходимое число испытаний n^* определяется формулой (14) § 5.2. Отсюда имеем

$$\begin{aligned} \frac{\mathbf{E}_0(\nu)}{n^*} &\approx -\frac{2}{(\zeta_\alpha+\zeta_\beta)^2}\left[(1-\alpha)\ln\frac{\beta}{1-\alpha} + \alpha\ln\frac{1-\beta}{\alpha}\right], \\ \frac{\mathbf{E}_1(\nu)}{n^*} &\approx \frac{2}{(\zeta_\alpha+\zeta_\beta)^2}\left[\beta\ln\frac{\beta}{1-\alpha} + (1-\beta)\ln\frac{1-\beta}{\alpha}\right] \end{aligned} \tag{11}$$

Если $\alpha=\beta$, то $\zeta_\alpha \quad \zeta_\beta$, $\ln\dfrac{\beta}{1-\alpha} = -\ln\dfrac{1-\beta}{\alpha}$ и из соотношений (11) следует, что

$$\frac{\mathbf{E}_0(\nu)}{n^*} \approx \frac{\mathbf{E}_1(\nu)}{n^*} \approx \frac{1-2\alpha}{2\zeta_\alpha^2}\ln\frac{1-\alpha}{\alpha} \equiv \psi(\alpha). \tag{12}$$

При $\alpha = 0{,}05$ квантиль $\zeta_\alpha = -1{,}645$ и $\psi(\alpha) = 0{,}49$, т. е. экономия наблюдений при последовательной процедуре составляет приблизительно 50 %.

Более определенные выводы о сравнении критериев Вальда и Неймана—Пирсона можно сделать при $\alpha=\beta\to 0$. В этом случае $\zeta_\alpha \to -\infty$ и, если воспользоваться известной асимптотической формулой для нормальной функции распределения $\Phi(x)$ [3, с. 10]:

$$\Phi(-x) = \frac{e^{-x^2/2}}{x\sqrt{2\pi}}(1+o(1)), \quad x\to\infty, \tag{13}$$

то на основании (13) можно записать асимптотическое равенство (вместо точного равенства $\Phi(\zeta_\alpha) = \alpha$)

$$\frac{1 + o(1)}{\zeta_\alpha \sqrt{2\pi}} e^{-\zeta_\alpha^2/2} = \alpha.$$

Решением этого уравнения, как нетрудно убедиться непосредственной подстановкой (проделайте это самостоятельно!), является

$$\zeta_\alpha^2 = 2 \ln \frac{1}{\alpha} - \ln \ln \frac{1}{\alpha} - \ln 4\pi + o(1).$$

Отсюда и из (12) следует (это мы оставляем читателю в качестве упражнения на асимптотические вычисления), что при $\alpha \to 0$

$$\psi(\alpha) = \frac{1}{4}\left(1 + \frac{\ln(\ln(1/\alpha)) - \ln 4\pi + o(1)}{2 \ln(1/\alpha)}\right). \tag{14}$$

Представление (14) говорит о том, что при ошибках $\alpha = \beta \to 0$ последовательная процедура в среднем приблизительно в четыре раза экономнее оптимальной процедуры с фиксированным объемом выборки. •



Последовательный критерий отношения правдоподобия А. Вальда (1943) стал важным открытием в статистике, позволившим (в типичных ситуациях) на 50 % уменьшить среднее число наблюдений (при тех же вероятностях ошибок). Насколько это важно для приложений говорит тот факт, что в годы второй мировой войны открытие Вальда было объявлено «секретным», и его основная книга «Последовательный анализ» была опубликована лишь в 1947 г. Год спустя Вальд и Дж. Волфовиц доказали, что методы, отличные от последовательного критерия отношения правдоподобия, не дают такого уменьшения числа элементов выборки. В настоящее время теория последовательного анализа Вальда является частным случаем общей теории стохастических процессов с остановкой, изложенной в [28].

Упражнения

Величайшая цель образования — не знание, а действие.

Герберт Спенсер [3]

1 Пусть $X = (X_1, \ldots, X_n)$ — выборка из показательного распределения $\Gamma(\theta, 1)$ (см. п. 3 § 1.2), о котором требуется проверить простую гипотезу H_0 $\theta = \theta_0$ против альтернативы H_1 $\theta = \theta_1$. Убедиться в том, что здесь наиболее мощный критерий выражается через достаточную статистику

$$T = \sum_{i=1}^{n} X_i$$

[3] Спенсер Герберт (1820–1903) — английский философ и социолог.

и имеет при $\theta_0 < \theta_1$ вид

$$\mathfrak{X}_{1\alpha}^{*+} = \left\{ T \geqslant \frac{\theta_0}{2} \chi^2_{1-\alpha,\,2n} \right\},$$

а при $\theta_0 > \theta_1$ — вид

$$\mathfrak{X}_{1\alpha}^{*-} = \left\{ T \leqslant \frac{\theta_0}{2} \chi^2_{\alpha,\,2n} \right\}$$

Получить также следующие выражения для мощностей:

$$W(\theta_1; \mathfrak{X}_{1\alpha}^{*+}) = 1 - F_{2n}\left(\frac{\theta_0}{\theta_1} \chi^2_{1-\alpha,\,2n} \right), \quad W(\theta_1, \mathfrak{X}_{1\alpha}^{*-}) = F_{2n}\left(\frac{\theta_0}{\theta_1} \chi^2_{\alpha,\,2n} \right),$$

где $F_{2n}(t)$ — функция распределения закона $\chi^2(2n)$.

◀ *Указание.* Воспользоваться соотношением $\mathcal{L}_\theta(2T/\theta) = \chi^2(2n)$, вытекающим из свойств гамма-распределения. ▶

2 Пусть для распределения Коши $\mathcal{K}(\theta)$ с плотностью

$$f(x; \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \quad x \in R^1$$

проверяется гипотеза H_0 $\theta = 0$ против альтернативы H_1 $\theta = 1$. Показать, что при уровне значимости

$$\alpha = \frac{1}{2} - \frac{1}{\pi} \operatorname{arctg} \frac{1}{2} \approx 0{,}352$$

наиболее мощный критерий по одному наблюдению имеет вид

$$\mathfrak{X}_{1\alpha}^{*} = \left\{ X > \frac{1}{2} \right\}$$

и его мощность равна

$$\frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} \frac{1}{2} \approx 0{,}648.$$

Если же

$$\alpha = \frac{1}{\pi}(\operatorname{arctg} 3 - \operatorname{arctg} 1) \approx 0{,}148,$$

то критерий имеет вид

$$\mathfrak{X}_{1\alpha}^{*} = \{1 \leqslant X \leqslant 3\}$$

и имеет мощность

$$\frac{1}{\pi} \operatorname{arctg} 2 \approx 0{,}352.$$

3 Убедиться в том, что наиболее мощный критерий для различения двух простых гипотез о симметричном относительно нуля распределении наблюдаемой случайной величины ξ H_0 $\mathcal{L}(\xi) = \mathcal{U}(-a, a)$ и H_1 $\mathcal{L}(\xi) = \mathcal{N}(0, \sigma^2)$ (a и σ известны) имеет для больших выборок следующую асимптотическую форму

$$\mathfrak{X}_{1\alpha}^{*} = \left\{ \sum_{i=1}^{n} X_i^2 \leqslant \frac{n}{3} a^3 + \zeta_\alpha \frac{2a^2}{3} \sqrt{\frac{n}{5}} \right\}, \quad \Phi(\zeta_\alpha) = \alpha.$$

◀ *Указание.* Воспользоваться центральной предельной теоремой при отыскании распределения тестовой статистики. ▶

4 В последовательности независимых испытаний с двумя исходами вероятность «успеха» равна p. Построить критерий проверки гипотезы H_0 $p = 0$ против альтернативы H_1 $p = 0{,}01$ и определить наименьший объем выборки, при котором вероятности ошибок 1-го и 2-го родов не превышают 0,01.

5 По схеме примера 4 § 5.2 построить и рассчитать критерий Неймана—Пирсона различения двух простых гипотез в биномиальной модели $\mathrm{Bi}(k, \theta)$. Показать, что для выборки $X = (X_1, \ldots, X_n)$ большого объема этот критерий в задаче (H_0 $\theta = \theta_0$, H_1^+ $\theta = \theta_1 > \theta_0$) имеет асимптотически вид

$$\mathfrak{X}_{1\alpha}^{*+} = \left\{ \sum_{i=1}^{n} X_i \geqslant kn\theta_0 - \zeta_\alpha \sqrt{kn\theta_0(1-\theta_0)} \right\}, \quad \Phi(\zeta_\alpha) = \alpha,$$

и его мощность $W_n(\theta_1)$ при «близких» альтернативах вида

$$\theta_1 = \theta_{1n} = \theta_0 + \frac{\delta}{\sqrt{n}}, \quad \delta > 0,$$

удовлетворяет соотношению

$$\lim_{n\to\infty} W_n(\theta_{1n}) = \Phi\left(\delta \sqrt{\frac{k}{\theta_0(1-\theta_0)}} + \zeta_\alpha \right).$$

◀ *Указание.* Воспользоваться теоремой Муавра—Лапласа. ▶

6 Чтобы проверить гипотезу $H_0 : \theta = \theta_0$ против альтернативы H_1 $\theta = \theta_1$ ($0 < \theta_0 < \theta_1 < 1$) в схеме Бернулли с неизвестной вероятностью «успеха» θ, осуществлен эксперимент, в котором наблюдается число «успехов», предшествующих первому «неуспеху». Построить наиболее мощный критерий при уровне значимости $\alpha = \theta_0^s$, где $s \geqslant 1$ — заданное целое число, и убедиться в том, что вероятность ошибки 2-го рода этого критерия $\beta = 1 - \theta_1^s$

◀ *Указание.* Использовать тот факт, что наблюдаемая случайная величина X имеет распределение $\mathcal{L}_\theta(X) = \overline{\mathrm{Bi}}(1, \theta)$ (см. п. 2 § 1.1). ▶

7 Доказать формулу (3) § 5.3 для мощности критерия обратного биномиального выбора.

8 Убедиться в том, что построенный в упр. 5 критерий Неймана—Пирсона для биномиальной модели $\mathrm{Bi}(k, \theta)$ является р. н. м. критерием в задаче (H_0 $\theta \leqslant \theta_0$, H_1^+ $\theta > \theta_0$).

◀ *Указание.* Воспользоваться свойством модели с монотонным отношением правдоподобия. ▶

9 Пусть в схеме Бернулли с неизвестной вероятностью «успеха» θ испытания продолжаются до получения r-го «неуспеха» и T_r — наблюдаемое число «успехов». Построить р. н. м. критерий проверки гипотезы H_0 $\theta \leqslant \theta_0$ против альтернативы H_1^+ $\theta > \theta_0$ и показать, что при $r \to \infty$ его асимптотическая форма есть

$$\mathfrak{X}_{1\alpha}^{*+} = \left\{ T_r \geqslant \frac{r\theta_0 - \zeta_\alpha \sqrt{r\theta_0}}{1-\theta_0} \right\}, \quad \Phi(\zeta_\alpha) = \alpha.$$

◀ *Указание.* Воспользоваться свойством модели с монотонным отношением правдоподобия и применить центральную предельную теорему к сумме $T_r = X_1 + \quad + X_r$, где $X_1, \ldots, X_r$ — независимые геометрические $\overline{\mathrm{Bi}}(1, \theta)$ случайные величины. ▶

10 Убедиться в том, что построенные в упр. 1 критерии Неймана—Пирсона $\mathfrak{X}^{*+}_{1\alpha}$ и $\mathfrak{X}^{*-}_{1\alpha}$ являются одновременно р. н. м.к. в задачах соответственно $(H_0 \quad \theta \leqslant \theta_0,$ $H_1^+ \quad \theta > \theta_0)$ и $(H_0 \quad \theta \geqslant \theta_0, H_1^- \quad \theta < \theta_0)$.

11 (*выборочный контроль*). Пусть партия из N изделий содержит неизвестное число θ дефектных изделий, $\theta \in \{0, 1, \ldots, N\}$ Чтобы проверить гипотезу $H_0 \quad \theta \leqslant \theta_0$ против альтернативы $H_1^+ \quad \theta > \theta_0$, берут на контроль случайную выборку из n $(< N)$ изделий и каждое из них проверяют. Основываясь на статистике T — обнаруженное в выборке число дефектных изделий, построить р. н. м. критерий.

◀ *Указание.* Убедиться в том, что распределение статистики T (гипергеометрическое распределение $H(\theta, N-\theta, n)$, см. п. 4 § 1.1) имеет монотонное отношение правдоподобия. ▶

12 Рассматривается задача $(H_0 \quad \theta = \theta_0, \ H_1 \quad \theta \neq \theta_0)$ для нормальной модели с неизвестной дисперсией $\mathcal{N}(\mu, \theta^2)$. Используя результаты примера 2 § 5.2 и прием объединения односторонних критических областей, показать, что р. н. м. несмещенный критерий имеет в данном случае вид

$$\widetilde{\mathfrak{X}}_{1\alpha} = \{T \leqslant \theta_0^2 \chi^2_{\alpha_1, n}\} \bigcup \{T \geqslant \theta_0^2 \chi^2_{1-\alpha_2, n}\},$$

где $t_1 = \chi^2_{\alpha_1, n}$ и $t_2 = \chi^2_{1-\alpha_2, n}$ определяются условиями (ниже $k_n(t) = F_n'(t)$ — плотность распределения $\chi^2(n)$)

$$t_1 k_n(t_1) = t_2 k_n(t_2), \quad \alpha_1 + \alpha_2 = \alpha.$$

Значения $(\chi^2_{\alpha_1, n}, \chi^2_{1-\alpha_2, n})$ для $\alpha = 0{,}05$ и $n = 2, 5, 10, 20, 30, 40, 60$ см. в [12, с. 275].

13 Рассматривается задача $(H_0 \quad \theta = \theta_0, \ H_1 \quad \theta \neq \theta_0)$ для показательной модели $\Gamma(\theta, 1)$. Используя результаты упр. 1 и прием объединения односторонних критических областей (см. также пример 3 § 5.3), показать, что р. н. м. несмещенный критерий имеет в данном случае вид

$$\widetilde{\mathfrak{X}}_{1\alpha} = \left\{T \leqslant \frac{\theta_0}{2} \chi^2_{\alpha_1, 2n}\right\} \bigcup \left\{T \geqslant \frac{\theta_0}{2} \chi^2_{1-\alpha_2, 2n}\right\},$$

где $t_1 = \chi^2_{\alpha_1, 2n}$ и $t_2 = \chi^2_{1-\alpha_2, 2n}$ определяются условиями (см. упр. 12)

$$t_1 k_{2n}(t_1) = t_2 k_{2n}(t_2), \quad \alpha_1 + \alpha_2 = \alpha.$$

14 Проверяется гипотеза $H_0 \quad \theta = \theta_0$ против общей альтернативы $H_1 \quad \theta \neq \theta_0$ для модели $\mathrm{Bi}(k, \theta)$. Показать, что для больших выборок л. н. м. несмещенный критерий (18) § 5.3 имеет вид

$$\left\{\frac{|T - kn\theta_0|}{\sqrt{kn\theta_0(1-\theta_0)}} \geqslant t_{\alpha/2}\right\}, \quad T = \sum_{i=1}^{n} X_i,$$

и его мощность при альтернативах вида $\theta = \theta_n = \theta_0 + \delta/\sqrt{n}$ удовлетворяет предельному соотношению

$$\lim_{n\to\infty} W(\theta_n) = \Phi\left(-\delta\sqrt{\frac{k}{\theta_0(1-\theta_0)}} - t_{\alpha/2}\right) + \Phi\left(\delta\sqrt{\frac{k}{\theta_0(1-\theta_0)}} - t_{\alpha/2}\right).$$

◀ *Указание.* См. табл. 1 в § 3.2. ▶

15 Убедиться в том, что в задаче ($H_0 \quad \theta = \theta_0$, $H_1 \quad \theta \neq \theta_0$) для пуассоновской модели $\Pi(\theta)$ л. н. м. несмещенный критерий для случая больших выборок имеет вид

$$\left\{ \frac{1}{\sqrt{n\theta_0}} \left| \sum_{i=1}^{n} X_i - n\theta_0 \right| \geqslant t_{\alpha/2} \right\},$$

а его мощность при локальных альтернативах вида $\theta = \theta_n = \theta_0 + \delta/\sqrt{n}$ удовлетворяет предельному соотношению

$$\lim_{n\to\infty} W(\theta_n) = \Phi\left(-\frac{\delta}{\sqrt{\theta_0}} - t_{\alpha/2}\right) + \Phi\left(\frac{\delta}{\sqrt{\theta_0}} - t_{\alpha/2}\right).$$

◀ *Указание.* Использовать общий результат (18) § 5.3 и табл. 1 в § 3.2. ▶

16 Основываясь на результате, полученном в примере 3 § 2.5 (соотношение (17)), убедиться в том, что для проверки гипотезы о равенстве дисперсий двух нормальных моделей можно использовать критерий

$$\mathcal{X}_{1\alpha} = \left\{ (\underline{x}, \underline{y}) : \frac{n(m-1)}{m(n-1)} \frac{S^2(\underline{x})}{S^2(\underline{y})} \leqslant F_{\frac{\alpha}{2}, n-1, m-1} \quad \text{либо} \quad \frac{n(m-1)}{m(n-1)} \frac{S^2(\underline{x})}{S^2(\underline{y})} \geqslant F_{1-} \qquad {}_{1} \right\}.$$

17 Убедиться в том, что в условиях упр. 63 к гл. 3 критерий проверки гипотезы однородности $H_0 \quad \tau = \theta_2/\theta_1 = 1$ имеет вид

$$\mathcal{X}_{1\alpha} = \left\{ (\underline{x}, \underline{y}) \quad \frac{\bar{x}}{\bar{y}} \leqslant F_{\alpha/2, 2n, 2m} \quad \text{либо} \quad \frac{\bar{x}}{\bar{y}} \geqslant F_{1-\frac{\alpha}{2}, 2n, 2m} \right\}$$

и вычислить его мощность при произвольной альтернативе $\tau \neq 1$.

18 Обращая построенный в упр. 64 к гл. 3 доверительный интервал, получить следующий критерий для гипотезы $H_0 \quad \theta = \theta_0$ при альтернативе $H_1 \quad \theta \neq \theta_0$:

$$\mathcal{X}_{1\alpha} = \left\{ \underline{x} \quad x_{(1)} \leqslant \theta_0 \quad \text{либо} \quad x_{(1)} \geqslant \theta_0 - \frac{1}{n} \ln \alpha \right\}$$

Вычислить функцию мощности этого критерия и убедиться в его несмещенности.

19 Используя результат примера 1 § 3.8, показать, что критерий проверки гипотезы $H_0 \quad \theta = \theta_0$ для равномерного распределения $\mathcal{U}(0, \theta)$ имеет вид

$$\mathcal{X}_{1\alpha} = \left\{ \underline{x} \quad x_{(n)} \leqslant \theta_0 \alpha^{1/n} \quad \text{либо} \quad x_{(n)} \geqslant \theta_0 \right\};$$

вычислить его функцию мощности и убедиться в том, что он несмещенный.

20 Обращая построенный в упр. 65 к гл. 3 доверительный интервал, получить, что критерий для проверки гипотезы $H_0 \quad \theta = \theta_0$ в модели Вейбулла $W(0, \lambda, \theta^{1/\lambda})$ имеет вид

$$\mathcal{X}_{1\alpha} = \left\{ T \leqslant \frac{\theta_0}{2} \chi^2_{\alpha_1, 2n} \right\} \bigcup \left\{ T \geqslant \frac{\theta_0}{2} \chi^2_{1-\alpha_2, 2n} \right\}, \quad \alpha_1 + \alpha_2 = \alpha.$$

Чтобы получить несмещенный критерий, величины $\chi^2_{\alpha_1, 2n}$ и $\chi^2_{1-\alpha_2, 2n}$ выбираются так же, как в упр. 13.

21 Обращая построенные в упр. 66 к гл. 3 доверительные интервалы для параметров двухпараметрической показательной модели $W(\theta_1, 1, \theta_2)$, построить критерии проверки гипотез $H_0 \quad \theta_1 = \theta_{10}$ и $H_0 \quad \theta_2 = \theta_{20}$.

22 Построить КОП для гипотезы $H_0 \quad \theta = \theta_0$ в биномиальной модели $\mathrm{Bi}(k, \theta)$ и убедиться в том, что его асимптотический (для больших выборок) вариант совпадает с локальным наиболее мощным критерием, построенным в упр. 14

◀ *Указание.* Воспользоваться общей теорией КОП для полиномиального распределения (пример 5 § 5.4). ▶

23 Построить КОП для гипотезы H_0 $\theta = \theta_0$ в пуассоновской модели $\Pi(\theta)$ и убедиться в том, что в случае больших выборок он совпадает с локальным наиболее мощным критерием, указанным в упр. 15.

◀ *Указание.* Воспользоваться тем, что при $n \to \infty$ предельные распределения при гипотезе H_0 статистик $-2\ln\lambda_n$ и $Q_n^{(2)} = V_n^2(\theta_0)/ni(\theta_0)$ в (18) § 5.4 совпадают. ▶

24 Пусть наблюдается реализация длины n простой однородной эргодической цепи Маркова с множеством $\mathcal{E}_0^2 = \{(ij)\}$ пар разрешенных переходов $i \to j$ и $d^0 = |\mathcal{E}_0^2|$. Показать, что КОП для простой гипотезы H^0, фиксирующей матрицу переходных вероятностей $\mathbf{P}^0 = \|p_{ij}^\circ\|_1^N$ где $p_{ij}^\circ > 0 \Leftrightarrow (ij) \in \mathcal{E}_0^2$, имеет асимптотически (при $n \to \infty$) вид (сравни с критерием (7) § 5.5)

$$\left\{2\sum_{(ij)\in\mathcal{E}_0^2}\nu_{ij}\ln\frac{\nu_{ij}}{\nu_{i*}p_{ij}^\circ} > \chi^2_{1-\alpha,\,d^0-N}\right\}.$$

25 Пусть гипотеза H^0 означает, что случайные величины $\xi_1, \ldots, \xi_n$ независимы и одинаково распределены на множестве $\mathcal{E} = \{1, 2, \ldots, N\}$, а при гипотезе $\mathcal{H}_1$ эти случайные величины связаны в простую однородную цепь Маркова, в которой возможны любые переходы $i \to j$. Показать, что при $n \to \infty$ КОП для проверки гипотезы независимости H^0 внутри гипотезы $\mathcal{H}_1$ задается критической областью

$$\left\{2\sum_{i,j=1}^{N}\nu_{ij}\ln\frac{n\nu_{ij}}{\nu_{i*}\nu_{*j}} > \chi^2_{1-\alpha,\,(N-1)^2}\right\}$$

(сравни с критерием (14) § 5.5).

Глава 6

(Специальная)

Линейная регрессия и метод наименьших квадратов

Статистика — это физика чисел.

П. Диаконис

Этой главой мы начинаем серию специальных глав, посвященных рассмотрению специфических задач, решаемых математической статистикой, которые объединяются тем или иным общим признаком (связанным чаще всего с областью их практических приложений), и выделяемых в отдельные разделы (теории). Настоящая глава посвящена *регрессионному анализу* (см. пример 3 общего Введения) — чрезвычайно важному прикладному разделу современной математической статистики. В ней рассматривается практически важный случай неодинаково распределенных наблюдений с одинаковыми дисперсиями, когда их средние имеют вид линейных функций от неизвестных параметров (так называемая *схема линейной регрессии*). Излагаются классический *метод наименьших квадратов* (*метод Гаусса—Маркова*) для оценивания параметров модели и его оптимальные свойства, рассматриваются вопросы его практического применения. В рамках нормальной модели дается полное решение задач доверительного оценивания произвольных линейных функций параметров регрессии и проверки линейных гипотез для них. Дается применение модели нормальной регрессии к задачам дисперсионного анализа. Излагаются элементы теории статистической регрессии и корреляции.

§ 6.1. Модель линейной регрессии

До сих пор рассматривались ситуации, когда исходные статистические данные (выборка) $\underline{X} = (X_1, \ldots, X_n)$ представляли собой результаты повторных независимых наблюдений над некоторой случайной величиной ξ (исключением является § 5.5, в котором наблюдения предполагались связанными в цепь Маркова), в такой случайной выборке компоненты X_i независимы и одинаково распределены: $F_{X_i} = F_\xi$, $i = 1, \ldots, n$. Однако в реальных ситуациях такие предположения выполняются далеко не всегда. Например, имеется обширный круг прикладных задач, в математической модели которых естественно предположение, что средние значения наблюдений X_i представляют собой линейные функции от неизвестных параметров $\underline{\beta} = (\beta_1, \ldots, \beta_k)$, а вторые

моменты удовлетворяют тем или иным дополнительным условиям. В таких случаях говорят о модели *линейной регрессии*.

Более конкретно, здесь речь идет об изучении таких случайных экспериментов, когда на их исход влияют некоторые неслучайные переменные (*факторы*) $z_1, \ldots, z_k$, значения которых могут меняться от опыта к опыту. Например, $z_1, \ldots, z_k$ могут быть входными характеристиками прибора, при задании которых наблюдается соответствующий «отклик» как исход эксперимента. В других случаях значения факторов для каждого опыта могут быть обусловлены не зависящими от исследователя причинами (как правило, это имеет место, например, в экономических исследованиях). Но в любом случае статистические данные состоят при этом из множества наблюдавшихся значений «откликов» $X_1, \ldots, X_n$ и соответствующих значений факторов, т. е. имеют вид $(x_i; z_1^{(i)}, \ldots, z_k^{(i)})$, $i = 1, \ldots, n$; при этом всегда предполагается, что $k < n$.

Часто можно считать, что факторы $z_1, \ldots, z_k$ оказывают влияние только на среднее значение «отклика», при этом математическое ожидание исхода i-го опыта имеет вид

$$\mathbf{E}X_i = \sum_{j=1}^{k} \beta_j z_j^{(i)} = \underline{z}^{(i)'}\underline{\beta}, \quad \underline{z}^{(i)} = (z_1^{(i)} \ldots z_k^{(i)}),$$

т. е. является линейной функцией неизвестных параметров $\underline{\beta} = (\beta_1, \ldots, \beta_k)$ (напомним, что при матричных операциях векторы понимаются как векторстолбцы). Другими словами, случайную величину X_i, описывающую результат i-го опыта, можно представить в виде

$$X_i = \underline{z}^{(i)'}\underline{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

где ε_i — случайная «ошибка», распределение которой не зависит от параметров $\underline{\beta}$ и $\mathbf{E}\varepsilon_i = 0$. Если ввести $(k \times n)$-матрицу $Z = \|\underline{z}^{(1)} \ldots \underline{z}^{(n)}\|$, составленную из векторов-столбцов $\underline{z}^{(i)}$ (она называется *матрицей плана*), и вектор «ошибок» $\underline{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$, то в матричных обозначениях предыдущие равенства принимают вид

Матрица плана

$$\underline{X} = Z'\underline{\beta} + \underline{\varepsilon}, \quad \mathbf{E}\underline{\varepsilon} = \underline{0}. \tag{1'}$$

Далее, обычно предполагают, что «ошибки» $\varepsilon_1, \ldots, \varepsilon_n$ (или, что то же самое, случайные наблюдения $X_1, \ldots, X_n$) некоррелированы и имеют одинаковые дисперсии $\mathbf{D}X_i = \mathbf{D}\varepsilon_i = \sigma^2 > 0$, $i = 1, \ldots, n$, где σ^2 обычно неизвестно. В этом случае матрица вторых моментов вектора наблюдений $\underline{X}$ имеет следующий вид:

$$\mathbf{D}(\underline{X}) = \mathbf{D}(\underline{\varepsilon}) = \mathbf{E}(\underline{\varepsilon}\,\underline{\varepsilon}') = \sigma^2 \mathbb{1}_n, \tag{2}$$

где $\mathbb{1}_n$ — единичная матрица порядка n. Соотношениями (1)–(2) и определяется *модель линейной регрессии*. Параметры $\beta_1, \ldots, \beta_k$ называются *коэффициентами регрессии*, а σ^2 — *остаточной дисперсией*.

В описанной схеме переменные $z_1, \quad , z_k$ могут быть функционально зависимы. Так, все z_i могут быть функциями одного переменного t, например, $z_i = a_i(t)$, где $a_i(t)$ — полином степени $i > 1$, — в этом случае говорят о *параболической регрессии*.

В более общем случае возможны корреляции между наблюдениями X_i, тогда вместо условия (2) вводится условие $\mathbf{D}(\underline{\varepsilon}) = \sigma^2 G$. Если матрица G известна и невырождена, то, положив, $\underline{Y} = G^{-1/2}\underline{X}$, получим, что так преобразованные данные удовлетворяют модели (1)–(2) (напомним, что при линейном преобразовании $\underline{Y} = L\underline{X}$ первые и вторые моменты случайного вектора преобразуются следующим образом: $\mathbf{E}(Y) = L\mathbf{E}(\underline{X})$, $\mathbf{D}(\underline{Y}) = L\mathbf{D}(\underline{X})L'$). Таким образом, самостоятельного изучения заслуживает только стандартная модель (1)–(2).

В излагаемой далее теории определяющую роль играет $(k \times k)$-матрица

$$A = ZZ' \tag{3}$$

Эта матрица всегда неотрицательно определена и положительно определена тогда и только тогда, когда $\operatorname{rank} Z = k$, т. е. когда ее строки линейно независимы. Действительно, квадратичная форма $\underline{t}'A\underline{t} = \underline{t}'ZZ'\underline{t} = (Z'\underline{t})'(Z'\underline{t}) \geqslant 0$, $\forall \underline{t} = (t_1, \quad , t_k)$, причем равенство нулю возможно только при $Z'\underline{t} = \underline{0}$. В свою очередь, это равенство эквивалентно равенству $t_1\underline{z}_1 + \quad + t_k\underline{z}_k = 0$, где $\underline{z}_1, \quad , \underline{z}_k$ — столбцы матрицы Z' Таким образом, равенство нулю $\underline{t}'A\underline{t}$ при некотором $\underline{t} \neq \underline{0}$ имеет место только в случае линейной зависимости векторов $\underline{z}_1, \ldots, \underline{z}_k$, т. е. при $\operatorname{rank} Z < k$. В дальнейшем мы будем всегда предполагать, что матрица A не вырождена (или, что эквивалентно, $\operatorname{rank} Z = k$). Случай вырожденной матрицы A более сложен для анализа и соответствующие результаты можно найти в [21, гл. 4]; полное изложение теории регрессионного анализа с ее многочисленными теоретическими и практическими нюансами имеется в [22].

§ 6.2. Оценивание параметров модели линейной регрессии

Рассмотрим прежде всего задачу точечного оценивания параметров $\beta_1, \quad , \beta_k$ и σ^2 модели (1)–(2) § 6.1.

1. Метод наименьших квадратов

Общим методом оценивания неизвестных коэффициентов регрессии $\beta_1, \ldots, \beta_k$ является разработанный К. Гауссом (1809) и А. Марковым (1900) (о нем см. в п. 5 § 1.1) *метод наименьших квадратов*, в соответствии с которым оценки этих параметров находят из условия обращения в минимум квадратичной формы

$$S(\underline{\beta}) = S(\underline{X}; \underline{\beta}) = (\underline{X} - Z'\underline{\beta})'(\underline{X} - Z'\underline{\beta}) = \underline{\varepsilon}'\underline{\varepsilon}, \tag{1}$$

представляющей собой сумму квадратов «ошибок» в (1) § 6.1. Точку $\underline{b} = (b_1, \ldots, b_k)$, удовлетворяющую равенству $S(\underline{b}) = \min\limits_{\underline{\beta}} S(\underline{\beta})$, называют *оцен-*

о. н. к.

кой наименьших квадратов (о. н. к.) параметра $\underline{\beta} = (\beta_1, \quad, \beta_k)$ и обозначают $\underline{b} = \widehat{\underline{\beta}}$.

Теорема 1. Если матрица A в (3) § 6.1 не вырождена, то о. н. к. $\underline{b}$ существует и определяется равенством

$$\widehat{\underline{\beta}} = \underline{b} = A^{-1} Z \underline{X}. \tag{2}$$

Доказательство. Пусть $\underline{\beta} = \underline{b} + \underline{\delta}$, тогда

$$S(\underline{\beta}) = (\underline{X} - Z'\underline{b} - Z'\underline{\delta})'(\underline{X} - Z'\underline{b} - Z'\underline{\delta}) = (X - Z'\underline{b})'(\underline{X} - Z'\underline{b}) + \underline{\delta}' A \underline{\delta},$$

поскольку

$$(\underline{X} - Z'\underline{b})' Z'\underline{\delta} + \underline{\delta}' Z(\underline{X} - Z'\underline{b}) = 2\underline{X}' Z'\underline{\delta} - 2\underline{b}' A \underline{\delta} = 0.$$

Таким образом,

$$S(\underline{\beta}) = S(\underline{b}) + \underline{\delta}' A \underline{\delta} \geqslant S(\underline{b}) \tag{3}$$

и равенство достигается лишь при $\underline{\delta} = \underline{0}$. Следовательно, $\underline{b}$ — единственная точка минимума квадратичной формы (1). ■

Замечание. Практически о. н. к. $\widehat{\underline{\beta}} = \arg\min_{\underline{\beta}} S(\underline{\beta})$ находят, решая систему уравнений

$$\frac{\partial S(\underline{\beta})}{\partial \beta_j} = 0, \quad j = 1, \quad, k,$$

Нормальные уравнения

которые называются *нормальными уравнениями* метода наименьших квадратов.

В ряде случаев интерес представляют не сами параметры $\beta_1, \quad, \beta_k$, а их некоторые линейные комбинации, т. е. новый параметрический вектор $\underline{t} = (t_1, \quad, t_m)$, $m \leqslant k$, связанный с $\underline{\beta}$ соотношением $\underline{t} = T\underline{\beta}$, где T — заданная $(m \times k)$-матрица. В этом случае о. н. к. $\widehat{\underline{t}}$ для $\underline{t}$ определяется равенством $\widehat{\underline{t}} = T\widehat{\underline{\beta}}$, где $\widehat{\underline{\beta}}$ определено в (2); представление $\widehat{\underline{t}}$ через исходные данные, следовательно, имеет вид

$$\widehat{\underline{t}} = TA^{-1} Z \underline{X}. \tag{4}$$

2. Оптимальность оценок наименьших квадратов

Исследуем свойства полученных в п. 1 оценок. В общем случае будем рассматривать задачу оценивания вектора $\underline{t} = T\underline{\beta}$ в классе линейных оценок, т. е. оценок вида $\underline{l} = L\underline{X}$, являющихся линейными функциями от наблюдений $\underline{X} = (X_1, \ldots, X_n)$.

Теорема 2. Пусть матрица A не вырождена. Тогда для произвольного вектора $\underline{t} = T\underline{\beta}$ о. н. к. $\widehat{\underline{t}}$, определенная равенством (4), является несмещенной оценкой с минимальной дисперсией в классе всех линейных несмещенных оценок $\underline{t}$; при этом

$$\mathbf{D}(\widehat{\underline{t}}) = \sigma^2 TA^{-1}T' \equiv \sigma^2 D. \tag{5}$$

Доказательство. Из (4), (3) и (1′) § 6.1 имеем

$$\mathbf{E}(\widehat{\underline{t}}) = TA^{-1}Z\mathbf{E}(\underline{X}) = TA^{-1}ZZ'\beta' = T\underline{\beta} = \underline{t},$$

т. е. $\widehat{\underline{t}}$ — линейная несмещенная оценка $\underline{t}$. Пусть $\underline{l} = L\underline{X}$ — произвольная линейная несмещенная оценка $\underline{t}$, т. е.

$$\mathbf{E}(\underline{l}) = L\mathbf{E}(\underline{X}) = LZ'\underline{\beta} = T\underline{\beta}.$$

Это равенство должно, по определению несмещенности, выполняться для любого $\underline{\beta}$, поэтому отсюда следует, что

$$LZ' = T \tag{6}$$

Из (2) § 6.1 находим

$$\mathbf{D}(\underline{l}) = L\mathbf{D}(\underline{X})L' = \sigma^2 LL' \tag{7}$$

Наша цель — минимизировать дисперсии оценок $l_1, \quad , l_m$, т. е. диагональные элементы матрицы LL'. Для этого запишем тождество, следующее из (6):

$$LL' = (TA^{-1}Z)(TA^{-1}Z)' + (L - TA^{-1}Z)(L - TA^{-1}Z)'$$

Здесь в правой части каждое слагаемое имеет вид HH', откуда следует неотрицательность диагональных элементов. Но от L зависит только второе слагаемое, поэтому диагональные элементы матрицы $\mathbf{D}(\underline{l})$ в (7) одновременно достигают минимума тогда и только тогда, когда $L = TA^{-1}Z$. Соответствующая оптимальная оценка имеет вид $\underline{t}^* = TA^{-1}Z\underline{X} = \widehat{\underline{t}}$, т. е. совпадает с о. н. к. (4). Наконец, формула (5) следует из (7), если подставить вместо L найденное оптимальное решение. ■

В качестве простого следствия теоремы 2 получаем, что $\mathbf{D}(\widehat{\underline{\beta}}) = \sigma^2 A^{-1}$, или

$$\operatorname{cov}(\widehat{\beta}_i, \widehat{\beta}_j) = \sigma^2 a^{ij}, \quad i, j = 1, \ldots, k, \quad \text{где} \quad \|a^{ij}\| = A^{-1} = \|a_{ij}\|^{-1} \tag{8}$$

Замечание. Если матрица T имеет вид $T = \Lambda A$, то формулы (4) и (5) принимают соответственно вид

$$\widehat{\underline{t}} = \Lambda Z\underline{X}, \quad \mathbf{D}(\widehat{\underline{t}}) = \sigma^2 \Lambda A \Lambda',$$

т. е. необходимость в вычислении обратной матрицы A^{-1} для нахождения о. н. к. и их вторых моментов отпадает.

Итак, теорема 2 позволяет решить задачу построения оптимальных линейных (по наблюдениям $\underline{X}$) оценок для произвольных линейных комбинаций коэффициентов регрессии — это оценки наименьших квадратов.

3. Оценивание остаточной дисперсии

Из (1) и (2) §6.1 имеем

$$\mathbf{E}S(\underline{\beta}) = \sum_{i=1}^{n} \mathbf{E}\varepsilon_i^2 = n\sigma^2$$

Далее, учитывая (8) и определение следа (tr) матрицы (см. п. 2 §2.5), найдем

$$\mathbf{E}(\underline{\widehat{\beta}} - \underline{\beta})' A(\underline{\widehat{\beta}} - \underline{\beta}) = \sum_{i,j=1}^{k} a_{ij}\mathbf{E}(\widehat{\beta}_i - \beta_i)(\widehat{\beta}_j - \beta_j) =$$

$$\sigma^2 \sum_{i,j=1}^{k} a_{ij}a^{ij} = \sigma^2 \operatorname{tr}(AA^{-1}) = \sigma^2 \operatorname{tr} \mathbb{1}_k = k\sigma^2$$

Отсюда и из представления (см. (3))

$$S(\underline{\beta}) = S(\underline{\widehat{\beta}}) + (\underline{\widehat{\beta}} - \underline{\beta})' A(\underline{\widehat{\beta}} - \underline{\beta}) \tag{9}$$

следует, что

$$\mathbf{E}S(\underline{\widehat{\beta}}) = (n-k)\sigma^2,$$

т. е. несмещенной оценкой для остаточной дисперсии σ^2 является статистика

$$\widetilde{\sigma}^2 = \frac{1}{n-k} S(\underline{\widehat{\beta}}) = \frac{(\underline{X} - Z'\underline{\widehat{\beta}})'(\underline{X} - Z'\underline{\widehat{\beta}})}{n-k} \tag{10}$$

Вектор $\underline{\mathcal{U}} = \underline{X} - \underline{Z'\widehat{\beta}}$ называют *остаточным вектором*, а его компоненты u_i, $i = 1, \ldots, n$, — *остатками*. Таким образом, полученный результат (10) можно сформулировать так: *оценка $\widetilde{\sigma}^2$ равна сумме квадратов остатков, поделенной на $n-k$* (разность между числом наблюдений и числом коэффициентов регрессии β_i).

Приведем также другое выражение для $\widetilde{\sigma}^2$, которое понадобится нам в дальнейшем. Используя представление (2), запишем остаточный вектор в виде

$$\underline{\mathcal{U}} = (\mathbb{1}_n - Z'A^{-1}Z)\underline{X} \equiv B\underline{X} = B(Z'\underline{\beta} + \underline{\varepsilon}) = B\underline{\varepsilon}. \tag{11}$$

Непосредственно проверяется, что матрица B в (11) симметрична и идемпотентна ($B^2 = B$); следовательно, вместо (10) можно использовать представление

$$\widetilde{\sigma}^2 = \frac{1}{n-k}\underline{X}'B\underline{X} = \frac{1}{n-k}\varepsilon' B\varepsilon, \tag{12}$$

показывающее явную зависимость оценки $\widetilde{\sigma}^2$ от наблюдений.

Наконец, из (11) непосредственно вытекают следующие свойства остаточного вектора $\underline{\mathcal{U}}$ (их проверку мы оставляем читателю в качестве упражнения):

$$\mathbf{E}(\underline{\mathcal{U}}) = \underline{0}, \quad \mathbf{D}(\underline{\mathcal{U}}) = \sigma^2 B = \mathbf{D}(\underline{X}) - \mathbf{D}(Z'\underline{\widehat{\beta}}), \quad \mathbf{E}\underline{\widehat{\beta}}\underline{\mathcal{U}}' = \underline{0} \tag{13}$$

(в последнем соотношении $\underline{0}$ обозначает нулевую $(k \times n)$-матрицу, т. е. состоящую сплошь из нулей, следовательно, о. н. к. $\widehat{\underline{\beta}}$ и остаточный вектор $\underline{\mathcal{U}}$ некоррелированы: $\operatorname{cov}(\widehat{\beta}_i, u_j) = 0$, $i = 1, \ldots, k$, $j = 1, \ldots, n$).

4. Условные о. н. к.

Выше мы рассмотрели случай, когда вектор $\underline{\beta} = (\beta_1, \quad, \beta_k)$ априори мог быть произвольной точкой евклидова пространства $\mathbb{R}^k$ (на его значения не накладывалось никаких ограничений). Однако в ряде практических задач допустимые значения $\underline{\beta}$ бывают ограничены теми или иными условиями. Типичная модель с ограничениями на $\underline{\beta}$ имеет вид

$$T\underline{\beta} = \underline{t}_0, \tag{14}$$

где T — некоторая заданная $(m \times k)$-матрица $(m \leqslant k)$ и $\underline{t}_0 = (t_{10}, \quad, t_{m0})$ — заданный вектор такой, что система (14) совместна. Другими словами, условие (14) означает, что допустимые значения коэффициентов регрессии β_1, β_k удовлетворяют m заданным линейным ограничениям

$$\underline{t}_j'\underline{\beta} = t_{j0}, \quad j = 1, \quad, m, \tag{14'}$$

где $\underline{t}_1', \quad, \underline{t}_m'$ — строки матрицы T. Естественно предполагать, что ограничения (14′) линейно независимы (иначе можно перейти к меньшему числу уравнений, исключив линейно зависимые). Таким образом, в дальнейшем будем считать, что $\operatorname{rank} T = m$ (случай $m = k$, однозначно фиксирующий вектор $\underline{\beta}$, в последующих рассуждениях формально допускается).

Рассмотрим задачу оценивания параметров $\underline{\beta}$ в этой усложненной ситуации. Обозначим

$$S_T = \min_{\underline{\beta}\,:\,T\underline{\beta}=\underline{t}_0} S(\underline{\beta}) \tag{15}$$

и назовем *условной оценкой наименьших квадратов* (у. о. н. к.) $\widehat{\underline{\beta}}_T$ то значение $\underline{\beta}$ (удовлетворяющее условию (14)), при котором этот условный минимум достигается: $S_T = S(\widehat{\underline{\beta}}_T)$.

Условная [redacted] о. н. к.

Теорема 3. Пусть $|A| \neq 0$ и $\operatorname{rank} T = m$. Тогда

$$\widehat{\underline{\beta}}_T = \widehat{\underline{\beta}} - A^{-1}T'D^{-1}(T\widehat{\underline{\beta}} - \underline{t}_0), \tag{16}$$

где $\widehat{\underline{\beta}}$ — обычная о. н. к. (2) и матрица D определена в (5).

Доказательство. Из (16) непосредственно видно, что $T\widehat{\underline{\beta}}_T = \underline{t}_0$, т. е. вектор $\widehat{\underline{\beta}}_T$ удовлетворяет условию (14). Запишем теперь разложение (см. (1))

$$S(\underline{\beta}) = [\underline{X} - Z'\widehat{\underline{\beta}}_T + Z'(\widehat{\underline{\beta}}_T - \underline{\beta})]'[\underline{X} - Z'\widehat{\underline{\beta}}_T + Z'(\widehat{\underline{\beta}}_T - \underline{\beta})] =$$

$$= S(\underline{\widehat{\beta}}_T) + 2(\underline{\widehat{\beta}}_T - \underline{\beta})'(Z\underline{X} - A\underline{\widehat{\beta}}_T) + (\underline{\widehat{\beta}}_T - \underline{\beta})'A(\underline{\widehat{\beta}}_T - \underline{\beta}).$$

Так как (см. (2)) $A\underline{\widehat{\beta}} = Z\underline{X}$, то из (16) имеем

$$Z\underline{X} - A\underline{\widehat{\beta}}_T = T'D^{-1}(T\underline{\widehat{\beta}} - \underline{t}_0).$$

Учитывая также, что $T(\underline{\widehat{\beta}}_T - \underline{\beta}) = \underline{t}_0 - T\underline{\beta}$, получаем окончательное представление для $S(\underline{\beta})$, справедливое для всех $\underline{\beta}$:

$$S(\underline{\beta}) = S(\underline{\widehat{\beta}}_T) + 2(\underline{t}_0 - T\underline{\beta})'D^{-1}(T\underline{\widehat{\beta}} - t_0) + (\underline{\widehat{\beta}}_T - \underline{\beta})'A(\underline{\widehat{\beta}}_T - \underline{\beta}). \qquad (17)$$

Пусть теперь $\underline{\beta}$ удовлетворяет условию (14). Тогда средний член в (17) обращается в 0, и мы получаем, что для таких $\underline{\beta}$

$$S(\underline{\beta}) = S(\underline{\widehat{\beta}}_T) + (\underline{\widehat{\beta}}_T - \underline{\beta})'A(\underline{\widehat{\beta}}_T - \underline{\beta}) \geqslant S(\underline{\widehat{\beta}}_T)$$

(сравни с (3)), причем равенство достигается только при $\underline{\beta} = \underline{\widehat{\beta}}_T$. ■

Замечание. Полагая в (9) $\underline{\beta} = \underline{\widehat{\beta}}_T$ и учитывая, что, в силу (16),

$$(\underline{\widehat{\beta}} - \underline{\widehat{\beta}}_T)'A(\underline{\widehat{\beta}} - \underline{\widehat{\beta}}_T) = (T\underline{\widehat{\beta}} - t_0)'D^{-1}TA^{-1}AA^{-1}T'D^{-1}(T\underline{\widehat{\beta}} - t_0) = (T\underline{\widehat{\beta}} - t_0)'D^{-1}(T\underline{\widehat{\beta}} - t_0),$$

получаем следующее представление условного минимума $S_T = S(\underline{\widehat{\beta}}_T)$ через абсолютный минимум $S(\underline{\widehat{\beta}})$:

$$S_T = S(\underline{\widehat{\beta}}) + Q_T, \quad Q_T = (T\underline{\widehat{\beta}} - \underline{t}_0)'D^{-1}(T\underline{\widehat{\beta}} - \underline{t}_0). \qquad (18)$$

5. Оптимальный выбор матрицы плана

Рассмотрим «активный» эксперимент, когда значения факторов $z_1, \quad , z_k$ для каждого опыта выбираются исследователем. Покажем, как в этом случае оптимально задать матрицу плана $Z = \|\underline{z}^{(1)} \quad \underline{z}^{(n)}\|$, где i-й столбец $\underline{z}^{(i)}$ есть комбинация значений факторов для i-го опыта ($i = 1, \ldots, n$). В качестве критерия оптимальности естественно использовать величину дисперсий оценок $\widehat{\beta}_i$, т. е. величины a^{ii} в (8); тогда задача сводится к выбору такой матрицы Z, чтобы диагональные элементы матрицы $A^{-1} = (ZZ')^{-1}$ были минимальны.

Если значения факторов выбирать произвольными, то все элементы матрицы A^{-1} можно сделать одновременно как угодно малыми, так как если Z заменить на aZ, то A^{-1} заменится на $A^{-1}/a^2 \to \underline{0}$ при $a \to \infty$. Чтобы исключить этот случай, предположим, что значения факторов можно менять в ограниченных областях, именно: наложим ограничения вида

$$\underline{z}_j'\underline{z}_j = \sum_{i=1}^{n} (z_j^{(i)})^2 = a_j^2, \quad j = 1, \ldots, k, \qquad (19)$$

где $\underline{z}_j = (z_j^{(1)}, \ldots, z_j^{(n)})$ — набор значений n уровней фактора z_j, $j = 1, \quad , k$ (строки матрицы Z), а $a_1^2, \ldots, a_k^2$ — заданные положительные константы.

Теорема 4. При ограничениях (19) имеют место неравенства

$$\mathbf{D}\widehat{\beta}_j \geqslant \frac{\sigma^2}{a_j^2}, \quad j = 1, \ldots, k,$$

и минимум достигается тогда и только тогда, когда векторы $\underline{z}_1, \ldots, \underline{z}_k$ ортогональны, т. е. $\underline{z}_i'\underline{z}_j = 0$ при $i \neq j$.

Доказательство. Пусть $j = 1$. Запишем матрицу $A = ZZ'$ в виде

$$A = \begin{Vmatrix} \underline{z}_1'\underline{z}_1 & \underline{b}' \\ \underline{b} & F \end{Vmatrix},$$

где $\underline{b} = (\underline{z}_2'\underline{z}_1, \ldots, \underline{z}_k'\underline{z}_1)$. Тогда, на основании формулы обращения матриц, $a^{11} = |F|/|A|$. Далее, умножая определитель $|A|$ справа на определитель, равный 1, получаем следующую формулу:

$$|A| = \begin{vmatrix} \underline{z}_1'\underline{z}_1 & \underline{b}' \\ \underline{b} & F \end{vmatrix} \begin{vmatrix} 1 & \underline{0} \\ -F^{-1}\underline{b} & \mathbb{1}_{k-1} \end{vmatrix} = \begin{vmatrix} \underline{z}_1'\underline{z}_1 - \underline{b}'F^{-1}\underline{b} & \underline{b}' \\ \underline{0} & F \end{vmatrix} = |F|(\underline{z}_1'\underline{z}_1 - \underline{b}'F^{-1}\underline{b}).$$

Отсюда, учитывая соотношения (19) и (8), имеем

$$\mathbf{D}\widehat{\beta}_1 = \sigma^2 a^{11} = \sigma^2 \frac{|F|}{|A|} = \frac{\sigma^2}{a_1^2 - \underline{b}'F^{-1}\underline{b}} \geqslant \frac{\sigma^2}{a_1^2},$$

так как $\underline{b}'F^{-1}\underline{b} \geqslant 0$. Знак равенства имеет место только при $\underline{b} = \underline{0}$ (поскольку подматрица F положительно определенной матрицы A также положительно определена), т. е. при $\underline{z}_i'\underline{z}_1 = 0$, $i = 2, \ldots, k$. Доказательство для других j можно получить простой перенумерацией факторов. ■

В заключение отметим, что для оптимальной матрицы плана Z матрица $A = ZZ'$ является диагональной с элементами $a_{jj} = a_j^2 = \underline{z}_j'\underline{z}_j$, $j = 1, \ldots, k$, поэтому в этом случае нет проблемы ее обращения: о. н. к. и их вторые моменты имеют вид

$$\widehat{\beta}_j = \frac{\underline{z}_j'\underline{X}}{\underline{z}_j'\underline{z}_j}, \quad \mathbf{D}\widehat{\beta}_j = \frac{\sigma^2}{\underline{z}_j'\underline{z}_j}, \quad \operatorname{cov}(\widehat{\beta}_i, \widehat{\beta}_j) = 0, \quad i \neq j. \tag{20}$$

6. Примеры применения метода наименьших квадратов

Пример 1 (Простая регрессия, оценивание параметров). Проиллюстрируем общую теорию на примере важного для практических приложений случая *простой регрессии*, когда число параметров $k = 2$, т. е. $\underline{\beta} = (\beta_1, \beta_2)$, а векторы $\underline{z}^{(i)}$ имеют вид $\underline{z}^{(i)} = (1, t_i)$, $i = 1, \ldots, n$. Тогда

$$\mathbf{E}X_i = \beta_1 + \beta_2 t_i, \quad i = 1, \ldots, n, \tag{21}$$

т. е. среднее значение наблюдений является линейной функцией одного фактора t Так, t может быть температурой, при которой производится эксперимент, дозой лечебного препарата, возрастом обследуемых лиц и т. д., и речь

идет об изучении связи между откликом (исходом эксперимента) и фактором t на основании результатов регистрации n пар измерений (X_i, t_i), $i = 1, \ldots, n$, где X_i наблюдается при значении t_i фактора t. Прямую $\varphi(t) = \beta_1 + \beta_2 t$, соответствующую (21), называют *линией регрессии*, а коэффициент β_2 — ее *наклоном*.

В данном случае матрицы Z, A и столбец $\underline{Y} = Z\underline{X}$ равны:

$$Z = \begin{Vmatrix} 1 & 1 \\ t_1 & t_n \end{Vmatrix}, \quad A = \begin{Vmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{Vmatrix} \quad \underline{Y} = \begin{pmatrix} \sum X_i \\ \sum t_i X_i \end{pmatrix}$$

Будем предполагать, что не все t_i одинаковы (тогда $\operatorname{rank} Z = 2$), тогда

$$|A| = n\sum t_i^2 - \left(\sum t_i\right)^2 = n\sum_{i=1}^{n}(t_i - \bar{t})^2 > 0,$$

где, как обычно, черта сверху означает арифметическое среднее, и

$$A^{-1} = \frac{1}{|A|}\begin{Vmatrix} \sum t_i^2 & -n\bar{t} \\ -n\bar{t} & n \end{Vmatrix},$$

$$A^{-1}\underline{Y} = \frac{n}{|A|}\begin{pmatrix} \overline{X}\sum t_i^2 - \bar{t}\sum t_i X_i \\ \sum t_i X_i - n\bar{t}\overline{X} \end{pmatrix} \quad \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix}$$

В результате несложных преобразований запишем оценки $\widehat{\beta}_1$ и $\widehat{\beta}_2$ в более удобном для вычислений виде (проверить!)

$$\widehat{\beta}_2 = \frac{\sum(t_i - \bar{t})(X_i - \overline{X})}{\sum(t_i - \bar{t})^2}, \quad \widehat{\beta}_1 = \overline{X} - \bar{t}\widehat{\beta}_2. \tag{22}$$

Вторые моменты этих оценок образуют матрицу $\sigma^2 A^{-1}$ (см. (8)), поэтому

$$\mathrm{D}\widehat{\beta}_1 = \frac{\sum t_i^2}{n\sum(t_i - \bar{t})^2}\sigma^2 \quad \mathrm{D}\widehat{\beta}_2 = \frac{\sigma^2}{\sum(t_i - \bar{t})^2},$$

$$\operatorname{cov}(\widehat{\beta}_1, \widehat{\beta}_2) = -\frac{\bar{t}}{\sum(t_i - \bar{t})^2}\sigma^2$$

Наконец, в силу (22),

$$S(\underline{\widehat{\beta}}) = \sum(X_i - \widehat{\beta}_1 - \widehat{\beta}_2 t_i)^2 = \sum\left[(X_i - \overline{X}) - \widehat{\beta}_2(t_i - \bar{t})\right]^2 =$$
$$= \sum(X_i - \overline{X})^2 - \widehat{\beta}_2^2\sum(t_i - \bar{t})^2 \tag{23}$$

и несмещенная оценка для остаточной дисперсии σ^2 вычисляется по формуле

$$\widetilde{\sigma}^2 = \frac{S(\widehat{\beta})}{n-2}. \quad \bullet$$

Пример 2 (Параболическая регрессия). Пусть j-й фактор z_j является полиномом $a_j(t)$ степени $j-1$ от общей переменной t; тогда

$$\underline{z}_j = (a_j(t_1), \quad , a_j(t_n)), \quad j = 1, . \quad , k,$$

— набор его значений для n уровней, а

$$\underline{z}^{(i)} = (a_1(t_i), . \quad , a_k(t_i)), \quad i = 1, \ldots, n,$$

— комбинация значений факторов $z_1, \ldots, z_k$ для i-го опыта. Таким образом, матрица плана $Z = \|\underline{z}^{(1)} \quad \underline{z}^{(n)}\|$ определяется выбором n точек $t_1, \quad , t_n$ значений общего фактора t для n опытов. В этом случае среднее значение i-го наблюдения X_i имеет вид

$$\mathbf{E}X_i = \varphi(t_i; \underline{\beta}) = \sum_{j=1}^{k} \beta_j a_j(t_i), \quad i = 1, \ldots, n. \tag{24}$$

Определенная здесь функция $\varphi(t; \underline{\beta})$ представляет собой полином (параболу) степени $k-1$ и называется *кривой параболической регрессии*.

Опишем типичную ситуацию, когда имеет место схема параболической регрессии. Предположим, что имеется теоретическая зависимость вида $x = \varphi(t; \underline{\beta})$ между переменными t и x, причем значения t и x получают из наблюдений. Задача состоит в определении по экспериментальным данным неизвестных параметров $\beta_1, \quad , \beta_k$, входящих в это уравнение. Каждое измерение x при заданном t дает уравнение, связывающее неизвестные параметры, и если бы измерения величины x производились без погрешностей, то для определения параметров $\beta_1, \quad , \beta_k$ было бы достаточно k измерений. Однако точные измерения на практике чаще всего невозможны, поэтому для заданных значений t соответствующие значения x известны с какими-то погрешностями, т. е. реально вместо точного значения $x_i = \varphi(t_i; \underline{\beta})$ мы имеем случайный результат $X_i = \varphi(t_i; \underline{\beta}) + \varepsilon_i$, где ε_i — ошибка измерения. Если условия эксперимента обеспечивают отсутствие систематической ошибки ($\mathbf{E}\varepsilon_i = 0$), равноточность ($\mathbf{D}\varepsilon_i = \sigma^2 \quad i = 1, \quad , n$) и некоррелированность ($\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, \ i \neq j$) результатов измерений, то приходим к схеме регрессии вида (24).

Чтобы исключить влияние погрешностей измерений, нужна дополнительная информация. Для этого проводят большое число измерений $n > k$ и полученные экспериментальные данные обрабатывают по методу наименьших квадратов. В результате находят не точные значения коэффициентов $\beta_1, \ldots, \beta_k$, а определяют их о. н. к. $\widehat{\beta}_1, \ldots, \widehat{\beta}_k$. В данном случае этому методу можно дать следующую геометрическую интерпретацию (см. рис. 1).

Нанесем на плоскости (t, x) наблюдавшиеся точки (t_i, x_i), $i = 1, \quad , n$. Тогда значения $\underline{\beta}^*$ неизвестных коэффициентов $\underline{\beta}$ подбирают так, чтобы со-

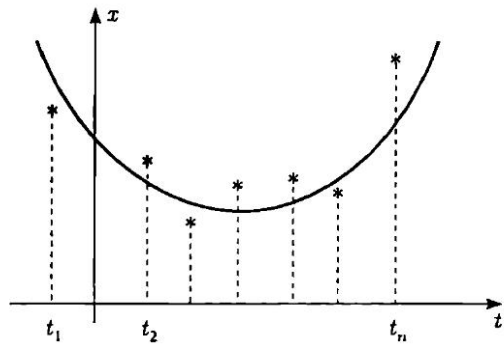


Рис. 1

ответствующий *эмпирический* график $x = \varphi(t; \underline{\beta}^*)$ «наилучшим образом» проходил около наблюдавшихся точек. Если в качестве критерия оптимальности взять величину

$$\sum_{i=1}^{n} (X_i - \varphi(t_i; \underline{\beta}^*))^2$$

— сумму квадратов отклонений эмпирического графика в точках $t_1, \ldots, t_n$ от наблюдавшихся значений $X_1, \ldots, X_n$, то приходим к методу наименьших квадратов с решением $\underline{\beta}^* = \widehat{\underline{\beta}}$.

Формально такой же вид имеет и классическая задача математического анализа приближений функций многочленами. В этом случае данные (t_i, X_i), $i = 1, \ldots, n$, интерпретируют как пары соответствующих абсцисс и ординат графика изучаемой функции и задача состоит в подборе *интерполяционного многочлена* вида $\varphi(t; \underline{\beta})$, который давал бы *наилучшее приближение в среднем*, т. е. минимизировал бы величину $\sum_{i=1}^{n} (X_i - \varphi(t; \underline{\beta}))^2$ •

Пример 3 (Численная иллюстрация [11]). Приведем пример вычисления оценок параметров $\underline{\beta}$. Пусть требуется определить неизвестные коэффициенты $(\beta_1, \beta_2, \beta_3)$ функциональной зависимости

$$x(t) = \beta_1 + t\beta_2 + t^3\beta_3.$$

Будем предполагать, что в точках $t_i = 2 + 3i/n$, $i = 1, \ldots, n$, измерены значения функции $x(t)$. Ошибки измерений ε_i, $i = 1, \ldots, n$, будем считать независимыми и нормально распределенными с $\mathbf{E}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = \sigma^2$. В этом случае имеем линейную модель

$$X_i = x(t_i) + \varepsilon_i = \beta_1 + t_i\beta_2 + t_i^3\beta_3 + \varepsilon_i, \quad i = 1, \ldots, n.$$

Пусть $\beta_1 = 3$, $\beta_2 = -1$, $\beta_3 = 1$, $\sigma^2 = 0{,}04$. Смоделируем ε_i для $n = 25$ и $n = 100$, используя алгоритмы п. 8 и 6 § 1.3, и найдем соответствующие X_i.

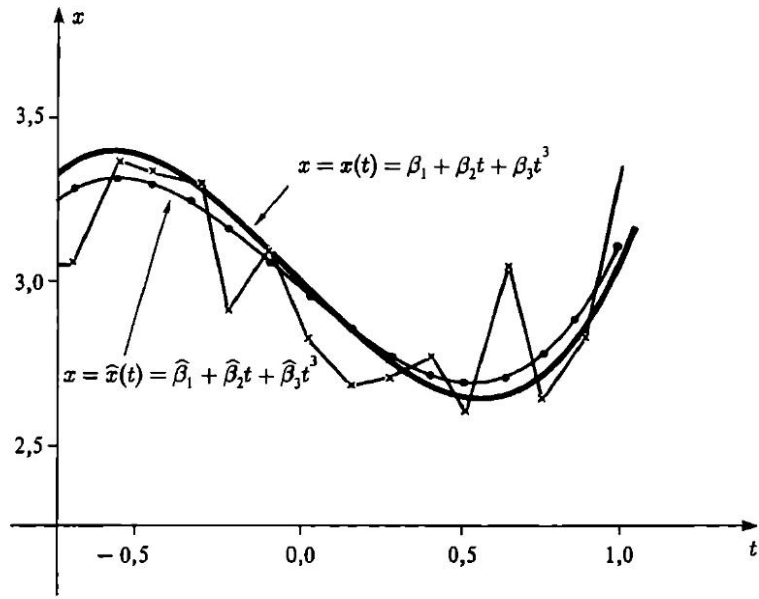


Рис. 2. ($n = 25$)

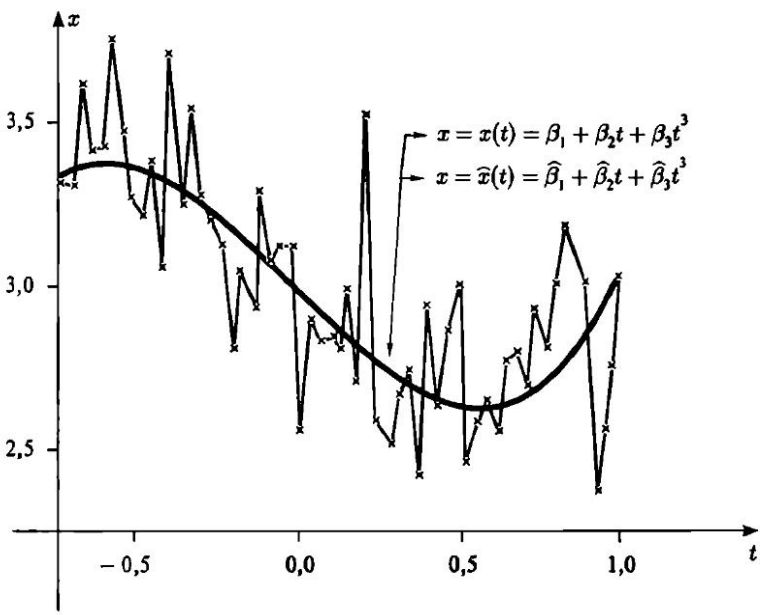


Рис. 3. ($n = 100$)

Решая систему нормальных уравнений

$$\frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} (X_i - \beta_1 - t_i\beta_2 - t_i^3\beta_3)^2 = 0, \quad j = 1, 2, 3,$$

найдем оценки $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$, а также $\widetilde{\sigma}^2$:

1) для $n = 25$: $\widehat{\beta}_1 = 2{,}983$; $\widehat{\beta}_2 = -0{,}828$; $\widehat{\beta}_3 = 0{,}895$; $\sigma^2 = 0{,}034$;

2) для $n = 100$: $\widehat{\beta}_1 = 2{,}992$; $\widehat{\beta}_2 = -1{,}007$; $\widehat{\beta}_3 = 1{,}001$; $\sigma^2 = 0{,}046$.

На рис. 2 и 3 приведены соответствующие графики $x(t)$ и $\widehat{x}(t) = \widehat{\beta}_1 + t\widehat{\beta}_2 + t^3\widehat{\beta}_3$, знаком $\otimes$ отмечены результаты измерений (t_i, X_i). •

7. Ортогональные многочлены Чебышева

В описанной в примере 2 схеме предполагалось, что многочлены $a_j(t)$ заданы. В частном случае $a_j(t) = t^{j-1}$ и тогда

$$\varphi(t; \underline{\beta}) = \sum_{j=1}^{k} \beta_j t^{j-1}$$

(это имеет место в примере 1). В других случаях (например, в задачах интерполяции) многочлены $a_j(t)$ можно выбрать достаточно произвольно, и тогда критерием их выбора является простота соответствующих вычислений. Как отмечалось в п. 5, наиболее просто решается задача вычисления о. н. к. $\widehat{\underline{\beta}}$ при диагональной матрице $A = ZZ'$: в этом случае (см. (20))

$$\widehat{\beta}_j = \frac{\sum_{i=1}^{n} a_j(t_i) X_i}{\sum_{i=1}^{n} a_j^2(t_i)}, \quad j = 1, \quad , k. \tag{25}$$

Условием диагональности матрицы A является ортогональность векторов $\underline{z}_j = (a_j(t_1), \ldots, a_j(t_n))$:

$$\underline{z}_j' \underline{z}_r = \sum_{i=1}^{n} a_j(t_i) a_r(t_i) = 0, \quad j \neq r \tag{26}$$

Многочлены, удовлетворяющие условию (26) называют *ортогональными многочленами Чебышева*.

Покажем, что при заданных $t_1, \ldots, t_n$ такие многочлены всегда можно построить. Без потери общности можно предполагать старший коэффициент $a_j(t)$ равным 1, таким образом, $a_1(t) \equiv 1$, $a_2(t) = t + c$, ... Из условия (26)

при $j = 1$, $\tau = 2$ имеем

$$0 = \sum_{i=1}^{n} a_2(t_i) = \sum_{i=1}^{n} t_i + nc,$$

т. е. $c = -\bar{t}$. Тем самым, второй многочлен Чебышева есть $a_2(t) = t - \bar{t}$.

Выведем теперь общую рекуррентную формулу, позволяющую вычислять следующий многочлен по двум предыдущим, — этим и будет доказано существование всех многочленов системы Чебышева. Предположим, что для некоторого $\tau \geqslant 2$ многочлены $a_1(t), \ldots, a_\tau(t)$ уже построены. Рассмотрим произвольный многочлен $\psi(t)$ степени, меньшей τ. Его можно однозначно представить в виде

$$\psi(t) = \sum_{s=1}^{\tau} \gamma_s a_s(t).$$

Действительно, из условия ортогональности (26) при $j = 1, \quad , \tau$ имеем

$$\sum_{i=1}^{n} \psi(t_i) a_j(t_i) = \gamma_j \sum_{i=1}^{n} a_j^2(t_i),$$

т. е. эти равенства однозначно определяют коэффициенты $\gamma_1, \ldots, \gamma_\tau$. Отсюда, в частности, следует, что если $\psi(t)$ — многочлен степени, меньшей $\tau - 1$, то

$$\sum_{i=1}^{n} \psi_j(t_i) a_\tau(t_i) = 0. \tag{27}$$

Рассмотрим теперь многочлен степени τ следующего вида

$$a(t) = (t + \gamma) a_\tau(t) + \beta a_{\tau-1}(t). \tag{28}$$

Если $j < \tau - 1$, то в силу (27)

$$\sum_{i=1}^{n} a_j(t_i) a(t_i) = \sum_{i=1}^{n} (t_i + \gamma) a_j(t_i) a_\tau(t_i) + \beta \sum_{i=1}^{n} a_j(t_i) a_{\tau-1}(t_i) = 0$$

(так как степень многочлена $(t+\gamma)a_j(t)$ равна $j < \tau - 1$), т. е. при любых постоянных β и γ многочлен (28) ортогонален многочленам $a_1(t), \ldots, a_{\tau-2}(t)$. Выберем эти постоянные так, чтобы он был ортогонален также $a_{\tau-1}(t)$ и $a_\tau(t)$. Для этого должны выполняться следующие условия:

$$\sum_{i=1}^{n} a_{\tau-1}(t_i) a(t_i) = \sum_{i=1}^{n} t_i a_{\tau-1}(t_i) a_\tau(t_i) + \beta \sum_{i=1}^{n} a_{\tau-1}^2(t_i) = 0,$$

$$\sum_{i=1}^{n} a_\tau(t_i) a(t_i) = \sum_{i=1}^{n} t_i a_\tau^2(t_i) + \gamma \sum_{i=1}^{n} a_\tau^2(t_i) = 0.$$

Отсюда находим β и γ:

$$\beta = -\sum_{i=1}^{n} \frac{t_i a_{\tau-1}(t_i) a_\tau(t_i)}{a_{\tau-1}^2}, \quad \gamma = -\sum_{i=1}^{n} \frac{t_i a_\tau^2(t_i)}{a_\tau^2}, \tag{29}$$

где

$$a_j^2 = \sum_{i=1}^{n} a_j^2(t_i), \quad j = 1, 2, \ldots$$

При таком выборе постоянных β и γ многочлен (28) ортогонален всем многочленам $a_1(t), \ldots, a_\tau(t)$, следовательно, он является следующим многочленом $a_{\tau+1}(t)$ системы Чебышева. Тем самым, получен алгоритм построения системы ортогональных многочленов Чебышева. В дополнение к указанным ранее двум первым многочленам $a_1(t)$ и $a_2(t)$ приведем явный вид третьего многочлена $a_3(t)$, который следует из (28) и (29):

$$a_3(t) = (t - \bar{t})\left(t - \frac{1}{a_2^2}\sum_{i=1}^{n} t_i (t_i - \bar{t})^2\right) - \frac{1}{n}\sum_{i=1}^{n} t_i (t_i - \bar{t}) =$$

$$= (t - \bar{t})\left(t - \bar{t} - \frac{1}{a_2^2}\sum_{i=1}^{n} (t_i - \bar{t})^3\right) - \frac{a_2^2}{n}. \tag{30}$$

Итак, если в схеме параболической регрессии (24) многочлены $a_j(t)$ являются ортогональными многочленами Чебышева, то о. н. к. $\widehat{\beta}_j$ вычисляются по формулам (25), а соответствующее значение $S(\underline{\widehat{\beta}}) = \min\limits_{\underline{\beta}} S(\underline{\beta})$ таково:

$$S(\underline{\widehat{\beta}}) = \sum_{i=1}^{n}\left(X_i - \sum_{j=1}^{k} \widehat{\beta}_j a_j(t_i)\right)^2$$

$$= \sum_{i=1}^{n} X_i^2 - 2\sum_{j=1}^{k} \widehat{\beta}_j \sum_{i=1}^{n} a_j(t_i) X_i + \sum_{j=1}^{k} a_j^2 \widehat{\beta}_j^2 =$$

$$= \sum_{i=1}^{n} X_i^2 - \sum_{j=1}^{k} a_j^2 \widehat{\beta}_j^2. \tag{31}$$

Отметим следующее важное обстоятельство. Как видно из формулы (25), о. н. к. $\widehat{\beta}_j$ определяется только многочленом $a_j(t)$ и не зависит от параметра k схемы (24). Кто позволяет упростить задачу построения интерполяционного многочлена более высокой степени, если требуется повысить точность интерполяции. Так, предположим, что для заданного k построен интерполяционный многочлен

$$\varphi(t; \widehat{\beta}_1, \ldots, \widehat{\beta}_k) = \sum_{j=1}^{k} \widehat{\beta}_j a_j(t),$$

однако значение $S(\widehat{\beta}_1, \ldots, \widehat{\beta}_k)$ в (31) еще велико, т. е. точность приближения исследуемой функции многочленом степени $k-1$ недостаточна. Тогда строят интерполяционный многочлен степени k вида

$$\varphi(t; \widehat{\beta}_1, \ldots, \widehat{\beta}_k, \widehat{\beta}_{k+1}) = \sum_{j=1}^{k+1} \widehat{\beta}_j a_j(t),$$

т. е. добавляя следующий, $(k+1)$-й многочлен Чебышева $a_{k+1}(t)$ с коэффициентом $\widehat{\beta}_{k+1}$, вычисляемым по формуле (25) (при этом предыдущие коэффициенты $\widehat{\beta}_1, \ldots, \widehat{\beta}_k$ остаются прежними). Далее вычисляют

$$S(\widehat{\beta}_1, \ldots, \widehat{\beta}_{k+1}) = S(\widehat{\beta}_1, \quad , \widehat{\beta}_k) - a_{k+1}^2 \widehat{\beta}_{k+1}^2, \tag{32}$$

и если точность приближения, достигнутая с помощью многочлена k-й степени недостаточна, то можно подбирать далее многочлен $(k+1)$-й степени и т. д., пока не достигнем желаемой точности.

Пример 4 (Численная иллюстрация [11]). В «Основах химии» Д. И. Менделеев[1] приводит следующие данные о количестве (x) азотно-натриевой соли $NaNO_3$, которое можно растворить в 100 г воды в зависимости от температуры (t):

t_i	0	4	10	15	21	29	36	51	68
x_i	66,7	71,0	76,3	80,6	85,7	92,9	99,4	113,6	125,1

Построим по этим данным приближенную эмпирическую формулу вида $x = \beta_1 + \beta_2 t$, описывающую зависимость между рассматриваемыми величинами. Здесь речь идет о линейной зависимости, поэтому это схема простой регрессии, рассмотренная в примере 1. Используя формулы (22), получаем $\widehat{\beta}_1 = 67{,}5$, $\widehat{\beta}_2 = 0{,}87$ Таким образом, линейная приближенная формула имеет вид

$$x = 67{,}5 + 0{,}87t. \tag{33}$$

При этом сумма квадратов отклонений, вычисленная по формуле (23), равна 11,1.

Установим, как повышается точность приближения, если в качестве интерполяционного многочлена использовать квадратичную параболу. Запишем искомый многочлен в каноническом виде:

$$\varphi(t; \underline{\beta}) = \beta_1 a_1(t) + \beta_2 a_2(t) + \beta_3 a_3(t),$$

где $a_j(t)$ — многочлены Чебышева (напомним, что $a_1(t) \equiv 1$, $a_2(t) = t - \bar{t} = t - 26$, а $a_3(t)$, указан в (30)). Результат (33) можно записать в каноническом

[1] Менделеев Дмитрий Иванович (1834–1907) — выдающийся русский химик, разносторонний ученый, педагог; открыл периодический закон химических элементов (1869) — один из основных законов естествознания.

виде так:

$$x = 90{,}12a_1(t) + 0{,}87a_2(t),$$

и достаточно вычислить по формуле (25) $\widehat{\beta}_3$. В данном случае

$$a_3(t) = (t - 26)(t - 40) - 451{,}1,$$

поэтому значения $a_3(t_i)$ таковы: 588,9; 340,9; 28,9; −176,1; −356,1; −484,1; −484; −491,1; −176,1; 724,9. Отсюда

$$a_3^2 = \sum a_3^2(t_i) \approx 1{,}6 \quad 10^6 \qquad \sum a_3(t_i)x_i \approx -2{,}2 \quad 10^3$$

и по формуле (25) $\widehat{\beta}_3 \approx -1{,}4 \quad 10^{-3}\cdot$ поправка в (32) $a_3^2\widehat{\beta}_3^2 \approx 3{,}0$. Таким образом, учитывая многочлен $a_3(t)$, мы незначительно увеличиваем точность аппроксимации, и в качестве удовлетворительной эмпирической зависимости можно принять формулу (33). •

Общее замечание. Метод наименьших квадратов применяют также в случаях, когда зависимость $\mathbf{E}X_i$ от $\underline{\beta}$ не является линейной. Пусть

$$X_i = f(t_i; \beta_1, \ldots, \beta_k) + \varepsilon_i, \quad i = 1, \quad , n,$$

где $\mathbf{E}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = \sigma^2$, $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$. Тогда о. н. к. $\widehat{\underline{\beta}}$ параметров $\underline{\beta}$ минимизируют по $\underline{\beta}$ выражение

$$Q(\underline{\beta}) = \sum_{i=1}^{n} \left(X_i - f(t_i; \beta_1, \ldots, \beta_k)\right)^2$$

и для их отыскания надо решить систему уравнений

$$\frac{\partial Q(\underline{\beta})}{\partial \beta_i} = 0, \quad i = 1, \ldots, k.$$

§ 6.3. Нормальная регрессия

1. Модель нормальной регрессии

До сих пор делались предположения только о первых и вторых моментах наблюдений (см. условия (1′) и (2) § 6.1). Развитая при этих минимальных предположениях теория наименьших квадратов позволяет получать только точечные оценки для параметров $\beta_1, \quad , \beta_k$ и их линейных комбинаций. Чтобы получать более сильные утверждения (строить доверительные интервалы и проверять различные гипотезы) необходимо сделать дополнительные предположения о виде распределения наблюдений $\underline{X} = (X_1, \quad , X_n)$ или, что то же, вектора «ошибок» $\underline{\varepsilon} = (\varepsilon_1, \quad , \varepsilon_n)$ (см. (1)–(1′) § 6.1). Чаще всего предполагается, что этот вектор имеет нормальное распределение, т. е. в дополнение к (1′) и (2) § 6.1 полагается, что

$$\mathcal{L}(\underline{\varepsilon}) = \mathcal{N}(\underline{0}, \sigma^2 \mathbb{1}_n). \tag{1}$$

В этом случае говорят о *нормальной регрессии*. Отметим, что условия (1′)–(2) § 6.1 и (1) можно объединить и записать в виде

$$\mathcal{L}(\underline{X}) = \mathcal{N}(Z'\underline{\beta}, \sigma^2 \mathbb{1}_n). \tag{2}$$

Общие свойства многомерного нормального распределения изложены в п. 2 § 1.2. Подчеркнем, в частности, что условие (2) означает, что компоненты вектора $\underline{X}$ независимы и имеют одинаковые дисперсии σ^2, но их средние, вообще говоря, различны.

2. Оценки максимального правдоподобия (о. м. п.) параметров нормальной регрессии

Нормальная модель (2) определяется $(k+1)$-мерным параметром

$$\theta = (\beta_1, \quad , \beta_k, \sigma^2),$$

область возможных значений которого представляет собой евклидово полупространство

$$\Theta = \{\theta \quad -\infty < \beta_j < \infty, \ j = 1, \quad , k, \ \sigma^2 > 0\}.$$

Если $\underline{x} = (x_1, \quad , x_n)$ — наблюдавшаяся реализация вектора $\underline{X}$, то функция правдоподобия этих данных имеет вид (см. соотношение (9) § 1.2)

$$L(\underline{x}; \theta) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} S(\underline{x}; \underline{\beta})\right\}, \tag{3}$$

где квадратичная форма $S(\underline{x}; \underline{\beta})$ определена в (1) § 6.2.

Найдем о. м. п. параметров θ, которые максимизируют функцию $L(\underline{x}, \theta)$. Из (3) следует, что при любом $\sigma^2 > 0$ максимизация $L(\underline{x}, \theta)$ по $\underline{\beta}$ эквивалентна минимизации по $\underline{\beta}$ квадратичной формы $S(\underline{x}; \underline{\beta})$. Таким образом, *для модели нормальной регрессии о. н. к.* $\widehat{\underline{\beta}}$ *совпадает с о. м. п. этих параметров.*

Замечание. Из теоремы 2 § 6.2 следует, что о. н. к. являются оптимальными в классе линейных оценок. Для схемы нормальной регрессии справедливо более сильное утверждение: *о. н. к. являются оптимальными в классе всех* (не только линейных) *несмещенных оценок рассматриваемых параметрических функций.*

Найдем теперь о. м. п. $\widehat{\sigma}^2$ для остаточной дисперсии σ^2 Подставив в (3) вместо $\underline{\beta}$ оценку $\widehat{\underline{\beta}}$ и прологарифмировав, получим, что $\widehat{\sigma}^2$ — это то значение σ^2, которое минимизирует выражение

$$\frac{S(\widehat{\underline{\beta}})}{\sigma^2} + n \ln \sigma^2$$

Отсюда имеем

$$\widehat{\sigma}^2 = \frac{S(\widehat{\underline{\beta}})}{n}. \tag{4}$$

Ранее было показано (см. п. 3 § 6.2), что в общем случае несмещенная оценка для σ^2 имеет вид $S(\widehat{\underline{\beta}})/(n-k)$, поэтому о. м. п. $\widehat{\sigma}^2$ оказывается смещенной и ее смещение

$$\mathbf{E}\widehat{\sigma}^2 - \sigma^2 = \mathbf{E}\left(\frac{n-k}{n}\widetilde{\sigma}^2\right) - \sigma^2 = \left(\frac{n-k}{n} - 1\right)\sigma^2 = -\frac{k}{n}\sigma^2$$

убывает по модулю с ростом числа наблюдений n. Таким образом, оценка (4) является лишь асимптотически несмещенной (что вообще характерно для о. м. п.).

3. Основная теорема теории нормальной регрессии

Все дальнейшие выводы теории нормальной регрессии базируются на следующей важной теореме.

Теорема 1. Случайные величины $\widehat{\underline{\beta}}$ и $S(\widehat{\underline{\beta}}) = S(\underline{X}; \widehat{\underline{\beta}})$, а также $S(\widehat{\underline{\beta}})$ и $Q = S(\underline{\beta}) - S(\widehat{\underline{\beta}})$ независимы; при этом

$$\mathcal{L}_\theta(\widehat{\underline{\beta}}) = \mathcal{N}(\underline{\beta}, \sigma^2 A^{-1}), \quad \mathcal{L}_\theta\left(\frac{S(\widehat{\underline{\beta}})}{\sigma^2}\right) = \chi^2(n-k), \quad \mathcal{L}_\theta\left(\frac{Q}{\sigma^2}\right) = \chi^2(k). \tag{5}$$

Доказательство. Введем нормированный вектор «ошибок» $\underline{\varepsilon}^* = \underline{\varepsilon}/\sigma \cdot \mathcal{L}(\underline{\varepsilon}^*) = \mathcal{N}(\underline{0}, \mathbb{1}_n)$ (см. (1)). Тогда (см. (1′) § 6.1)

$$\underline{X} = Z'\underline{\beta} + \sigma\underline{\varepsilon}^* \tag{6}$$

а равенство (2) § 6.2 примет вид

$$\widehat{\underline{\beta}} = \underline{\beta} + \sigma A^{-1} Z \underline{\varepsilon}^* \tag{7}$$

Из представления (12) § 6.2 следует, что

$$\frac{S(\widehat{\underline{\beta}})}{\sigma^2} = \underline{\varepsilon}^{*\prime} B \underline{\varepsilon}^* \tag{8}$$

Наконец, из тождества (3) § 6.2 имеем следующее представление для Q:

$$Q = S(\underline{\beta}) - S(\widehat{\underline{\beta}}) = (\widehat{\underline{\beta}} - \underline{\beta})' A (\widehat{\underline{\beta}} - \underline{\beta}). \tag{9}$$

Независимость $\widehat{\underline{\beta}}$ и $S(\widehat{\underline{\beta}})$ следует теперь из представлений (7) и (8), легко проверяемого факта $A^{-1}ZB = A^{-1}Z(\mathbb{1}_n - Z'A^{-1}Z) = 0$ и леммы 1 § 2.5.

Далее, из представления (9) имеем, что Q зависит от данных $\underline{X}$ лишь через $\widehat{\underline{\beta}}$; следовательно, в силу независимости $\widehat{\underline{\beta}}$ и $S(\widehat{\underline{\beta}})$ случайные величины Q и $S(\widehat{\underline{\beta}})$ также независимы. Первая часть теоремы доказана.

Докажем теперь утверждения (5). Так как $\widehat{\underline{\beta}}$ — линейная функция нормального вектора (см. (7)), то $\widehat{\underline{\beta}}$ — также нормальный вектор, а его первые и вторые моменты вычислены в п. 2 § 6.2 и имеют указанный в (5) вид.

Закон распределения Q/σ^2 устанавливается на основании представления (9), факта нормальности вектора $\widehat{\underline{\beta}}$ и замечания к теореме 1 § 2.5.

Наконец, закон распределения $S(\widehat{\underline{\beta}})/\sigma^2$ следует из представления (8) и теоремы 1 § 2.5, поскольку матрица B идемпотентна и

$$\operatorname{tr} B = \operatorname{tr} \mathbb{1}_n - \operatorname{tr}(Z'A^{-1}Z) = n - \operatorname{tr}(ZZ'A^{-1}) = n - \operatorname{tr} \mathbb{1}_k = n - k. \ \blacksquare$$

Следствие 1. *Из первого соотношения в* (5) *следует, что для любого* $j = 1, \ldots, k$

$$\mathcal{L}_\theta\left(\frac{\widehat{\beta}_j - \beta_j}{\sigma\sqrt{a^{jj}}}\right) = \mathcal{N}(0, 1) \tag{10}$$

и $\widehat{\beta}_j$ *не зависит от* $S(\widehat{\underline{\beta}})$. *Поэтому из определения закона Стьюдента* (*п.* 6 *§* 1.2) *и из теоремы* 1 *следует, что при любом* θ

$$\mathcal{L}_\theta\left(t_{n-k}^{(j)} \equiv (\widehat{\beta}_j - \beta_j)\sqrt{\frac{n-k}{a^{jj}S(\widehat{\underline{\beta}})}}\right) = S(n-k). \tag{11}$$

Следствие 2. *Из определения закона Снедекора* (*п.* 7 *§* 1.2) *и теоремы* 1 *имеем, что при любом* θ

$$\mathcal{L}_\theta\left(F_{k,n-k} \equiv \frac{n-k}{k}\frac{Q}{S(\widehat{\underline{\beta}})}\right) = S(k, n-k). \tag{12}$$

4. Доверительное оценивание параметров нормальной регрессии

Построим доверительный интервал для коэффициента регрессии β_j. Из (10) следует, что статистика $\widehat{\beta}_j$ имеет распределение $\mathcal{N}(\beta_j, \sigma^2 a^{jj})$, поэтому здесь мы имеем задачу оценивания неизвестного среднего нормального закона с неизвестной же дисперсией (так как σ^2 неизвестно) по наблюдению над случайной величиной $\widehat{\beta}_j$. На основании (11) стьюдентово отношение $t_{n-k}^{(j)}$ является центральной статистикой для β_j (см. п. 1 § 3.8), поэтому γ-доверительный интервал для β_j строится по схеме примера 1 § 2.5 и имеет вид (сравни с (12) § 2.5).

Доверительные интервалы для коэффициентов регрессии

$$\left(\widehat{\beta}_j \mp t_{(1+\gamma)/2,\,n-k}\sqrt{\frac{a^{jj}}{n-k}S(\widehat{\underline{\beta}})}\right). \tag{13}$$

Чтобы построить доверительный интервал для остаточной дисперсии σ^2, воспользуемся вторым соотношением в (5), из которого следует, что $S(\widehat{\underline{\beta}})/\sigma^2$ — соответствующая центральная статистика. Искомый γ — доверительный

Доверительные интервалы для остаточной дисперсии

интервал строится здесь по схеме примера 1 § 2.5 и имеет вид (сравни с (10) § 2.5)

$$\left(\frac{S(\widehat{\underline{\beta}})}{\chi^2_{(1+\gamma)/2,\,n-k}},\ \frac{S(\widehat{\underline{\beta}})}{\chi^2_{(1-\gamma)/2,\,n-k}}\right). \tag{14}$$

Иногда требуется оценить весь параметрический вектор $\underline{\beta} = (\beta_1, \quad, \beta_k)$, т. е. построить доверительную область $\mathcal{G}_\gamma$ в евклидовом пространстве $\mathbb{R}^k$ накрывающую точку $\underline{\beta}$ с заданной вероятностью γ. Такую область можно построить, основываясь на результате (12). Действительно, если $F_{\gamma,\,k,\,n-k}$ есть γ-квантиль распределения $S(k, n-k)$, то из (9) и (12) следует, что при любом θ

$$\gamma = \mathbf{P}_\theta\{F_{k,\,n-k} < F_{\gamma,\,k,\,n-k}\} = \mathbf{P}_\theta\{\underline{\beta} \in \mathcal{G}_\gamma(\underline{X})\}, \tag{15}$$

Доверительная область

где

$$\mathcal{G}_\gamma(\underline{X}) = \left\{\underline{\beta} \quad (\widehat{\underline{\beta}}-\underline{\beta})' A(\widehat{\underline{\beta}}-\underline{\beta}) < \frac{k}{n-k} S(\widehat{\underline{\beta}}) F_{\gamma,\,k,\,n-k}\right\}$$

Тем самым построена искомая γ-доверительная область для $\underline{\beta}$: это внутренность эллипсоида с центром в точке $\widehat{\underline{\beta}}$, граница которого задается уравнением

$$(\underline{\beta} - \widehat{\underline{\beta}})' A(\underline{\beta} - \widehat{\underline{\beta}}) = a^2_\gamma(\underline{X}) = \frac{k}{n-k} S(\widehat{\underline{\beta}}) F_{\gamma,\,k,\,n-k}. \tag{16}$$

Пример 1 (Простая регрессия, доверительное оценивание параметров). Построим γ-доверительный интервал для наклона β_2 простой регрессии (см. пример 1 § 6.2). Из (10) и (22)–(23) § 6.2 следует, что это интервал

$$\left(\widehat{\beta}_2 \mp t_{(1+\gamma)/2,\,n-2} \sqrt{\frac{S(\widehat{\underline{\beta}})}{(n-2)\sum_{i=1}^{n}(t_i - \bar{t})^2}}\right)$$

Аналогично, из (15)–(16) и формул примера 1 § 6.2 имеем, что внутренность эллипса

$$(\beta_1 - \widehat{\beta}_1)^2 + 2\bar{t}(\beta_1 - \widehat{\beta}_1)(\beta_2 - \widehat{\beta}_2) + (\beta_2 - \widehat{\beta}_2)^2 \frac{1}{n}\sum_{i=1}^{n} t_i^2 = \frac{2}{n(n-2)} S(\widehat{\underline{\beta}}) F_{\gamma,\,2,\,n-2}$$

в плоскости переменных (β_1, β_2) с вероятностью γ содержит неизвестную точку $\underline{\beta}$. •

5. Доверительная область для линейных комбинаций параметров $\underline{\beta}$

Пусть требуется оценить одновременно $m \leqslant k$ линейных комбинаций $\underline{t} = (t_1, \ldots, t_m)$: $\underline{t} = T\underline{\beta}$, где T — заданная $(m \times k)$-матрица и $\operatorname{rank} T = m$. Тогда из теоремы 1 следует, что о. н. к. $\widehat{\underline{t}} = T\widehat{\underline{\beta}}$ имеет распределение $\mathcal{N}(\underline{t}, \sigma^2 D)$, где матрица D определена в (5) § 6.2. Отсюда по теореме 1 § 2.5 (см. замечание к ней) следует, что квадратична форма

$$Q_T = (\widehat{\underline{t}} - \underline{t})' D^{-1} (\widehat{\underline{t}} - \underline{t}) \tag{17}$$

имеет следующее распределение:

$$\mathcal{L}_\theta\left(\frac{Q_T}{\sigma^2}\right) = \chi^2(m).$$

При этом Q_T как функция $\widehat{\underline{\beta}}$ не зависит от $S(\widehat{\underline{\beta}})$ (теорема 1). Следовательно, статистика Снедекора имеет в данном случае вид

$$F_{m, n-k} = \frac{n-k}{m} \frac{Q_T}{S(\widehat{\underline{\beta}})}$$

и при этом $\mathcal{L}_\theta(F_{m,n-k}) = S(m, n-k)$. Отсюда, как и в случае оценивания $\underline{\beta}$, получаем следующий γ-доверительный эллипсоид для $\underline{t}$:

Доверительный эллипсоид

$$\mathcal{G}_{T\gamma}(\underline{X}) = \left\{\underline{t} \;\; (\underline{t} - T\widehat{\underline{\beta}})' D^{-1} (\underline{t} - T\widehat{\underline{\beta}}) < \frac{m}{n-k} S(\widehat{\underline{\beta}}) F_{\gamma, m, n-k}\right\}. \tag{18}$$

При $T = \mathbb{1}_k$ полученное решение сводится к (15).

Отметим некоторые частные случаи общего решения (18). При $m = 1$ речь идет об оценивании одной линейной комбинации $\underline{\lambda}'\underline{\beta} = \lambda_1\beta_1 + \ldots + \lambda_k\beta_k$. В этом случае эллипсоид (18) вырождается в интервал

$$\left(\underline{\lambda}'\widehat{\underline{\beta}} \mp t_{(1+\gamma)/2, n-k} \sqrt{(\underline{\lambda}' A^{-1} \underline{\lambda}) \frac{S(\widehat{\underline{\beta}})}{n-k}}\right), \tag{19}$$

поскольку $F_{\gamma, 1, n-k} = t^2_{(1+\gamma)/2, n-k}$ (показать, используя упр. 51 (б) к гл. 1).

Пусть матрица T имеет следующую структуру: в каждой строке только один элемент отличен от нуля, при этом в r-й строке на месте j_r стоит единица, $r = 1, \ldots, m$, $j_1 < j_2 < \ldots < j_m$. Тогда $T\underline{\beta} = (\beta_{j_1}, \ldots, \beta_{j_m}) \equiv \underline{\beta}(m)$ и, следовательно, речь идет об одновременном оценивании части координат вектора $\underline{\beta}$. В этом случае матрица D представляет собой минор

$$A^{-1}(j_1, \ldots, j_m) \equiv A(m)$$

матрицы A^{-1} получающийся вычеркиванием всех строк и столбцов с номерами, отличными от $j_1, \ldots, j_m$, и из (18) получаем, что γ-доверительный

эллипсоид для подвектора $\underline{\beta}(m)$ имеет вид

$$\mathcal{G}_{T\gamma}(\underline{X}) = \left\{ \underline{\beta}(m) \quad (\underline{\beta}(m) - \underline{\widehat{\beta}}(m))' A^{-1}(m)(\underline{\beta}(m) - \underline{\widehat{\beta}}(m)) < \right.$$

$$\left. < \frac{m}{n-k} S(\widehat{\beta}) F_{\gamma, m, n-k} \right\}. \tag{20}$$

Пример 2 (Доверительный интервал для ординаты линии регрессии). Пусть в схеме простой регрессии (см. пример 1 § 6.2) требуется оценить значение линии регрессии $\varphi(t) = \beta_1 + \beta_2 t$ в произвольной точке t. Используя обозначение

$$s^2(\underline{t}) = \frac{1}{n} \sum_{i=1}^{n} (t_i - \bar{t})^2$$

из формул примера 1 § 6.2 получаем, что при $\underline{\lambda} = (1, t)$

$$\underline{\lambda}' A^{-1} \underline{\lambda} = \frac{1}{n} \left[1 + \frac{(t - \bar{t})^2}{s^2(\underline{t})} \right]$$

Кроме того, $\widehat{\varphi}(t) = \underline{\lambda}' \underline{\widehat{\beta}} = \overline{X} + (t - \bar{t})\widehat{\beta}_2$. Отсюда и из (19) окончательно имеем, что искомый γ-доверительный интервал имеет вид

$$\left(\overline{X} + (t - \bar{t})\widehat{\beta}_2 \mp t_{(1+\gamma)/2, n-2} \sqrt{\frac{1}{n(n-2)} S(\underline{\widehat{\beta}}) \left[1 + \frac{(t - \bar{t})^2}{s^2(\underline{t})} \right]} \right) \quad \bullet$$

6. Система совместных доверительных интервалов

В конкретных практических задачах построение доверительных эллипсоидов в (15) и (18) — трудная вычислительная задача. Поэтому предпочтительнее иметь результат, сводящий проблему к построению системы доверительных интервалов для отдельных компонент оцениваемого параметрического вектора, с заданной вероятностью *одновременно* накрывающих соответствующие компоненты. В таких случаях говорят о *системе совместных доверительных интервалов*. Расчет такой системы основывается на следующем интересном утверждении из теории квадратичных форм.

Теорема 2. Пусть B — положительно определенная матрица и $\underline{t}$, $\underline{\lambda}$ — вектор-столбцы. Тогда

$$\underline{t}' B \underline{t} = \max_{\underline{\lambda}} \frac{(\underline{\lambda}' \underline{t})^2}{\underline{\lambda}' B^{-1} \underline{\lambda}}. \tag{21}$$

Доказательство. Представим B в виде $B = HH'$ Это всегда можно сделать, выбрав $H = U\Lambda^{1/2}$, где U — ортогональная матрица, приводящая B к диагональному виду Λ. Положим теперь $\underline{X} = H'\underline{t}$, $\underline{Y} = H^{-1}\underline{\lambda}$; тогда

$$\underline{X}'\underline{Y} = \underline{t}'\underline{\lambda}, \quad \underline{X}'\underline{X} = \underline{t}' B \underline{t}, \quad \underline{Y}'\underline{Y} = \underline{\lambda}' B^{-1} \underline{\lambda}.$$

По неравенству Коши—Буняковского[2] $(\underline{X}'\underline{Y})^2 \leqslant (\underline{X}'\underline{X})(\underline{Y}'\underline{Y})$, причем знак «=» $\Leftrightarrow$ векторы $\underline{X}$ и $\underline{Y}$ линейно зависимы. Отсюда и из предыдущих соотношений имеем $(\underline{\lambda}'\underline{t})^2 \leqslant (\underline{t}'B\underline{t})(\underline{\lambda}'B^{-1}\underline{\lambda})$ или $\underline{t}'B\underline{t} \geqslant (\underline{\lambda}'\underline{t})^2/(\underline{\lambda}'B^{-1}\underline{\lambda})$. ■

Положим теперь в соотношении (21) $\underline{t} = \underline{\widehat{\beta}} - \underline{\beta}$, $B = A$; тогда соотношения (15)–(16) можно записать в виде

$$\gamma = \mathbf{P}_\theta\left\{ \max_{\underline{\lambda}} \frac{|\underline{\lambda}'(\underline{\widehat{\beta}} - \underline{\beta})|}{(\underline{\lambda}'A^{-1}\underline{\lambda})^{1/2}} < a_\gamma(\underline{X}) \right\} = \mathbf{P}_\theta\{|\underline{\lambda}'(\underline{\widehat{\beta}} - \underline{\beta})| < u_\gamma(\underline{X}; \underline{\lambda}), \forall\, \underline{\lambda}\},$$

где

$$u_\gamma(\underline{X}; \underline{\lambda}) = a_\gamma(X)(\underline{\lambda}'A^{-1}\underline{\lambda})^{1/2} \tag{22}$$

Таким образом, для любого θ выполняется соотношение

$$\mathbf{P}_\theta\{\underline{\lambda}'\underline{\widehat{\beta}} - u_\gamma(\underline{X}; \underline{\lambda}) < \underline{\lambda}'\underline{\beta} < \underline{\lambda}'\underline{\widehat{\beta}} + u_\gamma(\underline{X}; \underline{\lambda}),\ \forall\ \underline{\lambda}\} = \gamma. \tag{23}$$

Тем самым для всех (!) линейных функций $\underline{\lambda}'\underline{\beta}$ указана система совместных доверительных интервалов с доверительным уровнем γ. Если требуется оценить конечное число линейных комбинаций $\underline{\lambda}_1'\underline{\beta}, \quad , \underline{\lambda}_m'\underline{\beta}$, то из (23) имеем следующий результат:

$$\mathbf{P}_\theta\{\underline{\lambda}_r'\underline{\widehat{\beta}} - u_\gamma(\underline{X}; \underline{\lambda}_r) < \underline{\lambda}_r'\underline{\beta} < \underline{\lambda}_r'\underline{\widehat{\beta}} + u_\gamma(\underline{X}; \underline{\lambda}_r),\ r = 1, \ldots, m\} \geqslant \gamma. \tag{24}$$

Пример 3 (совместные доверительные интервалы для коэффициентов регрессии). Пусть $1 \leqslant m \leqslant k$ и $\underline{\lambda}_r = (0 \quad 010 \quad 0)$, где 1 стоит на месте j_r, $r = 1, \quad , m$. Тогда $\underline{\lambda}_r'\underline{\beta} = \beta_{j_r}$, $\underline{\lambda}_r'A^{-1}\underline{\lambda}_r = a^{j_r j_r}$ и на основании (24)

$$\left\{\left(\widehat{\beta}_{j_r} \mp a_\gamma(\underline{X})\sqrt{a^{j_r j_r}}\right),\ r = 1, \quad , m\right\}$$

— система совместных доверительных интервалов уровня $\geqslant \gamma$ для координат $\beta_{j_1}, \quad , \beta_{j_m}$ При $m = k$ получаем решение для всех коэффициентов регрессии $\beta_1, \ldots, \beta_k$. •

7. Доверительный интервал для отклика

На практике большой интерес представляет предсказание будущих значений отклика X при тех или иных заданных значениях факторов $z_1, \ldots, z_k$. В этом случае речь идет о построении доверительного интервала для значения X_* отклика X, соответствующего фиксированным значениям $z_{*1}, \quad , z_{*k}$ факторов:

$$X_* = \sum_{j=1}^{k} \beta_j z_{*j} + \varepsilon_* = \underline{z}_*'\underline{\beta} + \varepsilon_*. \tag{25}$$

Предположим, что мы уже оценили вектор $\underline{\beta}$ по n наблюдениям $\underline{X} = (X_1, \ldots, X_n)$ и что ε_* не зависит от $\underline{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$ и $\mathcal{L}(\varepsilon_*) = \mathcal{N}(0, \sigma^2)$.

[2] Буняковский Виктор Яковлевич (1804–1889) — русский математик, специалист по теории вероятностей, теории чисел и математическому анализу.

Тогда мы можем использовать оценку $\widehat{X}_* = \underline{z}_*'\widehat{\underline{\beta}}$, которая на основании теоремы 1 имеет нормальное распределение:

$$\mathcal{L}_\theta(\widehat{X}_*) = \mathcal{N}\left(\underline{z}_*'\underline{\beta}, \sigma^2 \underline{z}_*' A^{-1} \underline{z}_*\right). \tag{26}$$

Но также

$$\mathcal{L}_\theta(X_*) = \mathcal{N}(\underline{z}_*'\underline{\beta}, \sigma^2)$$

и X_* не зависит от $\widehat{X}_*$. Отсюда следует, что

$$\mathcal{L}_\theta(\widehat{X}_* - X_*) = \mathcal{N}\left(0, \sigma^2(1 + \underline{z}_*' A^{-1} \underline{z}_*)\right). \tag{27}$$

Это дает нам возможность построения стьюдентова отношения

$$t_{n-k} = \frac{\widehat{X}_* - X_*}{\sigma\sqrt{1 + \underline{z}_*' A^{-1} \underline{z}_*}} \Big/ \sqrt{\frac{S(\widehat{\underline{\beta}})}{\sigma^2(n-k)}} = \frac{(\widehat{X}_* - X_*)\sqrt{n-k}}{\sqrt{S(\widehat{\underline{\beta}})(1 + \underline{z}_*' A^{-1} \underline{z}_*)}}$$

Доверительный интервал для отклика

и получения следующего γ-доверительного интервала для X_*.

$$\left(\widehat{X}_* \mp t_{(1+\gamma)/2,\, n-k} \sqrt{S(\widehat{\underline{\beta}}) \frac{1 + \underline{z}_*' A^{-1} \underline{z}_*}{n-k}}\right) \tag{28}$$

Пример 4 (Прогнозирование в схеме простой регрессии). Пусть в схеме простой регрессии требуется построить γ-доверительный интервал для отклика X_* в точке t_*. Используя обозначения примера 2, находим, что интервал (28) принимает вид

$$\left(\overline{X} + (t_* - \bar{t})\widehat{\beta}_2 \mp t_{(1+\gamma)/2,\, n-2} \sqrt{\left(1 + \frac{1}{n}\left(1 + \frac{(t_* - \bar{t})^2}{s^2(\underline{t})}\right)\right) \frac{S(\widehat{\underline{\beta}})}{n-2}}\right) \tag{29}$$

Подчеркнем, что этот интервал отличен от соответствующего интервала для ординаты $\varphi(t_*)$ (см. пример 2): он шире, что естественно, так как случайная флуктуация ε_* вносит дополнительную неопределенность в проблему. •

8. Проверка адекватности модели

В практических ситуациях при подборке подходящей модели типа модели нормальной регрессии (2), важным является вопрос проверки адекватности модели, который может быть конкретизирован так: все ли факторы, оказывающие существенное влияние на отклик X, учтены в модели (2), или же действительное число факторов больше k? Если модель (2) справедлива, то, как показано в § 6.2, статистика $\widetilde{\sigma}^2$ (см. (10) § 6.2) является несмещенной оценкой для остаточной дисперсии σ^2 Если же реальная модель имеет на самом деле больше коэффициентов (факторов), то оценка $\widetilde{\sigma}^2$ будет смещенной и при этом $\mathbf{E}\widetilde{\sigma}^2 > \sigma^2$ В этом случае использование модели (2) может привести к неправильному предсказанию отклика. Поэтому при проверке адекватности

основное внимание уделяется недопущению модели с меньшим числом коэффициентов (факторов), чем требуется.

Чаще всего процедуры проверки адекватности модели состоят в сравнении оценки $\widetilde{\sigma}^2$ с независимой от нее другой оценкой $\overset{*}{\sigma}{}^2$ для остаточной дисперсии σ^2. Если $\widetilde{\sigma}^2$ соизмеримо с $\overset{*}{\sigma}{}^2$ влияние неадекватности можно считать незначительным и модель (2) принимается; если же $\widetilde{\sigma}^2$ существенно больше $\overset{*}{\sigma}{}^2$ то этим влиянием (неадекватности) нельзя пренебречь, и модель надо скорректировать, добавляя новые коэффициенты.

Чтобы получить независимую оценку $\overset{*}{\sigma}{}^2$ надо провести в неизменных условиях (т. е. при некотором фиксированном значении факторов $z_1, \quad , z_k$) серию независимых опытов, в результате которых получим выборку значений отклика $(Y_1, \quad , Y_m)$. В этом случае наблюдения $Y_1, \quad Y_m$ будут независимыми одинаково распределенными нормальными случайными величинами с дисперсией σ^2 и несмещенная оценка для σ^2 по этим данным имеет вид

$$\overset{*}{\sigma}{}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2 \qquad \overline{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i, \tag{30}$$

а ее распределение есть

$$\mathcal{L}\left(\frac{m-1}{\sigma^2}\overset{*}{\sigma}{}^2\right) = \chi^2(m-1). \tag{31}$$

Следовательно, тестовой статистикой для проверки гипотезы адекватности

$$H_0 \quad \mathbf{E}\widetilde{\sigma}^2 = \sigma^2 \tag{32}$$

является дисперсионное отношение Фишера

$$F_{n-k,\,m-1} = \frac{\widetilde{\sigma}^2}{\overset{*}{\sigma}{}^2} \tag{33}$$

Если гипотеза адекватности (32) истинна, то статистика $F_{n-k,\,m-1}$ имеет распределение Снедекора с $n-k$ и $m-1$ степенями свободы:

$$\mathcal{L}(F_{n-k,\,m-1}|H_0) = S(n-k, m-1). \tag{34}$$

Алгоритм проверки адекватности модели

Следовательно, алгоритм проверки адекватности регрессионной модели состоит в следующем:

1. После оценивания регрессионных коэффициентов $\underline{\beta} = (\beta_1, \quad , \beta_k)$, т. е. получения о. н. к. $\underline{\widehat{\beta}} = (\widehat{\beta}_1, \quad , \widehat{\beta}_k)$, вычисляют значение оценки $\widetilde{\sigma}^2$ в (10) § 6.2 для остаточной дисперсии σ^2
2. Проводят серию новых опытов $\underline{Y} = (Y_1, \quad , Y_m)$ и по (30) получают независимую оценку для σ^2
3. Вычисляют дисперсионное отношение (33).
4. При заданном уровне значимости α (напр., $\alpha = 0{,}05$) и числах степеней свободы $n-k$ и $m-1$ по таблицам распределения Снедекора определяется величина F_T:

$$\mathbf{P}\{F_{n-k,\,m-1} > F_T\} = \alpha.$$

5. Сравнивают величины $F_{n-k,m-l}$ (см. (33)) и F_T и делают один из двух выводов:

А. Если $F_{n-k,m-l} \leqslant F_T$, модель считается адекватной.

Б. В противном случае модель неадекватна и надо изменить ее структуру, введя дополнительные коэффициенты, и заново собрать данные. Например, если полином первой степени (простая регрессия) оказался неадекватным, то надо перейти к модели второй степени и т. д.

§ 6.4. Общая линейная гипотеза нормальной регрессии

1. Понятие линейной гипотезы

При применении схемы нормальной регрессии на практике часто возникает необходимость проверить те или иные гипотезы о значениях коэффициентов регрессии $\beta_1, \quad, \beta_k$. В общем виде задача формулируется следующим образом: по наблюдениям $\underline{X} = (X_1, \quad, X_n)$ требуется проверить гипотезу H_0, согласно которой коэффициенты $\beta_1, .., \beta_k$ удовлетворяют некоторым заданным ограничениям, локализующим их допустимые значения в некотором подмножестве $\mathcal{B}_0 \in R^k$. Если эти ограничения линейные, то говорят о *линейной* гипотезе. Для линейной гипотезы подмножество $\mathcal{B}_0$ является линейным подпространством вида

Линейная гипотеза

$$\mathcal{B}_0 = \{\underline{\beta} \quad T\underline{\beta} = \underline{t}_0\}, \tag{1}$$

где T — заданная $(m \times k)$-матрица $(m \leqslant k)$, $\operatorname{rank} T = m$, $\underline{t}_0$ — заданный вектор и предполагается, что система $T\underline{\beta} = \underline{t}_0$ совместна. Таким образом, в общем случае линейная гипотеза записывается в виде $H_0 \quad \underline{\beta} \in \mathcal{B}_0$. В любом случае это сложная гипотеза, так как она оставляет произвольным значение параметра σ^2. В терминах общего параметра модели $\theta = (\underline{\beta}, \sigma^2)$ можно записать

$$H_0 \quad \theta \in \Theta_0 = \{\theta \quad \underline{\beta} \in \mathcal{B}_0, \ \sigma^2 > 0\}. \tag{2}$$

Примером конкретной линейной гипотезы является утверждение

$$H_0 \quad \beta_2 = \beta_{20},$$

задающее конкретное значение наклона линии регрессии в модели простой регрессии (см. пример 1 § 6.2).

2. F-критерий для линейной гипотезы

Чтобы построить критерий проверки гипотезы H_0, воспользуемся уже известным нам решением задачи доверительного оценивания (см. п. 5 § 6.3) и принципом соответствия между проверкой гипотез и доверительным оцениванием (см. п. 4 § 5.3). Напомним общую схему этого соответствия.

Пусть в некоторой модели $\mathcal{F} = \{F(x;\theta), \theta \in \Theta\}$ построена система γ-доверительных областей $\{\mathcal{G}_\gamma(\underline{x}), \underline{x} \in \mathcal{X}\}$ для параметрической функции $g = g(\theta)$.

Тогда подмножество выборочного пространства, задаваемое условием

$$\mathfrak{X}_{0\alpha} = \{\underline{x} \quad g_0 \in \mathcal{G}_\gamma(\underline{x})\}, \quad \alpha = 1 - \gamma,$$

определяет область принятия гипотезы $H_0 \quad g = g_0$ с уровнем значимости α.

В нашем случае, согласно (17)–18) § 6.3, γ-доверительная область для $\underline{t} = T\underline{\beta}$ имеет вид

$$\mathcal{G}_{T\gamma} = \left\{\underline{t} \quad \frac{Q_T(\underline{X}, \underline{t})}{S(\widehat{\underline{\beta}}(\underline{X}))} < \frac{m}{n-k} F_{\gamma, m, n-k}\right\},$$

поэтому область принятия гипотезы $H_0 : \underline{t} = \underline{t}_0$ с уровнем значимости $\alpha = 1 - \gamma$ задается условием

$$\mathfrak{X}_{0\alpha} = \left\{\underline{x} \quad \frac{Q_T(\underline{x}, t_0)}{S(\widehat{\underline{\beta}}(\underline{x}))} < \frac{m}{n-k} F_{\gamma, m, n-k}\right\}.$$

Отсюда имеем, что *критерий для гипотезы* $H_0 : \underline{\beta} \in \mathcal{B}_0$ (см. (1)) *задается критической областью*

F-критерий

$$\mathfrak{X}_{1\alpha} = \left\{\underline{x} \quad \frac{n-k}{m} \frac{Q_T(\underline{x}, \underline{t}_0)}{S(\widehat{\underline{\beta}}(\underline{x}))} \geqslant F_{1-\alpha, m, n-k}\right\} \tag{3}$$

Выпишем явное выражение для тестовой статистики критерия (3). В соответствии с (17) § 6.3 она есть

$$F \equiv \frac{n-k}{m} \frac{Q_T(\underline{X}, \underline{t}_0)}{S(\widehat{\underline{\beta}})} = \frac{n-k}{m} \frac{(T\widehat{\underline{\beta}} - \underline{t}_0)' D^{-1} (T\widehat{\underline{\beta}} - \underline{t}_0)}{S(\widehat{\underline{\beta}})} \tag{4}$$

и большие ее значения являются критическими для проверяемой гипотезы. Построенный критерий называют *F-критерием.*

Приведем и другой подход к построению критерия (3), проливающий дополнительный свет на его существо. Рассмотрим о. н. к. $\widehat{\underline{t}} \quad T\widehat{\underline{\beta}}$ для $\underline{t}$ и составим вектор отклонений $T\widehat{\underline{\beta}} - \underline{t}_0$. Если гипотеза H_0 не верна, то весьма вероятно, что эти отклонения велики. Будем измерять степень отклонения величиной $Q_T(\underline{X}, \underline{t}_0)$. Тогда, поскольку математическое ожидание квадратичной формы можно преобразовать следующим образом (ниже $\Sigma = \mathbf{D}(\underline{Y}) = \|\operatorname{cov}(Y_i, Y_j)\|$):

$$\mathbf{E}(\underline{Y}' B \underline{Y}) = \sum_{i,j} b_{ij} \mathbf{E}(Y_i Y_j) = \sum_{i,j} b_{ij} \operatorname{cov}(Y_i, Y_j) + \sum_{i,j} b_{ij} \mathbf{E}Y_i \mathbf{E}Y_j =$$
$$= \operatorname{tr}(B\Sigma) + \mathbf{E}(\underline{Y}') B \mathbf{E}(\underline{Y}),$$

учитывая (5) § 6.2, имеем при $\underline{Y} = T\widehat{\underline{\beta}} - \underline{t}_0$, $B = D^{-1}$

$$\frac{1}{m}\mathbf{E}Q_T(\underline{X}, \underline{t}_0) = \frac{\sigma^2}{m} \operatorname{tr}(\mathbb{1}_m) + \frac{1}{m}\mathbf{E}(T\widehat{\underline{\beta}} - \underline{t}_0)' D^{-1} \mathbf{E}(T\widehat{\underline{\beta}} - \underline{t}_0) =$$
$$= \sigma^2 + \frac{1}{m}(T\underline{\beta} - \underline{t}_0)' D^{-1} (T\underline{\beta} - \underline{t}_0) \geqslant \sigma^2;$$

знак равенства имеет здесь место только при $T\underline{\beta} = \underline{t}_0$, т. е. при справедливости гипотезы H_0. С другой стороны, как мы знаем из п. 3 § 6.2, $S(\underline{\widehat{\beta}})/(n-k)$ — несмещенная оценка для σ^2 всегда, т. е. безотносительно к истинному значению $\underline{\beta}$. Поэтому у отношения F в (4) числитель в среднем больше знаменателя, когда гипотеза H_0 ложна, и потому большие значения F естественно рассматривать как свидетельствующие против гипотезы H_0. Из этих рассуждений также следует, что статистика F является отношением двух независимых и несмещенных (при гипотезе H_0) оценок для остаточной дисперсии σ^2. С такой

Дисперсионное отношение

дисперсионное отношение интерпретацией связан термин, используемый иногда для статистики F.

Для вычисления статистики F можно использовать также формулу

$$F = \frac{n-k}{m}\,\frac{S_T - S(\underline{\widehat{\beta}})}{S(\underline{\widehat{\beta}})}, \tag{5}$$

которая следует из (18) § 6.2. Таким образом, дисперсионное отношение F можно выразить непосредственно через условный S_T (при условии $T\underline{\beta} = \underline{t}_0$) и абсолютный $S(\underline{\widehat{\beta}})$ минимум исходной квадратичной формы $S(\underline{\beta})$. В конкретных задачах условный минимум S_T зачастую находят непосредственно, поэтому представление (5) более удобно для практических расчетов, нежели (4).

Пример 1 (Простая регрессия, гипотеза о наклоне ее линии регрессии). Рассмотрим гипотезу H_0 $\beta_2 = \beta_{20}$, фиксирующую значение наклона линии регрессии (см. пример 1 § 6.2). В данном случае $m = 1$ и условный минимум (при гипотезе H_0)

$$S_T = \min_{\beta_1} \sum_{i=1}^{n} (X_i - \beta_{20} t_i - \beta_1)^2$$

находится непосредственно: он достигается при $\beta_1 = \overline{X} - \beta_{20}\bar{t}$ (черта сверху, как обычно, означает среднее арифметическое) и его значение, учитывая соотношения (22)–(23) § 6.2, можно записать в виде (проверить самостоятельно!)

$$S_T = S(\underline{\widehat{\beta}}) + (\widehat{\beta}_2 - \beta_{20})^2 \sum_{i=1}^{n} (t_i - \bar{t})^2$$

По формуле (5) имеем, что в данном случае дисперсионное отношение равно

$$F = (n-2)(\widehat{\beta}_2 - \beta_{20})^2 \frac{1}{S(\underline{\widehat{\beta}})} \sum_{i=1}^{n} (t_i - \bar{t})^2, \tag{6}$$

а критическая граница в (3) выбирается равной $F_{1-\alpha,\,1,\,n-2}$. Так как

$$F_{1-\alpha,\,1,\,n-2} = t^2_{1-\alpha/2,\,n-2}$$

(см. замечание к формуле (19) § 6.3), то F-критерий в рассматриваемом случае сводится к критерию

$$H_0 \text{ отвергается} \iff |\widehat{\beta}_2 - \widehat{\beta}_{20}| \geqslant t_{1-\alpha/2,\,n-2} \sqrt{\frac{S(\widehat{\beta})}{(n-2)\sum_{i=1}^{n}(t_i - \bar{t})^2}}, \tag{7}$$

где $S(\widehat{\beta})$ определено в (23) § 6.2).

Отметим, что к этому результату можно прийти также и обращая доверительный интервал для β_2, построенный в примере 1 § 6.3).

В том случае, когда $\beta_{20} = 0$, речь идет о проверке значимости влияния фактора t на исход эксперимента, и если гипотеза H_0 $\beta_2 = 0$ отвергается, то говорят, что регрессия значима и влиянием фактора t пренебрегать нельзя.

Добавим к этому следующее. Как легко видеть из соотношений (22)–(23) § 6.2, можно записать также представление

$$S(\widehat{\beta}) = \sum (X_i - \bar{X})^2 - \frac{\left[\sum (X_i - \bar{X})(t_i - \bar{t})\right]^2}{\sum (t_i - \bar{t})^2} = (1 - R^2)\sum (X_i - \bar{X})^2 \tag{8}$$

где

$$R = \frac{\sum (X_i - \bar{X})(t_i - \bar{t})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (t_i - \bar{t})^2}}, \tag{9}$$

есть выборочный коэффициент корреляции между X и t.

Из (8) следует, что чем больше значение R^2 (чем ближе оно к 1), тем меньше $S(\widehat{\beta})$, т. е. тем лучше подобранная прямая регрессии соответствует наблюдениям. Тем самым, если гипотеза H_0 $\beta_2 = 0$ отвергается, то величина $1 - R^2$ является *мерой согласия* подобранной прямой регрессии данным наблюдений.

Наконец, F-статистика (6) при $\beta_{20} = 0$ с учетом (8)–(9) может быть записана в виде

$$F = \frac{(n-2)\widehat{\beta}_2^2 \sum (t_i - \bar{t})^2}{(1 - R^2)\sum (X_i - \bar{X})^2} = \frac{(n-2)R^2}{1 - R^2}, \tag{10}$$

т. е. может быть выражена через R^2, а критерий (7) для гипотезы $\beta_2 = 0$ может быть записан соответственно в виде

$$\left\{ \frac{|R|}{\sqrt{(1 - R^2)/(n-2)}} \geqslant t_{1-\alpha/2,\,n-2} \right\}. \tag{11}$$

•

Пример 2 (Критерий значимости регрессии). В приложениях теории линейной регрессии обычно первый коэффициент β_1 выделяется, т. е. полагается первый фактор $z_1 \equiv 1$ (как в схеме простой регрессии (21) § 6.2). В этом случае наблюдаемая переменная X зависит от $k-1$ фактора $z_2, \quad , z_k$ и часто бывает желательно удостовериться, является ли их влияние на X значимым. Итак, если априори предполагать, что модель имеет вид

$$X_i = \beta_1 + \beta_2 z_2^{(i)} + \quad + \beta_k z_k^{(i)} + \varepsilon_i, \quad i = 1, \quad , n, \tag{12}$$

то речь идет о проверке гипотезы $H_0 \quad \beta_2 = \quad = \beta_k = 0$. Если гипотеза H_0 отвергается, то говорят, что регрессия значима и факторами $z_2, \quad , z_k$ нельзя, вообще говоря, пренебречь (в то же время следует иметь в виду, что отклонение гипотезы H_0 вовсе не означает, что модель (12) действительно адекватна).

В данном случае гипотеза H_0 имеет вид $T\underline{\beta} = \underline{0}$, где $T = \|\underline{0} \quad \mathbb{1}_{k-1}\|$ — матрица размера $(k-1) \times k$ и ранга $k-1$, так что применима общая теория F-критерия с $m = k-1$ и

$$S_T = \min_{\beta_1} \sum_{i=1}^{n} (X_i - \beta_1)^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 = \underline{X}'\underline{X} - n\overline{X}^2$$

Кроме того, величину $S(\widehat{\underline{\beta}})$ в силу соотношений (10)–(12) § 6.2 можно записать также в виде

$$S(\widehat{\underline{\beta}}) = \underline{X}'(\mathbb{1}_n - Z'A^{-1}Z)\underline{X} = \underline{X}'\underline{X} - \widehat{\underline{\beta}}'Z\underline{X}. \tag{13}$$

Следовательно, F-статистика (5) имеет в данном случае вид

$$F = \frac{n-k}{k-1} \frac{\widehat{\underline{\beta}}'Z\underline{X} - n\overline{X}^2}{\underline{X}'\underline{X} - \widehat{\underline{\beta}}'Z\underline{X}}, \tag{14}$$

а F-критерий для гипотезы H_0 (т. е. критерий для «всей» регрессии в целом) задается критической областью $\{F \geqslant F_{1-\alpha, k-1, n-k}\}$, где α — заданный уровень значимости.

Полезной мерой степени соответствия аппроксимирующей регрессии $\widehat{X} = Z'\widehat{\underline{\beta}}$ имеющимся данным $\underline{X} = (X_1, \ldots, X_n)$ является *выборочный множественный коэффициент корреляции* R_{k-1}. Он определяется как коэффициент корреляции между данными $\{X_1, \quad , X_n\}$ и их оценками $\{\widehat{X}_1, \quad , \widehat{X}_n\}$:

$$R_{k-1} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(\widehat{X}_i - \overline{\widehat{X}})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2 \sum_{i=1}^{n} (\widehat{X}_i - \overline{\widehat{X}})^2}}. \tag{15}$$

Величину R^2_{k-1} обычно называют *коэффициентом детерминации.* Коэффициент детерминации

Покажем, что

$$S(\underline{\widehat{\beta}}) = (1 - R^2_{k-1}) \sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{16}$$

Доказательство. Действительно, рассмотрим первое уравнение системы нормальных уравнений $\partial S(\underline{\beta})/\partial \beta_1 = 0$: оно имеет вид

$$\sum_i \left(X_i - \widehat{\beta}_1 - \widehat{\beta}_2 z_2^{(i)} \qquad - \widehat{\beta}_k z_k^{(i)}\right) = 0$$

или $\sum_i (X_i - \widehat{X}_i) = 0$, т. е. $\overline{X} = \overline{\widehat{X}}_i$. С учетом этого имеем

$$\sum_i (X_i - \overline{X})^2 = \sum_i (X_i - \widehat{X}_i + \widehat{X}_i - \overline{X})^2 = \sum_i (X_i - \widehat{X}_i)^2 + \sum_i (\widehat{X}_i - \overline{X})^2 \tag{17}$$

поскольку (см. (2) § 6.2)

$$\sum_i (X_i - \widehat{X}_i)(\widehat{X}_i - \overline{X}) = \sum_i (X_i - \widehat{X}_i)\widehat{X}_i = (\underline{X} - \widehat{X})'\widehat{X} = \underline{X}'\widehat{X} - \widehat{X}'\widehat{X} =$$

$$= \underline{X}' Z' A^{-1} Z \underline{X} - \underline{X}' Z' A^{-1} Z Z' A^{-1} Z \underline{X} = 0.$$

Из (17) следует, что

$$S(\underline{\widehat{\beta}}) = \sum_i (X_i - \widehat{X}_i)^2 = \sum_i (X_i - \overline{X})^2 - \sum_i (\widehat{X}_i - \overline{X})^2 \tag{18}$$

Далее,

$$\sum_i (X_i - \overline{X})(\widehat{X}_i - \overline{\widehat{X}}) = \sum_i (X_i - \overline{X})(\widehat{X}_i - \overline{X}) =$$

$$= \sum_i (X_i - \widehat{X}_i + \widehat{X}_i - \overline{X})(\widehat{X}_i - \overline{X}) = \sum_i (\widehat{X}_i - \overline{X})^2,$$

поэтому (см. (15))

$$R^2_{k-1} = \frac{\sum_i (\widehat{X}_i - \overline{X})^2}{\sum_i (X_i - \overline{X})^2}. \tag{19}$$

Теперь соотношение (16) непосредственно следует из (18) и (19). ■

Из равенства (16) видно, что чем ближе значение R^2_{k-1} к 1, тем лучше аппроксимирующая поверхность $\widehat{X} = Z'\underline{\widehat{\beta}}$ соответствует данным наблюдений. Отметим, что при $k = 2$ (один фактор) мы получаем соответствующие

результаты примера 1. Так, коэффициент $R_1 = R$ (см. (9)), а равенство (16) соответствует равенству (8). Можно также показать, что аналогом представления (10) является здесь следующая формула для статистики (14):

$$F = (n-k)\frac{R^2_{k-1}}{1-R^2_{k-1}}. \tag{20}$$

•

§ 6.5. Некоторые применения теории нормальной регрессии

Изложенная в предыдущем параграфе теория проверки гипотез о коэффициентах регрессии имеет широкое применение. К схеме регрессии могут быть сведены многие гипотезы о математических ожиданиях нормальных совокупностей, которые зачастую имеют форму, на первый взгляд, отличную от рассмотренной выше. Некоторые из таких задач рассматриваются в этом параграфе.

1. Гипотеза о параллельности линий регрессии

Предположим, что имеется $r \geqslant 2$ групп независимых наблюдений $X_1^{(i)}$ $X_{n_i}^{(i)}$ $i = 1, \ldots, r$, распределенных нормально с общей (неизвестной) дисперсией σ^2 и средними

$$\mathbf{E}X_j^{(i)} = \mu_j^{(i)} = \beta_1^{(i)} + \beta_2^{(i)} t_j^{(i)} \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, r. \tag{1}$$

Другими словами, имеется r различных схем простой регрессии (пример 1 § 6.2) и проверяемая гипотеза H_0 состоит в том, что все линии регрессии (1) имеют один и тот же наклон, т. е.

$$H_0 \quad \beta_2^{(1)} = \quad = \beta_2^{(r)}$$

Такая гипотеза может встретиться, например, при проверке равенства нескольких скоростей роста, при сравнении нескольких способов обработки и т. д.

Покажем, что рассматриваемый случай можно свести к схеме общей линейной гипотезы. Положим

$$X = (X_1, \ldots, X_n) = \left(X_1^{(1)}, \quad, X_{n_1}^{(1)}, X_1^{(2)}, \quad, X_{n_2}^{(2)}, \ldots, X_1^{(r)} \quad, X_{n_r}^{(r)}\right),$$

$$n = n_1 + \quad + n_r,$$

$$\underline{\beta} = (\beta_1, \quad, \beta_{2r}) = 2\left(\beta_1^{(1)}, \beta_2^{(1)}, \beta_1^{(2)}, \beta_2^{(2)} \quad, \beta_1^{(r)}, \beta_2^{(r)}\right)$$

и введем матрицу Z размера $2r \times n$, составленную из векторов-столбцов $\underline{z}^{(i)}$ вида

$$Z = \left\| \underline{z}^{(1)} \ldots \underline{z}^{(n_1)} \underline{z}^{(n_1+1)} \ldots \underline{z}^{(n)} \right\| =$$

$$= \left\| \begin{array}{cccc} 1 & 1 & 0 & 0 \\ t_1^{(1)} & t_{n_1}^{(1)} & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & t_1^{(2)} & t_{n_2}^{(2)} \end{array} \right. \begin{array}{c} \\ \\ \\ \\ \end{array} \begin{array}{c} \\ \\ \\ \\ \end{array} \left. \begin{array}{cc} & \\ & \\ & \\ & \\ & \\ & \\ 1 & 1 \\ t_1^{(r)} & t_{(n_r)}^{(r)} \end{array} \right\|$$

В этих обозначениях вектор $\underline{X}$ распределен по нормальному закону $\mathcal{N}(Z'\underline{\beta}, \sigma^2 \mathbb{1}_n)$. Таким образом, имеем частный случай (при указанном выборе $\underline{z}^{(i)}$) схемы нормальной регрессии. Проверяемую гипотезу можно записать в виде $H_0 \quad \beta_4 - \beta_2 = \quad = \beta_{2r} - \beta_2 = 0$, т. е. она задается $(r-1)$-м линейным соотношением между коэффициентами регрессии и, следовательно, для ее проверки можно применить F-критерий § 6.4. Для этого надо найти абсолютный $S_1 = S(\widehat{\underline{\beta}})$ и условный S_T (т. е. при гипотезе H_0) минимумы квадратичной формы

$$S(\underline{\beta}) = \sum_{i=1}^{n} \left(X_i - \underline{z}^{(i)'}\underline{\beta}\right)^2 = \sum_{i,j} \left(X_j^{(i)} - \mu_j^{(i)}\right)^2 \tag{2}$$

Нахождение S_1 сводится к минимизации выражения

$$S(\underline{\beta}) = \sum_{i=1}^{r} S^{(i)},$$

где

$$S^{(i)} = \sum_{j=1}^{n_i} \left(X_j^{(i)} - \beta_1^{(i)} - \beta_2^{(i)} t_j^{(i)}\right)^2 \qquad i = 1, \quad , r,$$

т. е. достаточно найти минимум каждой квадратичной формы $S^{(i)}$ отдельно. В несколько отличных обозначениях эта задача решена в примере 1 § 6.2, поэтому из (22) и (23) § 6.2 имеем, что минимум $S^{(i)}$ достигается при

$$\beta_1^{(i)} = \widehat{\beta}_1^{(i)} \qquad \beta_2^{(i)} = \widehat{\beta}_2^{(i)},$$

где

$$\widehat{\beta}_2^{(i)} = \frac{\sum_{j=1}^{n_i} \left(X_j^{(i)} - \overline{X}^{(i)}\right)\left(t_j^{(i)} - \bar{t}^{(i)}\right)}{\sum_{j=1}^{n_i} \left(t_j^{(i)} - \bar{t}^{(i)}\right)^2}, \qquad \widehat{\beta}_1^{(i)} = \overline{X}^{(i)} - \bar{t}^{(i)}\widehat{\beta}_2^{(i)} \tag{3}$$

(здесь $\overline{X}^{(i)} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_j^{(i)}$ $\overline{t}^{(i)} = \frac{1}{n_i}\sum_{j=1}^{n_i} t_j^{(i)}$ и при этом предполагается, что для любого $i = 1, \ldots, r$ не все числа $t_1^{(i)}$ $, t_{n_i}^{(i)}$ одинаковы). Таким образом, в данном случае

$$S_1 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left[\left(X_j^{(i)} - \overline{X}^{(i)}\right)^2 - \widehat{\beta}_2^{(i)\,2}\left(t_j^{(i)} - \overline{t}^{(i)}\right)^2\right] \tag{4}$$

Введем следующие обозначения:

$$\begin{aligned} s_i^2(t) &= \sum_{j=1}^{n_i}\left(t_j^{(i)} - \overline{t}^{(i)}\right)^2 \\ s_i^2(X) &= \sum_{j=1}^{n_i}\left(X_j^{(i)} - \overline{X}^{(i)}\right)^2 \\ s_i(X,t) &= \sum_{j=1}^{n_i}\left(X_j^{(i)} - \overline{X}^{(i)}\right)\left(t_j^{(i)} - \overline{t}^{(i)}\right), \quad i = 1, \quad , r \end{aligned} \tag{5}$$

(с точностью до множителя — это выборочные дисперсии и ковариация множеств $(X_1^{(i)}, \quad , X_{n_i}^{(i)})$ и $(t_1^{(i)}, \ldots, t_{n_i}^{(i)})$). Тогда из (3) и (4) следует, что S_1 можно записать в следующем, удобном для вычислений, виде:

$$S_1 = \sum_{i=1}^{r}\left[s_i^2(X) - \frac{s_i^2(X,t)}{s_i^2(t)}\right]. \tag{6}$$

Для вычисления S_T надо минимизировать по $\beta_1^{(1)}$ $, \beta_1^{(r)}, \beta_2$ сумму

$$\sum_{i,j}\left(X_j^{(i)} - \beta_1^{(i)} - \beta_2 t_j^{(i)}\right)^2$$

т. е. приравнивая нулю все ее частные производные по этим параметрам. Непосредственным дифференцированием можно проверить (это мы оставляем читателю в качестве простого упражнения), что при любом фиксированном β_2 минимум достигается при $\beta_1^{(i)} = b_1^{(i)} = \overline{X}^{(i)} - \beta_2\overline{t}^{(i)}$ $i = 1, \quad , r$, и сумма квадратов при этом равна

$$\sum_{i,j}\left[\left(X_j^{(i)} - \overline{X}^{(i)}\right) - \beta_2\left(t_j^{(i)} - \overline{t}^{(i)}\right)\right]^2$$

Минимум же последнего выражения достигается при

$$\beta_2 = b_2 = \frac{\sum_{i,j}\left(X_j^{(i)} - \overline{X}^{(i)}\right)\left(t_j^{(i)} - \overline{t}^{(i)}\right)}{\sum_{i,j}\left(t_j^{(i)} - \overline{t}^{(i)}\right)^2} = \frac{s_1(X,t) + \quad + s_r(X,t)}{s_1^2(t) + \ldots + s_r^2(t)}$$

(см. (5)), и он равен

$$S_T = \sum_{i=1}^{r} s_i^2(X) - \frac{\left[\sum_{i=1}^{r} s_i(X, t)\right]^2}{\sum_{i=1}^{r} s_i^2(t)}. \tag{7}$$

В данном случае параметры k и m в (5) § 6.4 равны соответственно $2r$ и $r-1$, поэтому статистика F в силу соотношений (6) и (7) принимает вид

$$F = \frac{n-2r}{r-1} \cdot \frac{\sum_{i=1}^{r} \frac{s_i^2(X, t)}{s_i^2(t)} - \frac{\left[\sum_{i=1}^{r} s_i(X, t)\right]^2}{\sum_{i=1}^{r} s_i^2(t)}}{\sum_{i=1}^{r} s_i^2(X) - \sum_{i=1}^{r} \frac{s_i^2(X, t)}{s_i^2(t)}}. \tag{8}$$

В итоге имеем, что при заданном уровне значимости α F-критерий для проверяемой гипотезы H_0 задается критической областью $\{F \geqslant F_{1-\alpha, r-1, n-2r}\}$.

2. Гипотеза однородности для нескольких нормальных выборок

В примере 3 § 5.4 была сформулирована общая гипотеза о равенстве средних значений нескольких нормальных выборок. Такая гипотеза часто возникает в приложениях, например, при сравнении нескольких различных способов обработки или процедур, условий, размещений и т. п. с целью выяснения, влияют ли различия между ними на интересующий исследователя исход. Вообще подобная проблема имеет место каждый раз, когда интересуются влиянием на исход эксперимента какого-то одного фактора. Если число различных значений этого фактора (число уровней, на которых может находиться фактор) равно k, то n_j — число наблюдений, соответствующих j-му уровню фактора, а $X^{(j)} = (X_{j1}, \ldots, X_{jn_j})$ — сами наблюдения (результаты эксперимента). В этом случае говорят также *об одинарной классификации исходов*.

В примере 3 § 5.4 мы исследовали эту проблему (проверка гипотезы однородности) с позиций критерия отношения правдоподобия, но получили окончательное решение лишь для случая двух ($k = 2$) выборок (см. критерий (9) § 5.4). Общее же решение для произвольного числа выборок было получено лишь для случая больших выборок (см. критерий (29) § 5.4). Теперь же мы в состоянии получить общее решение для произвольного числа выборок и произвольных их объемов, если воспользуемся теорией F-критерия. Для этого сведем рассматриваемую схему к схеме регрессии.

Положим $\underline{X}$ $(X_1, \quad , X_n)$ $(\underline{X}^{(1)}, \quad , \underline{X}^{(k)})$, n $n_1 + \quad + n_k$, и введем матрицу плана

$$Z = \left\| \underline{z}^{(1)} \quad \underline{z}^{(n_1)} \underline{z}^{(n_1+1)} \quad \underline{z}^{(n)} \right\| = \begin{Vmatrix} 1 & 1 & 0 & & & 0 \\ 0 & 0 & 1 & & 0 & 0 \\ \cdot & & & & & \cdot \\ \cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot & & & & & \cdot \\ 0 & & & & 0 & 1 \end{Vmatrix}$$

где в j-й строке единицы стоят на местах $n_1 + \ldots + n_{j-1} + 1$, $n_1 + \ldots + n_j$, $j = 1, \ldots, k$ $(n_0 = 0)$. Тогда $X_1, \ldots, X_n$ независимы (по условию) и

$$\mathcal{L}(X_i) = \mathcal{N}\left(\underline{z}^{(i)'}\underline{\beta}, \theta_2^2\right), \quad i = 1, \ldots, n, \quad \underline{\beta} = (\theta_{11}, \quad , \theta_{1k}).$$

Таким образом, мы имеем здесь частный случай (с указанной матрицей плана Z) схемы нормальной регрессии, в которой требуется проверить гипотезу H_0 $\theta_{12} - \theta_{11} = \quad = \theta_{1k} - \theta_{11} = 0$ (H_0 задается $(k-1)$-м линейным соотношением между коэффициентами регрессии). Следовательно, можно применить теорию F-критерия. Для этого надо вычислить абсолютный S_1 и условный S_T (при гипотезе H_0) минимумы квадратичной формы

$$S = \sum_{i,j} (X_{ji} - \theta_{1j})^2$$

Так как

$$S = \sum_{i,j} (X_{ji} - \overline{X}_j)^2 + \sum_j n_j (\overline{X}_j - \theta_{1j})^2 \qquad \overline{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji},$$

то при отсутствии ограничений на $\theta_{11}, \ldots, \theta_{1k}$ минимум достигается при $\theta_{1j} = \overline{X}_j$, $j = 1, \quad , k$, и равен

$$S_1 = \sum_{i,j} (X_{ji} - \overline{X}_j)^2 \quad \sum_{j=1}^{k} s_j^2, \quad s_j^2 = \sum_{i=1}^{n_j} (X_{ji} - \overline{X}_j)^2 \tag{9}$$

Далее имеем

$$S_T = \min_{\theta_1} \sum_{i,j} (X_{ji} - \theta_1)^2$$

Здесь можем записать разложение

$$\sum_{i,j} (X_{ji} - \theta_1)^2 = \sum_{i,j} (X_{ji} - \overline{X})^2 + n(\overline{X} - \theta_1)^2$$

$$\overline{X} = \frac{1}{n} \sum_{i,j} X_{ji} = \frac{1}{n} \sum_{j=1}^{k} n_j \overline{X}^{(j)},$$

отсюда следует, что

$$S_T = \sum_{i,}(X_{ji} - \overline{X})^2 = S_1 + \sum_{j=1}^{k} n_j(\overline{X}^{(j)} - \overline{X})^2 \tag{10}$$

и минимум достигается при $\theta_1 = \overline{X}$.

Таким образом, в данном случае статистика F принимает вид (см. (5) § 6.4)

$$F = \frac{n-k}{k-1} \frac{\sum_{j=1}^{k} n_j(\overline{X}^{(j)} - \overline{X})^2}{\sum_{j=1}^{k} s_j^2}, \tag{11}$$

а F-критерий для гипотезы однородности H_0 задается при уровне значимости α критической областью $\{F \geqslant F_{1-\alpha,k-1,n-k}\}$. Нетрудно убедиться в том, что для случая двух выборок ($k = 2$) полученное решение сводится к (9) § 5.4.

[redacted] F-критерий однородности

Отметим, что соотношение (10) можно интерпретировать как разложение полной суммы квадратов отклонений наблюдений от общего среднего («полной изменчивости») на сумму квадратов отклонений каждой величины X_{ij} от соответствующего группового среднего значения («изменчивость внутри групп») и сумму квадратов отклонений групповых средних от общего среднего («изменчивость между группами»). При этом, если гипотеза H_0 истинна, то (см. формулу (11) § 2.2)

$$\mathbf{E}S_T = \mathbf{E}\sum_{i,j}(X_{ji} - \overline{X})^2 = (n-1)\theta_2^2,$$

$$\mathbf{E}S_1 = \sum_{j=1}^{k} \mathbf{E}s_j^2 = \sum_{j=1}^{k}(n_j - 1)\theta_2^2 = (n-k)\theta_2^2,$$

$$\mathbf{E}\sum_{j=1}^{k} n_j(\overline{X}^{(j)} - \overline{X})^2 = \mathbf{E}S_T - \mathbf{E}S_1 = (k-1)\theta_2^2.$$

Отсюда следует, что (при нулевой гипотезе H_0) каждая из статистик

$$Q = \frac{1}{n-1}\sum_{i,j}(X_{ji} - \overline{X})^2,$$

$$Q_1 = \frac{1}{k-1}\sum_{j} n_j(\overline{X}^{(j)} - \overline{X})^2,$$

$$Q_2 = \frac{1}{n-k}\sum_{i,j}(X_{ij} - \overline{X}^{(j)})^2$$

является несмещенной оценкой дисперсии θ_2^2, и F-критерий можно считать *критерием совместности независимых оценок* Q_1 и Q_2 (статистику F можно записать в виде $F = Q_1/Q_2$) Такое разложение изменчивости исходных данных и его интерпретация характерны для *дисперсионного анализа* — раздела математической статистики, объединяющего совокупность статистических приемов, широко применяемых в самых разнообразных экспериментах, в которых наблюдения тем или иным способом классифицируются (разбиваются на группы или классы) в соответствии со значениями некоторых сопутствующих факторов.

3. Двойная классификация. Дисперсионный анализ

Рассмотренный в предыдущем пункте случай является простейшей схемой дисперсионного анализа (один фактор), в которой решение получается с помощью методов регрессионного анализа. Рассмотрим применение регрессионного анализа еще к одной, имеющей большое практическое значение, схеме дисперсионного анализа, когда число факторов, влияющих на исход эксперимента, равно двум.

Предположим, что исход эксперимента зависит от значений двух факторов R и C, причем фактор R может находиться на r уровнях $R_1, \ldots, R_r$, а фактор C — на s уровнях $C_1, \ldots, C_s$. Пусть, далее, при каждой возможной комбинации уровней обоих факторов производится ровно одно наблюдение: X_{ij} при комбинации (R_i, C_j), $i = 1, \ldots, r$; $j = 1, \ldots, s$. При этом наблюдения $\{X_{ij}\}$ независимы и распределены нормально с одинаковыми дисперсиями σ^2 и средними значениями

$$\mathbf{E}X_{ij} = \xi_{ij} = \mu + \alpha_i + \beta_j, \tag{12}$$

где

$$\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{s} \beta_j = 0. \tag{13}$$

Условия (12) и (13) означают, что оба фактора действуют независимо (в этом случае говорят, что они *аддитивны*). Действительно, если обозначать через α' и β' действия факторов R и C соответственно, то аддитивность означает, что $\xi_{ij} = \alpha_i' + \beta_j'$. Если положить теперь $\mu = \bar{\alpha}' + \bar{\beta}'$, $\alpha_i = \alpha_i' - \bar{\alpha}'$, $\beta_j = \beta_j' - \bar{\beta}'$ (черта сверху, как обычно, означает среднее арифметическое), то получим соотношения (12)–(13).

Такую схему удобно представить в виде прямоугольной таблицы из r строк, соответствующих уровням $R_1, \ldots, R_r$ и s столбцов, соответствующих уровням $C_1, \ldots, C_s$. Случайную величину X_{ij} называют *реагирующей*: она описывает *реакцию*, вызываемую комбинацией (R_i, C_j) факторов R и C

Из (12) следует, что $\mathbf{E}X_{ij}$ при любых i и j оказывается при условии (13) линейной функцией от $r+s-1$ неизвестных коэффициентов μ, α_i, β_j Таким образом, здесь мы также имеем частный случай схемы нормальной регрессии

и, следовательно, можно применить развитую выше теорию. Все ее положения, как мы знаем, базируются на анализе квадратичной формы

$$S = \sum_{i=1}^{r} \sum_{j=1}^{s} (X_{ij} - \xi_{ij})^2 \tag{14}$$

В данном случае удобно преобразовать эту форму следующим образом. Обозначим

$$\overline{X}_{..} = \frac{1}{rs} \sum_{i,j} X_{ij}, \quad \overline{X}_{i.} = \frac{1}{s} \sum_{j=1}^{s} X_{ij}, \quad \overline{X}_{.j} = \frac{1}{r} \sum_{i=1}^{r} X_{ij}.$$

Тогда, учитывая (12), имеем

$$\begin{aligned} S &= \sum_{i,j} (X_{ij} - \mu - \alpha_i - \beta_j)^2 = \\ &= \sum_{i,j} \left[(X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..}) + (\overline{X}_{i.} - \overline{X}_{..} - \alpha_i) + \right. \\ &\quad \left. + (\overline{X}_{.j} - \overline{X}_{..} - \beta_j) + (\overline{X}_{..} - \mu) \right]^2 = \\ &= \sum_{i,j} (X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..})^2 + s \sum_{i} (\overline{X}_{i.} - \overline{X}_{..} - \alpha_i)^2 + \\ &\quad + r \sum_{j} (\overline{X}_{.j} - \overline{X}_{..} - \beta_j)^2 + rs(\overline{X}_{..} - \mu)^2, \end{aligned} \tag{15}$$

так как все попарные произведения после суммирования дадут нули.

Оценки наименьших квадратов для коэффициентов регрессии — это значения параметров, обращающие квадратичную форму (14) в минимум, поэтому из (15) сразу получаем, что в данном случае эти оценки таковы:

$$\widehat{\mu} = \overline{X}_{..}, \quad \widehat{\alpha}_i = \overline{X}_{i.} - \overline{X}_{..}, \quad \widehat{\beta}_j = \overline{X}_{.j} - \overline{X}_{..}. \tag{16}$$

Из (15) и (16) также имеем, что минимальное значение формы S равно

$$S_1 = \min S = \sum_{i,j} (X_{ij} - \widehat{\xi}_{ij})^2 = \sum_{i,j} (X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..})^2 \tag{17}$$

Найдем дисперсии полученных оценок (16). Так как

$$\overline{X}_{..} = \frac{1}{r} \sum_{l=1}^{r} \overline{X}_{l.} = \frac{1}{s} \sum_{t=1}^{s} \overline{X}_{.t},$$

то

$$\widehat{\alpha}_i = \frac{r-1}{r} \overline{X}_{i.} - \frac{1}{r} \sum_{l \neq i} \overline{X}_{l.}, \quad \widehat{\beta}_j = \frac{s-1}{s} \overline{X}_{.j} - \frac{1}{s} \sum_{t \neq j} \overline{X}_{.t}.$$

Но $\overline{X}_{l.}$, $l = 1, \ldots, r$, независимы и $\mathrm{D}\overline{X}_{l.} = \sigma^2/s$; аналогично, $\overline{X}_{.t}$, $t = 1, \ldots, s$, независимы и $\mathrm{D}\overline{X}_{.t} = \sigma^2/r$; следовательно,

$$\mathbf{D}\widehat{\alpha}_i = \left(\frac{(r-1)^2}{r^2} + \frac{r-1}{r^2}\right)\frac{\sigma^2}{s} = \frac{r-1}{rs}\sigma^2,$$
$$\mathbf{D}\widehat{\beta}_j = \frac{s-1}{rs}\sigma^2 \qquad \mathbf{D}\widehat{\mu} = \frac{\sigma^2}{rs}. \tag{18}$$

Доверительные интервалы

В соответствии с общей теорией (см. п. 4 § 6.3) отсюда находим, что γ-доверительными интервалами для параметров μ, α_i и β_j являются соответственно (в данном случае параметр $n-k$ в (13) § 6.3 равен $rs-(r+s-1)=(r-1)(s-1)$)

$$\left(\widehat{\mu} \mp t_{(1+\gamma)/2,(r-1)(s-1)}\sqrt{\frac{S_1}{rs(r-1)(s-1)}}\right),$$

$$\left(\widehat{\alpha}_i \mp t_{(1+\gamma)/2,(r-1)(s-1)}\sqrt{\frac{S_1}{rs(s-1)}}\right), \quad \left(\widehat{\beta}_j \mp t_{(1+\gamma)/2,(r-1)(s-1)}\sqrt{\frac{S_1}{rs(r-1)}}\right),$$

где $\widehat{\mu}$, $\widehat{\alpha}_i$, $\widehat{\beta}_j$ и S_1 определены в (16) и (17).

Отметим также, что $S_1/(r-1)(s-1)$ — несмещенная оценка σ^2

Проверка гипотез

Рассмотрим теперь задачу проверки наиболее интересных в исследуемой схеме гипотез. Пусть требуется проверить гипотезу $H_0^{(1)}$ о том, что фактор R не влияет на исход испытаний, т. е.

$$H_0^{(1)} \quad \alpha_1 = \qquad = \alpha_r = 0. \tag{19}$$

Так, например, главным интересующим исследователя фактором может быть фактор C, соответствующий, скажем, различным способам обработки, в то время как R соответствует классификации по месту, где производятся наблюдения. В этом случае гипотеза $H_0^{(1)}$ соответствует предположению о том, что эта вспомогательная классификация не влияет на результаты эксперимента и, следовательно, может не приниматься во внимание.

Следуя общей теории F-критерия (см. § 6.4), вычислим сначала S_T-минимальное значение квадратичной формы (14) при гипотезе $H_0^{(1)}$ Из (15) и (17) имеем

$$S_T = S_1 + s\sum_{i=1}^{r}(\overline{X}_{i.} - \overline{X}_{..})^2$$

поэтому статистика F (см. (5) § 6.4) имеет в данном случае вид (при $m = r-1$)

$$F = s(s-1)\frac{\sum_{i=1}^{r}(\overline{X}_{i.} - \overline{X}_{..})^2}{\sum_{i,j}(X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..})^2}, \tag{20}$$

а F-критерий уровня значимости α для гипотезы $H_0^{(1)}$ задается критической областью

$$\mathfrak{X}_{1\alpha}^{(1)} = \{F \geqslant F_{1-\alpha, r-1, (r-1)(s-1)}\}. \tag{21}$$

Аналогично рассматривается гипотеза

$$H_0^{(2)} \quad \beta_1 = \quad = \beta_s = 0. \tag{22}$$

о несущественности влияния на исходы испытаний второго фактора C. Здесь критическая область имеет вид

$$\mathfrak{X}_{1\alpha}^{(2)} = \left\{ r(r-1)\frac{\sum_{j=1}^{s}(\overline{X}_{\cdot j} - \overline{X}_{\cdot\cdot})^2}{\sum_{i,j}(X_{ij} - \overline{X}_{i\cdot} - \overline{X}_{\cdot j} + \overline{X}_{\cdot\cdot})^2} \geqslant F_{1-\alpha, s-\quad -1)} \right\} \tag{23}$$

Наконец, критерий уровня значимости α для гипотезы

$$H_0^{(3)} \quad \alpha_1 = \quad = \alpha_r = \beta_1 = \quad = \beta_s = 0$$

о независимости результатов испытаний от влияния обоих факторов задается критической областью

$$\mathfrak{X}_{1\alpha}^{(3)} = \left\{ \frac{(r-1)(s-1)}{r+s-2} \cdot \frac{s\sum_{i=1}^{r}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2 + r\sum_{j=1}^{s}(\overline{X}_{\cdot j} - \overline{X}_{\cdot\cdot})^2}{\sum_{i,j}(X_{ij} - \overline{X}_{i\cdot} - \overline{X}_{\cdot j} + \overline{X}_{\cdot\cdot})^2} \geqslant F_{1-\alpha, \atop r+s-2, \atop (r-1)(s-1)} \right\} \tag{24}$$

Доверительные области

Найдем, наконец, доверительные области для различных групп параметров модели (12). Пусть требуется построить γ-доверительную область $\mathcal{G}_\gamma^{(1)}$ для эффектов $\alpha_1, \quad, \alpha_r$ фактора R. Согласно принципу соответствия между проверкой гипотез и доверительным оцениванием (см. п. 4 § 5.3), чтобы построить $\mathcal{G}_\gamma^{(1)}$, надо рассмотреть соответствующую гипотезу

$$H_0 \quad \alpha_i = \alpha_{i0}, \quad i = 1, \ldots, r \quad \left(\sum_{i=1}^{r} \alpha_{i0} = 0\right).$$

Критическую область уровня значимости $\alpha = 1 - \gamma$ для нее можно получить из (20)–(21), заменяя X_{ij} на $X_{ij} - \alpha_{i0}$; она имеет вид

$$\left\{ \sum_{i=1}^{r}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot} - \alpha_{i0})^2 \geqslant F_{\gamma, r-1, (r-1)(s-1)} \frac{1}{s(s-1)} \sum_{i,j}(X_{ij} - \overline{X}_{i\cdot} - \overline{X}_{\cdot j} + \overline{X}_{\cdot\cdot})^2 \right\}.$$

Отсюда следует вывод, что

$$\mathcal{G}_\gamma^{(1)} = \left\{ (\alpha_1, \quad , \alpha_r) \ \sum_{i=1}^{r} \alpha_i = 0 \quad \text{и} \right.$$

$$\left. \sum_{i=1}^{r} (\alpha_i - \overline{X}_{i.} + \overline{X}_{..})^2 < F_{\gamma, r-1, (r-1)(s-1)} \frac{1}{s(s-1)} \sum_{i,j} (X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..})^2 \right\}$$

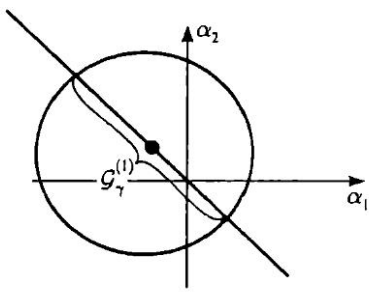


Рис. 1

Здесь неравенство определяет внутренность шара в пространстве параметров $(\alpha_1, \quad , \alpha_r)$ с центром $(\overline{X}_{1.} - \overline{X}_{..}, \ \overline{X}_{r.} - \overline{X}_{..})$ в гиперплоскости $\sum \alpha_i = 0$. Так, если $r = 2$, то доверительная область — отрезок биссектрисы $\alpha_2 = -\alpha_1$, являющийся диаметром некоторой окружности с центром на этой биссектрисе (см. рис. 1).

Аналогично строится доверительная область для эффектов $\beta_1, \ , \beta_s$ второго фактора.

Интерпретация В заключение дадим интерпретацию рассмотренной схемы, характерную для дисперсионного анализа. Положим в разложении (15) все α_i и β_j равными нулю, а $\mu = \overline{X}_{..}$. В результате получим соотношение

$$S_0 \equiv \sum_{i,j} (X_{ij} - \overline{X}_{..})^2 = s \sum_i (\overline{X}_{i.} - \overline{X}_{..})^2 + r \sum_j (\overline{X}_{.j} - \overline{X}_{..})^2 + S_1,$$

которое можно интерпретировать как разложение полной изменчивости S_0 на три компоненты:

$$S_{.0} = s \sum_i (X_{i.} - \overline{X}_{..})^2 \quad S_{0.} = r \sum_j (X_{.j} - \overline{X}_{..})^2 \quad \text{и} \quad S_{..} = S_1. \qquad (25)$$

Компонента $S_{.0}$ описывает изменчивость, обусловленную первым фактором R, компонента $S_{0.}$ вторым фактором C. Компонента же $S_{..}$ есть сумма квадратов величин с нулевыми средними:

$$\mathrm{E}(X_{ij} - \overline{X}_{i.} - \overline{X}_{.j} + \overline{X}_{..}) = \mu + \alpha_i + \beta_j - (\mu + \alpha_i) - (\mu + \beta_j) + \mu = 0,$$

поэтому она не может быть связана с факторами R и C. (Напомним, что величина $Q = S_{..}/(r-1)(s-1)$ является несмещенной оценкой дисперсии σ^2.) Величину $S_{..}$ принято называть «ошибкой», подчеркивая тем самым, что она связана со случайностью результатов наблюдений, а не с каким-либо расхождением в их средних.

Вычислим, наконец, $\mathbf{E}S_{.0}$ и $\mathbf{E}S_{0.}$. Из (16) и (18) имеем

$$\mathbf{E}(\overline{X}_{i.} - \overline{X}_{..})^2 = \mathbf{E}\widehat{\alpha}_i^2 = \mathbf{D}\widehat{\alpha}_i + (\mathbf{E}\widehat{\alpha}_i)^2 = \frac{r-1}{rs}\sigma^2 + \alpha_i^2.$$

Отсюда и из (25) следует, что

$$\mathbf{E}S_{.0} = (r-1)\sigma^2 + s\sum_{i=1}^{r}\alpha_i^2. \tag{26}$$

Аналогично находим

$$\mathbf{E}S_{0.} = (s-1)\sigma^2 + r\sum_{j=1}^{s}\beta_i^2. \tag{27}$$

Из (26) имеем, что при гипотезе (19) величина

$$Q_1 = \frac{S_{.0}}{r-1}$$

может служить несмещенной оценкой для σ^2. Аналогично, из (27) заключаем, что

$$Q_2 = \frac{S_{0.}}{s-1}$$

— несмещенная оценка σ^2, когда справедлива гипотеза (22).

Теперь F-критерий (21) для гипотезы $H_0^{(1)}$ (19) можно интерпретировать как критерий совместности двух независимых оценок Q_1 и Q для σ^2 (статистика (20) в этих обозначениях равна отношению Q_1/Q). Аналогично можно интерпретировать и критерий (23) для гипотезы $H_0^{(2)}$ (22) (его тестовой статистикой является Q_2/Q).

Обычно составляющие этой двухфакторной модели объединяют в следующую *таблицу дисперсионного анализа*:

Источник дисперсии	Степень свободы	Сумма квадратов	Среднее суммы квадратов	Отношение Снедекора
строки	$r-1$	$S_{.0}$	$Q_1 = \dfrac{S_{.0}}{r-1}$	$F_{.0} = \dfrac{Q_1}{Q}$
столбцы	$s-1$	$S_{0.}$	$Q_2 = \dfrac{S_{0.}}{s-1}$	$F_{0.} = \dfrac{Q_2}{Q}$
ошибка	$(r-1)(s-1)$	$S_{..}$	$Q = \dfrac{S_{..}}{(r-1)(s-1)}$	

Первое отношение Снедекора $F_{.0}$ служит для проверки гипотезы $H_0^{(1)}$ о том, что все α_i равны нулю, второе — для проверки гипотезы $H_0^{(2)}$ о равенстве нулю всех β_j.

§ 6.6. Статистическая регрессия и прогнозирование

Очень трудно что-либо предсказывать,
но особенно трудно предсказывать будущее.

Китайская пословица

1. Задачи статистического прогноза

В различных областях человеческой деятельности часто возникают ситуации, когда по имеющейся информации (данным) X требуется предсказать (спрогнозировать, оценить) некоторую величину Y, стохастически связанную с X (т. е. X и Y имеют некоторое совместное распределение $\mathcal{L}(X, Y)$), но которую непосредственно измерить невозможно (например, Y может относиться к будущему, а X — к настоящему). Так, например, может представлять интерес прогноз успеваемости первокурсников очередного набора по оценкам, полученным ими на вступительных экзаменах. Здесь X было бы средним баллом этих студентов на вступительных экзаменах, а Y — средним баллом по итогам, скажем, первой сессии; при этом совместное распределение $\mathcal{L}(X, Y)$ можно в принципе определить (оценить) по аналогичным данным за прошлые годы. Или: можно ли, зная рост отца (X), предсказать рост его сына (Y), если у выбранных случайно из большой популяции людей эти характеристики (X, Y) имеют известный (нормальный) закон распределения? Такой же характер имеют задачи прогнозирования погоды по результатам соответствующих атмосферных измерений, селекционирования новых видов растений и животных, определения возможностей индивидуумов в определенных областях с помощью соответствующей системы контрольных тестов и т. д. Особо важное значение прогнозирование имеет в таких областях, как индустрия, экономика, коммерция (прогнозирование экономических показателей, динамики цен на тот или иной продукт, курса акций на какое-то время вперед и т. д.). Во всех этих случаях речь идет о величинах, недоступных непосредственному наблюдению (измерению) в данный момент и которые надо оценивать (прогнозировать) с помощью доступных измерению сопутствующих величин.

В общем случае X означает некоторую совокупность $\{X_1, X_2, \ldots\}$ наблюдаемых случайных величин, которые в рассматриваемом контексте называются *предсказывающими* (или *прогнозными*) переменными, и задача состоит в построении такой функции $\varphi(X)$ от этих величин, которую можно было бы использовать в качестве оценки для прогнозируемой величины Y $\varphi(X) \approx Y$ (т. е. чтобы она была в каком-то смысле «близка» к Y); такие функции называются *предикторами* величины Y по X. Задачей разработки методов построения оптимальных (в том или ином смысле) предиктов занимается *теория статистической регрессии* (или, что то же, *стохастического прогнозирования*).

Для построения этой теории прежде всего требуется уточнить смысл приближенного равенства $\varphi(X) \approx Y$ Если $\varphi(X)$ используется для предсказания величины Y, то одной из разумных мер расхождения между ними является квадратическая ошибка $(\varphi(X) - Y)^2$, но так как величина Y неизвестна,

то для измерения точности предиктора φ используется *среднеквадратическая ошибка* (с. к. о.) $\Delta(\varphi) = \mathbf{E}(\varphi(X) - Y)^2$. Предиктор, минимизирующий с. к. о. в заданном классе предикторов Φ, называется *оптимальным предиктором*, или, кратко, *прогнозом* (в классе Φ), и обозначается символом φ^*:

Прогноз

$$\varphi^* = \arg\min_{\varphi \in \Phi} \Delta(\varphi); \tag{1}$$

его с. к. о. есть

$$\Delta^* = \Delta(\varphi^*) = \mathbf{E}(\varphi^*(X) - Y)^2 = \inf_{\varphi \in \Phi} \mathbf{E}(\varphi(X) - Y)^2 \tag{2}$$

Таким образом, в общем случае задача построения наилучшего (в среднеквадратическом смысле и в заданном классе предикторов Φ) предиктора для величины Y по X сводится к решению экстремальной проблемы (1)–(2). Как будет ясно из дальнейшего, основным аналитическим инструментом решения этой проблемы является аппарат условных распределений, поэтому мы предварительно напомним некоторые факты и формулы из теории вероятностей, которые будут использоваться далее.

2. Условное математическое ожидание

Напомним, что мы ограничиваемся рассмотрением лишь абсолютно непрерывных либо дискретных распределений $\mathcal{L}(X, Y)$ и используем общее обозначение $f_{XY}(x, y)$ для совместной плотности — в первом случае и совместной вероятности $\mathbf{P}\{X = x,\ Y = y\}$ — во втором, называя в любом случае функцию $f_{XY}(x, y)$ плотностью. По совместному распределению $\mathcal{L}(X, Y)$ можно определить условное распределение $\mathcal{L}(Y|X = x)$ величины Y при заданном значении X. Так, в абсолютно непрерывном случае это распределение задается условной плотностью

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)},$$

где

$$f_X(x) = \int f_{XY}(x, y)\, dy$$

— маргинальная плотность X (для дискретных распределений последний интеграл заменяется соответствующей суммой). В предположении существования соответствующих моментов мы можем вычислить среднее и дисперсию этого условного распределения: условное математическое ожидание

$$M(x) = \mathbf{E}(Y|X = x), \tag{3}$$

называемое *функцией регрессии* Y на X, и условную дисперсию

Функция регрессии

$$\mathbf{D}(Y|X = x) = \mathbf{E}\big[(Y - M(x))^2 \mid X = x\big]. \tag{4}$$

Так, для абсолютно непрерывных распределений

$$M(x) = \int y f_{Y|X}(y|x)\,dy = \int y f_{XY}(x,y)\,dy \frac{1}{f_X(x)},$$

$$\mathbf{D}(Y|X=x) = \int (y - M(x))^2 f_{Y|X}(y|x)\,dy;$$

в случае же дискретных распределений интегралы в этих формулах заменяются соответствующими суммами.

В соответствии с (3) и (4) можно рассматривать случайные величины $M(X)$ и $\mathbf{D}(Y|X)$, которые при $X = x$ принимают значения соответственно $M(x)$ и $\mathbf{D}(Y|X=x)$. Обозначим

$$\sigma_Y^2 = \mathbf{D}Y, \quad \sigma_M^2 = \mathbf{D}M(X), \quad \sigma_{YX}^2 = \mathbf{E}\mathbf{D}(Y|X). \tag{5}$$

Тогда имеют место следующие полезные формулы:

$$\mathbf{E}Y = \mathbf{E}[\mathbf{E}(Y|X)] = \mathbf{E}M(X), \tag{6}$$

$$\mathbf{D}Y = \mathbf{D}\mathbf{E}(Y|X) + \mathbf{E}\mathbf{D}(Y|X), \quad \text{или} \quad \sigma_Y^2 = \sigma_M^2 + \sigma_{YX}^2. \tag{7}$$

Соотношение (6) называют *формулой полного математического ожидания*, а (7) — *формулой разложения дисперсии*.

Формула (6) проверяется непосредственно: например, для абсолютно непрерывных распределений из определения $M(X)$ имеем

$$\mathbf{E}M(X) = \int M(x) f_X(x)\,dx = \int \left(\int y f_{Y|X}(y|x)\,dy \right) f_X(x)\,dx =$$

$$\int y \left(\int f_{XY}(x,y)\,dx \right) dy = \int y f_Y(y)\,dy = \mathbf{E}Y$$

Формула же (7) получается с использованием определения (4) и формулы полного математического ожидания (6) из цепочки равенств

$$\begin{aligned}
\mathbf{D}Y &= \mathbf{E}(Y - \mathbf{E}Y)^2 = \mathbf{E}\big[(Y - M(X)) + (M(X) - \mathbf{E}M(X))\big]^2 = \\
&= \mathbf{E}\Big\{ \mathbf{E}\big[(Y - M(X))^2 \mid X\big] \Big\} + \mathbf{D}M(X) + \\
&\quad + 2\mathbf{E}\Big\{ \mathbf{E}\big[(Y - M(X))(M(X) - \mathbf{E}M(X)) \mid X\big] \Big\} = \\
&= \mathbf{E}\mathbf{D}(Y|X) + \mathbf{D}M(X),
\end{aligned}$$

поскольку

$$\begin{aligned}
&\mathbf{E}\big[(Y - M(X))(M(X) - \mathbf{E}M(X)) \mid X\big] = \\
&= (M(X) - \mathbf{E}M(X))\mathbf{E}\big[(Y - M(X)) \mid X\big] = 0.
\end{aligned}$$

Отметим важное следствие формулы (7):

$$\sigma_M^2 = \mathbf{D}\mathbf{E}(Y|X) \leqslant \sigma_Y^2 = \mathbf{D}Y, \tag{8}$$

причем равенство в (8) выполняется лишь в случае, когда

$$\sigma^2_{YX} = \mathbf{E}\mathbf{D}(Y|X) = \mathbf{E}\Big\{\mathbf{E}\big[(Y - M(X))^2 \mid X\big]\Big\} = \mathbf{E}(Y - M(X))^2 = 0,$$

т. е. когда $Y = M(X)$ (Y является функцией X).

Примеры вычисления условных средних и дисперсий в многомерной нормальной модели указаны в упр. 65 и 66 к гл. 1; подчеркнем, что в этих примерах функции регрессии $M(x)$ являются линейными по x, а условные дисперсии от x не зависят. Приведем еще один иллюстративный пример вычисления условных среднего и дисперсии.

Пример 1. Пусть $\mathcal{U}_1$ и $\mathcal{U}_2$ — независимые показательные с параметром масштаба $a > 0$ случайные величины (см. п. 3 § 1.2). Положим

$$X = \mathcal{U}_2, \quad Y = \mathcal{U}_1 + \mathcal{U}_1\mathcal{U}_2.$$

Тогда

$$M(x) = \mathbf{E}(Y|X = x) = \mathbf{E}[(1 + x)\mathcal{U}_1] = a(1 + x),$$
$$\mathbf{D}(Y|X = x) = \mathbf{D}[(1 + x)\mathcal{U}_1] = a^2(1 + x)^2$$

Следовательно, $M(X) = a(1 + \mathcal{U}_2)$, $\mathbf{D}(Y|X) = a^2(1 + \mathcal{U}_2)^2$ и потому

$$\sigma^2_M = \mathbf{D}M(X) = a\mathbf{D}\,\mathcal{U}_2 = a^4,$$
$$\sigma^2_{YX} = \mathbf{E}\mathbf{D}(Y|X) = a^2\mathbf{E}(1 + \mathcal{U}_2)^2 = a^2(1 + 2a + 2a^2). \quad \bullet$$

3. Оптимальный предиктор и его свойства

Решение экстремальной проблемы (1)–(2) существенно зависит от задания класса Φ, в котором ищется прогноз. Если класс Φ есть класс всех функций $\varphi(X)$, то ответ на вопрос о существовании и виде оптимального предиктора (прогноза) $\varphi^*(X)$ дает следующее утверждение.

Теорема 1. Оптимальный предиктор φ^* в классе всех функций $\Phi = \{\varphi\}$ существует и совпадает с функцией регрессии Y на X, т. е.

$$\varphi^*(X) = M(X) = \mathbf{E}(Y|X); \tag{9}$$

этот предиктор имеет максимальную корреляцию с Y среди всех предикторов.

Доказательство. По формуле полного математического ожидания для любой функции $\varphi \in \Phi$

$$\mathbf{E}\big[(Y - M(X))\big(M(X) - \varphi(X)\big)\big] =$$
$$= \mathbf{E}\Big\{\mathbf{E}\big[(Y - M(X))\big(M(X) - \varphi(X)\big) \mid X\big]\Big\} =$$
$$= \mathbf{E}\Big\{\big(M(X) - \varphi(X)\big)\mathbf{E}\big[(Y - M(X)) \mid X\big]\Big\} =$$
$$= \mathbf{E}\Big\{\big(M(X) - \varphi(X)\big)\big(M(X) - M(X)\big)\Big\} = 0,$$

поэтому для с. к. о. предиктора φ можно записать следующую цепочку соотношений:

$$\Delta(\varphi) = \mathbf{E}(\varphi(X) - Y)^2 = \mathbf{E}\big[(\varphi(X) - M(X)) + (M(X) - Y)\big]^2 =$$
$$= \mathbf{E}\big(\varphi(X) - M(X)\big)^2 + \mathbf{E}(M(X) - Y)^2 \geqslant \mathbf{E}(M(X) - Y)^2 = \Delta(M).$$

Знак равенства достигается здесь лишь при $\varphi = M$, следовательно, $\varphi^* = M$.

Отсюда также следует, что минимальная ошибка предсказания Δ^* в (2) может быть записана в виде (см. (4)–(5))

$$\Delta^* = \Delta(M) = \mathbf{E}(Y - M(X))^2 = \mathbf{E}\Big\{\mathbf{E}\big[(Y - M(X))^2 \mid X\big]\Big\} = \mathbf{E}\mathbf{D}(Y \mid X) = \sigma^2_{Y|X}. \tag{10}$$

Переходя к доказательству второй части теоремы, прежде всего заметим, что для произвольного предиктора φ с учетом (6) имеем

$$\operatorname{cov}(\varphi(X), Y) = \mathbf{E}\big(\varphi(X) - \mathbf{E}\varphi(X)\big)(Y - \mathbf{E}Y) =$$
$$= \mathbf{E}\Big\{\mathbf{E}\big[(\varphi(X) - \mathbf{E}\varphi(X))(Y - \mathbf{E}Y) \mid X\big]\Big\} =$$
$$= \mathbf{E}\Big\{\big(\varphi(X) - \mathbf{E}\varphi(X)\big)\mathbf{E}\big[(Y - \mathbf{E}Y) \mid X\big]\Big\} =$$
$$= \mathbf{E}\big(\varphi(X) - \mathbf{E}\varphi(X)\big)\big(M(X) - \mathbf{E}M(X)\big) = \operatorname{cov}\big(\varphi(X), M(X)\big).$$

В дальнейшем для сокращения записи аргумент X будем опускать и для любых случайных величин ξ, ξ_1, ξ_2 будем использовать стандартные обозначения

$$\sigma^2_\xi = \mathbf{D}\xi, \quad \rho(\xi_1, \xi_2) = \operatorname{corr}(\xi_1, \xi_2) = \frac{\operatorname{cov}(\xi_1, \xi_2)}{\sigma_{\xi_1}\sigma_{\xi_2}}.$$

При $\varphi = M$ полученное выше равенство дает

$$\operatorname{cov}(M, Y) = \operatorname{cov}(M, M) = \sigma^2_M \geqslant 0,$$

откуда

$$\rho(M, Y) = \frac{\sigma_M}{\sigma_Y}$$

Используя это и то, что $|\rho(\xi_1, \xi_2)| \leqslant 1$, имеем

$$\rho^2(\varphi, Y) = \frac{\operatorname{cov}^2(\varphi, Y)}{\sigma^2_\varphi \sigma^2_Y} = \frac{\operatorname{cov}^2(\varphi, M)}{\sigma^2_\varphi \sigma^2_M}\frac{\sigma^2_M}{\sigma^2_Y} = \rho^2(\varphi, M)\rho^2(M, Y) \leqslant \rho^2(M, Y).$$

Таким образом, $\rho(M, Y) \geqslant |\rho(\varphi, Y)|$ для любого предиктора φ. ■

Корреляционное отношение

Квадрат максимального значения коэффициента корреляции $\rho^2(M, Y)$ имеет специальное обозначение η^2_{YX} и называется *корреляционным отношением*. Из предыдущего и (7), (10) для него можно записать следующие представления:

$$\eta^2_{YX} \equiv \rho(M, Y) = \sup_{\varphi \in \Phi} \rho^2(\varphi, Y) = \frac{\sigma^2_M}{\sigma^2_Y} = 1 - \frac{\sigma^2_{Y|X}}{\sigma^2_Y} = 1 - \frac{\Delta^*}{\sigma^2_Y}. \tag{11}$$

Отсюда следует, что $\eta^2_{YX} \to 1$, если ошибка прогноза $\Delta^* \to 0$, и $\eta^2_{YX} = 0$, если учет X не уменьшает ошибки прогноза ($\Delta^* = \sigma^2_{YX} = \sigma^2_Y$). Таким образом, корреляционное отношение η^2_{YX} представляет собой меру зависимости между Y и X (или меру точности прогноза, так как $\Delta^* = \sigma^2_Y(1 - \eta^2_{YX})$), и по этой мере можно сравнивать различные совокупности прогнозных переменных в конкретных задачах: значения этой меры принадлежат интервалу $[0, 1]$, чем ближе η^2_{YX} к 1, тем точнее прогноз Y по данному X, а $\eta^2_{YX} = 1$ тогда и только тогда, когда $Y = M(X)$, т. е. когда Y функционально связано с X.

Пример 2 (Прогноз в нормальной модели). Пусть совместное распределение пары (X, Y) есть нормальное распределение с $\mathbf{E}X = \mu_X$, $\mathbf{E}Y = \mu_Y$, $\mathbf{D}X = \sigma^2_X$, $\mathbf{D}Y = \sigma^2_Y$, $\operatorname{corr}(X, Y) = \rho$, $|\rho| < 1$. Тогда (см. упр. 65 к гл. 1) прогноз Y по X имеет вид

$$M(X) = \mathbf{E}(Y|X) = \mu_Y + \frac{\sigma_Y}{\sigma_X}\rho(X - \mu_X). \tag{12}$$

Его с. к. о. $\Delta^* = \sigma^2_{YX} = (1 - \rho^2)\sigma^2_Y < \sigma^2_Y$ (см. (10)), а корреляционное отношение $\eta^2_{YX} = 1 - \Delta^*/\sigma^2_Y = \rho^2$ (см. (11)), т. е. равно квадрату коэффициента корреляции X и Y. •

4. Из истории регрессии

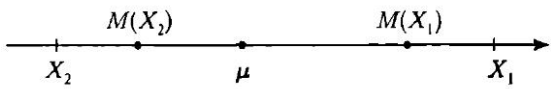
Термин «регрессия» был предложен Фрэнсисом Гальтоном [3] и основан на следующем наблюдении. Гальтон измерял рост отца (X) и сына (Y), выбранных случайно из большой популяции людей. Если предположить, что

$$\mathcal{L}(X, Y) = \mathcal{N}\left((\mu, \mu), \begin{Vmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{Vmatrix}\right) \qquad \rho > 0$$

(здесь μ — среднее роста мужчин по популяции, σ^2 — дисперсия роста по популяции), то в соответствии с (12), прогноз роста сына у отца, чей рост X, есть

$$M(X) = \mu + \rho(X - \mu) = (1 - \rho)\mu + \rho X \tag{13}$$

и совпадает с $\mathbf{E}(Y|X)$ — средним ростом сыновей, чьи отцы имеют рост X. Из (13) следует, что прогноз $M(X)$ ближе к среднему по популяции μ, чем рост отца X: если рост отца $X_1 > \mu$, то $M(X_1) < X_1$ — у высоких отцов сыновья будут, как правило, меньшего роста, т. е. наблюдается «регрессия к среднему». Но эта тенденция компенсируется «прогрессией к среднему» среди сыновей отцов небольшого роста: если $X_2 < \mu$, то $M(X_2) > X_2$ (см. диаграмму), и тем самым парадокса не получается — популяционное среднее остается постоянным.



[3] Гальтон Фрэнсис (1822–1911) — английский психолог и биолог, один из создателей евгеники.

При этом дисперсия предсказанного значения

$$\mathbf{D}M(X) = \mathbf{D}(\rho X) = \rho^2\sigma^2$$

меньше дисперсии σ^2 реальных отцов.

Здесь уместно сделать следующее замечание, имеющее общий характер. На практике (и у Гальтона) точное распределение $L(X, Y)$ обычно не известно и линия регрессии $M(X)$ строится приближенно по оценкам $\widehat{\mu}$ и $\widehat{\rho}$ на основе большой выборки (X_i, Y_i), $i = 1, \ldots, n$, из генеральной совокупности, и в итоге используют *эмпирический прогноз*

Эмпирический прогноз

$$\widehat{M}(X) = (1 - \widehat{\rho})\widehat{\mu} + \widehat{\rho}X$$

(детальнее об этом речь будет идти позже).

Итак, если коэффициент корреляции описывает зависимость между двумя случайными величинами одним числом, то регрессия выражает эту зависимость в виде функционального соотношения и поэтому дает более полную информацию. Вначале регрессионный анализ применялся в биологии; между 1920 и 1930 гг. он начал интенсивно использоваться в экономике, что привело к возникновению новой области науки — эконометрики; в настоящее время регрессионный анализ интенсивно используется практически во всех областях науки, так как анализ величины Y, определяемой по другой величине X, когда Y измерить трудно, а X достаточно легко, весьма важен.

5. Линейный прогноз

Итак, при минимальных предположениях о совместном распределении $\mathcal{L}(X, Y)$ величин X и Y (требуется лишь существование вторых моментов) мы имеем универсальное решение для прогнозирования Y по наблюдению над X, даваемое функцией регрессии (9). Однако здесь имеются две трудности:

1) для вычисления прогноза $M(X) = \mathbf{E}(Y|X)$ надо абсолютно точно знать закон распределения $\mathcal{L}(X, Y)$ и
2) явное вычисление $M(X)$ может оказаться трудной аналитической задачей и прогноз может оказаться слишком сложной функцией, мало пригодной для конкретных расчетов.

Однако обеих этих трудностей можно избежать, если сузить круг поисков наилучшего предиктора до класс L простейших, а именно, линейных предикторов $\varphi(\underline{X}) = \beta_0 + \underline{\beta}'\underline{X}$ (здесь мы считаем, что $\underline{X} = (X_1, \ldots, X_p)$ и $\underline{\beta} = (\beta_1, \ldots, \beta_p)$). Для этого случая решение экстремальной проблемы (1) дается следующим утверждением.

Теорема 2. Пусть $L = \{\varphi \quad \varphi(\underline{X}) = \beta_0 + \underline{\beta}'\underline{X}\}$ есть класс всех линейных предикторов Y по $\underline{X}$. Тогда оптимальный линейный предиктор (линейный

прогноз)

$$\varphi_L^* = \arg\min_{\varphi \in L} \Delta(\varphi)$$

Линейный прогноз дается соотношением

$$\varphi_L^*(\underline{X}) = \mathbf{E}Y + \underline{a}'\Sigma^{-1}(\underline{X} - \mathbf{E}\underline{X}) = \beta_0^* + \underline{\beta}^{*\prime}\underline{X}, \tag{14}$$

т. е.

$$\beta_0^* = \mathbf{E}Y - \underline{a}'\Sigma^{-1}\mathbf{E}\underline{X} = \mathbf{E}Y - \underline{\beta}^{*\prime}\mathbf{E}\underline{X}, \quad \underline{\beta}^* = \Sigma^{-1}\underline{a},$$

где $\underline{a} = \operatorname{cov}(Y, \underline{X}) = \big(\operatorname{cov}(Y, X_1), \quad , \operatorname{cov}(Y, X_p)\big)$, $\Sigma = \mathbf{D}(\underline{X})$ и предполагается, что $|\Sigma| \neq 0$. Этот предиктор имеет максимальную корреляцию с Y среди всех линейных предикторов.

Доказательство. Пусть φ — произвольный линейный предиктор. Тогда его с. к. о. $\Delta(\varphi)$ может быть записана в виде

$$\Delta(\varphi) = \mathbf{E}(Y - \beta_0 - \underline{\beta}'\underline{X})^2 = \mathbf{E}\big[(Y - \mathbf{E}Y) - b - \underline{\beta}'(\underline{X} - \mathbf{E}\underline{X})\big]^2 =$$
$$= \mathbf{D}Y + b^2 + \underline{\beta}'\Sigma\underline{\beta} - 2\underline{\beta}'\underline{a},$$

где обозначено $b = \beta_0 - \mathbf{E}Y + \underline{\beta}'\mathbf{E}\underline{X}$ (проверяется непосредственно, что мы оставляем читателю в качестве простого упражнения).

Пусть теперь $\underline{\beta} = \underline{\beta}^* + \underline{\delta}$, где $\underline{\beta}^*$ указано в (14) и $\underline{\delta}$ — произвольный вектор. Тогда (также проверяется непосредственно)

$$b^2 + \underline{\beta}'\Sigma\underline{\beta} - 2\underline{\beta}'\underline{a} = b^2 + \underline{\delta}'\Sigma\underline{\delta} - \underline{\beta}^{*\prime}\underline{a} \geqslant -\underline{\beta}^{*\prime}\underline{a}$$

и равенство здесь достигается лишь при $b = 0$, $\underline{\delta} = \underline{0}$ (поскольку невырожденная дисперсионная матрица является положительно определенной). Таким образом,

$$\Delta(\varphi) \geqslant \mathbf{D}Y - \underline{\beta}^{*\prime}\underline{a} = \sigma_Y^2 - \underline{a}'\Sigma^{-1}\underline{a} = \Delta(\varphi_L^*), \tag{15}$$

т. е. предиктор φ_L^* (14) обладает в классе L минимальной с. к. о. Первая часть теоремы доказана.

Далее, из определения векторов $\underline{a}$ и $\underline{\beta}^*$ следуют равенства

$$\begin{aligned} \operatorname{cov}(Y, \underline{\beta}'\underline{X}) &= \underline{\beta}'\underline{a} = \underline{\beta}'\Sigma\underline{\beta}^* \\ \operatorname{cov}(Y, \underline{\beta}^{*\prime}\underline{X}) &= \underline{\beta}^{*\prime}\Sigma\underline{\beta}^* = \mathbf{D}(\underline{\beta}^{*\prime}\underline{X}) \geqslant 0. \end{aligned} \tag{16}$$

Из последнего равенства также имеем

$$\sigma_Y^2\rho^2(Y, \underline{\beta}^{*\prime}\underline{X}) = \frac{\operatorname{cov}(Y, \underline{\beta}^{*\prime}\underline{X})}{\mathbf{D}(\underline{\beta}^{*\prime}\underline{X})} = \underline{\beta}^{*\prime}\Sigma\underline{\beta}^* \tag{17}$$

Воспользуемся неравенством Коши—Буняковского (для любых векторов $\underline{a}$ и $\underline{b}$ $(\underline{a}'\underline{b})^2 \leqslant (\underline{a}'\underline{a})(\underline{b}'\underline{b})$) в виде

$$(\underline{\beta}'\Sigma\underline{\beta}^*)^2 = (\underline{\beta}'\Sigma^{1/2}\Sigma^{1/2}\underline{\beta}^*)^2 \leqslant (\underline{\beta}'\Sigma\underline{\beta})(\underline{\beta}^{*\prime}\Sigma\underline{\beta}^*),$$

тогда из (16) и (17) получим

$$\sigma_Y^2\rho^2(Y,\underline{\beta}'\underline{X}) = \frac{\operatorname{cov}^2(Y,\underline{\beta}'\underline{X})}{\mathbf{D}(\underline{\beta}'\underline{X})} = \frac{(\underline{\beta}'\Sigma\underline{\beta}^*)^2}{\underline{\beta}'\Sigma\underline{\beta}} \leqslant \underline{\beta}^{*\prime}\Sigma\underline{\beta}^* = \sigma_Y^2\rho^2(Y,\underline{\beta}^{*\prime}\underline{X}).$$

Отсюда следует, что для любого предиктора $\varphi \in L$

$$\rho(Y,\underline{\beta}^{*\prime}\underline{X}) = \rho(Y,\varphi_L^*(\underline{X})) \geqslant \left|\rho(Y,\varphi(\underline{X}))\right|. \blacksquare$$

Множественный коэффициент корреляции

Из этих рассуждений (см. формулу (15)) следует, что в теории линейного прогнозирования особую роль играет величина

$$\rho^2_{Y\underline{X}} = \frac{\underline{a}'\Sigma^{-1}\underline{a}}{\sigma_Y^2}, \tag{18}$$

которая имеет специальное название *множественный коэффициент корреляции*. Эта величина зависит только от вторых моментов совокупности переменных $(\underline{X}, Y)$ и представляет собой обобщение квадрата обычного коэффициента корреляции на случай многомерного $\underline{X}$ (если $\underline{X}$ есть скалярная величина X, т. е. $p = 1$, то $\rho^2_{YX} = \rho^2(X,Y)$). Таким образом, $\rho^2_{Y\underline{X}}$ — это мера корреляционной зависимости Y от вектора $\underline{X}$. Через эту меру с. к. о. линейного прогноза φ_L^* можно записать в виде (см. (15))

$$\Delta_L^* = \Delta(\varphi_L^*) = \sigma_Y^2(1 - \rho^2_{Y\underline{X}}), \tag{19}$$

откуда следует, что чем ближе $\rho^2_{Y\underline{X}}$ к 1, тем точнее прогноз φ_L^*. Множественный коэффициент корреляции играет в теории линейного прогнозирования такую же роль, как корреляционное отношение η^2_{YX} (см. (11)) в общей теории прогнозирования. Для $\rho^2_{Y\underline{X}}$ можно записать аналогичную (11) цепочку соотношений:

$$\rho^2_{Y\underline{X}} = \frac{\underline{\beta}^{*\prime}\Sigma\underline{\beta}^*}{\sigma_Y^2} = \rho^2(Y,\varphi_L^*) = \sup_{\varphi\in L}\rho^2(Y,\varphi) = 1 - \frac{\Delta_L^*}{\sigma_Y^2}, \tag{20}$$

откуда, в частности, видно, что $\rho^2_{Y\underline{X}}$ *равен квадрату обычного коэффициента корреляции между* Y *и оптимальным линейным предиктором для* Y

Поскольку $L \subset \Phi$, то из (11) и (20) следует, что $\eta^2_{Y\underline{X}} \geqslant \rho^2_{Y\underline{X}}$, а в тех случаях, когда функция регрессии $M(\underline{X})$ линейна (тогда она является линейным прогнозом), ошибки Δ^* и Δ_L^* совпадают, следовательно, $\eta^2_{Y\underline{X}} \quad \rho^2_{Y\underline{X}}$. Поэтому в общем случае разность $\eta^2_{Y\underline{X}} - \rho^2_{Y\underline{X}}$ может служить мерой отклонения регрессии от линейности.

Итак, в условиях теоремы 2 линейный прогноз $\varphi_L^*(\underline{X})$ для Y также всегда может быть построен, при этом для его вычисления достаточно знать лишь первые и вторые моменты распределения $\mathcal{L}(\underline{X}, Y)$, т. е. минимальную информацию об этом распределении. На практике обычно используют линейные прогнозы, а необходимые для этого моменты оцениваются по результатам

прошлых наблюдений над $\underline{X}$ и Y. Заменив в (14) точные значения моментов (когда они неизвестны, — а так обычно и бывает) их оценками, как это описано в п. 5 § 2.3, мы получаем *эмпирический предиктор* $\widehat{\varphi}_L^*(\underline{X})$ (см. (16) § 2.3), который и используется для прогнозирования будущих значений Y. Также по прошлым наблюдениям оценивается множественный коэффициент корреляции $\rho^2_{Y\underline{X}}$ (см. формулу (17) § 2.3), что дает возможность оценить с. к. о. (19), т. е. точность прогноза.

Примерами линейного прогноза являются прогноз в нормальной модели (см. пример 2, соотношение (12), и упр. 25), а также прогнозы, указанные в примере 1 и упр. 23, 24: во всех этих случаях функция регрессии $M(X)$ линейна, следовательно, она, как абсолютно оптимальный предиктор (в соответствии с теоремой 1), и является линейным прогнозом. То, что функция регрессии не всегда оказывается линейной, демонстрирует следующий пример.

Пример 3. Пусть модель $\mathcal{L}(X, Y)$ задается следующим образом:

$$\mathcal{L}(Y) = \mathrm{Be}(a, b)$$

— бета-распределение (см. п. 4 § 1.2), а

$$\mathcal{L}(X|Y = y) = \overline{\mathrm{Bi}}(r, y)$$

— отрицательное биномиальное распределение (см. п. 2 § 1.1). Тогда по формуле Байеса (см. соотношение (3) в упр. 59 к гл. 1) условная плотность $f_{Y|X}(y|x)$ есть

$$f_{Y|X}(y|x) = \frac{f_Y(y) f_{X|Y}(x|y)}{f_X(x)} \cong y^{a-1}(1-y)^{b-1} y^x (1-y)^r = y^{a+x-1}(1-y)^{b+r-1},$$

т. е. $\mathcal{L}(Y|X = x) = \mathrm{Be}(a + x, b + r)$ и потому (см. (24) § 1.2)

$$M(X) = \mathbf{E}(Y|X) = \frac{a + X}{a + b + r + X}.$$

Это есть прогноз Y по X в классе всех функций $\varphi(x)$, и он не является линейным. •

6. Использование в прогнозе дополнительных переменных. Алгоритм обновления прогноза

При практической реализации теории линейного прогнозирования возникают следующие важные вопросы: 1) как из множества всех факторов $\{X_i\}$, влияющих на прогнозируемую величину Y, выделить наиболее существенные, или, другими словами, как упорядочить прогнозные переменные по степени их важности в прогнозе Y? 2) если решен 1-й вопрос, следовательно, мы имеем упорядоченный ряд факторов $X_1 \succ X_2 \succ \quad \succ X_{p-1} \succ X_p \succ$ то сколькими факторами ограничиться в построении предиктора φ_L^*? 3) наконец, как рационально действовать при расчете прогноза, чтобы избежать чрезмерных затрат?

> *Хорошо поставить вопрос — значит уже наполовину решить его.*
>
> **Д. И. Менделеев**

Чтобы ответить на эти вопросы, исследуем предварительно, как изменяется точность прогноза при увеличении числа прогнозных переменных. Условимся далее кодировать переменные X_i их индексом «i», $i = 0, 1, 2,$ при этом $X_0 = Y$, тогда $\sigma_Y^2 = \sigma_0^2$, $\rho_{Y\underline{X}}^2 = \rho_{0(1\cdots p)}^2$, а для с. к. о. Δ_L^* в (19) будем использовать обозначение $\sigma_{0(1\cdots p)}^2$. В этих обозначениях формула (19) примет вид

$$\sigma_{0(1\cdots p)}^2 = \sigma_0^2(1 - \rho_{0(1\cdots p)}^2).$$

Пусть по факторам $\underline{X}$ $(X_1, \ldots, X_{p-1})$ построен линейный прогноз $\varphi_L^*(\underline{X}) \equiv Y_{p-1}^*$ (см. (14), где $\underline{X}$ есть вектор размерности $p-1$); его с. к. о. есть $\sigma_{0(1\cdots p-1)}^2 = \sigma_0^2(1 - \rho_{0(1\cdots p-1)}^2)$. Если добавить в прогноз новое переменное X_p, то с. к. о. нового предиктора $Y_p^* = \varphi_L^*(X_1, \ldots, X_p)$ станет равной $\sigma_{0(1\cdots p)}^2 = \sigma_0^2(1 - \rho_{0(1\cdots p)}^2)$. Следовательно, относительное уменьшение ошибки прогноза при добавлении p-го переменного будет равно

$$\frac{\sigma_{0(1\cdots p-1)}^2 - \sigma_{0(1\cdots p)}^2}{\sigma_{0(1\cdots p-1)}^2} = \frac{\rho_{0(1\cdots p)}^2 - \rho_{0(1\cdots p-1)}^2}{1 - \rho_{0(1\cdots p-1)}^2} \equiv \rho_{0p(1\cdots p-1)}^2. \qquad (21)$$

Частная корреляция

Определенную в (21) величину $\rho_{0p(1\cdots p-1)}$ называют *часткой корреляцией между Y и X_p, исключающей влияние «промежуточных» перемекных $X_1, \ldots, X_{p-1}$* Из (21) следует, что

$$\sigma_{0(1\cdots p)}^2 = \sigma_{0(1\cdots p-1)}^2(1 - \rho_{0p(1\cdots p-1)}^2), \qquad (22)$$

и если частная корреляция $\rho_{0p(1\cdots p-1)}$ мала, то учет X_p мало что добавляет в прогноз. При $p =$ равенство (22) принимает вид $\sigma_{0(1)}^2 = \sigma_0^2(1 - \rho_{01}^2)$, т. е. наибольшую точность простейший предиктор Y_1^* имеет, если фактор X_1 имеет наибольшую корреляцию с Y.

Отсюда вытекает следующий принцип упорядочения факторов по степени их важности в прогнозе: *в качестве первого фактора X_1 выбирается фактор с наибольшей корреляцией с Y; в качестве следующего по важности фактора X_2 выбирается фактор с наибольшей частной корреляцией с Y, исключающей влияние X_1*, и т. д., т. е. частная корреляция $\rho_{0p(1\cdots p-1)}$ является инструментом упорядочения предсказывающих переменных по степени их важности в прогнозировании Y.

Отметим некоторые свойства частной корреляции. Обозначим

$$\rho_{ij} = \rho(X_i, X_j), \quad i, j = 0, 1, \ldots, p, \quad \mathbf{P} = \|\rho_{ij}\|_0^p \quad \text{и} \quad \mathbf{P}^{-1} = \|\rho^{ij}\|_0^p.$$

Тогда

$$\rho_{0p(1\cdots p-1)} = \frac{\rho^{0p}}{\sqrt{\rho^{00}\rho^{pp}}}\bigg|_{p=2} = \frac{\rho_{02} - \rho_{01}\rho_{12}}{\sqrt{(1-\rho_{01}^2)(1-\rho_{12}^2)}}. \tag{23}$$

Далее, построим линейный прогноз для X_p по переменным $\underline{X} = (X_1, \ldots X_{p-1})$:

$$X_p^* = \alpha_0^* + \underline{\alpha}^{*\prime}\underline{X}, \quad \underline{\alpha}^* = \Sigma^{-1}\underline{b}, \quad \underline{b} = \operatorname{cov}(X_p, \underline{X}),$$
$$\alpha_0^* = \mathbf{E}X_p - \underline{\alpha}^{*\prime}\mathbf{E}\underline{X}, \quad \Sigma = \mathbf{D}(\underline{X}),$$

и рассмотрим *остатки* $e_1 = Y - Y_{p-1}^*$ и $e_2 = X_p - X_p^*$. Тогда можно показать, что

$$\rho_{0p(1\cdots p-1)} = \rho(e_1, e_2), \tag{24}$$

т. е. *частная корреляция — это обычный коэффициент корреляции между остатками, получающимися вычитанием из Y и X_p их оптимальных линейных предикторов, основанных на $X_1, \quad , X_{p-1}$* (так исключается влияние общих факторов $X_1, \quad , X_{p-1}$ на переменные Y и X_p).

Остатки e, являющиеся ошибками соответствующих прогнозов, обладают следующими свойствами:

1) они имеют нулевое среднее, так как

$$\mathbf{E}e_1 = \mathbf{E}Y - \mathbf{E}Y_{p-1}^* = \mathbf{E}Y - \beta_0^* - \underline{\beta}^{*\prime}\mathbf{E}\underline{X} = 0 \quad \text{(см. (14))};$$

2) следовательно, $\mathbf{D}e_1 = \mathbf{E}(Y - Y_{p-1}^*)^2 = \Delta^* = \sigma^2_{0(1\cdots p-1)}$;

3) $\operatorname{cov}(\underline{\beta}'\underline{X}, e_1) = \operatorname{cov}(\underline{\beta}'\underline{X}, Y - \underline{\beta}^{*\prime}\underline{X}) = \underline{\beta}'\underline{a} - \underline{\beta}'\Sigma\underline{\beta}^* = \underline{\beta}'\underline{a} - \underline{\beta}'\underline{a} = 0$, т. е. они некоррелированы с прогнозными переменными. Это свойство можно проинтерпретировать также так: *оптимальный линейный предиктор концентрирует в себе всю корреляцию Y с прогнозными переменными.*

Алгоритм обновления прогноза

На изложенном выше основан следующий эффективный алгоритм обновления прогноза для Y при добавлении в прогноз дополнительного переменного. Пусть построен прогноз (см. (14))

$$Y_{p-1}^* \equiv \varphi_L^*(X_1, \ldots, X_{p-1}) = \beta_0^* + \underline{\beta}^{*\prime}\underline{X}, \quad \underline{X} = (X_1, \quad , X_{p-1}),$$

и будем искать новый предиктор Y_p^* при добавлении в прогноз новой переменной (фактора) X_p в виде линейной комбинации $\beta_0 + \underline{\beta}'\underline{X} + \beta_p e_2$. Тогда для определения коэффициентов надо минимизировать с. к. о.

$$\mathbf{E}((Y - \beta_0 - \underline{\beta}'\underline{X}) - \beta_p e_2)^2 =$$
$$= \mathbf{E}(Y - \beta_0 - \underline{\beta}'\underline{X})^2 + \beta_p^2\sigma_{e_2}^2 - 2\beta_p \operatorname{cov}(e_1, e_2) =$$
$$= \mathbf{E}(Y - \beta_0 - \underline{\beta}'\underline{X})^2 + \left(\beta_p - \frac{\operatorname{cov}(e_1, e_2)}{\sigma_{e_2}^2}\right)^2\sigma_{e_2}^2 - \frac{\operatorname{cov}^2(e_1, e_2)}{\sigma_{e_2}^2}$$

(поскольку

$$\operatorname{cov}(Y - \beta_0 - \underline{\beta}'\underline{X}, e_2) = \operatorname{cov}(e_1, e_2) + \operatorname{cov}(Y^*_{p-1} - \underline{\beta}'\underline{X}, e_2) = \operatorname{cov}(e_1, e_2)).$$

Отсюда следует, что оптимальные значения для β_0 и $\underline{\beta} = (\beta_1, \ldots, \beta_{p-1})$ — прежние (см. (14)), а оптимальный выбор β_p таков (см. (24)):

$$\beta^*_p = \frac{\operatorname{cov}(e_1, e_2)}{\sigma^2_{e_2}} = \rho(e_1, e_2)\frac{\sigma_{e_1}}{\sigma_{e_2}} = \rho_{0p(1\cdot\ \cdot p-1)}\frac{\sigma_{e_1}}{\sigma_{e_2}} \tag{25}$$

Таким образом, новый прогноз может быть вычислен по формуле

$$Y^*_p = Y^*_{p-1} + \beta^*_p e_2 = Y^*_{p-1} + \beta^*_p(X_p - X^*_p) =$$

$$= \beta^*_0 - \alpha^*_0\beta^*_p + \sum_{i=1}^{p-1}(\beta^*_i - \alpha^*_i\beta^*_p)X_i + \beta^*_p X_p, \tag{26}$$

т. е. предыдущие коэффициенты (перед переменными $X_1, \ldots, X_{p-1}$) пересчитываются, а коэффициент при новой переменной X_p равен β^*_p (см. (25)). При этом ошибка нового прогноза вычисляется по формуле (22).

Фактически это есть реккурентный алгоритм построения прогноза: мы начинаем с простейшего предиктора.

$$Y^*_1 = \beta^*_0 + \beta^*_1 X_1, \quad \beta^*_1 = \rho_{01}\frac{\sigma_0}{\sigma_1}, \quad \beta^*_0 = \mathbf{E}Y - \beta^*_1\mathbf{E}X_1, \tag{27}$$

его с. к. о. $\sigma^2_{e_1} = \sigma^2_{0(1)} = \sigma^2_0(1 - \rho^2_{01})$, и последовательно добавляем новые переменные $X_2, X_3, \ldots$ (по степени убывания их значимости) с пересчетом по формуле (26) на каждом шагу коэффициентов предыдущей линейной комбинации, пока не достигнем желаемой точности прогноза, что контролируется формулой (22) (см. также пример 1 в п. 5 § 2.3).

7. Прогнозирование стационарных временных рядов

Наиболее важные применения теории линейного прогноза связаны с прогнозированием *стационарных случайных процессов*. Под этим термином мы будем понимать последовательность случайных величин $\{X_t,\ t \in \mathbb{Z}\}$, $\mathbb{Z}$ — множество всех целых чисел, удовлетворяющих следующим условиям:

$$\mathbf{E}X_t = m = \text{const}, \quad \operatorname{cov}(X_{t+k}, X_t) = R_k, \quad \forall\, t. \tag{28}$$

Таким образом, для стационарной последовательности $\{X_t\}$ среднее любого ее члена остается постоянным, а ковариации между ее членами зависят лишь от разности их номеров. Параметр t обычно трактуется как (дискретное) время, а последовательность $\{X_t\}$ называют *временным рядом*. Величину R_k в (28) обычно называют *автоковариацией с задержкой (лагом) k*, а всю последовательность $\{R_k\}_{-\infty}^{\infty}$ — *автоковариационной функцией* или, кратко, *автоковариацией*. Отметим следующие очевидные

Временной ряд

ее свойства:

$$R_0 = \mathbf{D}X_t \equiv \sigma^2 = \text{const},$$
$$\operatorname{corr}(X_{t+k}, X_t) = \frac{R_k}{R_0} \equiv \rho_k,$$
$$|R_k| \leqslant R_0, \quad R_{-k} = R_k, \quad \rho_{-k} = \rho_k, \tag{29}$$
$$\sum_{i,j=1}^{n} R_{i-j} c_i c_j = \sum_{i,j=1}^{n} c_i c_j \operatorname{cov}(X_i, X_j) = \mathbf{D}\left(\sum_{i=1}^{n} c_i X_i\right) \geqslant 0, \quad \forall\, n.$$

Определенный в (29) коэффициент корреляции ρ_k обычно называют *автокорреляцией с задержкой* k, а всю последовательность $\{\rho_k\}_{-\infty}^{\infty}$ — *автокорреляционной функцией* или, кратко, *автокорреляцией* (в технической литературе используется также термин *коррелограмма*). Последнее свойство в (29) называется *свойством положительной определенности* автоковариации. В дальнейшем предполагается, что матрица $R(n) = \|R_{i-j}\|_1^n$ не вырождена при любом $n \geqslant 1$.

Сделаем еще одно дополнительное предположение, при котором модель временного ряда обладает хорошими свойствами регулярности. Именно, будем считать, что выполняется условие

$$\sum_{k=-\infty}^{\infty} |R_k| = R_0 + 2\sum_{k=1}^{\infty} |R_k| < \infty. \tag{30}$$

Если имеет место (30), то можно рассмотреть *преобразование Фурье*[4] автоковариации $\{R_k\}$:

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R_k \cos k\lambda = \frac{1}{2\pi} R_0 + \frac{1}{\pi} \sum_{k=1}^{\infty} R_k \cos k\lambda, \tag{31}$$
$$\lambda \in [-\pi, \pi].$$

Условие (30) обеспечивает равномерную сходимость ряда (31) к непрерывной функции, и его сумма $f(\lambda)$ называется *спектральной плотностью* временного ряда $\{X_t\}$. Коэффициенты сходящегося ряда Фурье, как известно, однозначно определяются его суммой по формулам

Спектральная плотность

$$R_k = \int_{-\pi}^{\pi} f(\lambda) \cos k\lambda \, d\lambda, \quad k \in \mathbb{Z}. \tag{32}$$

Соотношения (31)–(32) показывают, что спектральная плотность $f(\lambda)$ (когда она существует) и автоковариация $\{R_k\}$ находятся во взаимно-однозначном соответствии, поэтому стационарный временной ряд $\{X_t\}$ можно описывать в терминах любой из этих двух функций.

[4] Фурье Жан Батист Жозеф (1768–1830) — французский математик, один из основоположников математической физики, почетный член Петербургской АН (1829).

Модель стационарного временного ряда является удобной для исследования и в то же время подходящей для описания многих реальных процессов. В рамках этой модели, в частности, достаточно просто решаются задачи прогнозирования, что и является нашей целью в этом пункте. В общем виде задача ставится так: построить оптимальный линейный предиктор (прогноз) для значения X_t процесса в момент t, предполагая, что известно некоторое число $n \geqslant 1$ предыдущих значений $X_{t-1}, \ldots, X_{t-n}$. Покажем, как адаптируется разработанная выше общая теория линейного прогнозирования к модели стационарного временного ряда. Для простоты будем считать, что $m = \mathbf{E}X_t = 0$, в противном случае X_t нужно просто заменить на $X_t - m$. В этом случае по теореме 2 n-прогноз X_{tn}^* для X_t имеет вид

$$X_{tn}^* = \beta_{n1}^* X_{t-1} + \ldots + \beta_{nn}^* X_{t-n} = \sum_{j=1}^{n} \beta_{nj}^* X_{t-j}, \tag{33}$$

где β_{nj}^*, $j = 1, \ldots, n$, минимизируют относительно $\beta_1, \ldots, \beta_n$ среднеквадратическую ошибку прогноза $\mathbf{E}\left(X_t - \sum_{j=1}^{n} \beta_j X_{t-j}\right)^2$. Согласно (14), коэффициенты $\underline{\beta}_n^* = (\beta_{n1}^*, \ldots, \beta_{nn}^*)$ оптимального n-предиктора определяются в данном случае автоковариацией $\{R_k\}$ процесса $\{X_t\}$ по следующим формулам:

$$\underline{\beta}_n^* = R^{-1}(n)\underline{a}, \quad \underline{a} = \operatorname{cov}\big(X_t, (X_{t-1}, \ldots, X_{t-n})\big) = (R_1, \ldots, R_n). \tag{34}$$

Если обозначить $R^{-1}(n) = \|R^{ij}\|_1^n$, то в координатной записи решение имеет вид

$$\beta_{nj}^* = \sum_{i=1}^{n} R^{ij} R_i, \quad j = 1, \ldots, n. \tag{34'}$$

Наконец, с. к. о. $\Delta^*(n)$ прогноза X_{tn}^*, согласно соотношениям (18)–(19) и (29), есть

$$\Delta^*(n) = \mathbf{E}(X_t - X_{tn}^*)^2 = R_0 - \underline{a}' R^{-1}(n)\underline{a}. \tag{35}$$

Если воспользоваться следующей формулой для разложения определителя:

$$\begin{vmatrix} a & \underline{b}' \\ \underline{c} & A \end{vmatrix} = |A|(a - \underline{b}' A^{-1} \underline{c})$$

(здесь A — квадратная невырожденная матрица, $\underline{b}$ и $\underline{c}$ — векторы-столбцы и a — скаляр) и положить здесь $A = R(n)$, $a = R_0$, $\underline{b} = \underline{c} = \underline{a}$, то формуле (35) можно придать вид

$$\Delta^*(n) = \frac{|R(n+1)|}{|R(n)|}. \tag{35'}$$

Соотношениями (33)–(35′) полностью решается задача построения и расчета линейного прогноза в модели стационарного временного ряда.

Обновление прогноза

Рассмотрим, наконец, как конкретизируется в данном случае алгоритм обновления прогноза, если мы хотим добавить в прогноз дополнительное наблюдение X_{t-n-1} (см. соотношение (26)). Прежде всего мы должны построить еще один n-прогноз по тем же переменным $X_{t-1}, \ldots, X_{t-n}$ для величины X_{t-n-1} (так сказать, прогноз в прошлое). В силу специальной структуры стационарного временного ряда (корреляции между его членами зависят лишь от разности соответствующих моментов времени) этот «обратный» предиктор X^*_{t-n-1} имеет те же коэффициенты, что в (33), но переменные должны быть взяты в обратном порядке:

$$X^*_{t-n-1} = \beta^*_{n1} X_{t-n} + \quad + \beta^*_{nn} X_{t-1} \tag{36}$$

(в этом легко убедиться и формальными выкладками, повторив рассуждения, использованные при выводе (33), что мы рекомендуем читателю проделать самостоятельно). Далее мы рассматриваем остаток $e_n = X_{t-n-1} - X^*_{t-n-1}$, который некоррелирован с переменными $X_{t-1}, \quad, X_{t-n}$ и имеет дисперсию $\sigma^2_{e_n} = \Delta^*(n)$, указанную в (35′). Наконец, новый $(n+1)$-предиктор $X^*_{t,n+1}$ для X_t, учитывающий дополнительную, $(n+1)$-ю переменную X_{t-n-1}, имеет вид

$$X^*_{t,n+1} = X^*_{tn} + \beta^*_{n+1,n+1} e_n, \tag{37}$$

где (см. (25) и (36))

$$\beta^*_{n+1,n+1} = \frac{\operatorname{cov}(X_t, e_n)}{\sigma^2_{e_n}} = \frac{R_{n+1} - \beta^*_{n1} R_n - \quad - \beta^*_{nn} R_1}{\Delta^*(n)}; \tag{38}$$

при этом $\beta^*_{n+1,n+1}$ есть частная корреляция X_t с X_{t-n-1}, исключающая влияние всех промежуточных переменных $X_{t-1}, \ldots, X_{t-n}$.

Подставив в (37) выражения (33) и (36), получим явное представление $(n+1)$-прогноза $X^*_{t,n+1}$ через прогнозные переменные $X_{t-1}, \ldots, X_{t-n-1}$:

$$X^*_{t,n+1} = \sum_{j=1}^{n} (\beta^*_{nj} - \beta^*_{n+1,n+1} \beta^*_{n,n-j+1}) X_{t-j} + \beta^*_{n+1,n+1} X_{t-n-1}. \tag{39}$$

Этими соотношениями (36)–(39) и решается проблема обновления прогноза в модели временного ряда: мы видим, что добавление к n-прогнозу X^*_{tn} в (37) нового члена $\beta^*_{n+1,n+1} e_n$ вносит поправки в старые коэффициенты при переменных $X_{t-1}, \quad, X_{t-n}$. Таким образом, если записать $(n+1)$-прогноз в стандартной форме типа (33)

$$X^*_{t,n+1} = \sum_{j=1}^{n+1} \beta^*_{n+1,j} X_{t-j}, \tag{40}$$

то мы получаем, сравнивая (40) с (39), следующие рекуррентные уравнения для вычисления его коэффициентов:

$$\beta^*_{n+1,j} = \beta^*_{nj} - \beta^*_{n+1,n+1} \beta^*_{n,n-j+1}, \quad j = 1, \ldots, n, \tag{41}$$

где $\beta^*_{n+1,n+1}$ указано в (38).

Наконец, формула (22) для с. к. о. обновленного прогноза конкретизируется в данном случае к виду

$$\Delta^*(n+1) = \Delta^*(n)(1 - \beta^{*2}_{n+1,n+1}). \tag{42}$$

Если задать еще соответствующие величины для первого шага ($n = 1$):

$$\beta^*_{11} = \frac{R_1}{R_0}, \quad \Delta^*(1) = \frac{1}{R_0}\begin{vmatrix} R_0 & R_1 \\ R_1 & R_0 \end{vmatrix} = R_0(1 - \beta^{*2}_{11}) \tag{43}$$

(см. соотношения (34) и (35′) для $n = 1$), то мы имеем эффективный рекуррентный алгоритм построения прогноза в модели стационарного временного ряда. Именно: начиная с соотношений (43), далее для $n > 1$ по уравнениям (38) и (41) последовательно вычисляем коэффициенты прогноза до требуемого порядка, а уравнение (42) позволяет на каждом шагу контролировать ошибку соответствующего предиктора.

Замечание. Мы рассмотрели проблему прогнозирования на один шаг вперед (при известных значениях X_{t-j}, $j = 1, 2, \ldots$, оценивается X_t). Но, очевидно, на том же пути можно рассчитать соответствующий прогноз и на любое число шагов вперед, т. е. построить предиктор для X_{t+k} по наблюдениям над $X_{t-1}, \ldots, X_{t-n}$ для любого конечного $k > 0$: для этого в алгоритме вычисления предиктора (33) достаточно лишь заменить вектор $\underline{a} = \underline{a}_0 = (R_1, \ldots, R_n)$ на вектор $\underline{a}_k = (R_{k+1}, \ldots, R_{k+n})$.

Замечание. Представляет также интерес вопрос об асимптотическом (при $n \to \infty$) поведении ошибки (35′) прогноза (33), т. е. при неограниченно возрастающем числе прогнозных переменных. Если $\Delta^*(n) \to 0$ при $n \to \infty$, то прогноз (33) является асимптотически точным. Обозначим λ_{nj}, $j = 1, \ldots, n$, собственные числа матрицы ковариаций $R(n)$. Можно показать, что $\Delta^*(n) \to e^k$, где

$$k = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \ln \lambda_{nj}.$$

Таким образом, прогноз X^*_{tn} является асимптотически точным $\Leftrightarrow k = -\infty$. Если же $k > -\infty$, то, как установил А. Н. Колмогоров (1941),

$$\lim_{n\to\infty} \Delta^*(n) = \exp\left\{ \ln 2\pi + \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln f(\lambda)\, d\lambda \right\},$$

где $f(\lambda)$ — спектральная плотность процесса $\{X_t\}$, определенная в (31).

Мы ограничиваемся этим общим введением в теорию стохастического прогнозирования и отсылаем заинтересованного читателя к книге [4], где дано более детальное изложение этой теории и ее применений к анализу ряда важных для приложений конкретных моделей временных рядов (как стационарных, так и нестационарных), а также подробно описана и проиллюстрирована ее компьютерная реализация в системе STATISTICA.

Если знать, что ждет нас впереди, то стоит ли переживать это заново?

Конфуций [5)]

В заключение мы рассмотрим несколько примеров расчета прогнозов в конкретных физических моделях.

Пример 4 (Прогнозирование в схеме бесповторного выбора). Имеется урна с a шарами, a_1 из которых белые и $a_2 = a - a_1$ — черные. Шары извлекаются по одному последовательно, равновероятно и без возвращения. Пусть $\xi(n)$ обозначает число белых шаров в выборке объема n, $n < N$. Построим прогноз $\xi_n^*(k)$ для $\xi(n+k)$ при известном $\xi(n)$, $1 \leqslant k \leqslant a - n$. •

Решение. Если $\xi(n) = s$, то в урне после n извлечений осталось $a - n$ шаров, из которых $a_1 - s$ — белые, а остальные $a - n - a_1 + s$ — черные. Далее из этой остаточной урны должно быть извлечено еще k шаров, поэтому (см. п. 4 § 1.1) число $\xi(n+k) - s$ белых шаров в этой дополнительной выборке имеет гипергеометрическое распределение $H(a_1 - s, a_2 - n + s, k)$. По формуле (10) § 1.1 математическое ожидание этой случайной величины, т.е. $\mathbf{E}\big(\xi(n+k) - s \mid \xi(n) = s\big)$, равно $k\dfrac{a_1 - s}{a - n}$. Но прогноз

$$\xi_n^*(k) = \mathbf{E}\big(\xi(n+k) \mid \xi(n)\big),$$

поэтому отсюда имеем

$$\xi_n^*(k) = \xi(n) + \frac{k(a_1 - \xi(n))}{a - n} = \left(1 - \frac{k}{a-n}\right)\xi(n) + a_1\frac{k}{a-n} \in [\xi(n), a_1]. \blacksquare \tag{44}$$

Пример 5 (Прогнозирование в схеме повторного выбора). В урне имеется N шаров, занумерованных числами $1, 2, \ldots, N$. Шары извлекаются по одному последовательно, равновероятно и с возвращением. Пусть $X(n)$ обозначает число различных шаров в выборке объема n. Построим прогноз $X_n^*(k)$ для $X(n+k)$ при известном $X(n)$. •

Решение. Пусть $X(n) = s$. Обозначим $i_1, \ldots, i_{N-s}$ номера шаров, не извлекавшихся ни разу в первых n извлечениях, и положим $I_j = 1$, если шар с номером i_j будет извлечен хотя бы один раз при следующих k извлечениях, и $I_j = 0$ в противном случае. Тогда, очевидно,

$$X(n+k) = s + \sum_{j=1}^{N-s} I_j$$

[5)] Конфуций (551–479 до н.э.) — древнекитайский мыслитель.

и

$$\mathbf{E}(I_j \mid X(n) = s) = \mathbf{P}\{I_j = 1 \mid X(n) = s\} =$$
$$= 1 - \mathbf{P}\{I_j = 0 \mid X(n) = s\} = 1 - \left(1 - \frac{1}{N}\right)^k$$

откуда

$$\mathbf{E}\big(X(n+k) \mid X(n) = s\big) = s + (N - s)\left[1 - \left(1 - \frac{1}{N}\right)^k\right]$$

Таким образом, окончательно имеем

$$X_n^*(k) = \mathbf{E}\big(X(n+k) \mid X(n)\big) =$$
$$= X(n) + (N - X(n))\left[1 - \left(1 - \frac{1}{N}\right)^k\right] =$$
$$= \left(1 - \frac{1}{N}\right)^k X(n) + N\left[1 - \left(1 - \frac{1}{N}\right)^k\right] \in [X(n), N]. \blacksquare \qquad (45)$$

Предвидеть — значит управлять.

Б. Паскаль

Упражнения

Приступай к решению своих проблем правильно и начинай с ответов. Тогда, возможно, однажды ты найдешь главный вопрос.

Р. ван Гулик. Отшельник в журавлиных перьях (в книге «Убийства в китайском лабиринте»)

1 Убедиться в том, что система нормальных уравнений метода наименьших квадратов

$$\frac{\partial S(\underline{\beta})}{\partial \beta_j} = 0, \quad j = 1, \ldots, k,$$

имеет вид $A\underline{\beta} = Z\underline{X}$, где $A = ZZ'$

2 Получить следующие явные формулы для о. н. к. $\widehat{\underline{\beta}}$ (см. (2) § 6.2) при $k = 2$:

$$\widehat{\beta}_1 = \frac{\left[\sum_{i=1}^n z_2^{(i)2} \sum_{i=1}^n z_1^{(i)} X_i - \sum_{i=1}^n z_1^{(i)} z_2^{(i)} \sum_{i=1}^n z_2^{(i)} X_i\right]}{\left[\sum_{i=1}^n z_1^{(i)2} \sum_{i=1}^n z_2^{(i)2} - \left(\sum_{i=1}^n z_1^{(i)} z_2^{(i)}\right)^2\right]},$$

$$\widehat{\beta}_2 = \frac{\left[\sum_{i=1}^{n} z_1^{(i)2} \sum_{i=1}^{n} z_2^{(i)} X_i - \sum_{i=1}^{n} z_1^{(i)} z_2^{(i)} \sum_{i=1}^{n} z_1^{(i)} X_i\right]}{\left[\sum_{i=1}^{n} z_1^{(i)2} \sum_{i=1}^{n} z_2^{(i)2} - \left(\sum_{i=1}^{n} z_1^{(i)} z_2^{(i)}\right)^2\right]}.$$

Получить отсюда при $\underline{z}^{(i)} = (z_1^{(i)}, z_2^{(i)}) = (1, t_i)$ формулы (22) § 6.2

3 Проверить непосредственно несмещенность о. н. к. (22) § 6.2 в модели простой регрессии и найти условие их состоятельности, т. е. условие, при котором $\mathrm{D}\widehat{\beta}_j \to 0$ при $n \to \infty$ $(j = 1, 2)$.

4 Показать, что для состоятельности оценки $\widetilde{\sigma}^2$ в модели простой регрессии достаточно, чтобы $\mathrm{D}S(\underline{\widehat{\beta}})/n^2 \to 0$ при $n \to \infty$, где $S(\underline{\widehat{\beta}})$ дано в (23) § 6.2. Проверить, что в случае нормальной регрессии это условие выполняется.

5 Значения функции $x(t) = \beta_1 + \beta_2 t + \beta_3 t^2$ измерены в точках t_i $(i = 1, \ldots, n)$:

$$X_i = x(t_i) + \varepsilon_i, \quad \mathrm{E}\varepsilon_i = 0, \quad \mathrm{D}\varepsilon_i = \sigma^2 \quad \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j.$$

Найти: 1) о. н. к. $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$; 2) $\mathrm{E}\widehat{\beta}_i$, $\mathrm{D}\widehat{\beta}_i$, $\mathrm{cov}(\widehat{\beta}_i, \widehat{\beta}_j)$.

6 (*Продолжение*). Убедиться в том, что статистика

$$\widehat{J} = \int_a^b \widehat{x}(t)\, dt,$$

где $\widehat{x}(t) = \widehat{\beta}_1 + \widehat{\beta}_2 t + \widehat{\beta}_3 t^2$, является несмещениой оценкой интеграла

$$J = \int_a^b x(t)\, dt;$$

получить формулу

$$\mathrm{D}\widehat{J} = \sum_{i,j=1}^{3} \frac{(b-a)^{i+j}}{ij} \mathrm{cov}(\widehat{\beta}_i, \widehat{\beta}_j).$$

7 Смоделировать наблюдения $X_i = \beta_1 + \beta_2 t_i + \varepsilon_i$, $i = 1, \ldots, n$, если $n = 100$, $\beta_1 = 2$, $\beta_2 = 1$, $t_i = 2i/n$, ε_i — независимые равномерно распределенные на $[-1{,}386; 1{,}386]$ случайные величины. Построить графики функций $x(t) = 2 + t$, $\widehat{x}(t) = \widehat{\beta}_1 + \widehat{\beta}_2 t$ на отрезке $[0, 2]$; отметить также точки (t_i, X_i), $i = 1, \ldots, n$.

◀ *Указание.* См. замечание в п. 6 § 1.3. ▶

8 Выполнить такое же задание, как в упр. 7, но с ε_i распределениями нормально с $\mathrm{E}\varepsilon_i = 0$, $\mathrm{D}\varepsilon_i = 0{,}16$.

◀ *Указание.* Использовать алгоритмы п. 8 и 6 § 1.3. ▶

9 В предыдущем упражнении получить следующие общие формулы для γ-доверительных интервалов параметров β_1, β_2 и σ^2:

$$\left(\widehat{\beta}_1 \mp t_{(1+\gamma)/2,\, n-2} \sqrt{\frac{S(\underline{\widehat{\beta}}) \sum_{i=1}^{n} t_i^2}{n(n-2) \sum_{i=1}^{n} (t_i - \bar{t})^2}}\right)$$

$$\left(\widehat{\beta}_2 \mp t_{(1+\gamma)/2,\, n-2} \sqrt{\frac{S(\widehat{\beta})}{(n-2)\sum_{i=1}^{n}(t_i - \bar{t})^2}} \right)$$

$$\frac{S(\widehat{\beta})}{\chi^2_{(1+\gamma)/2,\, n-2}} < \sigma^2 < \frac{S(\widehat{\beta})}{\chi^2_{(1-\gamma)/2,\, n-2}}$$

и рассчитать их численные значения по результатам моделирования при $\gamma = 0{,}9$ и $\gamma = 0{,}95$ (выражение для $S(\widehat{\beta})$ указано в формуле (23) § 6.2). Построить также γ-доверительный эллипс $\mathcal{G}_\gamma(\underline{X})$ для вектора (β_1, β_2), указанный в примере 1 § 6.3.

10 По данным независимых равноточных измерений (X_i, t_i), $i = 1, \ldots, n$, значений некоторой линейной функции $x(t) = \beta_1 + \beta_2 t$ (погрешности измерений подчиняются нормальному распределению $\mathcal{N}(0, \sigma^2)$ с неизвестной дисперсией) построить доверительный интервал для величины

$$\int_{-a}^{a} x(t)\, dt \quad (a \text{ задано}).$$

Провести соответствующие вычисления для следующих данных: (2,96; −2), (3,20; −1), (3,41; 0), (3,63; 1), (3,79; 2) при $a = 2$ и доверительном уровне $\gamma = 0{,}95$.

◄ *Указание.* Воспользоваться предыдущим упр. ►

11 Координаты $a(t)$ движущейся равномерно и прямолинейно точки (т. е. $a(t) = a(0) + vt$) в моменты $t = 1, 2, 3, 4, 5$ оказались соответственно равны 12,98; 13,05; 13,32; 14,22; 13,97. Предполагая погрешности измерений независимыми и нормальными $\mathcal{N}(0, \sigma^2)$, построить 0,95-доверительный эллипс для вектора $(a(0), v)$, где v — скорость точки.

◄ *Указание.* Использовать результат примера 1 § 6.3. ►

12 Смоделировать наблюдения $X_i = \beta_1 + \beta_2 t + \beta_3 t^2 + \varepsilon_i$, $i = 1, \ldots, n$ с $\beta_1 = -8$, $\beta_2 = 10$, $\beta_3 = -2$, $n = 100$, $t_i = 1 + 2i/n$, ε_i — независимые равномерно распределенные на отрезке $[-1{,}386; 1{,}386]$ случайные величины. Построить графики функций $x(t) = \beta_1 + \beta_2 t + \beta_3 t^2$ и $\widehat{x}(t) = \widehat{\beta}_1 + \widehat{\beta}_2 t + \widehat{\beta}_3 t^2$ на отрезке $[1, 3]$.

13 Выполнить такое же задание, как в упр. 12, в случае нормально распределенных ε_i с $\mathrm{E}\varepsilon_i = 0$, $\mathrm{D}\varepsilon_i = 0{,}16$.

14 В предыдущем упр. построить γ-доверительные интервалы для параметров β_1, β_2, β_3 и σ^2, а также систему совместных доверительных интервалов уровня $\geqslant \gamma$ для $\beta_1, \beta_2, \beta_3$ (см. пример 3 § 6.3).

15 В четырехугольнике $ABCD$ результаты независимых и равноточных измерений углов ABD, DBC, ABC, BCD, CDB, BDA, CDA и DAB (в градусах) соответственно таковы: 50,78; 30,25; 78,29; 99,57; 50,42; 40,59; 88,87; 89,86. Считая, что ошибки измерений распределены нормально $\mathcal{N}(0, \sigma^2)$, найти о. н. к. углов $\beta_1 = ABD$, $\beta_2 = DBC$, $\beta_3 = CDB$ и $\beta_4 = BDA$. Построить 0,95-доверительный интервал для σ^2.

◀ *Указание.* В данном случае речь идет о следующей модели нормальной регрессии:

$$X_1 = \beta_1 + \varepsilon_1, \quad X_2 = \beta_2 + \varepsilon_2, \quad X_3 = \beta_1 + \beta_2 + \varepsilon_3, \quad X_4 = 180 - \beta_2 - \beta_3 + \varepsilon_4,$$
$$X_5 = \beta_3 + \varepsilon_5, \quad X_6 = \beta_4 + \varepsilon_6, \quad X_7 = \beta_3 + \beta_4 + \varepsilon_7, \quad X_8 = 180 - \beta_1 - \beta_4 + \varepsilon_8$$

(X_i — результаты указанных последовательных измерений). ▶

16 Используя представление (8) § 6.2), установить формулу

$$\sum_{j=1}^{k} \mathrm{D}\widehat{\beta}_j = \sigma^2 \operatorname{tr} A^{-1} = \sigma^2 \sum_{j=1}^{k} \lambda_j^{-1}$$

где λ_j — собственные числа матрицы A; вывести отсюда критерий состоятельности о. н. к. $\widehat{\beta}_j$: $\min\limits_j \lambda_j \to \infty$ при $n \to \infty$.

◀ *Указание.* Использовать следующие факты из теории матриц: собственные числа матрицы A^{-1} являются обратными значениями собственных чисел A и след A равен сумме ее собственных чисел. ▶

17 Убедиться в том, что интервалы

$$\left(\widehat{\beta}_i - \widehat{\beta}_j \mp \left[\frac{k}{n-k} S(\underline{\widehat{\beta}}) F_{\gamma, k, n-k}(a^{ii} - 2a^{ij} + a^{jj})\right]^{1/2}\right), \quad 1 \leqslant j < i \leqslant k,$$

образуют систему совместных доверительных интервалов уровня $\geqslant \gamma$ для разностей $\beta_i - \beta_j$, $i > j$.

18 Построить систему совместных доверительных интервалов для средних значений всех наблюдений $X_1, \quad , X_n$ в модели нормальной регрессии.

◀ *Указание.* Положить в соотношении (24) § 6.3) $\underline{\lambda}_r = \underline{z}^{(r)}$ $r = 1, \ldots, n$. ▶

19 Пусть имеются k предметов, веса которых $\beta_1, \ldots, \beta_k$ неизвестны. Для определения этих весов взвешиваются комбинации предметов: каждая операция (взвешивание) состоит в том, что несколько предметов кладут на одну чашу весов, несколько — на другую и добавляют разновес для приведения весов в равновесие. В результате получают соотношения

$$\beta_1 z_1^{(i)} + \quad + \beta_k z_k^{(i)} = x_i$$

(для i-го взвешивания, $i = 1, \quad , n$), где $z_j^{(i)} = 1, -1, 0$ в зависимости от того, лежит j-й предмет на левой чаше весов, на правой или вообще не участвует в данном (i-м) взвешивании, а x_i — добавляемый вес. Считав погрешности измерений независимыми и нормальными $\mathcal{N}(0, \sigma^2)$, оценить веса четырех ($k = 4$) предметов по данным следующей таблицы восьми ($n = 8$) взвешиваний:

β_1	1	1	1	1	1	1	1	1
β_2	1	−1	1	−1	1	−1	1	−1
β_3	1	1	−1	−1	1	1	−1	−1
β_4	1	−1	−1	1	1	−1	−1	1
x_i	20,2	8,1	9,7	1,9	19,9	8,3	10,2	1,8

(содержимое внутренней части таблицы образует матрицу плана $Z = \|\underline{z}^{(1)} \quad \underline{z}^{(8)}\|$ с ортогональными строками $\underline{z}_j = (z_j^{(1)}, \ldots, z_j^{(8)})$, $j = 1, 2, 3, 4$). Вычислить вторые

моменты оценок, а также найти оценку для σ^2. Построить систему совместных доверительных интервалов для $\beta_1, \ldots, \beta_4$ с доверительным уровнем $\geqslant 0{,}95$.

◀ *Указание.* Воспользоваться теоремой 4 § 6.2. ▶

20 Построить F-критерий (6) § 6.4, используя результат примера 1 § 6.3 и принцип соответствия между проверкой гипотез и доверительным оцениванием (см. п. 2 § 6.4 и п. 4 § 5.3).

21 Убедиться в том, что для случая двух выборок F-критерий однородности, построенный в п. 2 § 6.5, совпадает с критерием (9) § 5.4; в случае же произвольного числа выборок λ-статистика (7) § 5.4 и F-статистика (11) § 6.5 связаны соотношением

$$\lambda_{n_1 \ldots n_k} = \left(\frac{k-1}{n-k} F + 1 \right)^{-n/2}$$

22 Доказать формулу (20) § 6.4, выражающую F-статистику критерия значимости регрессии через коэффициент детерминации R^2_{k-1}.

23 Пусть $\mathcal{L}(X) = \Pi(\lambda)$ — распределение Пуассона с параметром λ, а $\mathcal{L}(Y|X = k) = \mathrm{Bi}(k, p)$ — биномиальное распределение с параметрами k и p. Показать, что

$$M(X) = \mathbf{E}(Y|X) = pX, \quad M(Y) = \mathbf{E}(X|Y) = \lambda q + Y \quad q = 1 - p,$$
$$\mathbf{D}(Y|X) = pqX, \quad \mathbf{D}(X \mid Y) = \lambda q$$

и, следовательно, $\sigma^2_{YX} = \lambda pq$, $\sigma^2_{XY} = \lambda q$.

24 (*Продолжение*). Показать, что $\mathcal{L}(Y) = \Pi(\lambda p)$ и потому $\sigma^2_Y = \lambda p$, $\sigma^2_X = \lambda$. Получить отсюда, что для обоих прогнозов $M(X)$ и $M(Y)$ их корреляционные отношения совпадают и равны p.

25 Используя результат упр. 66 к гл. 1, построить прогноз Y по $\underline{X}$ в общей нормальной модели и убедиться в том, что корреляционное отношение в этом случае имеет вид

$$\eta^2_{Y\underline{X}} = 1 - (\sigma_{rr}\sigma^{rr})^{-1}$$

26 Построить линейный прогноз $\varphi^*_L(X)$ в модели, описанной в примере 3 § 6.6.

27 Пусть величина Y имеет распределение Парето с параметрами $x_0 > 0$ и $\alpha > 2$ (см. п. 9 § 1.2), а $\mathcal{L}(X|Y = y) = \mathcal{U}(0, y)$ — равномерное распределение на $[0, y]$. Убедиться в том, что прогноз Y по X в классе всех функций $\varphi(X)$ имеет вид

$$M(X) = \mathbf{E}(Y|X) = \frac{\alpha + 1}{\alpha} \max(x_0, X) \quad \text{и его с. к. о.} \quad \Delta^* = \frac{x_0^2}{\alpha(\alpha - 2)}$$

◀ *Указание.* См. решение аналогичной задачи в примере 3 § 6.6. ▶

28 Доказать формулу (23) § 6.6 для случая $p = 2$.

29 Получить вид прогноза Y_2^*, указанный в примере 1 п. 5 § 2.3, как результат обновления прогноза Y_1^* (см. (27) § 6.6) по алгоритму (26) § 6.6.

> *Каждая истина проходит через три инстанции: «какая чушь», «в этом что-то есть» и «кто же этого не знает»!*
>
> **Триада Гумбольдта**[6)]

[6)] Гумбольдт Вильгельм (1767–1835) — немецкий ученый, философ и государственный деятель.

Глава 7

Элементы теории решений. Дискриминантный анализ

Дело не в том, что они не видят решения.
Дело в том, что они не видят проблемы.

Г. К. Честертон [1)]
Острие булавки (из сборника
«Скандальное происшествие с отцом Брауном»)

Эта небольшая глава представляет собой краткое введение в одно из важных направлений современных статистических исследований — теорию статистических решающих функций. Здесь излагаются основные идеи и некоторые результаты этой теории в рамках двух традиционных подходов к принятию решения — байесовского и минимаксного. Общие идеи иллюстрируются практически важным примером решения задачи классификации наблюдений по нескольким классам. Подробно исследуется ряд ситуаций, характеризующихся предположением о нормальном законе распределения наблюдений.

§ 7.1. Статистические решающие функции. Байесовское и минимаксное решения

1. Понятие решающей функции

Во многих случаях конечную цель статистического анализа можно выразить в форме решения о том или ином поведении или действии. Так, при выборочном контроле продукции следует принять одно из двух решений: принять партию изделий или забраковать ее; врач, анализируя симптомы болезни, должен отнести ее к одному из конечного числа стандартных видов, т. е. принять одно из конечного числа возможных решений; анализируя данные наблюдения за некоторым случайным процессом с постоянным, но неизвестным средним, требуется принять решение о величине воздействия на процесс (его коррекции) для приведения среднего, например, к нулю; это воздействие может быть выражено некоторым действительным числом t и, следовательно, здесь число решений бесконечно. Во всех этих случаях решение принимается на основании анализа наблюдения x над соответствующей случайной величиной X и, следовательно, представляет собой некоторую функцию $\delta(x)$

[1)] Честертон Гилберт Кит (1874–1936) — английский писатель.

на выборочном пространстве $\mathfrak{X} = \{x\}$, область значений которой — множество возможных в данной ситуации решений $D \quad \{d\}$. Таким образом, $\delta(x)$ — это правило, ставящее в соответствие каждому результату наблюдения $x \in \mathfrak{X}$ решение $d = \delta(x) \in D$, которое должно быть принято. Функцию $\delta(x)$ называют *решающей функцией* (или *решающим правилом*, или *процедурой*), и ее следует выбирать в соответствии с некоторыми требованиями оптимальности. Принципы решения этой задачи развивает *теория статистических решающих функций* (*теория решений*), разработанная А. Вальдом (1950). Ниже кратко излагаются некоторые основные идеи и результаты этой теории.

2. Функция риска и допустимые решающие правила

Пусть заданы класс распределений (модель) $\mathcal{F} = \{F(x; \theta),\ \theta \in \Theta\}$, которому, по предположению, принадлежит (неизвестное) распределение наблюдаемой случайной величины X, и множество решений $D \quad \{d\}$, которые можно принимать на основании наблюдения над X. Чтобы получить критерий выбора решающей функции, надо сравнить результаты использования различных правил δ. Введем для этого неотрицательную *функцию потерь* $L(d, \theta)$, определенную на прямом произведении $D \times \Theta$, где для каждых $d \in D$ и $\theta \in \Theta$ число $L(d, \theta) \geqslant 0$ интерпретируется как *убыток* или *потеря* от принятия решения d, когда X имеет распределение $F(x; \theta)$. Тогда для всякого правила δ можно определить *функцию риска* (или просто *риск*)

Риск [redacted]

$$R(\delta, \theta) = \mathbf{E}_\theta L(\delta(X), \theta), \tag{1}$$

представляющую собой средние потери от применения решающего правила δ, когда истинным распределением наблюдений является распределение $F(x, \theta)$.

Функция риска дает критерий сравнения различных решающих правил, а именно, если для двух правил δ' и δ выполняется неравенство

$$R(\delta', \theta) \leqslant R(\delta, \theta), \quad \forall\, \theta \in \Theta, \tag{2}$$

со строгим неравенством хотя бы для одного значения θ, то правило δ' предпочтительнее δ, так как использование δ' приводит в среднем к меньшим потерям. С другой стороны, решающие правила δ_1 и δ_2 могут оказаться несравнимыми по критерию (2), если $R(\delta_1, \theta) < R(\delta_2, \theta)$ для некоторых значений θ, а для остальных значений θ имеет место обратное неравенство. Чтобы выбрать в такой ситуации одно из двух правил, необходимо привлекать дополнительные соображения.

Недопустимые решающие правила

Таким образом функция риска позволяет лишь *частично упорядочить* все решающие правила $\delta : \mathfrak{X} \longrightarrow D$, тем не мене, она позволяет «отсеять» заведомо плохие правила. Назовем правило δ *недопустимым*, если существует правило δ' предпочтительнее δ в смысле (2). Решающее правило, не являющееся недопустимым, называется *допустимым*. Ясно, что все недопустимые правила должны быть отвергнуты, так как для каждого из них можно найти более оптимальное (в указанном смысле) правило. Если класс допустимых решающих

правил состоит из единственного правила, то оно и является оптимальным решением. Но обычно этот класс достаточно широк и никакие два решающие правила из множества допустимых несравнимы (по критерию (2)). Поэтому для дальнейшего упорядочения допустимых правил и выбора среди них лучшего привлекают дополнительные соображения. Для решения этой задачи в статистике традиционно применяют два подхода: *байесовский и минимаксный.*

3. Байесовское решение

При байесовском подходе дополнительно предполагается, что параметр θ — это случайная величина с некоторым известным распределением, задаваемым плотностью $\pi(\theta)$ и называемым *априорным распределением*. В этом случае можно усреднить риск (1) по этому априорному распределению и тем самым подсчитать полную среднюю потерю (усредненный риск)

Априорное распределение

$$r(\delta) = \int R(\delta, \theta)\pi(\theta)\, d\theta \tag{3}$$

(если θ принимает дискретные значения, то интеграл заменяется суммой $\sum_i R(\delta, \theta_i)\pi(\theta_i)$), которую мы несем при использовании решающего правила δ. Величина $r(\delta)$ в (3) называется *байесовским риском* правила δ. Таким образом, при таком подходе каждое решающее правило характеризуется одним числом и, следовательно, все решающие правила могут быть упорядочены в соответствии со значениями этой характеристики. Оптимальной является процедура, минимизирующая байесовский риск $r(\delta)$:

Байесовский риск

$$\delta^* = \arg\min_{\delta} \delta; \tag{4}$$

она называется *байесовским решением*. Подчеркнем, что δ^* зависит от выбора априорного распределения π: $\delta^* = \delta^*_\pi$, поэтому для различных априорных распределений параметра соответствующие байесовские решения будут, вообще говоря, различны.

Байесовское решение

Байесовское решение δ^* находят на основании *теоремы Байеса*, согласно которой *апостериорное распределение параметра* при условии $X = x$ задается условной плотностью

Теорема Байеса

$$\pi(\theta|x) = \frac{f(x;\theta)\pi(\theta)}{f(x)}, \tag{5}$$

где

$$f(x) = \mathbf{E} f(x;\theta) = \int f(x;\theta)\pi(\theta)\, d\theta,$$

или

$$f(x) = \sum_i f(x, \theta_i)\pi(\theta_i)$$

в дискретном случае. Тогда равенство (3) можно записать в виде (в абсолютно непрерывном случае)

$$r(\delta) = \int \left(\int L(\delta(x), \theta) f(x; \theta)\, dx \right) \pi(\theta)\, d\theta =$$

$$= \int \left(\int L(\delta(x), \theta) \pi(\theta|x)\, d\theta \right) f(x)\, dx =$$

$$= \int f(x) \mathrm{E}\big[L(\delta(X), \theta) \mid X = x\big] dx \qquad (6)$$

(в дискретном случае интегралы заменяются соответствующими суммами). Из соотношения (6) следует, что *оптимальной является такая процедура, которая при каждом x минимизирует средние потери относительно апостериорного распределения $\pi(\theta|x)$.*

Алгоритм [redacted] Таким образом, для заданного априорного распределения параметра $\pi(\theta)$ алгоритм нахождения байесовского решения имеет следующий вид:

а) для $X = x$ по формуле (5) находят апостериорное распределение $\pi(\theta|x)$;

б) после этого для каждого возможного решения $d \in D$ вычисляют среднюю потерю $\mathbf{E}[L(d, \theta) \mid X = x]$ относительно этого апостериорного распределения;

в) в качестве искомого выбирают решение с минимальной средней потерей.

Отметим, наконец, что байесовское решение, построенное для любого распределения $\pi(\theta) > 0$, является допустимым. Действительно, пусть δ^* — соответствующее байесовское решение и предположим, что существует процедура δ, для которой $R(\delta, \theta) \leqslant R(\delta^*, \theta)$ со строгим неравенством на множестве значений θ, имеющем положительную (относительно распределения π) вероятность. Тогда, очевидно, для соответствующих байесовских рисков (3) имеет место неравенство $r(\delta) < r(\delta^*)$, что противоречит определению байесовского решения.

4. Минимаксное решение

При отсутствии априорной информации о θ применяют прием упорядочения допустимых решающих правил по максимальному значению функции риска $R(\delta, \theta)$ (или *максимальному риску*)

$$m(\delta) = \sup_{\theta} R(\delta, \theta), \qquad (7)$$

и из двух правил предпочтительным считают правило с меньшим максимальным риском (отсюда — термин «минимаксное»). Правило $\widetilde{\delta}$, минимизирующее $m(\delta)$, называется *минимаксным решающим правилом*:

Минимакс

$$\widetilde{\delta} = \arg\min_{\delta} m(\delta). \qquad (8)$$

Таким образом, минимаксное правило избавляет от чрезмерных потерь: наихудший ожидаемый ущерб для него минимален. Но принцип минимакса не всегда благоразумен. Так, на рис. 1 приведены графики функций риска для двух правил δ_1 и δ_2, откуда видно, что правило δ_2 хуже δ_1 для большинства значений θ, но предпочтительнее δ_1 по принципу минимакса.

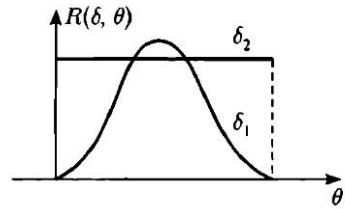


Рис. 1

В общем случае вопрос о существовании и строении минимаксного правила достаточно трудный и здесь не рассматривается. Отметим один случай, когда удается просто установить минимаксность некоторого решающего правила. Именно: предположим, что существует априорное распределение параметра $\pi(\theta) > 0$, для которого функция риска соответствующего байесовского правила δ_π^* постоянна: $R(\delta_\pi^*, \theta) = a = \text{const}$ (такое распределение π называется *наименее благоприятным априорным распределением*). Тогда δ_π^* — минимаксное решение.

Наименее благоприятное априорное распределение

Действительно, в противном случае существовала бы процедура δ, для которой максимальный риск $m(\delta) < a$. Но это означало бы, что $R(\delta, \theta) < R(\delta_\pi^*, \theta)$, $\forall\, \theta$, в противоречие с допустимостью байесовского решения.

Пример 1 (бернуллиевская модель, решающие правила для нее). Пусть X — бернуллиевская случайная величина, причем вероятность «успеха» θ может быть либо $\theta_1 = 1/3$, либо $\theta_2 = 1/2$. Таким образом, здесь $\mathfrak{X} = \{0, 1\}$, $\Theta = \{1/3, 1/2\}$ и $f(x; \theta) = \theta^x(1-\theta)^{1-x}$, $x = 0,1$. Пусть, далее, множество решений D состоит из двух элементов d_1 и d_2, а функция потерь $L(d, \theta)$ задается следующей таблицей:

	d_1	d_2
θ_1	0	2
θ_2	3	1

В данном случае для каждого $x \in \mathfrak{X}$ возможны только два решения, а множество $\mathfrak{X}$ содержит две точки, поэтому всего имеются четыре решающие функции δ_k, $k = 1, 2, 3, 4$, а именно: $\delta_1(0) = d_1$, $\delta_1(1) = d_1$; $\delta_2(0) = d_1$, $\delta_2(1) = d_2$; $\delta_3(0) = d_2$, $\delta_3(1) = d_1$; $\delta_4(0) = d_2$, $\delta_4(1) = d_2$. По формуле (1), которая в данном случае принимает вид

$$R(\delta_k, \theta) = L\big(\delta_k(0), \theta\big)(1-\theta) + L\big(\delta_k(1), \theta\big)\theta, \quad \theta = \theta_1, \theta_2,$$

находим четыре вектора риска $\big(R(\delta_k, \theta_1), R(\delta_k, \theta_2)\big)$, $k = 1, \ldots, 4$, числовые значения которых соответственно равны $(0, 3)$, $(2/3, 2)$, $(4/3, 2)$, $(2, 1)$. Здесь процедура δ_3 недопустима, так как δ_2 предпочтительнее, а среди допустимых процедур δ_1, δ_2 и δ_4 две последние обладают минимаксным свойством: $m(\delta_2) = m(\delta_4) = 2 < m(\delta_3) = 3$.

Пусть теперь на множестве $\Theta \quad \{\theta_1, \theta_2\}$ задано некоторое априорное распределение π: $\pi(\theta_1) = \alpha$, $\pi(\theta_2) = 1 - \alpha$, $0 \leqslant \alpha \leqslant 1$. Найдем соответствующее байесовское решение δ^*

1-й способ. В соответствии с определением (3) вычислим байесовские риски

$$r(\delta_i) = R(\delta_i, \theta_1)\alpha + R(\delta_i, \theta_2)(1 - \alpha)$$

для допустимых правил $(i = 1, 2, 4)$:

$$r(\delta_1) = 3(1 - \alpha), \quad r(\delta_2) = 2 - \frac{4}{3}\alpha, \quad r(\delta_4) = \alpha + 1.$$

Графики этих функций от α изображены на рис. 2, из которого видно, что, если $\alpha \leqslant 3/7$, то $\min\limits_i r(\delta_i) = r(\delta_4)$, т. е. в соответствии с (4) здесь $\delta^* = \delta_4$; если $3/7 < \alpha \leqslant 3/5$, то $\min\limits_i r(\delta_i) = r(\delta_2)$ т. е. здесь $\delta^* = \delta_2$; наконец, $\delta^* = \delta_1$ при $\alpha > 3/5$.

Жирной линией на рис. 2 изображен график байесовского риска $\rho(\alpha) = r(\delta^*)$ как функции α — это выпуклая ломаная линия, являющаяся нижней гранью семейства линейных функций $\{r(\delta_i),\ i = 1, 2, 4\}$ Такая картина характерна для байесовских решений в задачах, в которых параметр θ принимает ровно два значения, а множество D состоит лишь из конечного числа решений.

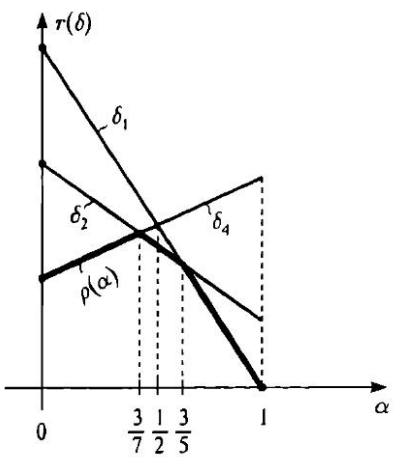


Рис. 2

2-й способ. Вычислим апостериорное распределение (см. (5))

$$\pi(\theta_i|x) = \frac{f(x; \theta_i)\pi(\theta_i)}{f(x)}.$$

Имеем

$$\pi(\theta_1|0) = \frac{\alpha(1 - \theta_1)}{f(0)},$$

$$\pi(\theta_2|0) = \frac{(1 - \alpha)(1 - \theta_2)}{f(0)},$$

$$\pi(\theta_1|1) = \frac{\alpha\theta_1}{f(1)},$$

$$\pi(\theta_2|1) = \frac{(1 - \alpha)\theta_2}{f(1)}.$$

Средняя потеря относительно этого апостериорного распределения при $x = 0$ для решения $\delta(0) = d_1$ равна

$$\mathbf{E}[L(d_1, \theta) \mid X = 0] = L(d_1, \theta_1)\pi(\theta_1|0) + L(d_1; \theta_2)\pi(\theta_2|0) =$$

$$= \frac{3(1 - \alpha)(1 - \theta_2)}{f(0)} = \frac{3(1 - \alpha)}{2f(0)},$$

а для решения $\delta(0) = d_2$ она равна

$$\frac{2\alpha(1-\theta_1)}{f(0)} + \frac{(1-\alpha)(1-\theta_2)}{f(0)} = \frac{(3+5\alpha)}{6f(0)}.$$

Сравнивая эти потери, находим, что при $\alpha \leqslant 3/7$ потери для решенияz d_2 меньше, т. е. $\delta^*(0) = d_2$, если же $\alpha > 3/7$, то $\delta^*(0) = d_1$

Аналогичный анализ для случая, когда $x = 1$, дает: если $\alpha \leqslant 3/5$, то $\delta^*(1) = d_2$, и $\delta^*(1)$ d_1 при $\alpha > 3/5$. Тем самым получены значения байесовского решения $\delta^*(x)$ в каждой точке $x = 0, 1$ при любом априорном распределении параметра.

Конечно, оба способа, хотя и в разной форме, дают один и тот же результат, так как (см. рис. 2) при $\alpha \leqslant 3/7$

$$\delta^*(0) = d_2 = \delta_4(0),$$

при $\alpha > 3/7$

$$\delta^*(0) = d_1 = \delta_2(0) = \delta_1(0),$$

и аналогично, при $\alpha \leqslant 3/5$

$$\delta^*(1) = d_2 = \delta_2(1) = \delta_4(1),$$

при $\alpha > 3/5$

$$\delta^*(1) = d_1 = \delta_1(1).$$

•

5. Оценивание параметров и проверка гипотез с позиций теории решений

Рассмотренные в предыдущих главах задачи оценивания и проверки гипотез также можно сформулировать в терминах принятия решений. Рассмотрим, например, задачу точечной оценки скалярного параметра θ. Выбор статистики $T(X)$, оценивающей θ, можно трактовать как решающее правило, согласно которому принимается решение d_t оценить параметр величиной $t = T(x)$, если наблюдается $X = x$. В этом случае функция потерь может быть, например, $L(d_t, \theta) = \omega(|t - \theta|)$, где ω — строго возрастающая функция ошибки $|t - \theta|$. Если, в частности, выбрать $\omega(z) = z^2$ то функция риска оценки T (квадратичный риск)

$$R(T, \theta) = \mathbf{E}_\theta (T(X) - \theta)^2 \tag{9}$$

совпадает со среднеквадратической ошибкой T. В главе 3 рассматривалась, в основном, проблема построения оценок, минимизирующих квадратичный риск (9) в классе несмещенных оценок, — это оценки с равномерно минимальной дисперсией. В п. 6 § 3.6 достаточно подробно обсуждался байесовский и минимаксный подходы в проблеме оценивания, где по ходу изложения обсуждались и некоторые дополнительные аспекты этой теории.

Рассмотрим теперь интерпретацию задачи проверки гипотез в терминах теории решений. Предположим, что требуется проверить гипотезу H_0 $\theta \in \Theta_0$ при альтернативе H_1 . $\theta \in \Theta_1 = \Theta \backslash \Theta_0$. Тогда любой критерий есть решающее

правило с двумя решениями: d_0 (принимать H_0) и d_1 (принимать H_1). Здесь естественно считать, что потеря равно 0, если принято правильное решение. Тогда функция потерь должна удовлетворять условиям $L(d_0, \theta) = 0, \forall\, \theta \in \Theta_0$, и $L(d_1, \theta) = 0, \forall\, \theta \in \Theta_1$. Если дополнительно принять, что потери от неправильного решения равны 1, т. е. $L(d_0, \theta) = 1, \forall\, \theta \in \Theta_1$, и $L(d_1, \theta) = 0$, $\forall\, \theta \in \Theta_0$, то при такой простой функции потерь для любого решающего правила δ функция риска будет иметь вид (см. (1))

$$R(\delta, \theta) = \begin{cases} \mathbf{P}_\theta\{\delta(X) = d_1\} = \mathbf{P}\{H_1|H_0\} & \text{при} \quad \theta \in \Theta_0, \\ \mathbf{P}_\theta\{\delta(X) = d_0\} = \mathbf{P}\{H_0|H_1\} & \text{при} \quad \theta \in \Theta_1, \end{cases}$$

т. е. ее значения совпадают с вероятностями ошибок 1-го и 2-го родов.

§ 7.2. Классификация наблюдений

1. Задача классификации

В этом параграфе мы рассмотрим один частный, но представляющий большой практический интерес случай, когда параметрическое множество модели $\mathcal{F} = \{F(x; \theta),\ \theta \in \Theta\}$ состоит из конечного числа точек: $\Theta = \{\theta_1, \quad, \theta_k\}$, т. е. имеется всего k распределений $F_i(x) = F(x; \theta_i)$, $i = 1, \quad, k$, одно из которых является истинным всякий раз, когда производится наблюдение над X с $\mathcal{L}(X) \in \mathcal{F}$ В этом случае задача состоит в том, чтобы по наблюдению над X решить, какое из этих k распределений истинно. С позиций теории решений в данном случае множество решений $D = \{d_1, \ldots, d_k\}$, где решение d_i означает, что в качестве истинного распределения выбирается F_i, $i = 1, \quad, k$.

Типичная ситуация, когда возникает подобная задача, может быть описана следующим образом. Пусть множество исследуемых объектов разбито на k классов или групп $H_1, \quad, H_k$. Каждый объект характеризуется набором X числовых параметров, которые непосредственно могут быть измерены. Предполагается, что X — случайная величина, а принадлежность объектов к разным классам выражается в том, что для объектов из класса H_i эта случайная величина имеет распределение F_i, $i = 1, \ldots, k$. Задача состоит в том, чтобы по наблюдению над X определить тот класс, к которому принадлежит соответствующий объект, или, что то же самое, какое из распределений $F_1, \quad, F_k$ истинно. Такие задачи называют задачами *классификации*, *идентификации* или *дискриминации* (от англ. discrimination — различение) и их решение составляет содержание *дискриминантного анализа* — одного из разделов современной математической статистики. Мы рассмотрим общие принципы решения таких задач с позиций теории решений и проиллюстрируем их на примерах наиболее важных моделей таких ситуаций.

2. Функция риска в задаче классификации

Пусть $\delta = \delta(x)$ — произвольное решающее правило в рассматриваемой задаче: $\delta : \mathfrak{X} \longrightarrow D$. Тогда оно порождает разбиение выборочного пространства $\mathfrak{X} = \{x\}$

на k взаимно непересекающихся областей $W_1, \ldots, W_k$, где

$$W_i = \{x \quad \delta(x) = d_i\}, \quad i = 1, \quad , k. \tag{1}$$

Таким образом, множество $W_i \subset \mathcal{X}$ включает все такие точки x, когда при наблюдении $X = x$ в качестве истинного выбирается распределение F_i.

Пусть, далее, на $D \times \Theta$ задана функция потерь $L(d, \theta)$, определяющая потери, от неправильной классификации, т. е. заданы числа $L(d_j, \theta_i) = l(j|i)$, $i, j = 1, \quad , k$, где $l(j|i)$ — убыток, который имеет место в случае, когда объект i-го класса отнесен к j-му ($j \neq i$). В данной задаче естественно считать, что потеря равна нулю, если выбрано правильное решение, т. е. $l(i|i) = 0$, $i = 1, \ldots, k$. Тогда функция риска (1) § 7.1 представляет собой k-мерный вектор $\underline{R}(\delta) = \big(R_1(\delta), \ldots, R_k(\delta)\big)$, где

$$R_i(\delta) = R(\delta, \theta_i) = \mathbf{E}_{\theta_i} L(\delta(X), \theta_i) = \sum_{j=1}^{k} l(j|i) p(j|i), \tag{2}$$

где $p(j|i) = \mathbf{P}_{\theta_i}\{X \in W_j\}$ — вероятность того, что объект i-го класса отнесен к j-му классу. Можно сказать, что $R_i(\delta)$ — средние потери, которые имеют место при классификации по правилу δ произвольного объекта i-го класса. Задача состоит в построении оптимального (т. е. с наименьшими потерями) решающего правила δ.

3. Байесовское решение

Пусть задано априорное распределение $\underline{\pi} = (\pi_1, \quad , \pi_k)$ на $\Theta = \{\theta_1 \quad , \theta_k\}$, т. е. известно, что произвольный наблюдаемый объект принадлежит к классу H_i с вероятностью π, $i = 1, \quad , k$ $(\pi_1 + \quad + \pi_k = 1)$. Тогда байесовский риск (3) § 7.1 на основании (2) равен

$$r(\delta) = \sum_{i=1}^{k} R_i(\delta)\pi_i = \sum_{j=1}^{k} \sum_{i=1}^{k} l(j|i) p(j|i) \pi_i. \tag{3}$$

В соответствии с общим алгоритмом построения байесовского решения (см. п. 3 § 7.1), т. е. правила δ, минимизирующего риск (3), по формуле (5) § 7.1 находим апостериорные вероятности классов при условии $X = x$:

$$\pi_i(x) = \frac{f_i(x)\pi_i}{\sum\limits_{s=1}^{k} \pi_s f_s(x)}, \quad i = 1, \quad , k. \tag{4}$$

Далее, если принять решение отнести объект с характеристикой x к j-му классу (т. е. если $\delta(x) = d_j$), то для такого правила средние условные (при

условии $X = x$) потери равны

$$\mathbf{E}\big[L(\delta(X), \theta) \mid X = x\big] = \sum_{i=1}^{k} L(d_j, \theta_i)\pi_i(x) = \frac{\sum_{i=1}^{k} l(j|i)\pi_i f_i(x)}{\sum_{s=1}^{k} \pi_s f_s(x)} \qquad (5)$$

Наконец, решение d_j должно быть выбрано так, чтобы минимизировать (5).

Алгоритм

В итоге мы приходим к следующей формулировке оптимального способа действия: *если наблюдалось* $X = x$, *то надо определить минимальную сумму*

$$h_j(x) = \sum_{i=1}^{k} l(j|i)\pi_i f_i(x), \quad j = 1, \quad , k, \qquad (6)$$

и отнести объект к классу с номером, равным номеру этой минимальной суммы (если минимум достигается при нескольких значениях j, то можно взять любое из них, например, минимальное).

Итак, имеет место следующее утверждение.

Теорема 1. Оптимальное (байесовское) решающее правило δ^* в задаче классификации определяется следующим разбиением выборочного пространства $\mathfrak{X} = W_1^* \cup \quad \cup W_k^*$:

$$W_i^* = \big\{x \quad h_i(x) = \min_{1 \leqslant j \leqslant k} h_j(x)\big\}, \quad i = 1, \quad , k, \qquad (7)$$

где функции $h_j(x)$ определены в (6) и i — минимальное значение индекса, удовлетворяющее указанному условию.

Рассмотрим частный случай, когда $l(j|i) = 1$, $j \neq i$. Тогда

$$h_j(x) = \sum_{i \neq j} \pi_i f_i(x) = -\pi_j f_j(x) + \sum_{i=1}^{k} \pi_i f_i(x)$$

и условие в (7) принимает вид

$$\pi_i f_i(x) = \max_{1 \leqslant j \leqslant k} \pi_j f_j(x). \qquad (8)$$

Величины $\pi_j f_j(x)$ пропорциональны апостериорным вероятностям классов (см. (4)), поэтому в данном случае байесовский принцип сводится к следующему: *относить объект с характеристикой* x *к тому классу, апостериорная вероятность которого максимальна.* Такой принцип называют *принципом максимума апостериорной вероятности*. Этот принцип

Принцип максимума апостериорной вероятности

используют во всех случаях, когда потери $l(j|i)$ либо неизвестны, либо их трудно оценить числом.

Выделим случай двух классов ($k = 2$), т. е. когда объекты классифицируются по альтернативному признаку. Здесь

$$h_1(x) = l(1|2)\pi_2 f_2(x), \quad h_2(x) = l(2|1)\pi_1 f_1(x),$$

следовательно, решение d_1 (отнести объект с характеристикой x к первому классу) принимается тогда и только тогда, когда $l(1|2)\pi_2 f_2(x) \leqslant l(2|1)\pi_1 f_1(x)$. Таким образом, байесовское правило δ^* имеет в данном случае следующий вид:

$$\delta^*(x) = \begin{cases} d_1 & \text{при} \quad x \in W_1^* \\ d_2 & \text{при} \quad x \in \overline{W}_1^* \end{cases} \qquad W_1^* = \left\{ x \quad \frac{f_2(x)}{f_1(x)} \leqslant \frac{\pi_1}{\pi_2}\frac{l(2|1)}{l(1|2)} \right\} \tag{9}$$

а соответствующий ему вектор риска $\big(R_1(\delta^*), R_2(\delta^*)\big)$ в соответствии с (2) равен

$$\big(l(2|1)\mathbf{P}_{\theta_1}\{X \in \overline{W}_1^*\}, l(1|2)\mathbf{P}_{\theta_2}\{X \in W_1^*\}\big). \tag{10}$$

4. Минимаксное решение

Если априорные вероятности классов неизвестны, то для построения решающего правила используют минимаксный подход (см. п. 4 § 7.1), в соответствии с которым ищется правило $\tilde{\delta}$, минимизирующее $\max\limits_{1 \leqslant i \leqslant n} R_i(\delta)$ (см. (2)). В ряде случаев минимаксное правило $\tilde{\delta}$ удается построить, определив наименее благоприятное априорное распределение $\underline{\pi} = (\pi_1, \quad, \pi_k)$, при котором для соответствующего байесовского решения δ^* все компоненты вектора риска $\underline{R}(\delta^*)$ одинаковы, — в этом случае $\tilde{\delta}$ совпадает с решением δ^*

Например, в случае двух классов из (10) следует, что π_1 ($\pi_2 = 1 - \pi_1$) надо определить из условия

$$l(2|1)\mathbf{P}_{\theta_1}\{X \in \overline{W}_1^*\} = l(1|2)\mathbf{P}_{\theta_2}\{X \in W_1^*\}. \tag{11}$$

Отсюда видно, что даже в этом простейшем случае для произвольных распределений F_i задача построения минимаксного решения достаточно сложна. В следующих параграфах эти вопросы рассмотрены детальнее для случая нормальных распределений.

§ 7.3. Классификация наблюдений в случае двух нормальных классов

Предположим теперь, что в эксперименте наблюдается нормальный случайный вектор $X = (X_1, \quad, X_r)$, распределенный для объектов из класса H_1 по закону $\mathcal{N}(\underline{\mu}^{(1)}, \Sigma)$, а для объектов из класса H_2 — по закону $\mathcal{N}(\underline{\mu}^{(2)}, \Sigma)$, причем векторы $\underline{\mu}^{(1)}$, $\underline{\mu}^{(2)}$ и матрица Σ известны ($|\Sigma| \neq 0$). Таким образом,

здесь предполагается, что имеет место случай двух нормальных классов, различающихся только средними значениями, и задача состоит в том, чтобы по наблюдению над вектором $\boldsymbol{X}$ решить, какое из этих двух распределений истинно.

Найдем байесовское и минимаксное правила в данной ситуации.

1. Байесовский подход

В этом случае предполагается, что заданы потери $l(1|2)$ и $l(2|1)$ от принятия неправильного решения и априорные вероятности π_1 и π_2 $(\pi_1+\pi_2=1)$ классов $\boldsymbol{H}_1$ и $\boldsymbol{H}_2$ соответственно, тем самым определена константа

$$c = \ln\left[\frac{\pi_2 l(1|2)}{\pi_1 l(2|1)}\right].$$

В соответствии с (9) область W_1^* наилучшей классификации в пользу класса $\boldsymbol{H}_1$ определяется неравенством $\ln\left(f_1(\underline{x})/f_2(\underline{x})\right) \geqslant c$, которое в соответствии с соотношением (9) § 1.2 принимает в данном случае вид

$$\frac{1}{2}\left[\left(\underline{x}-\underline{\mu}^{(2)}\right)'\Sigma^{-1}\left(\underline{x}-\underline{\mu}^{(2)}\right) - \left(\underline{x}-\underline{\mu}^{(1)}\right)'\Sigma^{-1}\left(\underline{x}-\underline{\mu}^{(1)}\right)\right] \geqslant c$$

или, если положить $\underline{a} = \Sigma^{-1}\left(\underline{\mu}^{(1)}-\underline{\mu}^{(2)}\right)$,

$$\underline{a}'\underline{x} - \frac{1}{2}\underline{a}'\left(\underline{\mu}^{(1)}+\underline{\mu}^{(2)}\right) \geqslant c. \tag{1}$$

Дискриминантная функция

Линейную функцию $\varphi(\underline{x}) = \underline{a}'\underline{x}$ называют *дискриминантной функцией*. Таким образом, области наилучшей классификации определяются в данном случае дискриминантной функцией $\varphi(\underline{x})$: наблюдение $\boldsymbol{X}=\underline{x}$ относится к классу $\boldsymbol{H}_1$ тогда и только тогда, когда

$$\varphi(\underline{x}) \geqslant c_1 = c + \frac{1}{2}\underline{a}'\left(\underline{\mu}^{(1)}+\underline{\mu}^{(2)}\right).$$

Если потери $l(i|j)$ неизвестны или их трудно оценить числом, то применяют указанное правило классификации с $l(1|2)=l(2|1)=1$, т. е. в этом случае (см. (1))

$$W_1^* = \left\{\underline{x} \;\; \underline{a}'\underline{x} \geqslant \ln\left(\frac{\pi_2}{\pi_1}\right) + \frac{1}{2}\underline{a}'\left(\underline{\mu}^{(1)}+\underline{\mu}^{(2)}\right)\right\} \tag{2}$$

Замечание. Ранее (см. пример 3 в § 5.2 и его продолжение там же) мы уже рассматривали задачу различения двух нормальных распределений $\mathcal{N}(\underline{\mu}^{(i)}, \Sigma)$ с позиций критерия Неймана—Пирсона, когда минимизируется вероятность ошибки 2-го рода при фиксированной вероятности ошибки 1-го рода (критерий (18) § 5.2), и с позиций минимизации суммы вероятностей этих ошибок (критерий (21) § 5.2). Теперь же мы получили новый, более сильный, результат с позиций теории решений, учитывающий как наличие дополнительной априорной информации о гипотетических распределениях (априорные вероятности π_1 и π_2 классов $\boldsymbol{H}_1$ и $\boldsymbol{H}_2$), так и более детализированное описание последствий от ошибочных решений (задание потерь $l(j|i)$). В частности, полученное здесь решение (2) соответствует критерию (21) § 5.2 и сводится к нему при $\pi_1=\pi_2=1/2$, т. е. в случае априори равновероятных классов $\boldsymbol{H}_1$ и $\boldsymbol{H}_2$.

2. Минимаксный подход

Найдем теперь минимаксное решение $\tilde{\delta}$, которое используется в случае неизвестных априорных вероятностей классов. Для этого вычислим вероятности ошибочных классификаций $p(j|i) = \mathbf{P}\{H_j|H_i\}$, $i, j = 1, 2$, $i \neq j$, для произвольного байесовского правила, задаваемого условием (1) (константа c теперь произвольна, так как априорное распределение $\underline{\pi} = (\pi_1, \pi_2)$ произвольно), и из условия (11) § 7.2 найдем наименее благоприятное априорное распределение. Соответствующее этому априорному распределению байесовское решение (1) и является искомым минимаксным решением $\tilde{\delta}$.

Для этого можно воспользоваться вычислениями, проведенными в примере 3 § 5.2. Именно, если ввести случайную величину

$$Y = \underline{a}'X - \frac{1}{2}\underline{a}'\left(\underline{\mu}^{(1)} + \underline{\mu}^{(2)}\right),$$

то из формул (16) § 5.2 получим, что

$$\mathcal{L}(Y|H_1) = \mathcal{N}\left(\frac{\rho}{2}, \rho\right), \quad \mathcal{L}(Y|H_2) = \mathcal{N}\left(-\frac{\rho}{2}, \rho\right), \tag{3}$$

где

$$\rho = \left(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}\right)'\Sigma^{-1}\left(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}\right) \tag{4}$$

есть расстояние Махаланобиса между распределениями

$$\mathcal{N}\left(\underline{\mu}^{(1)}, \Sigma\right) \quad \text{и} \quad \mathcal{N}\left(\underline{\mu}^{(2)}, \Sigma\right).$$

Отсюда находим вероятность ошибочной классификации:

$$p(2|1) = \mathbf{P}\{H_2|H_1\} = \mathbf{P}\{Y < c|H_1\} = \Phi\left(\frac{c - \rho/2}{\sqrt{\rho}}\right),$$

$$p(1|2) = \mathbf{P}\{H_1|H_2\} = \mathbf{P}\{Y \geqslant c|H_2\} = \Phi\left(-\frac{c + \rho/2}{\sqrt{\rho}}\right).$$

Следовательно, уравнение (11) § 7.2 для определения наименее благоприятного апроирного распределения $(\pi_1, 1-\pi_1)$, или, что эквивалентно, константы c, имеет вид

$$l(2|1)\Phi\left(\frac{c - \rho/2}{\sqrt{\rho}}\right) = l(1|2)\Phi\left(-\frac{c + \rho/2}{\sqrt{\rho}}\right). \tag{5}$$

Итак, если априорные вероятности классов неизвестны, то минимаксное правило классификации определяется разбиением $\mathfrak{X} = W_1^* \cup W_2^*$, где

$$W_1^* = \left\{\underline{x} : \underline{a}'\underline{x} - \frac{1}{2}\underline{a}'\left(\underline{\mu}^{(1)} + \underline{\mu}^{(2)}\right) \geqslant c\right\}, \quad W_2^* = \overline{W}_1^*, \tag{6}$$

и константа c выбирается из условия (5).

Отметим, что если $l(2|1) = l(1|2)$, что решением уравнения (5) является $c = 0$, а так как

$$c = \ln\left[\frac{\pi_2 l(1|2)}{\pi_1 l(2|1)}\right],$$

то в этом случае имеем, что наименее благоприятное априорное распределение является равномерным: $\pi_1 = \pi_2 = 1/2$. Для этого случая области наилучшей классификации W_1^* (6) и (2) совпадают и, как отмечено в замечании п. 1, мы приходим к критерию (21) § 5.2, — так смыкаются подходы байесовский и минимаксный с подходом, основанным на минимизации суммы вероятностей ошибок 1-го и 2-го рода. Отметим также, что в этом случае вероятность ошибочной классификации произвольного объекта равна $\Phi(-\sqrt{\rho}/2)$, следовательно, чем более далекими (в метрике ρ Махаланобиса) являются «гипотезы» H_1 и H_2, тем меньше вероятность ошибочной классификации правила (6).

§ 7.4. Классификация нормальных наблюдений. Общий случай

В этом параграфе мы рассмотрим применение изложенной в § 7.2 теории к общему случаю нескольких классов, заданных многомерными нормальными распределениями. Предположим, что эти распределения различаются только своими средними, и пусть $\mathcal{N}(\underline{\mu}^{(i)}, \Sigma)$ — распределение наблюдений $\boldsymbol{X}$ для объектов из класса H_i $(i = 1, \quad , k)$. Кроме того, все цены ошибочных классификаций будем считать равными единице: $l(j|i) = 1$, $j \neq i$.

1. Байесовский подход

Пусть заданы априорные вероятности $\pi_i = \mathbf{P}\{H_i\}$, $i = 1, \ldots, k$ $(\pi_1 + \ldots + \pi_k = 1)$ принадлежности произвольно выбранного объекта соответствующим классам. Тогда байесовское решение можно получить с помощью принципа максимума апостериорной вероятности (см. (8) § 7.2), в соответствии с которым области классификации имеют вид

$$W_i^* = \left\{ \underline{x} \;\; \frac{f_i(\underline{x})}{f_j(\underline{x})} \geqslant \frac{\pi_j}{\pi_i}, \; j = 1, \quad , k, \; j \neq i \right\}, \quad i = 1, \ldots \quad , k,$$

где (см. (9) § 1.2)

$$f_i(\underline{x}) = \frac{1}{(2\pi)^{r/2}|\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2}\left(\underline{x} - \underline{\mu}^{(i)}\right)' \Sigma^{-1} \left(\underline{x} - \underline{\mu}^{(i)}\right) \right\}, \quad i = 1, \quad , k.$$

Отсюда, как и в п. 1 § 7.3, получаем, что

$$W_i^* = \left\{ \underline{x} \;\; u_{ij}(\underline{x}) \geqslant c_i - c_j, \; j = 1, \quad k, \; j \neq i \right\}, \tag{1}$$

Классификационные функции где $c_i = \ln(1/\pi_i)$, $i = 1, \quad , k$, и

$$u_{ij}(\underline{x}) = \ln\left(\frac{f_i(\underline{x})}{f_j(\underline{x})}\right) = \underline{a}_{ij}' \underline{x} - \frac{1}{2}\underline{a}_{ij}' \left(\underline{\mu}^{(i)} + \underline{\mu}^{(j)}\right), \tag{2}$$

$$\underline{a}_{ij} = \Sigma^{-1}\left(\underline{\mu}^{(i)} - \underline{\mu}^{(j)}\right), \quad i, j = 1, \ldots, k, \quad i \neq j. \tag{3}$$

Отметим, что каждая *классификационная функция* $u_{ij}(\underline{x})$ связана только с i-й и j-й совокупностями, $u_{ij}(\underline{x}) = -u_{ji}(\underline{x})$, при этом все они — линейные функции наблюдений $\underline{x}$. Следовательно, области W_i^* ограничены гиперплоскостями.

Вычислим, наконец, вектор риска $\underline{R} = (R_1, \ldots, R_k)$ этого байесовского правила. В соответствии с формулой (2) § 7.2 в данном случае

$$R_i = \sum_{j \neq i} p(j|i) = 1 - p(i|i) = 1 - \int\limits_{W_i^*} f_i(\underline{x}) d\underline{x} =$$

$$= 1 - \mathbf{P}\{u_{ij}(\boldsymbol{X}) \geqslant c_i - c_j,\ j = 1, \ldots, k,\ j \neq i \mid H_i\}, \qquad (4)$$

поэтому надо знать распределение случайного вектора

$$\underline{U}_i = \big(u_{ij}(\boldsymbol{X}),\ j = 1, \ldots, k,\ j \neq i\big)$$

при гипотезе H_i. Чтобы найти это распределение, введем матрицу

$$\Sigma_i = \|\underline{a}_{i1} \ \ldots \underline{a}_{ik}\|,$$

составленную из векторов-столбцов $\underline{a}_{ij}$, $j = 1, \ldots, k$, $j \neq i$, определенных в (3). Тогда из формулы (2) следует, что вектор $\underline{U}_i$ получается из $\boldsymbol{X}$ с помощью линейного преобразования вида $\underline{U}_i = \Sigma_i' \boldsymbol{X} + \underline{b}_i$. Следовательно, по свойству 2) в п. 2 § 1.2 (сохранение нормальности при линейных преобразованиях)

$$\mathcal{L}(\underline{U}_i | H_i) = \mathcal{N}\big(\Sigma_i' \underline{\mu}^{(i)} + \underline{b}_i,\ \Sigma_i' \Sigma \Sigma_i\big). \qquad (5)$$

Отсюда имеем, что j-я координата вектора средних $\mathbf{E}(\underline{U}_i | H_i)$ равна

$$\underline{a}_{ij}' \underline{\mu}^{(i)} - \frac{1}{2} \underline{a}_{ij}' \big(\underline{\mu}^{(i)} + \underline{\mu}^{(j)}\big) = \frac{1}{2} \underline{a}_{ij}' \big(\underline{\mu}^{(i)} - \underline{\mu}^{(j)}\big) = \frac{\rho_{ij}}{2},$$

где $\rho_{ij} \quad \big(\underline{\mu}^{(i)} - \underline{\mu}^{(j)}\big)' \Sigma^{-1} \big(\underline{\mu}^{(i)} - \underline{\mu}^{(j)}\big)$ — расстояние Махаланобиса между i-м и j-м классами (или между распределениями $\mathcal{N}(\underline{\mu}^{(i)}, \Sigma)$ и $\mathcal{N}(\underline{\mu}^{(j)}, \Sigma)$), а (j, s)-й элемент дисперсионной матрицы $\Sigma_i' \Sigma \Sigma_i$, равный скалярному произведению j-й строки матрицы Σ_i' (т. е. вектора $\underline{a}_{ij}'$) на s-й столбец матрицы $\Sigma \Sigma_i$ (т. е. $\Sigma \underline{a}_{is} = \big(\underline{\mu}^{(i)} - \underline{\mu}^{(s)}\big)$), имеет вид

$$\rho_{ijs} = \big(\underline{\mu}^{(i)} - \underline{\mu}^{(j)}\big)' \Sigma^{-1} \big(\underline{\mu}^{(i)} - \underline{\mu}^{(s)}\big).$$

Тем самым распределение (5) полностью определено. Оно является невырожденным ($|\Sigma_i' \Sigma \Sigma_i| \neq 0$) тогда и только тогда, когда Σ_i — матрица полного ранга, т. е. когда $\operatorname{rank} \Sigma_i = k - 1$ (это имеет место, в частности, когда векторы средних значений $\underline{\mu}^{(1)}, \ldots, \underline{\mu}^{(k)}$ линейно независимы и $r = \dim \boldsymbol{X} \geqslant k - 1$). В этом случае существует плотность $g_i(\underline{u})$ нормального распределения (5) и формулу (4) можно записать в виде

$$R_i = 1 - p(i|i) = 1 - \int\limits_{c_i - c_1} \quad \int\limits_{c_i - c_k} g_i(\underline{u})\, d\underline{u}, \quad i = 1, \ldots, k \qquad (6)$$

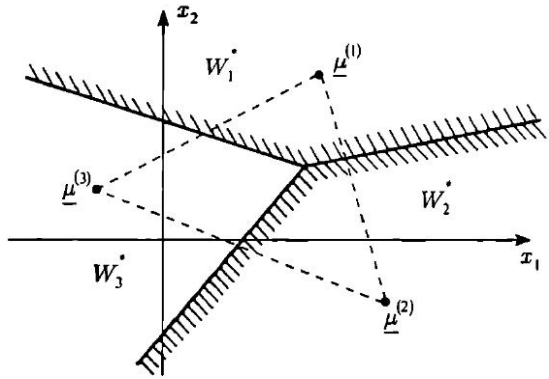


Рис. 1

(здесь интеграл имеет кратность $k-1$, так как среди разностей $c_i - c_j$ отсутствует разность $c_i - c_i$).

В качестве примера рассмотрим случай трех классов ($k = 3$), задаваемых двумерными нормальными распределениями ($r = 2$).

В данном случае области W_i^* $i = 1, 2, 3$, имеют соответственно вид

$$W_1^* = \{\underline{x} = (x_1, x_2) \quad u_{12}(\underline{x}) \geqslant c_1 - c_2,\ u_{13}(\underline{x}) \geqslant c_1 - c_3\},$$
$$W_2^* = \{\underline{x} = (x_1, x_2) \quad u_{12}(\underline{x}) < c_1 - c_2,\ u_{23}(\underline{x}) \geqslant c_2 - c_3\},$$
$$W_3^* = \{\underline{x} = (x_1, x_2) \quad u_{13}(\underline{x}) < c_1 - c_3,\ u_{23}(\underline{x}) < c_2 - c_3\}$$

Эти области должны исчерпывать всю плоскость, поэтому линии, задаваемые уравнениями $u_{12}(\underline{x}) = c_1 - c_2$, $u_{13}(\underline{x}) = c_1 - c_3$ и $u_{23}(\underline{x}) = c_2 - c_3$, должны пересечься в одной точке (см. рис. 1).

2. Минимаксный подход

Если априорные вероятности $\pi_1, \ldots, \pi_k$ классов неизвестны, то области классификации ищутся в виде (1), где неопределенные константы $c_i > 0$, $i = 1, \ldots, k$, следует выбирать из условия равенства всех компонент вектора риска $\underline{R} = (R_1, \ldots, R_k)$, которое в силу (6) приводит к равенствам

$$p(1|1) = p(2|2) = \quad = p(k|k). \tag{7}$$

Эти равенства однозначно определяют разности $c_i - c_j$ (по крайней мере в принципе), а тем самым и области (1) минимаксного правила классификации.

3. Классификация наблюдений в случае неизвестных параметров

Выше предполагалось, что все допустимые распределения наблюдений полностью известны. Однако в приложениях эти распределения часто бывают известны лишь с точностью до значений некоторых параметров (например, $\underline{\mu}^{(i)}$, $i = 1, \ldots, k$, или Σ, или одновременно всех этих параметров). В таких

случаях можно решать задачи классификации при наличии дополнительных, как говорят, *обучающих выборок* $\left(\underline{X}_1^{(i)}, \ldots, \underline{X}_{n_i}^{(i)}\right)$ из соответствующих распределений $\mathcal{N}(\underline{\mu}^{(i)}, \Sigma)$, $i = 1, \ldots, k$. Эти выборки можно использовать для оценки соответствующих неизвестных параметров и, заменив неизвестные параметры их оценками, поступать далее, как и в случае полностью известных распределений, получая в итоге *эмпирические* (приближенные) правила классификации.

Обучающие выборки

Если неизвестны средние $\underline{\mu}^{(i)}$ то их заменяют оценками

$$\widehat{\underline{\mu}}^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} \underline{X}_j^{(i)}$$

— средними арифметическими выборок. Дисперсионная матрица Σ (когда она неизвестна) оценивается выборочной дисперсионной матрицей (см. п. 2 § 2.3); при этом так как матрица Σ, по определению, — общая для всех классов, то для ее оценивания следует использовать информацию, доставляемую всеми выборками. Эту информацию объединяют следующим образом. Построим по i-й выборке $\left(\underline{X}_1^{(i)} \ldots, \underline{X}_{n_i}^{(i)}\right)$ выборочную дисперсионную матрицу

$$\widehat{\Sigma}^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(\underline{X}_j^{(i)} - \widehat{\underline{\mu}}^{(i)}\right)\left(\underline{X}_j^{(i)} - \widehat{\underline{\mu}}^{(i)}\right)', \quad i = 1, \ldots, k,$$

тогда (см. п. 2 § 2.3) матрица $\dfrac{n_i}{(n_i - 1)} \widehat{\Sigma}^{(i)}$ для любого i будет несмещенно оценивать Σ. Отсюда

$$\mathbf{E} \sum_{i=1}^{k} n_i \widehat{\Sigma}^{(i)} = \sum_{i=1}^{k} (n_i - 1)\Sigma = (n - k)\Sigma, \quad n = n_1 + \quad + n_k,$$

и, следовательно, выборочная матрица

$$\widehat{\Sigma} = \frac{1}{n - k} \sum_{i=1}^{k} n_i \widehat{\Sigma}^{(i)},$$

построенная с учетом всех данных, является также несмещенной оценкой матрицы Σ.

Построив эти оценки неизвестных параметров, далее можно ввести оценки классификационных функций $u_{ij}(\underline{x})$ (см. (2))

$$\widehat{u}_{ij}(\underline{x}) = \widehat{\underline{a}}_{ij}' \underline{x} - \frac{1}{2} \widehat{\underline{a}}_{ij}' \left(\widehat{\underline{\mu}}^{(i)} + \widehat{\underline{\mu}}^{(j)}\right), \quad \widehat{\underline{a}}_{ij} = \widehat{\Sigma}^{-1}\left(\widehat{\underline{\mu}}^{(i)} - \widehat{\underline{\mu}}^{(j)}\right),$$

и использовать их для построения эмпирических областей классификации

$$\widehat{W}_i = \left\{\underline{x} : \widehat{u}_{ij}(\underline{x}) \geqslant c_i - c_j,\ j = 1, \quad , k,\ j \neq i\right\}, \quad i = 1, \ldots, k,$$

заменяющих в данном случае области (1).

В качестве обоснования этой методики можно привести следующие «асимптотические» рассуждения. Предположим, что объемы обучающих выборок велики ($n_i \longrightarrow \infty$, $i = 1, \ldots, k$). Тогда оценки $\widehat{\underline{\mu}}^{(i)}$ $i = 1,$ $, k$, и $\widehat{\Sigma}$, как мы знаем (см. п. 3 § 2.3), сходятся по вероятности соответственно к $\underline{\mu}^{(i)}$ $i = 1,$ $, k$, и Σ. Отсюда следует, что $\widehat{\underline{a}}'_{ij}$ сходится по вероятности к $\underline{a}_{ij}$, а $\widehat{\underline{a}}'_{ij}\left(\widehat{\underline{\mu}}^{(i)} + \widehat{\underline{\mu}}^{(j)}\right)$ — к $\underline{a}'_{ij}\left(\underline{\mu}^{(i)} + \underline{\mu}^{(j)}\right)$ (см. теорему 1 § 2.2). Следовательно, предельное распределение $\widehat{u}_{ij}(\boldsymbol{X})$ совпадает с распределением $u_{ij}(\boldsymbol{X})$, поэтому для достаточно больших обучающих выборок эмпирические классификационные функции $u_{ij}(\underline{x})$ можно использовать так же, как если бы были точно известны распределения совокупностей.

Упражнения

Никогда не ставьте задачу, решение которой вам неизвестно.

Правило Берке

Знание ответов — великое преимущество.

Арабская пословица

1 Пусть $\mathcal{L}(X) = \mathrm{Bi}(1, \theta)$, $\Theta = \{\theta_1 = 1/3, \theta_2 = 2/3\}$, множество решений $D = \{d_1, d_2\}$ и функция потерь задана таблицей

	d_1	d_2
θ_1	0	1
θ_2	2	0

1) Определить все допустимые решающие правила в данной ситуации и убедиться в том, что $\widetilde{\delta} = \left(\widetilde{\delta}(0), \widetilde{\delta}(1)\right) = (d_1, d_2)$ — минимаксная процедура.

2) Убедиться в том, что для априорного распределения $\pi(\theta_1) = \alpha$, $\pi(\theta_2) = 1 - \alpha$, $\alpha \in [0, 1]$, байесовское решение имеет следующий вид: $\delta^*(0) = d_2$ при $\alpha \leqslant 1/2$, $\delta^*(0) = d_1$ при $\alpha > 1/2$ и $\delta^*(1) = d_2$ при $\alpha \leqslant 4/5$, $\delta^*(1) = d_1$ при $\alpha > 4/5$. Построить график байесовского риска $\rho(\alpha) = r(\delta^*)$ как функции α.

◀ *Указание.* См. решение в примере 1 § 7 1. ▶

2 Пусть $\mathcal{L}(X) = \mathrm{Bi}(1, \theta)$, $\Theta = \{\theta_1 = 1/4, \theta_2 = 3/4\}$, $D = \{d_1, d_2, d_3\}$ и функция потерь задана таблицей

	d_1	d_2	d_3
θ_1	0	1	1/2
θ_2	4	0	1/2

1) Убедиться в том, что байесовское решение $\delta^* = \left(\delta^*(0), \delta^*(1)\right)$ при априорном распределении $\pi(\theta_1) = \alpha \in [0, 1]$, $\pi(\theta_2) = 1 - \alpha$, имеет вид: $\delta^* = (d_2, d_2)$ при

$\alpha \leqslant 1/4$; $\delta^*(d_3, d_2)$ при $1/4 < \alpha \leqslant 7/10$; $\delta^* = (d_1, d_2)$ при $7/10 < \alpha \leqslant 3/4$; $\delta^* = (d_1, d_3)$ при $3/4 < \alpha \leqslant 21/22$; $\delta^* = (d_1, d_1)$ при $\alpha > 21/22$.

2) Показать, что байесовский риск $\rho(\alpha) = r(\delta^*)$ есть

$$\rho(\alpha) = \begin{cases} \alpha & \text{при } 0 \leqslant \alpha \leqslant \dfrac{1}{4}, \\ \dfrac{1+4\alpha}{8} & \text{при } \dfrac{1}{4} < \alpha \leqslant \dfrac{7}{10}, \\ \dfrac{4-3\alpha}{4} & \text{при } \dfrac{7}{10} < \alpha \leqslant \dfrac{3}{4}, \\ \dfrac{11-10\alpha}{8} & \text{при } \dfrac{3}{4} < \alpha \leqslant \dfrac{21}{22}, \\ 4(1-\alpha) & \text{при } \dfrac{21}{22} < \alpha \leqslant 1. \end{cases}$$

Построить график функции $\rho(\alpha)$, $\alpha \in [0, 1]$.

◀ *Указание.* Сравнить средние потери $\mathbf{E}\big(L(d, \theta)|x\big)$ относительно апостериорного распределения, указанные в следующей таблице. ▶

	d_1	d_2	d_3
$x = 0$	$\dfrac{1-\alpha}{f(0)}$	$\dfrac{3\alpha}{4f(0)}$	$\dfrac{1+2\alpha}{8f(0)}$
$x = 1$	$\dfrac{3(1-\alpha)}{f(1)}$	$\dfrac{\alpha}{4f(1)}$	$\dfrac{3-2\alpha}{8f(1)}$

3 Пусть $\mathcal{L}(X) = \mathrm{Bi}(1, \theta)$, $\Theta = \{\theta_1, \theta_2\}$, $D = \{d_1, d_2\}$ и функция потерь $L(d, \theta)$ задана таблицей $(a, b > 0)$.

	d_1	d_2
θ_1	0	a
θ_2	b	0

Рассмотреть два случая: 1) θ_1 $2/3$, $\theta_2 = 1/2$; 2) $\theta_1 = 3/4$, $\theta_2 = 1/2$. Убедиться в том, что в обоих случаях множества допустимых решающих правил совпадают, но для случая 2) при любом апроиорном распределеиии параметра байесовское решение предпочтительнее.

◀ *Указание.* Воспользоваться решением упр. 1. ▶

4 Пусть $\mathcal{L}(X) = \mathrm{Bi}\big(3, \theta\big)$, $\Theta = \{\theta_1 = 10^{-2}\ \ \theta_2 = 10^{-1}\}$, $D = \{d_1, d_2\}$ и функция потерь $L(d, \theta)$ задана таблицей

	d_1	d_2
θ_1	0	2
θ_2	1	0

Рассмотреть три решающие правила $\delta_i = (\delta_i(0), \delta_i(1), \delta_i(2), \delta_i(3))$: $\delta_1 = (d_1, d_2, d_2, d_2)$, $\delta_2 = (d_1, d_1, d_2, d_2)$, $\delta_3 = (d_1, d_1, d_1, d_2)$ и убедиться в том, что минимаксное решение $\bar{\delta} = \delta_1$ и $m(\bar{\delta}) = 0{,}792$.

5 Пусть $\mathcal{L}(X) = \overline{\mathrm{Bi}}(1, \theta)$, $\Theta = \{\theta_1, \theta_2\}$, $D = \{d_1, d_2\}$ и функция потерь задана таблицей, указанной в упр. 3. Рассмотреть решающие функции

$$\delta_i(x) = \begin{cases} d_1 & \text{при} \quad x = 0, 1, \ldots, i-1, \\ d_2 & \text{при} \quad x = i, i+1, \end{cases} \qquad i = 1, 2, \ldots,$$

и убедиться в том, что при $a\theta_1 \leqslant b(1-\theta_2)$ минимаксное решение $\bar{\delta} = \delta_1$. Построить $\bar{\delta}$ при $a\theta_1 > b(1-\theta_2)$.

◀ *Указание.* Получить сначала вектор риска для i-й процедуры:

$$\underline{R}_i = \big(R(\delta_i, \theta_1), R(\delta_i, \theta_2)\big) = \big(a\theta_1^i,\ b(1-\theta_2^i)\big). \ \blacktriangleright$$

6 Убедиться в том, что если в условии упр. 5 заменить $\mathcal{L}(X)$ на пуассоновский закон $\Pi(\theta)$, то $\bar{\delta} = \delta_1$ при $a(1-e^{-\theta_1}) \leqslant be^{-\theta_2}$, а соответствующий вектор риска есть $\big(a(1-e^{-\theta_1}),\ be^{-\theta_2}\big)$.

7 Предположим, что наблюдается случайная величина X, распределенная по нормальному закону с неизвестным средним θ и известной дисперсией σ^2 множество решений $D = \{d_1, d_2, d_3\}$ и функция потерь $L(d, \theta)$ задана таблицей

θ \ d	d_1	d_2	d_3
$\theta < 0$	0	1	2
$\theta = 0$	1	0	1
$\theta > 0$	2	1	0

Рассмотрим решающие функции вида

$$\delta_{ab}(x) = \begin{cases} d_1 & \text{при} \quad x < a, \\ d_2 & \text{при} \quad a \leqslant x \leqslant b, \\ d_3 & \text{при} \quad x > b, \end{cases} \qquad a < 0 < b.$$

Убедиться в том, что функция риска имеет вид

$$R(\delta_{ab}, \theta) = \begin{cases} \Phi\left(\dfrac{\theta - a}{\sigma}\right) + \Phi\left(\dfrac{\theta - b}{\sigma}\right) & \text{при} \quad \theta < 0, \\ \Phi\left(\dfrac{a}{\sigma}\right) + \Phi\left(-\dfrac{b}{\sigma}\right) & \text{при} \quad \theta = 0, \\ \Phi\left(\dfrac{a - \theta}{\sigma}\right) + \Phi\left(\dfrac{b - \theta}{\sigma}\right) & \text{при} \quad \theta > 0, \end{cases}$$

и построить ее график при $b = -a$.

8 Предположим, что $\Theta = \{0, 1\}$, $D = \{d\} = [0, 1]$ и функция потерь имеет вид $L(d, \theta) = |d - \theta|^a$, $a \geqslant 1$. Рассмотрим класс решающих функций вида $\delta(x) = \mathrm{const}$ (т. е. решение принимается без предварительных наблюдений). Убедиться в том, что

в этом классе байесовское решение d^*, соответствующее априорному распределению $\pi(0) = \alpha \in [0, 1]$, $\pi(1) = 1 - \alpha$, при $a > 1$ имеет вид

$$d^* = \left[1 + \left(\frac{\alpha}{1-\alpha}\right)^{1/(a-1)}\right]^{-1}$$

если же $a = 1$, то $d^* = 0$ при $\alpha > 1/2$, $d^* = 1$ при $\alpha < 1/2$, а при $\alpha = 1/2$ в качестве d^* может быть взято любое $d \in [0, 1]$.

◀ *Указание.* Установить сначала, что байесовский риск есть

$$r(d) = \alpha d^a + (1 - \alpha)(1 - d)^a \quad \blacktriangleright$$

9 Рассматривается задача оценивания неизвестной вероятности «успеха» θ по наблюдению числа «успехов» X в n испытаниях Бернулли (таким образом, здесь $\mathcal{L}(X) = \mathrm{Bi}(n, \theta)$, $\Theta = D = [0, 1]$). Пусть функция потерь имеет вид

$$L(d, \theta) = \frac{(d-\theta)^2}{\theta(1-\theta)},$$

а априорное распределение параметра является равномерным $U(0, 1)$. Показать, что байесовское решение есть $\delta^*(x) = x/n$, и оно же является и минимаксным, при этом $r(\delta^*) = 1/n$.

10 Предположим, что по наблюдению X оценивается параметр θ равномерного распределения $U(0, \theta)$, при этом параметр случаен и имеет априорную плотность $\pi(\theta) = \theta e^{-\theta}$ $\theta > 0$. Убедиться в том, что байесовская оценка при квадратичной функции потерь имеет вид $\delta^*(X) = X + 1$, а ее риск $r(\delta^*) = 1$.

◀ *Указание.* Записать среднюю апостериорную потерю для решения d в виде интеграла

$$\mathbf{E}[L(d, \theta) \mid X = x] = \int_0^\infty (d-\theta)^2 \pi(\theta|x)\, d\theta \cong \int_x^\infty (d-\theta)^2 e^{-\theta}\, d\theta \cong \int_0^\infty (d - x - t)^2 e^{-t}\, dt$$

и продифференцировать по d. ▶

11 Рассмотрим задачу оценивания скалярного параметра θ при функции потерь $L(d, \theta) = |d - \theta|$, $d, \theta \in R^1$. Доказать, что при наблюдении $X = x$ байесовское решение $\delta^*(x)$ при любом априорном распределении параметра $\mathcal{L}(\theta)$ есть медиана апостериорного распределения $L(\theta|x)$, т. е. такое число, что

$$\mathbf{P}\{\theta \leqslant d^*|x\} \geqslant \frac{1}{2}, \quad \mathbf{P}\{\theta \geqslant d^*|x\} \geqslant \frac{1}{2}.$$

◀ *Указание.* Установить неравенство $\mathbf{E}(|\theta - d||x) \geqslant \mathbf{E}(|\theta - d^*||x)$, $\forall d \in R^1$ ▶

12 (*Продолжение*). Пусть оценивается среднее модели $\mathcal{N}(\theta, b^2)$, когда $\mathcal{L}(\theta) = \mathcal{N}(\mu, \sigma^2)$. Используя упр. 64 (5) к гл. 1 и введенные там обозначения, убедиться в том, что байесовская оценка по выборке $X = (X_1, \ldots, X_n)$ из $\mathcal{N}(\theta, b^2)$ имеет вид

$$\delta^*(X) = \sigma_1^2\left(\frac{\mu}{\sigma^2} + \frac{n\overline{X}}{b^2}\right)$$

и ее риск равен $r(\delta^*) = \mathbf{E}|\theta - \delta^*(X)| = \sigma_1\sqrt{2/\pi}$ (сравни с упр. 83 к гл. 3).

◀ *Указание.* Воспользоваться формулой полного математического ожидания (см. (6) § 6.6). ▶

Глава 8

Факторный анализ

Если кто-то способен предсказать, чем закончатся его исследования, то эта проблема не очень глубока и, можно сказать, практически не существует.

А. Шильд

Эта глава, как и последующая, посвящена изложению методов исследования и объяснения структуры корреляционных матриц систем случайных величин. Такие исследования важны при анализе и интерпретации результатов эксперимента. Их целью в первом случае является выявление небольшого числа скрытых общих факторов, оказывающих решающее влияние на результаты эксперимента, во втором — определение небольшого числа линейных комбинаций наблюдаемых переменных, объясняющих их суммарную дисперсию. В обоих случаях такое исследование позволяет лучше понять природу случая, порождающего данные, и достичь более экономного их представления, что важно при обработке больших массивов информации.

§ 8.1. Постановка задач факторного анализа

При исследовании сложных объектов собирают статистику обычно по большому числу различных характеристик объекта, стремясь при этом учесть как можно больше. Однако это стремление приводит, как правило, к обратному эффекту, именно, с одной стороны, получается огромный массив информации, который трудно (а практически зачастую и невозможно) проанализировать математическими методами, а с другой стороны, получается «засорение» данных большим количеством несущественных характеристик, которые не несут в себе никакой существенной информации об исследуемом объекте, но по которым собирается большая (и часто дорогостоящая) статистика. В этих ситуациях естественно встает задача отделить существенные характеристики от несущественных, отыскать те общие факторы, наличием которых можно объяснить те или иные закономерности, действующие в множестве существенных характеристик, и тем самым «сжать» информацию, т. е. записать ее в терминах меньшего числа переменных. Конечно, все эти и аналогичные задачи можно математически строго формализовать и решить лишь в рамках определенной модели. Одной из таких моделей, достаточно хорошо действующей во многих реальных ситуациях, является модель *факторного анализа*.

Основной задачей факторного анализа является экономное описание экспериментальных данных. Он «объясняет» корреляционную матрицу системы

случайных величин (далее пишем кратко — с. в.) $x_1, \ldots, x_r$ (т. е. наличие корреляций между ними) наличием небольшого числа общих гипотетических переменных или *факторов*, от которых зависят $x_1, \quad, x_r$. В общем виде модель факторного анализа выглядит следующим образом:

$$x_1 = \varphi_1(f_1, \ldots, f_k) + e_1,$$

$$x_r = \varphi_r(f_1, \quad, f_k) + e_r, \tag{1}$$

где φ_i — известные функции k переменных, $f_1, \ldots, f_k$ и $e_1, \quad, e_r$ — с. в., причем $e_1, \quad, e_r$ независимы как между собой, так и от $f_1, \quad, f_k$.

Общие и частные факторы

Если в представлении (1) число k значительно меньше r, то с. в. $x_1, \ldots, x_r$ связаны между собой лишь через небольшое число с. в. $f_1, \quad, f_k$. С. в. $e_1, \ldots, e_r$ не участвуют в этой связи и каждая из них влияет лишь на соответствующую компоненту вектора $\underline{x} = (x_1, \ldots, x_r)$. Переменные $f_1, \ldots, f_k$ называются *общими факторами* (они отражают те причины-факторы, которые обуславливают зависимость между $x_1, \ldots, x_r$), а переменные $e_1, \ldots, e_r$ называются *частными факторами* (они отражают локальные причины-факторы). Определение общих факторов и выявление их смысла в конкретной ситуации является важной прикладной задачей. Решение ее позволяет: 1) объяснить природу зависимости наблюдаемых с. в. $x_1, \ldots, x_r$, 2) отделить существенные переменные (зависящие от $f_1, \ldots, f_k$) от несущественных (не зависящих от $f_1, \quad, f_k$) 3) более экономно записать информацию в терминах $f_1, \ldots, f_k$, а не самих с. в. $x_1, \quad, x_r$ (при $k \ll r$).

Линейная модель

Дополнительно к (1) в факторном анализе обычно принимаются следующие допущения.

1. Модель (1) линейна, т. е.

$$x_1 = l_{11}f_1 + \quad + l_{1k}f_k + e_1,$$

$$x_r = l_{r1}f_1 + \quad + l_{rk}f_k + e_r, \tag{2}$$

   или в матричном виде (напомним, что при матричных операциях векторы понимаются как вектор-столбцы)

$$\underline{x} = L\underline{f} + \underline{e},$$

   где $L = \|l_{ij}\|$ — прямоугольная матрица размера $r \times k$; $\underline{f} = (f_1, \ldots, f_k)$, $\underline{e} = (e_1, \quad, e_r)$.
2. Факторы $f_1, \ldots, f_k$ независимы.
3. Все факторы, как $f_1, \ldots, f_k$, так и $e_1, \ldots, e_r$ распределены по нормальному закону, при этом будем предполагать, что f_i распределено по стандартному закону $\mathcal{N}(0, 1)$, $i = 1, \quad, k$, а e_j — по закону $\mathcal{N}(0, v_j)$, $j = 1, \ldots, r$.

Матрица нагрузок

Коэффициент l_{ij} в (2) называется *нагрузкой i-й переменной на j-й фактор* (или нагрузкой j-го фактора в i-й переменной). Матрица $L = \|l_{ij}\|$ называется *матрицей нагрузок*.

Обычно нагрузки l_{ij} и *остаточные дисперсии* v_i являются неизвестными параметрами модели (2), так же, как и число общих факторов k, и их надо оценить по наблюдениям над с. в. $\underline{x} = (x_1, \ldots, x_r)$. Иногда еще ставится вопрос оценки значений самих факторов $f_1, \ldots, f_k$ по результатам наблюдений. Эти задачи мы и рассмотрим далее.

Исторически факторный анализ возник из нужд психологии. Он использовался для того, чтобы определить относительно небольшое число тестов, ответы на которые давали бы как можно более полное описание способностей человека. Однако в последнее время факторный анализ все более широко используется в социологии, экономике, медицине, геологии и многих других областях. Широкому распространению методов факторного анализа в последнее время способствовало развитие средств вычислительной техники. Это связано с тем обстоятельством, что практическая реализация положений теории факторного анализа связана, как правило, с весьма трудоемкими вычислениями, что будет ясно из последующего изложения.

§ 8.2. Неоднозначность решения в факторном анализе

Задачи факторного анализа имеют ту специфику, что ни нагрузки, ни факторы в модели (2) § 8.1 не могут быть определены однозначно. Действительно, пусть H — произвольная ортогональная матрица размера $k \times k$. Рассмотрим случайный вектор $\underline{f}^{(1)} = H\underline{f}$. В силу основных предположений модели (2) § 8.1 (компоненты вектора $\underline{f}$ независимы и распределены каждая по закону $\mathcal{N}(0, 1)$) и в силу ортогональности H имеем (см. упр. 31 к гл. 1), что компоненты вектора $\underline{f}^{(1)}$ также независимы и распределены каждая по закону $\mathcal{N}(0, 1)$. В этом случае представление $\underline{x} = L\underline{f} + \underline{e}$ можно записать в виде

$$\underline{x} = (LH')(H\underline{f}) + \underline{e} = L_1\underline{f}^{(1)} + \underline{e},$$

где $L_1 = LH'$. Отсюда следует, что компоненты вектора $\underline{f}^{(1)}$ также являются общими факторами модели (2) § 8.1, но с другой матрицей нагрузок.

Ортогональное преобразование H определяет некоторое вращение в пространстве факторов, и таким образом, (общие) факторы определяются с точностью до вращения; при этом соответственно изменяются и нагрузки на факторы.

Это свойство часто используется для преобразования факторов, полученных в каком-нибудь практическом исследовании. Вращение обычно подбирают так, чтобы коэффициенты нагрузок удовлетворяли тем или иным свойствам, например, чтобы определенные переменные имели как можно большую

нагрузку на один фактор и нулевые или почти нулевые нагрузки на другие факторы.

Во всяком случае в факторном анализе возникает еще одна задача — выбор вращения (ортогональной матрицы H) факторов и интерпретация факторов, получаемых после такого вращения (по коэффициентам нагрузок на факторы).

Выбор вращения факторов производится совместно со специалистом в той области, из которой возникла конкретная задача (*статистике часто принадлежит первое слово, но никогда — последнее!*). Иногда можно избежать произвол во вращении факторов, если из дополнительных соображений известно, что некоторые элементы матрицы нагрузок L должны быть равны нулю (т. е. некоторые переменные x_i не зависят от некоторых факторов f_j).

§ 8.3. Вывод уравнений максимального правдоподобия

Обозначим через $C = \|c_{ij}\|_1^r$ матрицу вторых моментов с. в. $\underline{x} = (x_1, \quad , x_r)$. Тогда из (2) § 8.1 следует, что

$$c_{ii} = \mathrm{D}x_i = \sum_{j=1}^{k} l_{ij}^2 + v_i,$$
$$c_{ij} = \mathrm{cov}(x_i, x_j) = \mathrm{E}x_i x_j = \sum_{s=1}^{k} l_{is} l_{js}, \quad i \neq j, \tag{1}$$

или в матричном виде

$$C = LL' + V, \tag{1'}$$

где $V = \begin{Vmatrix} v_1 & & 0 \\ & \ddots & \\ 0 & & v_r \end{Vmatrix}$ — диагональная матрица остаточных дисперсий. Матрица C обычно неизвестна и ее оценивают по выборке.

Пусть $\underline{x}^{(1)}, \quad , \underline{x}^{(n)}$, $n \geqslant r$, — независимые наблюдения над с. в. $\underline{x}$. Построим для этой системы с. в. выборочную ковариационную матрицу (см. п. 2. § 2.3)

$$A = \|a_{ij}\|_1^r = \frac{1}{n} \sum_{t=1}^{n} \underline{x}^{(t)} \underline{x}^{(t)'}$$

(напомним, что у нас вектор средних значений известен и равен нулю; если бы вектор средних был неизвестен, то в качестве A надо было бы взять матрицу

$$A = \|a_{ij}\|_1^r = \frac{1}{n} \sum_{t=1}^{n+1} \left(\underline{x}^{(t)} - \overline{\underline{x}}\right)\left(\underline{x}^{(t)} - \overline{\underline{x}}'\right),$$

и тогда с учетом этих обозначений все последующие выводы сохраняют свою силу). Наша цель — использовать информацию, содержащуюся в A, для получения оценок параметров l_{ij} и v_i. Для этого мы воспользуемся методом

максимального правдоподобия. Согласно этому методу оценки l_{ij} и v_i определяются из условия, чтобы совместная плотность элементов выборочной ковариационной матрицы, вычисленная в наблюденной «точке» $A = \|a_{ij}\|_1^r$, т. е. функция правдоподобия, имела наибольшее значение. Воспользуемся следующим результатом.

Если случайный вектор $\underline{x} = (x_1, \quad , x_r)$ распределен по нормальному закону $\mathcal{N}(\underline{0}, C)$, то совместное распределение элементов матрицы $A = \|a_{ij}\|$,

Распределение Уишарта

называемое *распределением Уишарта* и обозначаемое $W(r, n, C)$, имеет плотность вероятностей

$$\gamma(r, n)|C|^{-n/2}|A|^{(n-r-1)/2} \exp\left\{-\frac{n}{2}\sum_{i,j} a_{ij}c^{ij}\right\}, \tag{2}$$

где $\|c^{ij}\| = C^{-1}$ и $\gamma(r, n)$ — множитель, зависящий только от r и n. Отметим, что это распределение сосредоточено на множестве положительно определенных матриц $A = \|a_{ij}\|$ размерности $r \times r$ Сумма в (2) может быть записана также как $\operatorname{tr}(AC^{-1})$. Отметим также, что $W(1, n, \sigma^2)$ есть не что иное, как распределение $(\sigma^2/n)\chi^2(n)$ (см. п. 3 § 1.2), так что распределение Уишарта есть многомерный аналог распределения χ^2

Опуская члены, не зависящие от C, из (2) получаем, что о. м. п. для l_{ij} и v_i — это те значения l_{ij} и v_i, которые доставляют минимум функции

$$L = \ln|C| + \sum_{i,j} a_{ij}c^{ij} \tag{3}$$

Чтобы получить соответствующие уравнения, мы должны продифференцировать (3) по l_{ij} и v_i и приравнять все частные производные нулю.

Будем обозначать большими буквами C_{ij} алгебраические дополнения элементов c_{ij} в матрице C. Тогда поскольку

$$|C| = \sum_j c_{ij}C_{ij} = \sum c_{ij}C_{ij},$$

то

$$\frac{\partial|C|}{\partial c_{ij}} = C_{ij}.$$

Если при этом матрица C симметрична, то из $C_{ij} = C_{ji}$ следует, что при $i \neq j$

$$\frac{\partial|C|}{\partial c_{ij}} = 2C_{ij}.$$

Отсюда также следует, что если элементы c_{ij} являются функциями от $z_1, \quad , z_s$, то

$$\frac{\partial|C|}{\partial z_t} = \sum_{i,j}\frac{\partial|C|}{\partial c_{ij}}\frac{\partial c_{ij}}{\partial z_t} = \sum_{i,j} C_{ij}\frac{\partial c_{ij}}{\partial z_t}$$

Используя эти формулы и учитывая соотношение (1), получим:

$$\frac{\partial \ln |C|}{\partial v_i} = \frac{C_{ii}}{|C|} = c^{ii} \quad i = 1, \ldots, r,$$

$$\frac{\partial \ln |C|}{\partial l_{ij}} = \sum_{t,s} \frac{C_{ts}}{|C|} \frac{\partial c_{ts}}{\partial l_{ij}} = \sum_{t<s} c^{ts} \frac{\partial c_{ts}}{\partial l_{ij}} + \sum_{t} c^{tt} \frac{\partial c_{tt}}{\partial l_{ij}} = 2 \sum_{s} c^{si} l_{sj}, \tag{4}$$

$$i = 1, \ldots, r, \quad j = 1, \quad , k,$$

поскольку

$$\frac{\partial c_{tt}}{\partial l_{ij}} = \begin{cases} 0 & \text{при } t \neq i, \\ 2l_{ij} & \text{при } t = i, \end{cases}$$

и при $t \neq s$

$$\frac{\partial c_{ts}}{\partial l_{ij}} = \begin{cases} 0 & \text{при } t \neq i, \ s \neq i, \\ l_{sj} & \text{при } t = i, \\ l_{tj} & \text{при } s = i. \end{cases} \tag{5}$$

Вычислим теперь частные производные по l_{ij} и v_i от $c^{ts} = C_{ts}/|C|$.

Так как дифференцировать $|C|$ мы уже умеем (см. (4)), то достаточно найти производные от C_{ts}. Эти производные можно вычислить, воспользовавшись разложением C_{ts} по элементам некоторого столбца $(c_{1v}, c_{2v}, \ldots, c_{rv})$ без элемента c_{tv} (здесь $v \neq s$):

$$C_{ts} = \sum_{u} c_{uv} C_{ts,\,uv},$$

где $C_{ts,\,uv}$ обозначает алгебраическое дополнение элемента c_{uv} в C_{ts}. Отсюда

$$\frac{\partial C_{ts}}{\partial v_i} = C_{ts,\,ii}, \quad \text{если} \quad t \neq i, \ s \neq i,$$

$$\frac{\partial C_{ts}}{\partial l_{ij}} = \sum_{\substack{u,\,v \\ u \neq t,\, v \neq s}} \frac{\partial C_{ts}}{\partial c_{uv}} \frac{\partial c_{uv}}{\partial l_{ij}} = \sum_{\substack{u,\,v \\ u \neq t,\, v \neq s}} C_{ts,\,uv} \frac{\partial c_{uv}}{\partial l_{ij}}. \tag{6}$$

Если воспользоваться теперь соотношением

$$C_{ts,\,uv}|C| = C_{ts}C_{uv} - C_{tv}C_{us},$$

формулами (5) и (4), то из (6) можно получить, что

$$\frac{\partial c^{ts}}{\partial v_i} = \left(\frac{\partial C_{ts}}{\partial v_i} - C_{ts} \frac{\partial \ln |C|}{\partial v_i} \right) \frac{1}{|C|} = -c^{it}c^{si}$$

и аналогично

$$\frac{\partial c^{ts}}{\partial l_{ij}} = 2 \sum_{u \neq s} l_{uj} \frac{C_{ts,\,iu}}{|C|} - 2C_{ts} \sum_{u} l_{uj} \frac{C_{iu}}{|C|^2} = -2 \sum_{u} l_{uj} c^{ut} c^{si}, \quad t \neq i.$$

Таким образом, окончательно получаем, что

$$\frac{1}{2}\frac{\partial L}{\partial l_{ij}} = \sum_{u} l_{uj}c^{ui} - \sum_{t,s,u} l_{uj}c^{ut}a_{ts}c^{si} \tag{7}$$

$$\frac{\partial L}{\partial v_i} = c^{ii} - \sum_{t,s} c^{it}a_{ts}c^{si} \tag{8}$$

Нетрудно заметить, что правая часть (7) есть элемент j-й строки и i-го столбца матрицы

$$L'C^{-1} - L'C^{-1}AC^{-1},$$

а правая часть (8) есть i-й диагональный элемент матрицы

$$C^{-1} - C^{-1}AC^{-1}$$

Таким образом, уравнения максимального правдоподобия для определения оценок параметров l_{ij} и v_i имеют вид

$$L'C^{-1} - L'C^{-1}AC^{-1} = 0, \tag{9}$$

$$\operatorname{diag}(C^{-1} - C^{-1}AC^{-1}) = 0, \tag{10}$$

где $\operatorname{diag} M$ обозначает матрицу, содержащую только диагональную часть матрицы M При этом C имеет вид $(1')$.

Выше мы видели (§ 8.2), что матрица L определяется с точностью до умножения справа на произвольную ортогональную матрицу (это также видно и из уравнений (9)–(10), поскольку при умножении L справа на произвольную ортогональную матрицу величина C остается неизменной). Чтобы зафиксировать L (зафиксировать направление в пространстве факторов), мы добавим к (9)–(10) еще одно условие. Именно, мы будем требовать, чтобы матрица

$$J = L'V^{-1}L \tag{10'}$$

была диагональной с диагональными элементами, расположенными в порядке убывания.

Уравнения (9)–(10) можно значительно упростить. Для этого умножим сначала (9) справа на C. В результате мы придем к уравнению

$$L' - L'C^{-1}A = 0. \tag{11}$$

Далее, поскольку матрица $V = C - LL'$ диагональная с положительными элементами и поскольку в силу (11)

$$V(C^{-1} - C^{-1}AC^{-1}) = \mathbb{1}_r - AC^{-1} - LL'C^{-1} + LL'C^{-1}AC^{-1} = \mathbb{1}_r - AC^{-1}$$

то уравнение (10) эквивалентно уравнению

$$\operatorname{diag}(\mathbb{1}_r - AC^{-1}) = 0. \tag{12}$$

Аналогично имеем:

$$(\mathbb{1}_r - AC^{-1})V = C - A - LL' + AC^{-1}LL' = C - A$$

и потому (12) эквивалентно уравнению

$$\operatorname{diag}(C - A) = 0. \tag{13}$$

Из (13) следует, что для всех i $c_{ii} = a_{ii}$ и в силу (1)

$$v_i = a_{ii} - \sum_{j=1}^{k} l_{ij}^2. \tag{14}$$

Таким образом, оценки для остаточных дисперсий v_i могут быть найдены по формуле (14), коль скоро будут получены оценки для факторных нагрузок l_{ij}.

Приведем уравнение (11) к виду, более удобному для практического решения. Воспользуемся для этого тождеством

$$C^{-1} = V^{-1} - V^{-1}L(\mathbb{1}_k + J)^{-1}L'V^{-1} \tag{15}$$

Чтобы убедиться в его справедливости, умножим правую часть на $C = LL' + V$:

$$\begin{aligned}\mathbb{1}_r &= \mathbb{1}_r + V^{-1}LL' - V^{-1}L(\mathbb{1}_k + J)^{-1}L' - V^{-1}L(\mathbb{1}_k + J)^{-1}L'V^{-1}LL' = \\ &= \mathbb{1}_r + V^{-1}LL' - V^{-1}L\left[(\mathbb{1}_k + J)^{-1} + (\mathbb{1}_k + J)^{-1}J\right]L' = \mathbb{1}_r,\end{aligned}$$

что и требовалось. Умножив далее (15) слева на L', получим:

$$\begin{aligned}L'C^{-1} &= L'V^{-1} - L'V^{-1}L(\mathbb{1}_k + J)^{-1}L'V^{-1} = \\ &= \left[\mathbb{1}_k - J(\mathbb{1}_k + J)^{-1}\right]L'V^{-1} = \\ &= \left[(\mathbb{1}_k + J)(\mathbb{1}_k + J)^{-1} - J(\mathbb{1}_k + J)^{-1}\right]L'V^{-1} = \\ &= (\mathbb{1}_k + J)^{-1}L'V^{-1}\end{aligned} \tag{16}$$

Подставив (16) в (11), придем к уравнению

$$L' = (\mathbb{1}_k + J)^{-1}L'V^{-1}A.$$

Умножив теперь обе части этого равенства слева на $J^{-1}(\mathbb{1}_k + J)$, будем иметь:

$$L' + J^{-1}L' = J^{-1}L'V^{-1}A,$$

или

$$L' = J^{-1}L'V^{-1}(A - V). \tag{17}$$

§ 8.4. Итерационный метод нахождения факторных нагрузок

Уравнение (17) § 8.3 можно записать в виде

$$JL' = L'V^{-1}(A - V). \tag{1}$$

Отсюда следует, что элементы J являются собственными значениями, а строки L' — соответствующими собственными векторами матрицы $V^{-1}(A - V)$. На этом и основан итерационный метод решения уравнения (17) § 8.3.

Этот метод описывается следующим образом. Обозначим через $\underline{l}'_i$ i-ю строку матрицы L' и предположим, что у нас есть начальные приближения $L_{(1)}$ и $V_{(1)}$ для L и V соответственно. Пусть $\underline{l}'_{i(1)}$ обозначает i-ю строку $L'_{(1)}$. Тогда на первом шагу процедуры мы находим

вектор-строку $\underline{w}'_1 = \underline{l}'_{1(1)} V_{(1)}^{-1}$

вектор-строку $\underline{u}'_1 = \underline{w}'_1 A - \underline{l}'_{1(1)}$,

число $h_1 = \underline{u}'_1 \underline{w}_1$

и в качестве 2-го приближения к $\underline{l}'_1$ принимаем вектор-строку

$$\underline{l}'_{1(2)} = \frac{1}{\sqrt{h_1}} \underline{u}'_1.$$

Если у нас более одного фактора ($k \geqslant 2$), мы делаем второй шаг: находим последовательно

вектор-строку $\underline{w}'_2 = \underline{l}'_{2(1)} V_{(1)}^{-1}$

число $j_{21} = \underline{w}'_2 \underline{l}_{1(2)}$,

вектор-строку $\underline{u}'_2 = \underline{w}'_2 A - \underline{l}'_{2(1)} - j_{21} \underline{l}'_{1(2)}$,

число $h_2 = \underline{u}'_2 \underline{w}_2$

и в качестве второго приближения к $\underline{l}'_2$ принимаем вектор-строку

$$\underline{l}'_{2(2)} = \frac{1}{\sqrt{h_2}} \underline{u}'_2.$$

Для третьего фактора ($k \geqslant 3$) последовательно находим:

вектор-строку $\underline{w}'_3 = \underline{l}'_{3(1)} V_{(1)}^{-1}$

два числа $j_{31} = \underline{w}'_3 \underline{l}_{1(2)}$ и $j_{32} = \underline{w}'_3 \underline{l}_{2(2)}$,

вектор-строку $\underline{u}'_3 = \underline{w}'_3 A - \underline{l}'_{3(1)} - j_{31} \underline{l}'_{1(2)} - j_{32} \underline{l}'_{2(2)}$,

число $h_3 = \underline{u}'_3 \underline{w}_3$

и в качестве второго приближения к $\underline{l}'_3$ принимаем вектор-строку

$$\underline{l}'_{3(2)} = \frac{1}{\sqrt{h_3}} \underline{u}'_3.$$

и т. д. Если у нас k факторов, то делаем k таких шагов. Таким образом, в конце этого цикла мы получим второе приближение $L_{(2)}$ для L. Матрицу $V_{(2)}$ мы находим, подставляя элементы матрицы $L_{(2)}$ в уравнение (14) § 8.3.

Матрицы $L_{(2)}$ и $V_{(2)}$ можно использовать как начальные в новом цикле и получить следующие приближения $L_{(3)}$ и $V_{(3)}$ и т. д. Установить условия сходимости описанного итерационного процесса не представляется возможным, однако, на практике обычно процесс сходится и притом довольно быстро (хотя известны примеры, когда сходимость очень медленная или вообще отсутствует — последнее происходит, как правило, из-за плохого выбора начального приближения). Вопроса о выборе начального приближения мы еще коснемся несколько позже (в § 8.6).

§ 8.5. Проверка гипотезы о числе факторов

В предыдущей теории предполагалось, что нам точно известно число k общих факторов. Однако, на практике это число обычно неизвестно и всякое предположение о его значении должно быть подвергнуто проверке как статистическая гипотеза.

Итак, предположил, что мы из некоторых соображений приняли, что число общих факторов есть некоторое фиксированное число k (например, $k = 1$). Обозначим эту гипотезу H_0. Другими словами, H_0 есть предположение, что ковариационная матрица C (размером $r \times r$) может быть представлена в виде суммы диагональной матрицы V с положительными элементами и матрицы LL' ранга k (см. (1) § 8.3). Для дальнейшего потребуем, чтобы выполнялось соотношение

$$(r-k)^2 > r+k. \tag{1}$$

Общая же гипотеза H, внутри которой мы должны проверить H_0, не делает никаких предположений о матрице C (кроме, конечно, положительной определенности). Для проверки H_0 внутри H мы применим λ-метод (метод отношения правдоподобия, см. § 5.4).

Согласно λ-методу мы должны вычислить статистику λ_n, равную отношению максимума функции правдоподобия (2) § 8.3 по множеству параметров Θ_0, составляющему гипотезу H_0, к максимуму этой же функции по множеству параметров Θ, составляющему гипотезу H Если при этом объем выборки $n \to \infty$, то при гипотезе H_0

$$\mathcal{L}(-2\ln\lambda_n) \longrightarrow \chi^2(t),$$

где $t = \dim\Theta - \dim\Theta_0$ (см. п. 4 § 5.4).

В нашем случае максимум функции правдоподобия при гипотезе H_0 достигается при $\widehat{C} = \widehat{L}\widehat{L}' + \widehat{V}$, где оценки $\widehat{L}$ и $\widehat{V}$ определяются уравнениями (17) и (14) § 8.3. При гипотезе же H, очевидно, наилучшей оценкой для C является выборочная матрица ковариаций A, поэтому из (2) § 8.3 следует, что

$$\lambda_n = \left(\frac{|\widehat{C}|}{|A|}\right)^{-n/2} \exp\left\{\frac{rn}{2} - \frac{n}{2}\operatorname{tr}(A\widehat{C}^{-1})\right\},$$

или

$$-2\ln\lambda_n = n\left(\ln\frac{|\widehat{C}|}{|A|} + \operatorname{tr}(A\widehat{C}^{-1}) - r\right). \tag{2}$$

Вычислим теперь число степеней свободы t в предельном законе χ^2 Число независимых параметров (элементов матрицы C) при гипотезе H равно, очевидно, $r(r+1)/2$, следовательно,

$$\dim\Theta = \frac{r(r+1)}{2}.$$

Далее, число всех параметров в представлении (1′) § 8.3 равно $r + rk$, однако, вследствие неопределенности, возникающей из возможности вращения факторов, мы в § 8.3 накладывали еще условие (10′), фиксирующее однозначно матрицу L. Это условие накладывает еще $k(k-1)/2$ ограничений на элементы L и V, и потому действительное число независимых параметров при гипотезе H_0 равно

$$\dim \Theta_0 = r + rk - \frac{k(k-1)}{2}.$$

Таким образом, число степеней свободы

$$t = \frac{r(r+1)}{2} - r(k+1) + \frac{k(k-1)}{2} = \frac{1}{2}[(r-k)^2 - (r+k)]. \qquad (3)$$

В силу предположения (1) число t положительно.

Итак, если объем выборки n велик, то статистика (2) распределена приближенно по закону $\chi^2(t)$, где t дано в (3).

Этот результат позволяет стандартным образом организовать проверку согласия экспериментальных данных и гипотезы H_0 о наличии точно k общих факторов.

Критерий для гипотезы о числе факторов

Проверка гипотезы осуществляется следующим образом. Задаваясь вероятностью α отклонить гипотезу H_0, когда она верна (уровень значимости критерия), по таблицам распределения $\chi^2(t)$, где t дано в (3), определяем квантиль $\chi^2_{1-\alpha,t}$ из условия

$$\mathbf{P}\{\chi^2(t) > \chi^2_{1-\alpha,t}\} = \alpha.$$

Далее, для данного k методом § 8.3 вычисляем $\widehat{C} = \widehat{L}\widehat{L}' + \widehat{V}$ и определяем значение статистики $-2\ln\lambda_n$ в (2). Если это значение оказалось больше, чем $\chi^2_{1-\alpha,t}$, то мы отвергаем гипотезу H_0 и говорим, что в факторной модели необходимо не менее $k+1$ факторов. Если же $-2\ln\lambda_n \leqslant \chi^2_{1-\alpha,t}$ то, как обычно, можно лишь сказать, что опыт не опровергает гипотезу.

Практические рекомендации

На практике число факторов и оценки $\widehat{L}$ и $\widehat{V}$ находят так. Из априорных соображений определяют предполагаемое число факторов k и для этого числа методом § 8.3 определяют $\widehat{L}$ и $\widehat{V}$. После этого проверяют гипотезу о наличии ровно k факторов. Если гипотеза отвергается, то заново вычисляют $\widehat{L}$ и $\widehat{V}$ уже для $k+1$ фактора и снова проверяют гипотезу о наличии $k+1$ фактора и т. д., пока не получим впервые $-2\ln\lambda_n < \chi^2_{1-\alpha,t}$. Параметры $\widehat{L}$, $\widehat{V}$ и число факторов в момент остановки и принимаются за параметры нашей факторной модели (2) § 8.1.

В заключение приведем некоторые соображения, несколько улучшающие описанный критерий и упрощающие его практическое применение.

Прежде всего рекомендуется заменить множитель n в (2) на

$$n' = n - \frac{1}{6}(2r+5) - \frac{2}{3}k. \qquad (4)$$

Было замечено, что при этом распределение статистики (2) более точно аппроксимируется χ^2-распределением.

Далее, если бы уравнения для оценок были решены *точно*, то тогда согласно (12) § 8.3

$$\operatorname{tr}(A\widehat{C}^{-1}) = \operatorname{tr}\mathbb{1}_r = r$$

и, следовательно, статистику (2) можно было бы заменить на

$$n' \ln \frac{|\widehat{C}|}{|A|} = -n' \ln |\widehat{C}^{-1}A|. \tag{5}$$

Используя теперь (11) и (15) § 8.3, можно получить, что

$$\widehat{C}^{-1}A = \mathbb{1}_r + \widehat{C}^{-1}(A - \widehat{C}) = \mathbb{1}_r + \widehat{V}^{-1}(A - \widehat{C}).$$

Действительно, поскольку из (16) и (11) § 8.3 («крышки» для краткости опускаем)

$$V^{-1}L(\mathbb{1}_r + J)^{-1}L'V^{-1}A = V^{-1}L(\mathbb{1}_r + J)^{-1}L'C,$$

то в силу (15) § 8.3 имеем:

$$(V^{-1} - C^{-1})A = (V^{-1} - C^{-1})C,$$

или

$$V^{-1}(A - C) = C^{-1}(A - C).$$

Обозначив через

$$\|t_{ij}\| = T = \widehat{V}^{-1}(A - \widehat{C}), \tag{6}$$

статистику (5) можно представить теперь в виде

$$-n' \ln |\mathbb{1}_r + T|. \tag{7}$$

Поскольку матрица $\widehat{V}^{-1}$ диагональная, а диагональные элементы матрицы $A - \widehat{C}$ согласно (13) § 8.3 равны нулю, то диагональные элементы матрицы T также равны нулю. Далее, при больших n разности $a_{ij} - \widehat{c}_{ij}$, как правило, невелики, и потому в разложении определителя $|\mathbb{1}_r + T|$ можно опустить члены, содержащие произведения более двух этих разностей. С учетом этих соображений мы получаем следующее приближенное равенство

$$|\mathbb{1}_r + T| \approx 1 - \sum_{1 \leqslant i < j \leqslant r} t_{ij}t_{ji} = 1 - \sum_{1 \leqslant i < j \leqslant r} \frac{(a_{ij} - \widehat{c}_{ij})^2}{\widehat{v}_i\widehat{v}_j}.$$

Наконец, статистику (7) можно приближенно заменить статистикой

$$n' \sum_{1 \leqslant i < j \leqslant r} \frac{(a_{ij} - \widehat{c}_{ij})^2}{\widehat{v}_i\widehat{v}_j}. \tag{8}$$

Статистикой (8), вместо (2), обычно пользуются на практике. Вычисление выражения (8) неизмеримо проще, чем вычисление (2), и использование статистики (8), как правило, приводит к хорошей аппроксимации, даже когда сделано сравнительно мало итераций и точные о. м. п. $\widehat{L}$ и $\widehat{V}$ еще не найдены.

§ 8.6. Центроидный метод

В этом параграфе мы опишем один из приближенных методов определения факторных нагрузок, который часто используется на практике. Этот метод по существу своему является итерационным и носит название *центроидного метода*. Оценки, получаемые этим методом, обычно бывают близки к о. м. п. и достаточны для многих практических целей. Но как правило, рекомендуется использовать эти оценки в качестве начального приближения в методе максимального правдоподобия (см. § 8.4).

Факторные дисперсии

Введем понятие факторных дисперсий. Назовем *факторными дисперсиями* диагональные элементы матрицы LL' Следовательно, факторными дисперсиями с. в. $x_1, \quad , x_r$ являются те части общих дисперсий, которые вносят факторы $f_1, \quad , f_k$. Из (1) § 8.3 следует, что факторной дисперсией с. в. x_i является число $l_{i1}^2 + \quad + l_{ik}^2$.

Дадим описание центроидного метода, одновременно иллюстрируя его на примере. В этом методе обычно рассматривают выборочную корреляционную матрицу. Перед началом анализа надо знать начальные значения факторных дисперсий. В качестве таковых выбираются максимальные по модулю значения элементов в столбцах корреляционной матрицы (без учета диагональных) и ими заменяют единицы на диагонали этой матрицы. Так получаемую матрицу обозначим A_0.

В качестве иллюстративного примера мы будем рассматривать следующую корреляционную матрицу (для $r = 6$), в которой уже проделана отмеченная выше первоначальная процедура

	1	2	3	4	5	6
1	0,439	0,439	0,410	0,288	0,329	0,248
2	0,439	0,439	0,351	0,354	0,320	0,329
3	0,410	0,351	0,410	0,164	0,190	0,181
4	0,288	0,354	0,164	0,595	0,595	0,470
5	0,329	0,320	0,190	0,595	0,595	0,464
6	0,248	0,329	0,181	0,470	0,464	0,470

В данном случае все корреляции положительны, но если бы это было не так, то прежде чем действовать дальше, мы должны были бы изменением знака у одной или более переменных $x_1, \quad , x_r$ постараться минимизировать число отрицательных корреляций (при изменении знака у x_i изменяются знаки у элементов i-й строки и i-го столбца корреляционной матрицы; при

этом диагональный элемент остается неизменным). Этим мы достигаем того, чтобы суммы по столбцам были бы как можно больше.

После этой операции с изменением знаков (в нашем примере этого делать не надо) вычисляют суммы по столбцам и полную сумму всех элементов матрицы A_0. Числа, получаемые делением сумм в соответствующих столбцах на квадратный корень из полной суммы, и являются первыми оценками для нагрузок на первый фактор.

При этом, если бы мы меняли знаки у некоторых переменных, то у соответствующих нагрузок мы должны были бы изменить знаки на обратные. В примере

Суммы	2,153	2,232	1,706	2,466	2,493	2,162	$\Sigma = 13{,}212$
Нагрузки на 1-й фактор	0,592	0,614	0,469	0,678	0,686	0,595	$\sqrt{\Sigma} = 3{,}635$

Обозначим через $\underline{l}'_1 = (l_{11}, l_{21}, \quad , l_{r1})$ так полученную строку нагрузок и вычислим первую остаточную матрицу $A_0 - \underline{l}_1\underline{l}'_1$ (исключим влияние первого фактора из первоначальной матрицы). В примере $A_0 - \underline{l}_1\underline{l}'_1$ имеет следующий вид:

	1	2	3	4	5	6
1	0,086	0,076	0,132	−0,113	−0,077	−0,104
2	0,076	0,062	0,063	−0,062	−0,101	−0,036
3	0,132	0,063	0,190	−0,154	−0,132	−0,098
4	−0,113	−0,062	−0,154	0,135	0,130	0,067
5	−0,077	−0,101	−0,132	0,130	0,124	0,056
6	−0,104	−0,036	−0,098	0,067	0,056	0,116

В матрице $A_0 - \underline{l}_1\underline{l}'_1$ уже обязательно будут отрицательные элементы, поскольку суммы по столбцам в ней всегда равны нулю. Поэтому перед следующим этапом в $A_0 - \underline{l}_1\underline{l}'_1$ надо проделать описанную выше процедуру с изменением знаков как минимум у одной переменной (минимизируя число отрицательных знаков). В нашем примере число отрицательных знаков в $A_0 - \underline{l}_1\underline{l}'_1$ можно сделать нулем, если изменить знак у переменных x_4, x_5 и x_6. Затем снова в диагональные клетки $A_0 - \underline{l}_1\underline{l}'_1$ ставим максимальные по модулю значения элементов соответствующих столбцов (но это лишь на первом цикле итерационного процесса; на последующих же циклах замены диагональных элементов не производятся). Так получаемую матрицу обозначим A_1. В примере A_1 имеет вид:

0,132	0,076	0,132	0,113	0,077	0,104
0,076	0,101	0,063	0,062	0,101	0,036
0,132	0,063	0,190	0,154	0,132	0,098
0,113	0,062	0,154	0,154	0,130	0,067
0,077	0,101	0,132	0,130	0,132	0,056
0,104	0,036	0,098	0,067	0,056	0,116

Далее с матрицей A_1 проделывается то же, что и с матрицей A_0: находятся суммы по столбцам и полная сумма, извлекается квадратный корень из полной суммы и на это число делятся суммы в столбцах. Получаемые числа (с учетом восстановления знаков) и есть первые оценки нагрузок на второй фактор:

$$\underline{l}_2' = (l_{12}, l_{22}, \ldots, l_{r2}).$$

В примере мы должны взять нагрузки для x_4, x_5, x_6 со знаком минус.

Суммы	0,634	0,439	0,769	0,680	0,628	0,477	$\Sigma = 3{,}627$
Нагрузки на 2-й фактор	0,333	0,231	0,404	−0,357	−0,330	−0,251	$\sqrt{\Sigma} = 1{,}904$

После этого рассматривается матрица $A_1 - \underline{l}_2\underline{l}_2'$, с ней проделывается та же процедура, что и с матрицей $A_0 - \underline{l}_1\underline{l}_1'$ и т. д. пока не исчерпаем всех факторов и не получим полностью матрицу $L = \|l_{ij}\|$ $(i = 1, \quad, r;\ j = 1, \quad, k)$. Здесь так же, как и в методе максимального правдоподобия (§ 8.5), действует правило, ограничивающее число факторов k условием $r + k < (r - k)^2$ Так, в нашем примере мы должны ограничиться вторым этапом, так как значение $k = 3$ уже не удовлетворяет этому условию.

Суммируя теперь по строкам квадраты элементов этой матрицы:

$$d_i = \sum_{j=1}^{k} l_{ij}^2,$$

получим новые оценки для факторных дисперсий. На этом первый цикл итерационного процесса заканчивается. Только нужно у переменных, менявших знак, вернуться к первоначальным знакам. Полученными факторными дисперсиями можно воспользоваться для следующей итерации: в i-ю диагональную клетку исходной выборочной корреляционной матрицы надо подставить полученное число d_i, $i = 1, \quad, k$, и повторить тот же процесс и т. д. Сходимость на практике описанного процесса очень быстрая.

Для нашего примера первые оценки матрицы нагрузок и факторные дисперсии имеют вид

$$L = \begin{Vmatrix} 0{,}592 & 0{,}333 \\ 0{,}614 & 0{,}231 \\ 0{,}469 & 0{,}404 \\ 0{,}678 & -0{,}357 \\ 0{,}686 & -0{,}330 \\ 0{,}595 & -0{,}251 \end{Vmatrix} \qquad \begin{aligned} d_1 &= 0{,}461 \\ d_2 &= 0{,}430 \\ d_3 &= 0{,}383 \\ d_4 &= 0{,}587 \\ d_5 &= 0{,}580 \\ d_6 &= 0{,}417 \end{aligned}$$

Геометрическая интерпретация

Описанная процедура отыскания факторных нагрузок имеет наглядную геометрическую интерпретацию. Отождествим наши наблюдаемые переменные $x_1, \ldots, x_r$ с векторами, выходящими из начала координат r-мерного пространства, косинусы углов между которыми равны соответствующим корреляциям, а длины — соответствующим стандартным отклонениям. Далее, если направления, которые приписаны переменным, уже найдены, то мы изменением, если необходимо, их знаков сделаем как можно больше корреляций положительными; тогда векторы будут иметь тенденцию к группировке в одном направлении в пучок и их сумма будет иметь наибольшую длину. Эта сумма и принимается за первый фактор, а в качестве нагрузок переменных $x_1, \ldots, x_r$ на этот фактор принимаются длины проекций соответствующих векторов на суммарный вектор. После этого исключается влияние первого фактора (как это описано выше) и для первой остаточной матрицы строится новая геометрическая модель и т. д.

Критерий χ^2

Для центроидного метода, так же как и для метода максимального правдоподобия, существует критерий типа «χ^2 для больших выборок» для проверки гипотезы о наличии ровно k факторов, т. е. *критерий значимости остатков после исключения k факторов*. Приближенная статистика этого критерия, которая обычно используется на практике, имеет вид

$$n\left(\frac{1}{2}\sum_i x_{ii}^2 + \sum_{i<j} x_{ij}x_{ji}\right),$$

где $\|x_{ij}\| = X = V^{-1}(A_0 - LJ^{-1}L'V^{-1}A_0)$. Здесь $A_0 = A - V$, $J = L'V^{-1}L$ (вообще говоря, не диагональная). Число степеней свободы в предельном распределении χ^2 здесь равно $1/2\left((r-k)^2 - r - k\right)$ Эта гипотеза проверяется точно так же, как и в случае получения оценок методом максимального правдоподобия.

Общая схема факторного анализа

Общая схема применения факторного анализа на практике имеет следующий вид. Сначала с помощью центроидного метода получают первоначальные факторные нагрузки и факторные дисперсии. С помощью критерия для центроидного метода уточняют необходимое число факторов. После этого центроидные нагрузки используют в методе максимального правдоподобия

в качестве первого приближения. Вообще, методу максимального правдоподобия следует отдать предпочтение, так как в настоящее время он является единственным методом, дающим эффективные оценки факторных нагрузок.

§ 8.7. Оценка значений факторов

Решение вопроса о числе общих факторов и о получении нагрузок на факторы по переменным составляет основу факторного анализа. Однако на практике часто представляет интерес и задача приближенной оценки значений самих гипотетических факторов по наблюденным значениям переменных $x_1, \quad , x_r$, после того, как каким-либо методом уже решена первая задача. Здесь мы рассмотрим один из простых методов решения этой задачи.

В этом методе оценки $\widehat{f}_1, \quad , \widehat{f}_k$ факторов находятся из условия обращения в минимум величины

$$\sum_{i=1}^{r} \frac{e_i^2}{v_i},$$

которую в силу (2) § 8.1 можно записать также в виде

$$\sum_{i=1}^{r} \left(x_i - \sum_{j=1}^{k} l_{ij} f_j \right)^2 \frac{1}{v_i}. \tag{1}$$

Метод минимизации остатков

Таким образом, здесь минимизируется взвешенная сумма квадратов остатков в модели (2) § 8.1, поэтому метод называется *методом минимизации остатков*.

Дифференцируя (1) по f_j и приравнивая частные производные нулю, получаем следующую систему уравнений:

$$\sum_{i=1}^{r} \sum_{s=1}^{k} \frac{l_{is} l_{ij}}{v_i} f_s = \sum_{i=1}^{r} \frac{l_{ij}}{v_i} x_i, \quad j = 1, \quad , k,$$

или в матричной форме

$$L' V^{-1} L \underline{f} = L' V^{-1} \underline{x}. \tag{2}$$

Полагая

$$J = L' V^{-1} L,$$

находим, что оценки факторов по данному методу равны

$$\widehat{\underline{f}} = J^{-1} L' V^{-1} \underline{x}. \tag{3}$$

Таким образом, оценки факторов являются линейными функциями от наблюдаемых переменных.

Интерпретация факта может оказаться важнее самого факта.



Заключительные замечания. В заключение отметим, что в факторном анализе в силу неоднозначности решения (см. § 8.2) весьма существенную роль играет проблема интерпретации результатов исследования (бывает: *мысль, достойная оваций, погибнуть может от интерпретаций*). В качестве иллюстрации этого приведем результаты одного конкретного исследования в области социологии [1], описанные в книге [18, с. 98–99]. Известный специалист в области факторного анализа Кэттел исследовал показатели, характеризующие уровень культурного развития наций. Было выбрано 72 количественно измеряемых показателя для анализа культурного уровня 69 наций. В их число входили все показатели, которые, по мнению автора, характеризуют культуру нации. Среди них оказались число лауреатов Нобелевской премии, число зарегистрированных проституток, число секретных договоров, число революций за 100 лет, минимальный возраст вступления в брак, потребление сахара на душу населения, процент негров, отношение разводов к бракам, процент самоубийств и т. д. Анализу было подвергнуто 2556 корреляционных связей. Окончательные результаты были представлены 12 факторами. Автор пытался дать осмысленную интерпретацию каждому фактору, рассматривая его нагруженность исходными переменными. При этом использовались весьма вычурные формулировки для обозначения факторов, например: «просвещенное изобилие в противовес удручающей бедности», «осмысленное трудолюбие в противовес эмоциональности», «консервативность и патриархальная стабильность в противовес безрассудной богеме» и т. д.

«Результаты такого исследования, может быть, и любопытны в каком-то смысле, но нам кажется, что они все же не имеют большого эвристического значения, — вряд ли, исходя из подобных исследований, можно выдвинуть содержательные гипотезы о развитии культуры. Нам представляется неудачной сама попытка создания модели, всесторонне и полно описывающей такую сложную и плохо организованную систему. Видимо, факторный анализ имеет смысл применять к всестороннему описанию более простых систем, где легче выдвигать гипотезы, подлежащие дальнейшему обсуждению» [18]. И еще: «Если говорить о критической оценке факторного анализа, то следует вспомнить довольно старое, очень резкое, но, вероятно, не совсем справедливое замечание о том, что, пользуясь этим методом, получишь то, что туда внесешь». Но, хотя этот метод и не дает однозначного решения, он все же помогает исследователю формулировать реалистичные гипотезы об исследуемой системе, обостряет его интуицию, упрощает итоговый алгоритм проведения эксперимента, помогая отсеивать множество «шумовых» показателей и выделять наиболее существенные характеристики изучаемого объекта или процесса. В качестве литературы по теории и приложениям факторного анализа можно указать книги [17, 19].

Писанье книг — занятье без конца, их чтенье — утомительно для тела.

[1] *Cattell R. B.* The Dimentions of Cultural Patterns by Factorization of National Characters // J. Abnor. and Social Psychology. 1949. V. 44. P. 443–469.

Глава 9

Компонентный анализ

Не дороги нам дни, не жаль нам их нимало,
Чтоб то, чем дорожим, росло и созревало —
Цветок ли пестуем в саду, где сладко дышим,
Ребенка ли растим иль книжечку мы пишем.

Фридрих Рюккерт [1]

§ 9.1. Постановка задач компонентного анализа

Компонентный анализ или, иначе, *метод главных компонент* так же, как и факторный анализ, исследует структуру ковариационных (или корреляционных) матриц систем случайных величин (с. в.). Однако лишь только в этом и состоит их формальное сходство. Если для факторного анализа важны корреляции между переменными, которые он объясняет наличием некоторого числа общих для всех переменных факторов, то в компонентном анализе интересуются исключительно дисперсиями переменных и их линейных комбинаций. При этом в компонентном анализе не делается фактически никаких предположений о наблюдаемых с. в. $x_1, \ldots, x_r$, в то время, как для факторного анализа существенным является линейность и нормальность его модели (см. (2) § 8.1), которая, если и не верна, то хотя бы приближенно отражает суть дела.

В компонентном анализе ищется такое линейное преобразование

$$\begin{aligned} x_1 &= l_{11} f_1 + \quad + l_{1r} f_r, \\ &\qquad\qquad\qquad\qquad\qquad (1) \\ x_r &= l_{r1} f_1 + \quad + l_{rr} f_r, \end{aligned}$$

или в матричной форме

$$\underline{x} = L\underline{f}, \qquad (1')$$

где $\underline{x} = (x_1, \quad, x_r)$, $\underline{f} = (f_1, \quad, f_r)$ — векторы-столбцы и $L = \|l_{ij}\|$ квадратная матрица размером $r \times r$, в котором с. в. $f_1, \quad, f_r$ некоррелированы и нормированы ($\mathbf{E}f_i = 0$, $\mathbf{D}f_i = 1$, $i = 1, \quad, r$; всегда для простоты предполагается, что $\mathbf{E}x_i = 0$, $i = 1, \quad, r$).

Представление исходных с. в. $x_1, \quad, x_r$ в виде (1) имеет ряд преимуществ. Действительно, в этом случае их дисперсии весьма просто выражаются

[1] Рюккерт Фридрих (1788–1866) — немецкий поэт, драматург.

через коэффициенты l_{ij}:

$$\mathbf{D}x_i = l_{i1}^2 + \quad + l_{ir}^2, \quad i = 1, \quad , r. \tag{2}$$

Ввиду (2) суммарная дисперсия с. в. $x_1, \quad , x_r$ равна

$$\sum_{i=1}^{r} \mathbf{D}x_i = \sum_{i=1}^{r} l_{i1}^2 + \ldots + \sum_{i=1}^{r} l_{ir}^2, \tag{3}$$

поэтому величину l_{ij}^2 можно назвать вкладом компоненты f_j в дисперсию переменной x_i, а величину $l_{ij}^2 + \ldots + l_{rj}^2$ — вкладом компоненты f_j в суммарную дисперсию. Далее, предположим, что вклад компоненты f_r в суммарную дисперсию мал. Тогда можно пренебречь влиянием компоненты f_r, и вектор $\underline{x} = (x_1, \quad , x_r)$ характеризовался бы в этом случае не r случайными числами $x_1, \quad , x_r$ а лишь $r - 1$ случайными числами $f_1, \quad , f_{r-1}$. Вообще, если бы суммарный вклад компонент $f_{k+1}, \quad , f_r$ в суммарную дисперсию, равный в силу (2) и (3)

$$\sum_{j=k+1}^{r} \sum_{i=1}^{r} l_{ij}^2,$$

был мал настолько, что им можно было бы пренебречь, то это означало бы, что влиянием компонент $f_{k+1}, \quad , f_r$ можно пренебречь, и следовательно, вектор $\underline{x}$ характеризовался бы не r параметрами, а лишь k параметрами $f_1, \quad , f_k$. Иногда такое исследование помогает лучше понять природу случая, порождающего случайный вектор $\underline{x}$. Кроме этого, при этом удается «сжать» информацию, т. е. достичь экономии в описании изучаемого объекта (описывать его меньшим числом параметров).

Как будет показано ниже, отыскание представления (1) эквивалентно определению r таких нормированных линейных комбинаций $y_1, \quad , y_r$ переменных $x_1, \quad , x_r$ (т. е. сумма квадратов коэффициентов равна 1), что для каждого $k = 1, \quad r$ y_k имеет наибольшую дисперсию среди всех нормированных линейных комбинаций при условии некоррелированности с предыдущими комбинациями $y_1, \quad , y_{k-1}$. Такие линейные комбинации $y_1, \ldots, y_r$ называются *главными компонентами* системы с. в. $x_1, \quad , x_r$.

Главные компоненты

Следовательно, первой главной компонентой переменных $x_1, \quad , x_r$ называется нормированная линейная комбинация этих переменных, обладающая максимальной дисперсией.

Второй главной компонентой называется нормированная линейная комбинация, обладающая максимальной дисперсией при условии некоррелированности с первой главной компонентой.

Третьей главной компонентой называется нормированная линейная комбинация с. в. $x_1, \ . \quad , x_r$, обладающая максимальной дисперсией при условии некоррелированности с первой и второй главными компонентами и т. д.

Для практики обычно наибольший интерес представляют те переменные или их линейные комбинации, которые имеют большие дисперсии, поскольку именно такие комбинации и несут в себе максимум информации о степени изменчивости того или иного показателя или признака. В практических исследованиях, когда число рассматриваемых переменных, которые требуется обработать, слишком велико, стремятся выделить относительно небольшое число их линейных комбинаций с большими дисперсиями, а линейные комбинации с малыми дисперсиями отбрасываются, как несущественные. Таким образом, достигается сокращение числа исследуемых переменных.

Ввиду этого отыскание главных компонент и выделение из них тех, которые удовлетворительно объясняют суммарную дисперсию, представляет большой практический интерес. Метод главных компонент был предложен еще в 1901 г. К. Пирсоном и позднее вновь открыт и детально разработан Хотеллингом [2] (1933 г.). Этот метод хорошо изложен в книге [1].

§ 9.2. Решение основных уравнений. Главные компоненты

В этом параграфе мы получим представление (1) § 9.1 для любой системы с. в. $x_1, \ldots, x_r$. Предположим, что известна ковариационная матрица A этих с. в. (если ковариационная матрица неизвестна, но даны независимые наблюдения над с. в. $\underline{x} = (x_1, \quad, x_r)$, то во всех нижеследующих рассуждениях надо под A понимать выборочную ковариационную матрицу).

Обозначим через U ортогональную матрицу, с помощью которой A приводится к диагональной матрице

$$\Lambda = \begin{Vmatrix} \lambda_1 & & 0 \\ & & \\ 0 & & \lambda_r \end{Vmatrix}, \tag{1}$$

в которой элементы расположены в порядке убывания, т. е. $\lambda_1 > \lambda_2 > \quad > \lambda > 0$ (мы предполагаем для простоты, что все λ_i различны и положительны, что для большинства практических ситуаций обычно выполняется). Таким образом,

$$\Lambda = U'AU. \tag{2}$$

Фактически λ_k есть k-е по величине собственное значение матрицы A, а k-й столбец $\underline{u}_k$ матрицы U — соответствующий собственный вектор, так что

$$A\underline{u}_k = \lambda_k \underline{u}_k, \quad k = 1, \quad, r, \tag{3}$$

что является эквивалентной записью соотношения (2).

Рассмотрим теперь новые переменные $\underline{y} = (y_1, \quad, y_r)$, задаваемые равенством

$$\underline{y} = U'\underline{x}. \tag{4}$$

[2] Хотеллинг Гаральд (1895–1973) — американский статистик и экономист.

Вычислим их ковариационную матрицу B. Поскольку $\mathbf{E}x_i = 0$, то $\mathbf{E}y_1 = \quad = \mathbf{E}y_r = 0$ и потому

$$B = \mathbf{E}(\underline{y}\,\underline{y}') = \mathbf{E}(U'\underline{x}\,\underline{x}'U) = U'\mathbf{E}(\underline{x}\,\underline{x}')U = U'AU = \Lambda.$$

Из этого соотношения следует, что с. в. $y_1, \quad , y_r$ некоррелированы и $\mathbf{D}y_k = \lambda_k$, $k = 1, \quad , r$.

Покажем теперь, что y_1 есть нормированная комбинация с наибольшей дисперсией. Пусть

$$y^* = c_1x_1 + \quad + c_rx_r = \underline{c}'\underline{x}, \quad \underline{c} = (c_1, \quad , c_r), \quad \underline{c}'\underline{c} = c_1^2 + \quad + c_r^2 = 1,$$

есть произвольная нормированная линейная комбинация x-ов. Тогда из (4) следует, что

$$y^* = \underline{c}'U\underline{y} = \underline{c}^{*\prime}\underline{y} = c_1^*y_1 + \quad + c_r^*y_r, \quad \underline{c}^* = U'\underline{c}.$$

При этом $\underline{c}^{*\prime}\underline{c}^* = 1$, поскольку при ортогональном преобразовании длина вектора не меняется. Отсюда

$$\mathbf{D}y^* = \sum_{i=1}^{r} c_i^{*2}\lambda_i = \lambda_1 + \sum_{i=2}^{r} c_i^{*2}(\lambda_i - \lambda_1),$$

поскольку

$$c_1^{*2} = 1 - \sum_{i=2}^{r} c_i^{*2}$$

Это выражение, очевидно, достигает максимума при $c_2^* = \quad = c_r^* = 0$, так как $\lambda_1 > \lambda_i$ при $i \geqslant 2$.

Аналогично, y_2 представляет собой нормированную линейную комбинацию, некоррелированную с y_1 и имеющую наибольшую дисперсию (из того, что $y^* = \sum c_i^*y_i$ некоррелирована с y_1, следует $c_1^* = 0$, а далее применяются предыдущие рассуждения) и т. д.

Таким образом,

$$y_1 = \underline{u}_1'\underline{x} = u_{11}x_1 + u_{21}x_2 + \quad + u_{r1}x_r$$

есть первая главная компонента переменных $x_1, \quad , x_r$ и ее дисперсия равна первому собственному значению λ_1 (здесь и далее $U = \|u_{ij}\|$ и $\underline{u}_k'$ – k-я строка матрицы U'),

$$y_2 = \underline{u}_2'\underline{x} = u_{12}x_1 + u_{22}x_2 + \quad + u_{r2}x_r$$

есть вторая главная компонента и ее дисперсия равна второму собственному значению λ_2 и т. д. Вообще

$$y_k = \underline{u}_k'\underline{x} = u_{1k}x_1 + u_{2k}x_2 + \quad + u_{rk}x_r \tag{5}$$

есть k-я главная компонента и ее дисперсия равна k-му по величине собственному значению λ_k.

Главные компоненты (вид)

Итак, нахождение главных компонент и их дисперсий сводится, по существу, к нахождению собственных значений и собственных векторов ковариационной матрицы A. При этом k-я главная компонента имеет дисперсию, равную k-му по величине собственному значению, а коэффициенты ее линейной комбинации являются компонентами соответствующего собственного вектора A.

Чтобы получить теперь представление (1) § 9.1, достаточно пронормировать с. в. $y_1, \quad , y_r$ в (4).

Положим

$$f_k = \frac{1}{\sqrt{\lambda_k}} y_k = \frac{1}{\sqrt{\lambda_k}} \underline{u}_k' \underline{x}, \quad k = 1, \quad r \tag{6}$$

Тогда с. в. $f_1, \quad , f_r$ будут некоррелированы и нормированы ($\mathbf{D} f_k = 1$). В матричной записи (6) имеет вид

$$\underline{f} = \Lambda^{-1/2} \underline{y} = \Lambda^{-1/2} U' \underline{x}. \tag{7}$$

Это равенство выражает $\underline{f}$ через $\underline{x}$. Чтобы выразить $\underline{x}$ через $\underline{f}$, умножим (7) слева на $U\Lambda^{1/2}$ Поскольку $UU' = \mathbb{1}_r$, то это дает

$$\underline{x} = U\Lambda^{1/2} \underline{f} \tag{8}$$

Сравнивая (8) с (1) § 9.1, получаем, что

$$L = U\Lambda^{1/2} \tag{9}$$

Тем самым получено представление (1) § 9.1.

Итак, с. в. $f_1, \quad , f_r$ в (1) § 9.1 — это суть нормированные главные компоненты переменных $x_1, \quad , x_r$. При этом k-й столбец $\underline{l}_k$ матрицы L равен

$$\underline{l}_k = \sqrt{\lambda_k} \underline{u}_k$$

и при этом

$$\underline{l}_k' \underline{l}_k = \lambda_k \underline{u}_k' \underline{u}_k = \lambda_k = l_{1k}^2 + \quad + l_{rk}^2.$$

Таким образом, вклад компоненты f_k в суммарную дисперсию с. в. $x_1, \quad , x_r$ равен как раз λ_k. Следовательно, зная собственные значения матрицы A, мы легко можем найти и долю, вносимую каждой компонентой в суммарную дисперсию, равную в силу (3) § 9.1 $\lambda_1 + \quad + \lambda_r = \mathrm{tr}(\Lambda) = \mathrm{tr}(A)$.

§ 9.3. Метод Хотеллинга

Как следует из § 9.2 в компонентном анализе основная задача состоит в вычислении собственных значений и собственных векторов ковариационной матрицы A. В настоящее время для этой цели используются ЭВМ. Имеется несколько способов решения этой задачи. Мы приведем здесь итерационный метод, разработанный Хотеллингом.

Пусть $\underline{a}_{(0)}$ — любой вектор, не ортогональный к первому собственному вектору $\underline{u}_1$ (обычно в качестве компонент начального вектора $\underline{a}_{(0)}$ рекомендуется брать числа, приблизительно пропорциональные суммам элементов соответствующих столбцов матрицы A). Определяем последовательно

$$\underline{b}_i = \frac{1}{\sqrt{\underline{a}'_{(i)}\underline{a}_{(i)}}}\underline{a}_{(i)}, \quad i = 0, 1, 2, \ldots \tag{1}$$

$$\underline{a}_{(i)} = A\underline{b}_{(i-1)}, \quad i = 1, 2, \tag{2}$$

Тогда имеют место следующие соотношения:

$$\lim_{i\to\infty} \underline{a}'_{(i)}\underline{a}_{(i)} = \lambda_1^2, \tag{3}$$

$$\lim_{i\to\infty} \underline{b}_{(i)} = \pm\underline{u}_1. \tag{4}$$

Докажем (3) и (4). Поскольку $A = U\Lambda U'$, то $A^i = U\Lambda^i U'$ Полагая далее $s_i = \left(\underline{a}'_{(i)}\underline{a}_{(i)}\right)^{-1/2}$, из (1) и (2) получим, что

$$\underline{b}_{(i)} = s_i A\underline{b}_{(i-1)}. \tag{5}$$

Применяя последовательно несколько раз (5), найдем, что

$$\underline{b}_{(i)} = t_i A^i \underline{a}_{(0)} = t_i U\Lambda^i U'\underline{a}_{(0)}, \tag{6}$$

где $t_i = s_0 s_1 \ldots s_i$. Из (1) и (6) следует, что

$$1 = \underline{b}'_{(i)}\underline{b}_{(i)} = t_i^2 \underline{a}'_{(0)} U\Lambda^{2i}U'\underline{a}_{(0)}. \tag{7}$$

Соотношение (6) можно записать также в виде

$$\underline{b}_{(i)} = t_i\lambda_1^i U\left(\frac{1}{\lambda_1}\Lambda\right)^i U'\underline{a}_{(0)}. \tag{8}$$

Но

$$\lim_{i\to\infty}\left(\frac{1}{\lambda_1}\Lambda\right)^i = \lim_{i\to\infty}\begin{Vmatrix} 1 & & 0 \\ & \left(\frac{\lambda_2}{\lambda_1}\right)^i & \\ & & \\ 0 & & \left(\frac{\lambda_r}{\lambda_1}\right)^i \end{Vmatrix} = \begin{Vmatrix} 1 & & 0 \\ & 0 & \\ 0 & & 0 \end{Vmatrix} = C,$$

так как $\lambda_j/\lambda_1 < 1$ при $j > 1$. Таким образом,

$$\lim_{i\to\infty} U\left(\frac{1}{\lambda_1}\Lambda\right)^i U'\underline{a}_{(0)} = UCU'\underline{a}_{(0)} = \|\underline{u}_1\underline{0}\ldots\underline{0}\|U'\underline{a}_{(0)} = \underline{u}_1\underline{u}'_1\underline{a}_{(0)} = \left(\underline{u}'_1\underline{a}_{(0)}\right)\underline{u}_1 \tag{9}$$

(здесь символом $\underline{0}$ обозначается нулевой столбец матрицы).

С учетом (9) из (7) имеем:

$$1 = \lim_{i\to\infty}\left(t_i\lambda_1^i\right)^2 \lim_{i\to\infty}\underline{a}'_{(0)}U\left(\frac{1}{\lambda_1}\Lambda\right)^{2i}U'\underline{a}_{(0)} = \left(\underline{u}'_1\underline{a}_{(0)}\right)^2\lim_{i\to\infty}\left(t_i\lambda_1^i\right)^2$$

Таким образом,

$$\lim_{i\to\infty}\left(t_i\lambda_1^i\right)^2 = \frac{1}{\left(\underline{u}'_1\underline{a}_{(0)}\right)^2}, \tag{10}$$

поскольку, по предположению, $\underline{u}'_1\underline{a}_{(0)} \neq 0$. Из (8) с учетом (9) и (10) следует, что

$$\lim_{i\to\infty}\underline{b}_i = \pm\frac{1}{\underline{u}'_1\underline{a}_{(0)}}\left(\underline{u}'_1\underline{a}_{(0)}\right)\underline{u}_1 = \pm\underline{u}_1.$$

Далее из (2) с учетом этого соотношения и соотношения (3) §9.2 следует, что

$$\lim_{i\to\infty}\underline{a}_{(i)} = \pm A\underline{u}_1 = \pm\lambda_1\underline{u}_1.$$

Отсюда

$$\lim_{i\to\infty}\underline{a}'_{(i)}\underline{a}_{(i)} = \lambda_1^2\underline{u}'_1\underline{u}_1 = \lambda_1^2.$$

Таким образом, соотношения (3) и (4) доказаны.

Чтобы найти второе собственное число и соответствующий собственный вектор, рассмотрим матрицу

$$A_2 = A - \lambda_1\underline{u}_1\underline{u}'_1.$$

Тогда, если $i \neq 1$, то

$$A_2\underline{u}_i = A\underline{u}_i - \lambda_1(\underline{u}'_1\underline{u}_i)\underline{u}_1 = \lambda_i\underline{u}_i,$$

поскольку $\underline{u}'_1\underline{u}_i = 0$, и

$$A_2\underline{u}_1 = A\underline{u}_1 - \lambda_1(\underline{u}'_1\underline{u}_1)\underline{u}_1 = \lambda_1\underline{u}_1 - \lambda_1\underline{u}_1 = 0,$$

поскольку $\underline{u}'_1\underline{u}_1 = 1$.

Таким образом, $\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_r$ являются собственными векторами матрицы A_2, а соответствующие собственные значения равны $0, \lambda_2, \ldots, \lambda_r$. Следовательно, λ_2 — наибольшее собственное значение матрицы A_2, а $\underline{u}_2$ — соответствующий собственный вектор. Для определения $\underline{u}_2$ и λ_2 можно, следовательно, применить к A_2 описанный выше процесс последовательных приближений. После этого можно определить матрицу

$$A_3 = A_2 - \lambda_2\underline{u}_2\underline{u}'_2$$

и методом последовательных приближений определить λ_3 и $\underline{u}_3$ как ее наибольшее собственное значение и соответствующий собственный вектор и т. д.

Практические рекомендации

Имеется несколько способов, позволяющих сократить объем работы в методе последовательных приближений. Один из них состоит в том, что матрицу A возводят

в некоторую степень, а затем уже применяют метод последовательных приближений. Так, можно пользоваться матрицей A^2, положив

$$\underline{a}_{(i)} = A^2\underline{b}_{(i-1)}, \quad \underline{b}_{(i)} = \frac{1}{\sqrt{\underline{a}'_{(i)}\underline{a}_{(i)}}}\underline{a}_{(i)}.$$

Этот процесс сходится вдвое быстрее, чем процесс, определяемый формулами (1) и (2). При использовании $A^4 = A^2A^2$ сходимость получается вчетверо более быстрая и т. д. Отметим при этом, что A^2 — симметрическая матрица, поэтому при ее вычислении достаточно определить лишь $r(r+1)/2$ ее элементов. Для контроля вычислений можно использовать равенства $\mathrm{tr}\,(A) = \mathrm{tr}\,(\Lambda)$ и $A = LL'$, где L определено в (9) § 9.2.

Пример Джефферса



Приведем один конкретный пример применения метода главных компонент. В работе Джефферса [3] этот метод был применен в следующей задаче классификации. Объектом исследования служили летающие тли. Нужно было разбить этот вид насекомых на подгруппы по варьируемости их морфологических признаков. Было измерено 19 различных признаков, которые оказались весьма сильно коррелированными между собой; так коэффициенты парных корреляций по многим признакам достигали значений 0,90 ÷ 0,99. Компонентный анализ показал, что можно ограничиться двумя первыми главными компонентами, на которые приходится 85,5 % от общей дисперсии. Первая компонента задается различием в размерах насекомых, вторая в значительной степени связана с числом яйцекладок.

Графически результаты представлены на рис. 1 (результаты наблюдений представлены на плоскости, задаваемой двумя первыми компонентами). Здесь ясно видно, что обследованных насекомых можно разбить на четыре хорошо различимые группы.

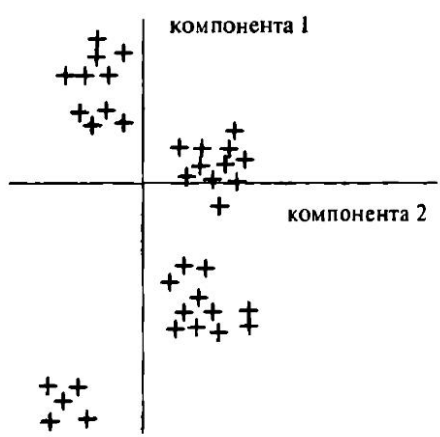


Рис. 1

[3] *Jeffers J. N. R.* Two Case Studies in the Application of Principal Component Analysis // Appl. Stat. 1967 16. № 3. P. 225–236.

Геометрическая интерпретация

Вообще, геометрически нахождение главных компонент сводится к переходу к новой ортогональной системе координат: первая координатная ось ищется так, чтобы соответствующая ей линейная форма извлекала возможно большую дисперсию, далее ищется ортогональная ей ось, которая делает то же самое с оставшейся дисперсией и т. д.

Заключительные замечания

В заключение отметим, что как факторный, так и компонентный анализ — это математические (статистические) методы, основанные на линейных моделях и (в первом случае) нормальном законе распределения, и служащие лишь средствами, облегчающими исследователю проникновение в тайны природы. Природа же надежно охраняет свои тайны:

> *Природа с красоты своей*
> *Покрова снять не позволяет,*
> ***И ты машинами не вынудишь у ней,***
> ***Чего твой дух не угадает.***
>
> **В. С. Соловьев** [4]
> **(Перевод из трагедии Гёте «Фауст»)**

Чтобы проникнуть в них, надо или ставить активный эксперимент, что всегда трудно, или научиться их отгадывать. Факторный анализ и метод главных компонент и служат этой цели. Возникнув сто лет назад в основном из нужд психологии и социологии, эти методы в настоящее время находят все более широкое применение как некоторые вспомогательные приемы анализа данных в задачах самого разного характера.

> *...нет ничего лучше, как наслаждаться человеку делами своими: потому что это — доля его; ибо кто приведет его посмотреть на то, что будет после него?*
>
> **Екклесиаст (Библия. Ветхий Завет. 3:22)**

> *Пусть не корят меня, что я не сказал ничего нового: ново уже само расположение материала; игроки в мяч бьют по одному и тому же мячу, но не с одинаковой меткостью.*
>
> **Б. Паскаль**

[4] Соловьев Владимир Сергеевич (1853–1900) — русский религиозный философ и поэт.

Приложение

Что виделось вчера как цель глазам твоим, —
Для завтрашнего дня — оковы;
Мысль — только пища мыслей новых,
Но голод их неутолим.

Эмиль Верхарн [1]

1. Нормальное распределение

Квантили распределения: $p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\zeta_p} e^{-x^2/2}\, dx = \Phi(\zeta_p)$

p	ζ_p	p	ζ_p	p	ζ_p
0,50	0,000	0,68	0,468	0,86	1,080
0,51	0,025	0,69	0,496	0,87	1,126
0,52	0,050	0,70	0,524	0,88	1,175
0,53	0,075	0,71	0,553	0,89	1,227
0,54	0,100	0,72	0,583	0,90	1,282
0,55	0,126	0,73	0,613	0,91	1,341
0,56	0,151	0,74	0,643	0,92	1,405
0,57	0,176	0,75	0,674	0,93	1,476
0,58	0,202	0,76	0,706	0,94	1,555
0,59	0,228	0,77	0,739	0,95	1,645
0,60	0,253	0,78	0,772	0,96	1,751
0,61	0,279	0,79	0,806	0,97	1,881
0,62	0,305	0,80	0,842	0,98	2,054
0,63	0,332	0,81	0,878	0,99	2,326
0,64	0,358	0,82	0,915	0,999	3,090
0,65	0,385	0,83	0,954	0,9999	3,720
0,66	0,412	0,84	0,994	0,99999	4,265
0,67	0,440	0,85	1,036		

По тексту используются также обозначения $c_\gamma = \Phi^{-1}\big((1+\gamma)/2\big)$, т. е. $c_\gamma = \zeta_{(1+\gamma)/2}$, и $t_\alpha = -\Phi^{-1}(\alpha)$ т. е. $t_\alpha = -\zeta_\alpha = \zeta_{1-\alpha}$; при этом $t_{\alpha/2} = c_{1-\alpha}$.

[1] Верхарн Эмиль (1855–1916) — бельгийский поэт и драматург.

2. Распределение Пуассона

Значения функции $\sum_{k=x}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$

x \ λ	0,1	0,2	0,3	0,4	0,5	0,6
0	1,000	1,000	1,000	1,000	1,000	1,000
1	0,095	0,181	0,259	0,330	0,394	0,451
2	005	018	037	062	090	122
3		001	003	008	014	023
4				001	002	003

x \ λ	0,7	0,8	0,9	1,0	2,0	3,0
0	1,000	1,000	1,000	1,000	1,000	1,000
1	0,503	0,551	0,593	0,632	0,865	0,950
2	156	191	228	264	594	801
3	034	047	063	080	323	577
4	006	009	014	019	143	353
5	001	001	002	004	053	185
6				001	018	084
7					005	034
8					001	012

x \ λ	4,0	5,0	6,0	7,0	8,0	9,0
0	1,000	1,000	1,000	1,000	1,000	1,000
1	0,982	0,993	0,998	0,999	1,000	1,000
2	908	960	983	924	0,997	0,999
3	762	875	938	970	986	994
4	567	735	849	918	958	979
5	371	560	715	827	900	945
6	215	384	554	699	809	884
7	111	238	394	550	687	793
8	051	133	256	401	547	676
9	021	068	153	271	408	544
10	008	032	084	170	283	413
11	003	014	043	099	184	294
12	001	005	020	053	112	197
13		002	008	027	068	124
14		001	004	013	034	074
15			001	006	017	042

x \ λ	4,0	5,0	6,0	7,0	8,0	9,0
16			001	002	008	022
17				001	004	011
18					002	005
19					001	002
20						001

3. Биномиальное распределение

95 % доверительные пределы (θ_1, θ_2) для параметра θ:

$$P_\theta(\theta_1 < \theta < \theta_2) = 0{,}95.$$

k \ $n-k$	0	1	2	3	4	5	6	7	8	9	10	11	12
0	—	0,98	0,84	0,71	0,60	0,52	0,46	0,41	0,37	0,34	0,31	0,29	0,27
	—	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1	1,00	0,99	0,91	0,81	0,72	0,64	0,58	0,53	0,48	0,45	0,41	0,39	0,36
	0,03	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2	1,00	0,99	0,93	0,85	0,78	0,71	0,65	0,60	0,56	0,52	0,48	0,45	0,43
	0,16	0,09	0,07	0,05	0,04	0,04	0,03	0,03	0,03	0,02	0,02	0,02	0,02
3	1,00	0,99	0,95	0,88	0,82	0,76	0,70	0,65	0,61	0,57	0,54	0,51	0,48
	0,29	0,19	0,15	0,12	0,10	0,09	0,08	0,07	0,06	0,06	0,05	0,05	0,04
4	1,00	1,00	0,96	0,90	0,84	0,79	0,74	0,69	0,65	0,61	0,58	0,55	0,52
	0,40	0,29	0,22	0,18	0,16	0,14	0,12	0,11	0,10	0,09	0,08	0,08	0,07
5	1,00	1,00	0,96	0,92	0,86	0,81	0,77	0,72	0,68	0,65	0,62	0,59	0,56
	0,48	0,36	0,29	0,25	0,21	0,19	0,17	0,15	0,14	0,13	0,12	0,11	0,10
6	1,00	1,00	0,97	0,93	0,88	0,83	0,79	0,75	0,71	0,68	0,65	0,62	0,59
	0,54	0,42	0,35	0,30	0,26	0,23	0,21	0,19	0,18	0,16	0,15	0,14	0,13
7	1,00	1,00	0,97	0,93	0,89	0,85	0,81	0,77	0,73	0,70	0,67	0,64	0,62
	0,59	0,47	0,40	0,35	0,31	0,28	0,25	0,23	0,21	0,20	0,18	0,17	0,16
8	1,00	1,00	0,98	0,94	0,90	0,86	0,82	0,79	0,75	0,72	0,69	0,67	0,64
	0,63	0,52	0,44	0,39	0,35	0,32	0,29	0,27	0,25	0,23	0,22	0,20	0,19
9	1,00	1,00	0,98	0,95	0,91	0,87	0,84	0,80	0,77	0,74	0,71	0,69	0,66
	0,66	0,56	0,48	0,43	0,39	0,35	0,32	0,30	0,28	0,26	0,24	0,23	0,22

k \ $n-k$	0	1	2	3	4	5	6	7	8	9	10	11	12
10	1,00	1,00	0,98	0,95	0,92	0,88	0,85	0,82	0,79	0,76	0,73	0,70	0,68
	0,69	0,59	0,52	0,46	0,42	0,38	0,35	0,33	0,31	0,29	0,27	0,26	0,24
11	1,00	1,00	0,98	0,95	0,92	0,89	0,86	0,83	0,80	0,77	0,74	0,72	0,69
	0,72	0,62	0,57	0,49	0,45	0,41	0,38	0,36	0,34	0,32	0,30	0,28	0,27
12	1,00	1,00	0,98	0,96	0,93	0,90	0,87	0,84	0,81	0,78	0,76	0,73	0,71
	0,74	0,64	0,57	0,52	0,48	0,44	0,41	0,38	0,36	0,34	0,32	0,31	0,29

Примечание. Значения θ_2 набраны в первых строках, значения θ_1 — во вторых.

4. Распределение $\chi^2(n)$

Квантили распределения: $p = \dfrac{1}{2^{n/2}\Gamma(n/2)} \displaystyle\int_0^{x_p^2,\,n} x^{n/2-1} e^{-x/2}\,dx$

n \ p	0,1	0,3	0,5	0,7	0,9	0,95	0,99	0,999
1	0,016	0,148	0,455	1,07	2,71	3,84	6,63	10,8
2	0,211	0,713	1,39	2,41	4,61	5,99	9,21	13,8
3	0,584	1,42	2,37	3,67	6,25	7,82	11,3	16,3
4	1,06	2,20	3,36	4,88	7,78	9,49	13,3	18,5
5	1,61	3,00	4,35	6,06	9,24	11,1	15,1	20,5
6	2,20	3,83	5,35	7,23	10,6	12,6	16,8	22,5
7	2,83	4,67	6,35	8,38	12,0	14,1	18,5	24,3
8	3,49	5,53	7,34	9,52	13,4	15,5	20,1	26,1
9	4,17	6,39	8,34	10,7	14,7	16,9	21,7	27,9
10	4,87	7,27	9,34	11,8	16,0	18,3	23,2	29,6
11	5,58	8,15	10,3	12,9	17,3	19,7	24,7	31,3
12	6,30	9,03	11,3	14,0	18,5	21,0	26,2	32,9
13	7,04	9,93	12,3	15,1	19,8	22,4	27,7	34,5
14	7,79	10,08	13,3	16,2	21,1	23,7	29,1	36,1
15	8,55	11,7	14,3	17,3	22,3	25,0	30,6	37,7
16	9,31	12,6	15,3	18,4	23,5	26,3	32,0	39,3
17	10,09	13,5	16,3	19,5	24,8	27,6	33,4	40,8
18	10,9	14,4	17,3	20,6	26,0	28,9	34,8	42,3
19	11,7	15,4	18,3	21,7	27,2	30,1	36,2	43,8
20	12,4	16,3	19,3	22,8	28,4	31,4	37,6	45,3

n \ p	0,1	0,3	0,5	0,7	0,9	0,95	0,99	0,999
21	13,2	17,2	20,3	23,9	29,6	32,7	38,9	46,8
22	14,0	18,1	21,3	24,9	30,8	33,9	40,3	48,3
23	14,8	19,0	22,3	26,0	32,0	35,2	41,6	49,7
24	15,7	19,9	23,3	27,1	33,2	36,4	43,0	51,2
25	16,5	20,9	24,3	28,2	34,3	37,7	44,3	52,6
26	17,3	21,8	25,3	29,2	35,6	38,9	45,6	54,1
27	18,1	22,7	26,3	30,3	36,7	40,1	47,0	55,5
28	18,9	23,6	27,3	31,4	37,9	41,3	48,3	56,9
29	19,8	24,6	28,3	32,5	39,1	42,6	49,6	58,3
30	20,6	25,5	29,3	33,5	40,3	43,8	50,9	59,7

Примечание. При больших n ($n > 30$) для вычисления квантилей $\chi^2_{p,n}$ можно использовать приближенное соотношение $\chi^2_{p,n} \approx \left(\zeta_p + \sqrt{2n-1}\right)^2/2$, где ζ_p дано в табл. 1.

5. Распределение Стьюдента $S(n)$

Квантили распределения: $p = \dfrac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\dfrac{n}{2}\right)} \displaystyle\int_{-\infty}^{t_{p,n}} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} dx$

n \ p	0,900	0,950	0,975	0,990	0,995
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921

n \ p	0,900	0,950	0,975	0,990	0,995
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750

Примечания.

1. Для малых значений p квантили $t_{p,n}$ находятся по формуле $t_{p,n} = -t_{1-p,n}$.
2. При $n > 30$ можно использовать приближенное соотношение $t_{p,n} \approx \zeta_p$, где ζ_p дано в табл. 1.

6. Распределение Снедекора $S(n_1, n_2)$

Значения функции F_{p,n_1,n_2}:

$$p = \left(\frac{n_1}{n_2}\right)^{n_1/2} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \int_0^{F_{p,n_1,n_2}} x^{n_1/2-1}\left(1+\frac{n_1}{n_2}x\right)^{-(n_1+n_2)/2} dx$$

при $p = 0{,}95$ и $p = 0{,}99$. «Левые» границы доверительных интервалов находятся из условия $F_{1-p,n_1,n_2} = F^{-1}_{p,n_2,n_1}$

n_1 \ n_2	1	2	3	4	6	8	10	12	20	50	100
1	161	200	216	225	234	239	242	244	248	252	253
	4052	4999	5403	5625	5859	5981	6056	6106	6208	6302	6334
2	18,51	19,00	19,16	19,25	19,33	19,37	19,39	19,41	19,44	19,47	19,49
	98,49	99,01	99,17	99,25	99,33	99,36	99,40	99,42	99,45	99,48	99,49
3	10,13	9,55	9,28	9,12	8,94	8,84	8,78	8,74	8,66	8,58	8,56
	34,12	30,81	29,46	28,71	27,91	27,49	27,23	27,05	26,69	26,35	26,23

n_1 \ n_2	1	2	3	4	6	8	10	12	20	50	100
4	7,71	6,94	6,59	6,39	6,16	6,04	5,96	5,91	5,80	5,70	5,66
	21,20	18,00	16,69	15,98	15,21	14,80	14,54	14,37	14,02	13,69	13,57
5	6,61	5,79	5,41	5,19	4,95	4,82	4,74	4,68	4,56	4,44	4,40
	16,26	13,27	12,06	11,39	10,67	10,27	10,05	9,89	9,55	9,24	9,13
6	5,99	5,14	4,76	4,53	4,28	4,15	4,06	4,00	3,87	3,75	3,71
	13,74	10,92	9,78	9,15	8,47	8,10	7,87	7,72	7,39	7,09	6,99
8	5,32	4,46	4,07	3,84	3,58	3,44	3,34	3,28	3,15	3,03	2,98
	11,26	8,65	7,59	7,01	6,37	6,03	5,82	5,67	5,36	5,06	4,96
10	4,96	4,10	3,71	3,48	3,22	3,07	2,97	2,91	2,77	2,64	2,59
	10,04	7,56	6,55	5,99	5,39	5,06	4,85	4,71	4,41	4,12	4,01
12	4,75	3,88	3,49	4,26	3,00	2,85	2,76	2,69	2,54	2,40	2,35
	9,33	6,93	5,95	5,41	4,82	4,50	4,30	4,16	3,86	3,56	3,46
20	4,35	3,49	3,10	2,87	2,60	2,45	2,35	2,28	2,12	1,96	1,90
	8,10	5,85	4,94	4,43	3,87	3,56	3,37	3,23	2,94	2,63	2,53
30	4,17	3,32	2,92	2,69	2,42	2,27	2,16	2,09	1,93	1,76	1,69
	7,56	5,39	4,51	4,02	3,47	3,17	2,98	2,84	2,55	2,24	2,13
50	4,03	3,18	2,79	2,56	2,29	2,13	2,02	1,95	1,78	1,60	1,52
	7,17	5,06	4,20	3,72	3,18	2,88	2,70	2,56	2,26	1,94	1,82
100	3,94	3,09	2,70	2,46	2,19	2,03	1,92	1,85	1,68	1,48	1,39
	6,90	4,82	3,98	3,51	2,99	2,69	2,51	2,36	2,06	1,73	1,59
200	3,89	3,04	2,65	2,41	2,14	1,98	1,87	1,80	1,62	1,42	1,32
	6,76	4,71	3,88	3,41	2,90	2,60	2,41	2,28	1,97	1,62	1,48
1000	3,85	3,00	2,61	2,38	2,10	1,95	1,84	1,76	1,58	1,36	1,26
	6,66	4,62	3,80	3,34	2,82	2,53	2,34	2,20	1,89	1,54	1,38

Примечание. Значения $F_{0,95;\,n_1,n_2}$ набраны в первых строках, значения $F_{0,99;\,n_1,n_2}$ — во вторых.

7. Критерий Колмогорова

Значения функции λ_p: $p = \mathrm{P}\left(D_n = \sup\limits_x |\widehat{F}_n(x) - F(x)| > \lambda_p\right)$

n \ p	0,10	0,05	0,01	n \ p	0,10	0,05	0,01
1	0,950	0,975	0,995	19	0,271	0,301	0,361
2	776	842	929	20	265	294	352

n \ p	0,10	0,05	0,01	n \ p	0,10	0,05	0,01
3	636	708	829	25	238	264	317
4	565	624	734	30	218	242	290
5	509	563	669	35	202	224	269
6	468	519	617	40	189	210	252
7	436	483	576	45	179	198	238
8	410	454	542	50	170	188	226
9	387	430	513	55	162	180	216
10	369	409	489	60	155	172	207
11	352	391	468	65	149	166	199
12	338	375	449	70	144	160	192
13	325	361	432	75	139	154	185
14	314	349	418	80	135	150	179
15	304	338	404	85	131	145	174
16	295	327	392	90	127	141	169
17	286	318	381	95	124	137	165
18	279	309	371	100	121	134	161

8. Критерий Смирнова

Значения вероятности $P(D_{nn} \leqslant k/n)$, где $D_{nn} = \sup_x \left|\widehat{F}_{1n}(x) - \widehat{F}_{2n}(x)\right|$.

n \ k	1	2	3	4	5	6	7	8	9	10	11
1	1,000										
2	0,667	1,000									
3	400	0,900	1,000								
4	229	771	0,971	1,000							
5	127	643	921	0,992	1,000						
6	069	526	857	974	0,998	1,000					
7	037	425	788	947	992	0,999	1,000				
8	020	340	717	913	981	998	1,000	1,000			
9	011	270	648	874	966	993	0,999	1,000	1,000		
10	006	213	582	832	948	988	998	1,000	1,000	1,000	
11	003	167	521	789	925	979	996	0,999	1,000	1,000	1,000
12	002	131	464	744	900	969	992	999	1,000	1,000	1,000
13	001	102	412	700	874	956	987	997	1,000	1,000	1,000
14	000	079	365	657	845	941	981	995	0,999	1,000	1,000
15	000	062	322	614	816	925	974	992	998	1,000	1,000
16	000	048	284	574	785	907	965	989	997	1,000	1,000
17	000	037	249	535	755	888	955	984	995	0,999	0,999
18	000	028	219	497	725	868	944	979	993	998	999

n \ k	1	2	3	4	5	6	7	8	9	10	11
19	000	022	192	462	694	847	932	973	991	997	999
20	000	017	168	429	664	825	919	966	988	996	999
21	000	013	147	397	635	804	905	959	984	995	998
22	000	010	128	368	606	782	891	951	980	993	998
23	000	008	112	340	578	759	876	942	975	991	997
24	000	006	098	314	551	737	860	932	970	988	996
25	000	004	085	290	525	715	844	922	964	985	994
26	000	003	074	267	499	693	828	911	958	982	993
27	000	003	064	246	474	671	811	900	952	978	991

9. Равномерно распределенные случайные числа

65 48	11 76	74 17	46 85	09 50	58 04	77 69	74 45
80 12	43 56	35 17	72 70	80 15	45 31	82 23	74 45
74 35	09 98	17 77	40 27	72 14	43 23	60 02	10 19
69 91	62 68	03 66	25 22	91 48	36 93	68 72	03 37
09 89	32 05	05 14	22 56	85 14	46 42	75 67	88 93
34 07	27 68	50 36	69 73	61 70	65 81	33 98	85 56
45 57	18 24	06 35	30 34	26 14	86 79	90 74	39 82
02 05	16 56	92 68	66 57	48 18	73 05	38 52	47 89
05 32	54 70	48 90	55 35	75 48	28 46	82 87	09 75
06 52	96 47	78 35	80 83	42 82	60 93	52 03	44 76
22 10	94 05	58 60	97 09	34 33	50 50	07 39	98 23
50 72	56 82	48 29	40 52	42 01	52 77	56 78	51 98
13 74	67 00	78 18	47 54	06 10	68 71	17 78	17 49
36 76	66 79	51 90	36 47	64 93	29 60	91 10	62 42
91 82	60 89	28 93	78 56	13 68	23 47	83 41	13 29
68 47	92 76	86 46	16 28	35 54	94 75	08 99	23 45
26 94	03 68	58 70	29 73	41 35	53 14	03 33	40 45
85 15	74 79	54 32	97 92	65 75	57 60	04 08	81 19
11 10	00 20	40 12	86 07	46 97	96 64	48 94	39 37

16 50	53 44	84 40	21 95	25 65	43 65	17 70	82 93
10 09	73 25	33 76	52 01	35 86	34 67	35 48	76 07
37 54	20 48	05 64	89 47	42 96	24 80	52 40	37 32
08 42	26 89	53 19	64 50	93 03	23 20	90 25	60 47
99 01	90 25	29 09	37 67	07 15	38 31	13 11	65 09
12 80	79 99	70 80	15 73	61 47	64 03	23 66	53 53
80 95	90 91	17 39	29 27	49 45	66 06	57 47	17 56
20 63	61 04	02 00	82 29	16 65	31 06	01 08	05 82
15 95	33 47	64 35	08 03	36 06	85 26	97 76	02 89
88 67	67 43	97 04	43 62	76 59	63 57	33 21	35 75
98 95	11 68	77 12	17 17	68 33	73 79	64 57	53 76

Примечания.

1. Приведенные в таблице цифры можно рассматривать как реализации независимых одинаково распределенных случайных величин, принимающих значения $0, 1, \ldots, 9$ с одинаковой вероятностью 10^{-1}
2. Если взять из этой таблицы непересекающиеся группы из, скажем, четырех цифр a_1, a_2, a_3, a_4 и образовать из них числа вида $0{,}a_1a_2a_3a_4$, то эти числа можно рассматривать в качестве значений (с точностью 10^{-4}) равномерно распределенных на $[0, 1]$ случайных величин.

10. Нормально распределенные $\mathcal{N}(0, 1)$ случайные числа

−0,486	0,856	−0,491	−1,983	−1,787	−0,261
−0,256	−0,212	0,219	0,779	−0,105	−0,357
0,065	0,415	−0,169	0,313	−1,339	1,827
1,147	−0,121	1,096	0,181	1,041	0,535
−0,199	−0,246	1,239	−2,574	0,279	−2,056
1,237	1,046	−0,508	−1,630	−0,146	−0,392
−1,384	0,360	−0,992	−0,116	−1,698	−2,832
−0,959	0,424	0,969	−1,141	−1,041	0,362
0,731	1,377	0,983	−1,330	1,620	−1,040
0,717	−0,873	−1,096	−1,396	1,047	0,089
−1,805	−2,008	−1,633	0,542	0,250	−0,166
−1,186	1,180	1,114	0,882	1,265	−0,202
0,658	−1,141	1,151	−1,210	−0,927	0,425
−0,439	0,358	−1,939	0,891	−0,227	0,602
−1,399	−0,230	0,385	−0,649	−0,577	0,237

0,032	0,079	0,199	0,208	−1,083	−0,219
0,151	−0,376	0,159	0,272	−0,313	0,084
0,290	−0,902	2,273	0,606	0,606	−0,747
0,873	−0,437	0,041	−0,307	0,121	0,790
−0,289	0,513	−1,132	−2,098	0,921	0,145
−0,291	1,221	1,119	0,004	0,768	0,079
−2,828	−0,439	−0,792	−1,275	0,375	−1,656
0,247	1,291	0,063	−1,793	−0,513	−0,344
−0,584	0,541	0,484	−0,986	0,292	−0,521
0,446	−1,661	1,045	−1,363	1,026	2,990
−0,034	−2,127	0,665	0,084	−0,880	−1,473
0,234	−0,656	0,340	−0,086	−0,158	−0,851
−0,736	1,041	0,008	0,427	−0,831	0,210
−1,206	−0,899	0,110	−0,528	−0,813	1,266
−0,491	−1,114	1,297	−1,433	−1,345	−0,574
−1,334	1,278	−0,568	−0,109	−0,515	−0,566
−0,287	−0,144	−0,254	0,574	−0,451	−1,181
0,161	−0,886	0,921	−0,509	1,410	−0,518
−1,346	0,193	−1,202	0,394	−1,045	0,843
1,250	−0,199	−0,288	1,810	1,378	0,584
2,923	0,500	0,630	−0,537	0,782	0,060
−1,190	−0,318	0,375	−1,941	0,247	−0,491
0,192	−0,432	−1,420	0,489	−1,711	−1,186
0,942	1,045	−0,151	−0,243	−0,430	−0,762
1,216	0,733	−0,309	0,531	0,416	−1,541
0,499	−0,431	1,705	1,164	0,424	−0,444
0,665	−0,135	−0,145	−0,498	0,593	0,658
0,754	−0,732	−0,066	1,006	0,862	−0,885
0,298	1,049	1,810	2,885	0,235	−0,628
1,456	2,040	−0,124	0,196	−0,853	0,402
0,593	0,993	−0,106	0,116	0,484	−1,272
−1,127	−1,407	−1,579	−1,616	1,458	1,262
−0,142	−0,504	0,532	1,381	0,022	−0,281
−0,023	−0,463	−0,899	−0,394	−0,538	1,707
0,777	0,833	0,410	−0,349	−1,094	0,580
0,241	−0,957	−1,885	0,371	−2,830	−0,238
0,022	0,525	−0,255	−0,702	0,953	−0,869
−0,853	−1,865	−0,423	−0,973	−1,016	−1,726
−0,501	−0,273	0,857	−0,465	−1,691	0,417
0,439	−0,035	−0,260	0,120	−0,558	0,056

−0,627	0,561	0,471	−1,029	−2,015	−0,594
−1,108	−2,357	−0,310	0,479	−0,623	−1,047
−1,726	1,956	0,610	2,709	−0,699	−1,347
0,524	−0,281	−0,220	−0,057	0,481	0,996
−0,573	0,932	0,738	−0,300	−0,586	−1,023
−0,579	0,551	0,359	0,326	0,884	−0,298
−0,120	0,418	−0,094	1,114	0,457	1,064
0,191	0,074	1,501	1,068	−0,798	0,162
0,071	0,524	0,031	0,772	−0,768	−0,129
−3,001	0,479	0,402	0,226	0,023	−1,204
1,066	1,097	−1,752	−0,329	−1,256	0,318
0,736	−0,916	−0,291	0,085	1,701	−1,087
−0,342	1,222	−0,933	0,130	0,674	0,899
−0,188	−1,153	−0,450	−0,244	0,072	1,028
1,395	1,298	0,512	−0,882	0,490	−1,304
1,531	0,349	−0,059	0,415	−1,084	−0,958
−0,443	−0,292	−0,539	−1,382	0,318	−0,243
1,409	−0,883	0,229	0,129	0,367	−0,095
1,730	−0,056	−0,078	−2,361	−0,992	−1,488
−0,266	0,757	0,194	−1,078	0,529	−0,361

Законы статистики везде одинаковы, — продолжал Николай Петрович солидно. Утром, например, гостей бывает меньше, потому что публика еще исправна; но чем больше солнце поднимается к зениту, тем наплыв делается сильнее. И, наконец, ночью, по выходе из театров — это почти целая оргия!

И заметьте, — пояснил Семен Иванович, — каждый день, в одни и те же промежутки времени, цифры всегда одинаковые. Колебаний — никаких! Такова незыблемость законов статистики!

М. Е. Салтыков-Щедрин [2)]

Случай всегда именует себя Богом, когда он не хочет подписываться своим именем.

А. Франс [3)]

[2)] Салтыков-Щедрин Михаил Евграфович (1826–1889) — велнкий русский писатель-сатирик.

[3)] Франс Анатоль (наст. имя Анатоль Франсуа Тибо) (1844–1924) — французский писатель, нобелевский лауреат (1921).

Литература

1. *Андерсон Т.* Введение в многомерный статистический анализ. М.: Физматгиз, 1963.
2. *Бикел П., Доксам К.* Математическая статистика. Т. 2. М.. ФиС, 1983.
3. *Большев Л. Н., Смирнов Н. В.* Таблицы математической статистики. М.: Наука, 1983.
4. *Боровиков В. П. Ивченко Г. И.* Прогнозирование в системе STATISTICA в среде Windows. М.. ФиС, 2006.
5. *Боровков А. А.* Математическая статистика. М.: Издательство Физико-математической литературы, 2007.
6. *Вальд А.* Последовательный анализ. М.: Физматгиз, 1960.
7. *Гаек Я., Шидак З.* Теория ранговых критериев. М.: Наука, 1971.
8. *Гнеденко Б. В.* Курс теории вероятностей. 9-е изд. М.: Издательство ЛКИ/URSS, 2007.
9. *Ивченко Г. И., Глибоченко А. Ф., Иванов В. А., Медведев Ю. И.* Статистический анализ дискретных случайных последовательностей. М.. МИЭМ, 1984.
10. *Ивченко Г. И. Медведев Ю. И.* Математическая статистика. 2-е изд. М.: Высшая школа, 1992.
11. *Ивченко Г. И., Медведев Ю. И. Чистяков А. В.* Задачи с решениями по математической статистике. М.: Дрофа, 2007.
12. *Кендалл М., Стьюарт А.* Статистические выводы и связи. М.: Наука, 1973.
13. *Кнут Д.* Искусство программирования. М.: Мир, 2004.
14. *Колчин В. Ф., Севастьянов Б. А. Чистяков В. П.* Случайные размещения. М.: Наука, 1976.
15. *Крамер Г.* Математические методы статистики. М.. Мир, 1975.
16. *Леман Э.* Проверка статистических гипотез. М.: Наука, 1979.
17. *Лоули Д., Максвелл А.* Факторный анализ как статистический метод. М.: Мир, 1967.
18. *Налимов В. В.* Теория эксперимента. М.: Наука, 1971.
19. *Окунь Я.* Факторный анализ. М.: Статистика, 1974.
20. *Пойа Д.* Математика и правдоподобные рассуждения. М.: Книжный дом «Либроком»/URSS, 2010.
21. *Рао К. Р.* Линейные статистические методы и их применения. М.: Наука, 1968.
22. *Себер Дж.* Линейный регрессионный анализ. М.: Мир, 1980.
23. *Секей Г* Парадоксы в теории вероятностей и математической статистике. М.: Мир, 1990.
24. *Уилкс С.* Математическая статистика. М.: Наука, 1967.
25. *Феллер В.* Введение в теорию вероятностей и ее приложения. Т. 1, 2. М.: Книжный дом «Либроком»/URSS, 2010.
26. *Холлендер М., Вульф Д.* Непараметрические методы статистики. М.: ФиС, 1983.
27. *Чистяков В. П.* Курс теории вероятностей. 7-е изд. М.: Дрофа, 2007
28. *Ширяев А. Н.* Статистический последовательный анализ. М.: Наука, 1976.
29. *Billingsley P.* Statistical inference for Markov processes. The University of Chicago Press, 1961.

Путеводитель по моделям в примерах и задачах

Смелость — начало дела, но случай — хозяин конца.

Демокрит [1)]

Назначение данного путеводителя — помочь читателю быстро найти в тексте нужное место, где доказывается либо формулируется тот или иной конкретный факт (результат), относящийся к конкретной модели (в скобках указаны соответствующие координаты, при этом, например, «упр. 3.10» означает упражнение 10 к гл. 3). Модели, как правило, занумерованы в той последовательности, в которой они встречаются в тексте, также в такой последовательности перечисляются и относящиеся к ним факты.

1. Бернуллиевская модель Bi(1, p)

- определение и свойства (§ 1.1, п. 1, упр. 1.1)
- реализация в урновой схеме (§ 1.1, прим. 1 и 2)
- моделирование (§ 1.3, п. 2)
- эмпирическая функция распределения для нее (§ 2.1, прим. 2)
- оценивание доли белых шаров в урне (§ 2.2, прим. 3)
- эксперимент Бюффона с бросанием монеты (упр. 2.1, § 4.2, прим. 1)
- обработка эмпирических данных (упр. 2.2)
- оценивание параметрических функций (§ 3.1, прим. , упр. 3.10, 3.11)
- оптимальность выборочного среднего при оценивании параметра (§ 3.1, прим. 5)
- оптимальная оценка доли белых шаров в урне (§ 3.1, прим. 6)
- эффективность выборочного среднего (§ 3.2, п. 2)
- достаточная статистика для нее (§ 3.2, табл. 2)
- о. м. п. параметра (§ 3.2, табл. 2)
- асимптотический доверительный интервал для параметра (§ 3.5, прим. 17, упр. 3.70)
- байесовская оценка параметра (§ 3.6, прим. 11, упр. 3.77)
- минимаксная оценка параметра (§ 3.6, прим. 12)
- доверительны интервал для параметра (§ 3.8, прим. 4)
- критерий согласия χ^2 для нее (упр. 4.1, 4.2)
- критерий Неймана—Пирсона (§ 5.2, прим. 4, упр. 5.4)
- р. н. м. к. при односторонней альтернативе (§ 5.3, п. 1)
- КОП для гипотезы однородности (§ 5.4, прим. 8)
- решающие правила в ней (§ 7.1, прим. 1, упр. 7.1, 7.2, 7.3).

[1)] Демокрит (ок. 460 г. до н. э. – год смерти неизв.) — древнегреческий философ-атомист.

2. Биномиальная модель Bi(n, p)

— определение и свойства (§ 1.1, п. 1, упр. 1.2, 1.3, 1.4, 2.25)
— реализация в урновой схеме (§ 1.1, прим. 1 и 2)
— дни рождения, эмпирические данные (§ 1.1, прим. 3)
— моделирование (§ 1.3, п. 2)
— аппроксимация распределением Пуассона (упр. 1.12)
— условные распределения компонент полиномиального вектора (упр. 1.18)
— моменты (упр. 1.19)
— нормальная аппроксимация, теорема Муавра—Лапласа (упр. 1.33)
— связь с распределением Маркова—Пойа (упр. 1.44)
— эмпирические данные в опыте с бросанием костей (§ 2.2, прим. 2)
— обработка эмпирических данных (упр. 2.3, 2.8, 2.9)
— информация Фишера для нее (§ 3.2, табл. 1)
— эффективная оценка для параметра (§ 3.2, табл. 2)
— достаточная статистика для нее (§ 3.2, табл. 2, упр. 3.31)
— оценивание параметрических функций (§ 3.4, прим. 7, упр. 3.12, 3.31)
— о. м. п. параметра (§ 3.2, табл. 2)
— преобразование о. м. п., стабилизирующее дисперсию (упр. 3.55)
— проверка гипотезы согласия (§ 4.2, прим. 4, упр. 4.7, 4.15)
— р. н. м. к. при односторонней альтернативе (§ 5.3, прим. 1, упр. 5.8)
— критерий Неймана—Пирсона (упр. 5.5)
— л. н. м. несмещенный критерий (упр. 5.14)
— КОП для простой гипотезы (упр. 5.22)
— прогноз в ней (упр. 6.23, 6.24)
— решающие правила в ней (упр. 7.4, 7.9)
— 95%-ные доверительные интервалы для параметра (прилож. табл. 3).

3. Отрицательная биномиальная модель $\overline{\text{Bi}}(r, p)$

— определение и свойства (§ 1.1, п. 2, упр. 1.5, 1.6, 1.7, 1.8, 1.14, 1.41, 2.25)
— как частный случай модели степенного ряда (§ 1.1, п. 8)
— моделирование (§ 1.3, п. 3)
— аппроксимация распределением Пуассона (упр. 1.13)
— оценивание параметрических функций (§ 3.1, прим. 3, § 3.4, прим. 2)
— информация Фишера для нее (§ 3.2, табл. 1)
— эффективная оценка для параметрической функции (§ 3.2, табл. 2)
— достаточная статистика для нее (§ 3.2, табл. 2)
— о. м. п. параметра (упр. 3.45)
— асимптотический доверительный интервал для параметра (упр. 3.71)
— байесовская оценка параметра (упр. 3.78)
— р. н. м. к. при одиосторонней альтернативе (§ 5.3, прим. 2, упр. 5.9).

4. Геометрическая модель $\overline{\mathrm{Bi}}(1, p)$

— определение и свойства (§ 1.1, п. 2)
— моделирование (§ 1.3, п. 3)
— достаточная статистика для нее (§ 3.2, табл. 2)
— о. м. п. параметра и основанный на ней доверительный интервал (§ 3.5, прим. 16)
— улучшенная оценка параметра (§ 3.7, прим. 2)
— оценивание параметрической функции (упр. 3.15)
— р. н. м. к. при односторонней альтернативе (§ 5.3, прим. 2)
— критерий Неймана—Пирсона (упр. 5.6)
— решающие правила в ней (упр. 7.5).

5. Пуассоновская модель $\Pi(\lambda)$

— определение и свойства (§ 1.1, п. 3, упр. 1.9, 1.10, 1.11, 1.13, 1.14, 1.41, 2.25)
— дни рождения, аппроксимация (§ 1.1, прим. 3)
— усечение в нуле (§ 1.1, п. 8)
— как частный случай модели степенного ряда (§ 1.1, п. 8)
— моделирование (§ 1.3, п. 5)
— теорема Пуассона для биномиального распределения (упр. 1.12)
— аппроксимация для полиномиального распределения (упр. 1.34, 1.35)
— оценивание параметрических функций (§ 3.1, прим. 2, § 3.4, прим. 1, упр. 3.13, 3.14, 3.32, 3.33)
— информация Фишера для нее (§ 3.2, табл. 1)
— эффективность выборочного среднего (§ 3.2, табл. 2)
— достаточная статистика для нее (§ 3.2, табл. 2, упр. 3.32)
— о. м. п. параметра (§ 3.2, табл. 2)
— о. м. п. для усеченного в нуле распределения (§ 3.5, прим. 9)
— доверительные интервалы для параметра (§ 3.5, прим. 15, § 3.8, прим. 5)
— преобразование о. м. п., стабилизирующее дисперсию (упр. 3.55)
— метод моментов для двойного распределения Пуассона (упр. 3.75)
— байесовская оценка параметра (упр. 3.79, 3.80)
— проверка гипотезы согласия для нее (§ 4.2, прим. 5, упр. 4.13, 4.14)
— гипотеза одородности для нее (упр. 4.21, § 5.4, прим. 9)
— критерий Неймана—Пирсона (§ 5.2, прим. 5)
— р. н. м. к. при односторонней альтернативе (§ 5.3, п. 1)
— л. н. м. несмещенный критерий (упр. 5.15)
— КОП для простой гипотезы (упр. 5.23)
— прогноз в ней (упр. 6.23, 6.24)
— решающее правило в ней (упр. 7.6)
— значения функции $\mathrm{P}\{\xi \geqslant x\}$ (прилож. табл. 2).

6. Гипергеометрическая модель $H(a_1, a_2, n)$

— определение и свойства, реализация в урновой схеме (§ 1.1, п. 4)
— многомерный вариант (§ 1.1, п. 7)
— моменты (упр. 1.15)
— аппроксимация биномиальным распределением (упр. 1.16)
— о. м. п. для параметра $a = a_1 + a_2$ (упр. 1.17)
— р. н. м. к. при односторонней альтернативе (упр. 5.11)
— прогноз в ней (§ 6.6, прим. 4).

7. Модель Маркова—Пойа МП(n; a_1, a_2, c)

— определение и свойства, реализация в урновой схеме (§ 1.1, п. 5)
— многомерный вариант (§ 1.1, п. 7, упр. 1.62)
— как смесь биномиальных распределений (упр. 1.44)
— оценивание параметра $p = a_1/(a_1 + a_2)$ (упр. 3.2)
— оценивание многомерного параметра (упр. 3.3)

8. Полиномиальная модель $M(n; p_1, \ldots, p_N)$

— определение и свойства (§ 1.1, п. 6, упр. 1.18, 1.21)
— моделирование (§ 1.3, п. 4)
— моменты (упр. 1.19, 1.20)
— связь с пуассоновским распределением (упр. 1.22)
— случайное число испытаний (упр. 1.23)
— центральная предельная теорема для нее (упр. 1.33)
— пуассоновская аппроксимация (упр. 1.34, 1.35)
— связь с распределением Маркова—Пойа (упр. 1.62)
— эмпирическая функция распределения для нее (§ 2.1, прим. 3)
— обработка эмпирических данных (упр. 2.4)
— информационная матрица для нее (§ 3.2, прим. 10)
— оценивание параметрического вектора (§ 3.2, прим. 10, упр. 3.7)
— оценивание параметрических функций (§ 3.3, прим. 8)
— о. м. п. параметрического вектора (§ 3.5, прим. 12)
— асимптотическая доверительная область для параметров (§ 3.5, прим. 18)
— байесовская оценка параметров (§ 3.6, прим. 13)
— минимаксная оценка параметров (§ 3.6, прим. 13)
— о. м. п. параметра в модели $M\left(n; p_1(\theta), \ldots, p_N(\theta)\right)$ (упр. 3.46)
— проверка гипотезы в эксперименте Менделя (§ 4.2, прим. 2)
— проверка гипотезы о равновероятности классов (§ 4.2, прим. 3)
— критерий согласия χ^2 для нее (упр. 4.3, 4.4, 4.5, 4.6)
— критерий согласия χ^2 для сложной гипотезы (упр. 4.11, 4.12)
— критерий Неймана—Пирсона (§ 5.2, прим. 6)
— КОП для простой гипотезы (§ 5.4, прим. 5).

9. Модель степенного ряда

- определение и свойства (§ 1.1, п. 8)
- производящая функция и моменты (упр. 1.24, 1.25)
- оценивание параметрических функций (§ 3.4, п. 1)
- о. м. п. параметра (§ 3.5, прим. 9)
- асимптотический доверительный интервал для параметра (§ 3.5, прим. 14)

10. Конечная совокупность

- занумерованные элементы, оценивание параметрических функций (§ 3.4, прим. 8 и 9)
- оценивание среднего, оценки Горвица—Томпсона (§ 3.9, п. 1)
- выборочный контроль (§ 3.9, п. 2)
- оценивание состава (§ 3.9, п. 2)
- оценивание доли белых шаров в урне (упр. 3.1)
- прогнозирование в схеме повторного выбора (§ 6.6, прим. 5).

11. Цепь Маркова

- определение и моделирование (§ 1.3, п. 10, § 4.5, п. 2)
- случайное блуждание (§ 1.3, прим. 3 и 4)
- обобщенный критерий χ^2 для контроля марковской зависимости (§ 4.5, п. 2)
- проверка гипотез (§ 5.5)
- критерий χ^2 для простой гипотезы (§ 5.5, п. 2)
- критерий χ^2 для сложной гипотезы (§ 5.5, п. 3)
- гипотеза о порядке цепи Маркова (§ 5.5, прим. 1)
- гипотеза о циклическом случайном блуждании (§ 5.5, прим. 2 и 3)
  КОП для общих параметрических гипотез (§ 5.5, п. 4)
- критерий однородности (§ 5.5, п. 5)
- критерии однородности в модели циклического блуждания (§ 5.5, прим. 4)
- оценивание порядка цепи (§ 5.5, п. 6).

12. Нормальная модель $\mathcal{N}(\mu, \sigma^2)$

- определение и свойства (введение к гл. 1, § 1.2, п. 1, упр. 1.26, 1.27, 1.28)
- моделирование (§ 1.3, п. 6 и 8)
- независимость выборочных среднего и дисперсии (упр. 1.28)
- связь с распределением Коши (упр. 1.48)
- сопряженность к нормальному распределению (упр. 1.64(5))
- распределение выборочной медианы в больших выборках (§ 2.4, прим. 2)
- распределение верхнего экстремума в больших выборках (§ 2.4, прим. 4)
- доверительное оценивание среднего (§ 2.5, прим. 1, § 3.8, прим. 2, упр. 3.58, 3.60, § 5.3, прим. 7)
- доверительное оценивание дисперсии (§ 2.5, прим. 1, § 3.8, прим. 3, упр. 3.61, § 5.3, прим. 8)
- доверительная область для параметров (§ 2.5, прим. 1)
- задача сравнения двух средних (§ 2.5, прим. 2, упр. 3.62)

— доверительный интервал для отношения дисперсий (§ 2.5, прим. 3)
— таблица доверительных интервалов (§ 3.8, п. 1)
— доверительный интервал для параметра θ модели $\mathcal{N}(\theta, \theta^2)$ (упр. 3.57)
— доверительный интервал для параметра θ модели $\mathcal{N}(\mu, \theta^2)$ (упр. 3.59)
— асимптотические доверительные интервалы в модели $\mathcal{N}(\mu, \theta^2)$ (упр. 3.68)
— обработка эмпирических данных (упр. 2.6)
— распределение статистики $\eta = (X_1 - \overline{X})/(\sqrt{n-1}S)$ (упр. 2.42)
— асимптотическая нормальность статистики $T_n = \sqrt{n}(\overline{X} - \mu)/S$ (упр. 2.44)
— оценивание дисперсии модели $\mathcal{N}(\theta_1, \theta_2^2)$ (§ 3.1, прим. 4)
— информация Фишера для нее (§ 3.2, табл. 1)
— эффективность выборочного среднего (§ 3.2, прим. 1, табл. 2)
— эффективная оценка дисперсии (§ 3.2, прим. 2 и 3, табл. 2)
— асимптотическая эффективность выборочной медианы (§ 3.2, прим. 4)
— оптимальная оценка для θ^2 в модели $\mathcal{N}(\theta, \sigma^2)$ (§ 3.2, прим. 7)
— оптимальное оценивание среднего модели $\mathcal{N}(\theta_1, \theta_2^2)$ (§ 3.2, прим. 8)
— оптимальное оценивание дисперсии модели $\mathcal{N}(\theta_1, \theta_2^2)$ (§ 3.2, прим. 9)
— информационная матрица для модели $\mathcal{N}(\theta_1, \theta_2^2)$ (§ 3.2, прим. 9)
— достаточные статистики для нее (§ 3.2, табл. 2, § 3.3, прим. 1)
— оценивание теоретической функции распределения (§ 3.3, прим. 6 и 7)
— оценивание параметрических функций в модели $\mathcal{N}(\theta, \sigma^2)$ (§ 3.4, прим. 6)
— оценивание параметрических функций в модели $\mathcal{N}(\mu, \theta^2)$ (§ 3.4, прим. 12, упр. 3.17–3.19, 3.34)
— оценивание параметрических функций в модели $\mathcal{N}(\theta, \theta_2^2)$ (§ 3.4, прим. 13, упр. 3.20, 3.34, 3.36)
— о. м. п. параметров (§ 3.5, прим. 1)
— о. м. п. функции распределения (§ 3.5, прим. 5)
— о. м. п. среднего и дисперсии логнормального распределения (§ 3.5, прим. 7)
— сверхэффективная оценка параметра модели $\mathcal{N}(\theta, 1)$ (§ 3.5, прим. 10)
— асимптотический доверительный интервал для функции распределения (§ 3.5, прим. 13)
— метод моментов для нее (§ 3.6, прим. 1)
модально несмещенная оценка среднего (§ 3.6, прим. 4)
— оценка Питмена (эквивариантная) среднего (§ 3.6, прим. 8)
— минимаксность выборочного среднего (§ 3.6, прим. 14)
— оценивание по цензурированным данным для нее (§ 3.6, прим. 15 и 16)
— объединение оценок (§ 3.7, прим. 1, упр. 3.56)
— оценка «складного ножа» дисперсии (§ 3.7, прим. 3)
— состоятельная оценка параметра модели $\mathcal{N}(\theta, \theta^2)$, $\theta > 0$ (упр. 3.5)
— сравнение оценок дисперсии в модели $\mathcal{N}(\mu, \theta^2)$ (упр. 3.16)
— достаточная статистика в модели $\mathcal{N}(\theta, \gamma\theta^2)$ (упр. 3.35)
— о. м. п. параметра модели $\mathcal{N}(\mu, \theta^2)$, асимптотические свойства (упр. 3.41)

- о. м. п. параметра модели $\mathcal{N}(\theta, 2\theta)$ (упр. 3.42)
- преобразование о. м. п., стабилизирующее дисперсию, в модели $\mathcal{N}(\mu, \theta^2)$ (упр. 3.55)
- прогнозирование будущих наблюдений (упр. 3.67)
- байесовская оценка параметра θ модели $\mathcal{N}(\theta, b^2)$ (упр. 3.83, 7.12)
- критерий согласия χ^2 по группированным данным (§ 4.2, прим. 6)
- критерий Колмогорова (упр. 4.10)
- критерий Неймана—Пирсона для среднего модели $\mathcal{N}(\theta, \sigma^2)$ (§ 5.2, прим. 1)
- критерий Неймана—Пирсона для дисперсии модели $\mathcal{N}(\mu, \theta^2)$ (§ 5.2, прим. 2)
- р. н. м. к. при односторонних альтернативах (§ 5.3, п. 1)
- р. н. м. и л. н. м. несмещенные критерии при двусторонней альтернативе для модели $\mathcal{N}(\theta, \sigma^2)$ (§ 5.3, прим. 4 и 5)
- р. н. м. несмещенный критерий при двусторонней альтернативе для модели $\mathcal{N}(\mu, \theta^2)$ (упр. 5.12)
- р. н. м. несмещенный критерий для среднего модели $\mathcal{N}(\theta_1, \theta_2^2)$ при двусторонней альтернативе (§ 5.3, прим. 9)
- р. н. м. несмещенный критерий для дисперсии модели $\mathcal{N}(\theta_1, \theta_2^2)$ при двусторонней альтернативе (§ 5.3, прим. 9)
- гипотеза о равенстве средних двух моделей (§ 5.3, прим. 10)
- КОП для среднего модели $\mathcal{N}(\theta_1, \theta_2^2)$ (§ 5.4, прим. 1)
- КОП для дисперсии модели $\mathcal{N}(\theta_1, \theta_2^2)$ (§ 5.4, прим. 2)
- КОП для гипотезы о равенстве средних (§ 5.4, прим. 3 и 6)
- КОП для гипотезы о равенстве дисперсий (§ 5.4, прим. 4 и 7)
- критерий Вальда (последовательный) для модели $\mathcal{N}(\theta, \sigma^2)$ (§ 5.6, прим. 1)
- гипотеза о равенстве дисперсий двух моделей (упр. 5.16)
- F-критерий однородности для средних (§ 6.5, п. 2, упр. 6.21)
- метод наименьших квадратов (упр. 6.8–6.11, 6.13–6.15, 6.17–6.19)
- решающие правила для нее (упр. 7.7)
- квантили распределения $\mathcal{N}(0, 1)$ (прилож. табл. 1)
- реализации случайных $\mathcal{N}(0, 1)$ чисел (прилож. табл. 10).

13. Двумерная нормальная модель

$$\mathcal{N}\left((\mu_1, \mu_2), \left\| \begin{matrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{matrix} \right\|\right)$$

- условное маргинальное распределение в ней (упр. 1.65)
- асимптотический доверительный интервал для ρ (§ 2.3, п. 4)
- гипотеза независимости $H_0 \quad \rho = 0$ (§ 2.3, п. 4)
- обработка эмпирических данных (упр. 2.29)
- распределение выборочных моментов (упр. 2.43)
- о. м. п. параметров модели $\mathcal{N}\left((0, 0), \left\| \begin{matrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{matrix} \right\|\right)$ (§ 3.5, прим. 6)

— оценивание коэффициента корреляции модели $\mathcal{N}\left((0,0), \begin{Vmatrix} 1 & \rho \\ \rho & 1 \end{Vmatrix}\right)$ (упр. 3.43)

— прогноз в ней (§ 6.6, прим. 2).

14. Многомерная нормальная модель $\mathcal{N}(\underline{\mu}, \Sigma)$

— определение и свойства (§ 1.2, упр. 1.30)
— моменты (упр. 1.29)
— линейные преобразования (упр. 1.31, 1.32)
— центральная предельная теорема для полиномиальной модели (упр. 1.33)
— распределение квадратичных форм (упр. 1.40, 2.38–2.41)
— условное маргинальное распределение в ней (упр. 1.66)
— о. м. п. ее параметров (§ 3.5, прим. 2)
— оценка Джеймса—Стейна вектора средних (§ 3.6, прим. 14)
— критерий Неймана—Пирсона для вектора средних (§ 5.2, прим. 3)
— критерий для вектора средних, минимизирующий сумму вероятностей ошибок (§ 5.2, продолжение прим. 3)
нормальная регрессия, о. м. п. ее параметров (§ 6.3, п. 2, упр. 6.4)
— доверительный интервал для наклона простой регрессии (§ 6.3, прим. 1)
— доверительный эллипс для параметров простой регрессии (§ 6.3, прим. 1)
— доверительный интервал для ординаты линии регрессии (§ 6.3, прим. 2)
— совместные доверительные интервалы для коэффициентов регрессии (§ 6.3, прим. 3)
— гипотеза о наклоне линии регрессии (§ 6.4, прим. 1, упр. 6.20)
— критерий значимости регрессии (§ 6.4, прим. 2, упр. 6.22)
— гипотеза о параллельности линий регрессии (§ 6.5, п. 1)
— дисперсионный анализ при двойной классификации (§ 6.5, п. 3)
— прогноз в ней (упр. 6.25)
— классификация наблюдений по двум классам (§ 7.3)
— классификация наблюдений по $k \geqslant 2$ классам (§ 7.4)
— распределение Уишарта (выборочной ковариационной матрицы) (§ 8.3).

15. Гамма-модель $\Gamma(\alpha, \lambda)$

— определение и свойства (§ 1.2, п. 3, упр. 1.36, 1.37, 1.41, 1,43, 2.25)
— связь с распределением Дирихле (упр. 1.60)
— сопряженность к распределению Пуассона (упр. 1.64(3))
— сопряженность к показательному распределению (упр. 1.64(4))
— информация Фишера для нее (§ 3.2, табл. 1)
— эффективная оценка для параметра α (§ 3.2, табл. 2)
— достаточная статистика для нее (§ 3.2, табл. 2)
— оценивание параметрических функций (§ 3.4, прим. 10, 11, упр. 3.21, 3.29, 3.34)
— метод моментов для нее (§ 3.6, прим. 3)
— модально несмещенная оценка параметра α (§ 3.6, прим. 7)
— медианно несмещенная оценка параметра α (§ 3.6, прим. 7)
— о. м. п. в ней (упр. 3.52)

— преобразование о. м. п. стабилизирующее дисперсию (упр. 3.55)
— асимптотические доверительные интервалы в ней (упр. 3.69)

16. Показательная (экспоненциальная) модель $\Gamma(a, 1)$

— определение и свойства (§ 1.2, п. 3)
моделирование (§ 1.3, п. 7)
— распределение нижнего экстремума (упр. 1.39)
— связь с распределением Снедекора (упр. 1.54)
— связь с распределением Парето (упр. 1.58)
порядковые статистики и спейсинги (§ 2.4, прим. 1)
— распределение верхних экстремумов в больших выборках (§ 2.4, прим. 3, упр. 2.36)
— обработка эмпирических данных (упр. 2.5)
приведенное к отрезку $[\theta, b]$ распределение $\Gamma(1, 1)$, оценивание параметрических функций (§ 3.4, прим. 5)
— оценивание функции надежности (§ 3.4, прим. 11)
— о. м. п. параметра (§ 3.2, табл. 2)
— модально несмещенная оценка параметра (§ 3.6, прим. 6)
медианно несмещенная оценка параметра (§ 3.6, прим. 7)
— оценивание параметра по цензурированным данным (§ 3.6, прим. 17)
— состоятельная оценка параметра (упр. 3.5)
— доверительный интервал для отношения параметров двух моделей (упр. 3.63)
— доверительный интервал для параметра (упр. 3.65)
— байесовская оценка параметра (упр. 3.81)
— критерий согласия χ^2 для нее (упр. 4.9)
— р. н. м. к. при односторонней альтернативе (§ 5.3, прим. 3, упр. 5.10)
— критерий Неймана—Пирсона (упр. 5.1)
— р. н. м. несмещенный критерий при двусторонней альтернативе (упр. 5.13)
— гипотеза однородности (упр. 5.17)
функция регрессии (§ 6.6, прим. 1).

17. Двухпараметрическая показательная модель $W(a, 1, b)$

— определение и свойства (§ 1.2, п. 8)
— достаточная статитстика для нее (§ 3.3, прим. 2)
— оценивание параметрических функций в модели $W(\theta, 1, 1)$ (§ 3.4, прим. 4)
— о. м. п. параметров (§ 3.5, прим. 4, упр. 3.54)
— оптимальные несмещенные оценки параметров (упр. 3.39)
— доверительный интервал для параметра модели $W(\theta, 1, 1)$ (упр. 3.64)
— доверительные интервалы для параметров (упр. 3.66)
— оценивание параметров по цензурированным данным (упр. 3.76)
— несмещенный критерий при двусторонней альтернативе для модели $W(\theta, 1, 1)$ (упр. 5.18)
критерии проверки гипотез $H_0\ \ \theta_1 = \theta_{10}$ и $H_0\ \ \theta_2 = \theta_{20}$ для модели $W(\theta_1, 1, \theta_2)$ (упр. 5.21).

18. Распределение хи-квадрат $\chi^2(n) = \Gamma(2, n/2)$

— определение и свойства (§ 1.2, п. 3, упр. 1.38)
— центральная предельная теорема для него (упр. 1.38)
— как предельное распределение статистики хи-квадрат К. Пирсона (§ 4.2, п. 2)
квантили (прилож. табл. 4).

19. Равномерная модель $U(a, b)$

— определение и свойства (§ 1.2, п. 5, упр. 1.45)
— преобразование к показательному распределению (§ 1.2, прим. 1)
— преобразование к нормальному распределению (§ 1.2, прим. 2)
— моделирование (§ 1.3, п. 6)
— совместное распределение и моменты экстремумов (упр. 1.46)
— связь с распределением Коши (упр. 1.47)
— распределение верхнего экстремума в больших выборках (§ 2.4, прим. 6)
— обработка эмпирических данных (упр. 2.7, 2.10)
распределение порядковых статистик и их моменты (упр. 2.32, 2.33)
— сверхэффективная оценка параметра модели $U(0, \theta)$ (§ 3.2, прим. 5)
— достаточная статистика для модели $U(0, \theta)$ (§ 3.3, прим. 3)
— достаточная статистика для модели $U(a(\theta), b(\theta))$ (§ 3.3, прим. 4)
— оценивание параметрических функций в модели $U(0, \theta)$ (§ 3.4, прим. 3)
— оценивание параметрических функций в модели $U(\theta_1, \theta_2)$ (упр. 3.24, 3.38)
— о. м. п. параметра в моделях $U(0, \theta)$ и $U(\theta, \theta + 1)$ (§ 3.5, прим. 3)
— о. м. п. параметра в моделях $U(\theta - 1/2, \theta + 1/2)$ (упр. 3.49)
— асимптотические свойства о. м. п. параметра в модели $U(0, \theta)$
(§ 3.5, прим. 11, упр. 3.48)
— оценки параметров по методу моментов (§ 3.6, прим. 2, упр. 3.72, 3.73)
— модально несмещенная оценка параметра модели $U(0, \theta)$ (§ 3.6, прим. 5)
— оценка Питмена (эквивариантная) параметра модели $U(\theta - 1/2, \theta + 1/2)$
(§ 3.6, прим. 9)
— оценка Питмена (эквивариантная) параметра модели $U(0, \theta)$ (§ 3.6, прим. 10)
— доверительный интервал для параметра модели $U(0, \theta)$ (§ 3.8, прим. 1)
— оценивание параметра модели $U(\theta, 2\theta)$ (упр. 3.22)
— сравнение оценок параметра в модели $U(0, \theta)$ (упр. 3.23, 3.37)
— байесовская оценка параметра модели $U(0, \theta)$ (упр. 3.82, 7.10)
— критерии согласия для нее (упр. 4.8)
— гипотеза однородности для нее (упр. 4.19)
— гипотеза случайности для нее (упр. 4.25)
— критерий Неймана—Пирсона различения равномерной и нормальной моделей
(упр. 5.3)
— несмещенный критерий при двусторонней альтернативе в модели $U(0, \theta)$
(упр. 5.19)
— метод наименьших квадратов для нее (упр. 6.7, 6.12)
— реализации случайных чисел (прилож. табл. 9).

20. Бета-распределение Be(a, b)

определение и свойства (§ 1.2, п. 4)
— как частный случай распределения Дирихле (§ 1.2, п. 10)
— смеси (§ 1.3, прим. 2, п. 9)
— моменты (упр. 1.42)
— связь с гамма-распределением (упр. 1.43)
— связь с распределением Маркова—Пойа (упр. 1.44)
— связь с распределением Стьюдента (упр. 1.50)
— связь с распределением Снедекора (упр. 1.52, 1.53)
— сопряженность к биномиальному распределению (упр. 1.64(2))
— оценивание параметрических функций в модели Be(θ, 1) (§ 3.4, прим. 14)
— прогноз (§ 6.6, прим. 3, упр. 6.26)

21. Распределение Стьюдента $S(n)$

— определение и свойства (§ 1.2, п. 6)
— моменты (упр. 1.49)
— связь с бета-распределением (упр. 1.50)
— квантили (прилож. табл. 5).

22. Распределение Коши $\mathcal{K}(a)$

— определение и свойства (§ 1.2, п. 6, упр. 1.48(б), 2.26, 3.26)
— моделирование (§ 1.3, п. 6)
— распределение верхнего экстремума в больших выборках (упр. 2.35)
— информация Фишера для него (§ 3.2, табл. 1)
— достаточная статистика для него (§ 3.3, прим. 5)
— о. м. п. параметра, метод накопления (§ 3.5, прим. 8)
— асимптотическая эффективность выборочной медианы (упр. 3.47)
— л. н. м. несмещенный критерий при двусторонней альтернативе (§ 5.3, прим. 6)
— критерий Неймана—Пирсона (упр. 5.2).

23. Распределение Снедекора (F-распределение) $S(n, m)$

— определение и свойства (§ 1.2, п. 7, упр. 1.51(а))
— связь с распределением Стьюдента (упр. 1.51(б))
— связь с бета-распределением (упр. 1.52, 1.53)
— связь с показательным распределением (упр. 1.54)
— квантили (прилож. табл. 6).

24. Распределение Вейбулла $W(a, \alpha, b)$

— определение и свойства (§ 1.2, п. 8, упр. 1.55)
— распределение нижнего экстремума (упр. 1.56)
— сверхэффективная оценка параметра (§ 3.2, прим. 6)
— оценивание параметрических функций (§ 3.4, прим. 15, упр. 3.40)
— о. м. п. параметра θ в модели $W(\theta, \alpha, b)$ (упр. 3.50)
— о. м. п. параметра θ в модели $W(0{,}2, \sqrt{\theta})$ (упр. 3.51)

— доверительный интервал для параметра θ модели $W(0, \alpha, \theta^{1/\alpha})$ (упр. 3.65)
— несмещенный критерий при двусторонней альтернативе для модели $W(0, \alpha, \theta^{1/\alpha})$ (упр. 5.20).

25. Распределение Парето

— определение и свойства (§ 1.2, п. 9, упр. 1.57)
— связь с показательным распределением (упр. 1.58)
— сопряженность к равномерному распределению (упр. 1.53, 1.64(6))
— распределение верхнего экстремума в больших выборках (§ 2.4, прим. 5) оценивание параметрических функций (§ 3.4, прим. 16)
— прогноз (упр. 6.27).

26. Распределение Дирихле $D(\alpha_1, \ldots, \alpha_k)$

— определение и свойства (§ 1.2, п. 10, упр. 1.61)
— как смешивающая мера в модели Маркова—Пойа (упр. 1.62)
— сопряженность к полиномиальному распределению (упр. 1.63).

27. Распределение Паскаля (§ 1.1, п. 2)

28. Распределение Эрланга (§ 1.1, п. 3)

29. Распределение логарифмическое (§ 1.1, п. 8)

30. Распределение логнормальное (§ 1.2, п. 1)

31. Распределение Релея (§ 1.2, п. 8)

32. Распределение Колмогорова (§ 2.1, п. 3, прилож. табл. 7)

33. Распределение дважды экспоненциальное (§ 2.4, п. 4)

34. Распределение Лапласа (§ 3.4, прим. 17, упр. 3.53, 3.74)

35. Распределение логистическое (упр. 3.25)

36. Распределение обратное гауссовское (упр. 3.30)

37. Распределение Кептайна (упр. 3.44)

38. Распределение Уишарта (§ 8.3)

Только терпеливый закончит дело, а торопливый упадет.

Саади [2)]

Это не конец, и даже не начало конца. Но, возможно, это — конец начала.

У. Черчилль [3)]

[2)] Саади (1210–1292) — великий персидский поэт.

[3)] Черчилль Уинстон (1874–1965) — английский политический деятель.

Именной указатель

Аристотель 239

Большев Логин Николаевич 27
Байес Томас 85, 251, 511, 527
Басу Дебабрата 196
Бернулли Даниил 317
Бернулли Якоб 25, 31–34, 36, 37, 39, 64, 74, 91, 92, 107, 114, 150, 164, 173, 252, 253, 298, 309, 368, 424, 453, 545
Бернштейн Сергей Натанович 69
Бертран Жозеф Луи Франсуа 21
Блекуэлл Дэвид 191
Боровков Александр Алексеевич 27
Борткевич Л. 36
Буняковский Виктор Яковлевич 171, 176, 481, 509
Бхаттачария 181, 182, 184, 300, 301
Бэбингтон Смит Б. 360, 364
Бюффон Жорж Луи Леклерк 150, 323, 586

Вальд Абрахам 26, 348, 372, 443–451, 526, 593
Ван дер Варден Бартел Лендерт 311
Вейбулл Валодди 57, 84, 179, 189, 218, 302, 304, 455, 598
Вейерштрасс Карл Теодор Вильгельм 69
Верхарн Эмиль 573
Вийон Франсуа 28
Вилкоксон Фрэнк 349, 350
Волфовиц Джекоб 451

Гаек Я. 333
Гальтон Фрэнсис 507, 508
Гаусс Карл Фридрих 25, 46, 457, 459
Генри О. (Уильям Сидни Портер) 25
Гессе Герман 14
Гёте Иоганн Вольфганг 17, 572
Гливенко Валерий Иванович 99, 263
Гнеденко Борис Владимирович 26
Горвиц 290, 291, 590
Грасиан Бальтасар 29
Гулик Роберт ван 520
Гумбольдт Вильгельм 524

Дейвид Ф. 338
Декарт Рене 89
Демокрит 586
Джеймс 258, 594
Джеффрис 571
Диаконис Перси 457
Дирихле Петер Густав 57, 58, 85, 86, 255, 594, 597, 598

Йенсен Иоганн Людвиг 337, 386, 393
Йинг 263

Каплан 262
Кельвин (Томсон Уильям) 19
Кендалл Морис Джордж 26, 358–360, 364
Кентайн 303, 599
Кетле Ломбер Адольф Жак 164
Кнут Дональд 365
Колмогоров Андрей Николаевич 26, 37, 98, 100, 191, 195, 197, 317–320, 341, 369, 518, 579, 592, 599
Конфуций 163, 519
Коши Огюстен Луи 55, 56, 67, 82, 154, 156, 169, 171, 175, 176, 179, 190, 226, 240, 243, 300, 303, 405, 452, 481, 509, 591, 596, 597
Крамер Габриэль 26, 173, 175–184, 220, 232, 317
Кэттел Рэймонд Бернард 563

Лагранж Жозеф Луи 267, 270, 335, 434
Лаплас Пьер Симон 25, 46, 80, 95, 151, 215, 218, 304, 308, 317, 333, 345, 390, 453, 587, 599
Ле Кам 233
Лей 263
Лейбниц Готфрид Вильгельм 28
Леман Эрих 26, 168, 219
Линден-Бэлл 263
Линник Юрий Владимирович 27
Ломоносов Михаил Васильевич 199
Ляпунов Александр Михайлович 26

Манн Генри 349, 350
Марков Андрей Андреевич 26, 37–42, 70–72, 81, 85, 86, 297, 367, 372, 427–431, 433–435, 440, 442, 456, 457, 459, 587, 589–591, 597, 598
Махаланобис Прасанта Чандра 383, 537–539

Мейер 262
Менделеев Дмитрий Иванович 473, 512
Мендель Грегор Иоганн 323, 324, 590
Мизес Гаральд Рихард фон 317
Муавр Абрахам де 25, 46, 52, 80, 95, 151, 308, 333, 345, 390, 453, 587
Мэндервил К. 25

Нейман Джон фон 317, 364
Нейман Юрий 26, 372–384, 386, 389–391, 394, 395, 401, 409, 444, 445, 449, 450, 453, 454, 536, 587–590, 592, 594, 595, 597, 598
Ницше Фридрих 159
Ньютон Исаак 11, 32, 106, 297

Октавиан Август 62

Парето Вильфредо 57, 84, 86, 87, 140, 218, 310, 524, 595, 598
Паскаль Блез 35, 324, 520, 572, 599
Пирсон Карл 26, 239, 241, 317, 320, 321, 366, 368, 420, 566, 596
Пирсон Эгон Шарп 26, 317, 372–384, 386, 389, 390, 394, 395, 401, 409, 444, 445, 449, 450, 453, 454, 536, 587–590, 592, 594, 595, 597, 598
Питмен 247–250, 592, 596
Платон 14
Пойа Дьердь 37–42, 44, 81, 85, 86, 172, 297, 587, 589, 597, 598
Прохоров Юрий Васильевич 13, 27
Пуанкаре Анри 44
Пуассон Симеон Дени 31, 35, 36, 42, 43, 75, 76, 78, 80, 227, 243, 308, 391, 524, 574, 587, 588, 594
Пушкин Александр Сергеевич 164

Рао Кальямпуди Радхакришна 26, 173, 175–184, 191, 220, 232, 233
Рассел Бертран Артур Уильям 318
Рейнольдс Дж. 164
Романовский Всеволод Иванович 27
Русас Дж. 45, 317
Руставели Шота 12, 368
Рыбников Константин Алексеевич 30
Рюккерт Фридрих 564

Саади Муслихат-Дин 599
Садовничий Виктор Антонович 11
Салтыков-Щедрин Михаил Евграфович 584
Севастьянов Борис Александрович 13, 27
Секей Габор 163, 317
Сенека 74
Слуцкий Евгений Евгеньевич 27
Смирнов Николай Васильевич 27, 99, 138, 152, 278, 317, 320, 334, 340, 341, 370, 580
Снедекор Джордж 56, 83, 149, 150, 306, 307, 415, 416, 477, 479, 483, 501, 578, 595, 597, 598
Соловьев Владимир Сергеевич 572
Спенсер Герберт 451
Спирмен 357, 358, 360, 361
Стейн К. 257, 258, 594
Стилтьес Томас Ян (мл.) 162
Стирлинг Джеймс 52, 134
Стьюарт Алан 26
Стьюдент (Уильям Д. Госсет) 55, 83, 124, 147, 149, 307, 410, 414, 477, 577, 597, 598

Твен Марк (Сэмюэль Ленгхорн Клеменс) 318
Тейлор Брук 122, 225, 231, 403, 417
Типпет Леонард 364
Томпсон В. 24, 290, 291, 590

Уилкс Самюэль Стэнли 26
Уитни 349, 350
Уишарт 550, 594, 599

Феллер Вильям 211
Фишер Роальд Э. 26, 53, 55–57, 122, 145–149, 167, 174, 175, 182, 186, 217, 219, 225, 265, 269, 278, 299, 317, 328, 333, 369, 404, 411, 412, 416, 483, 587, 588, 591, 594, 597
Франс Анатоль (Анатоль Франсуа Тибо) 584
Фурье Жан Батист Жозеф 515

Хадсон 150
Хотеллинг Гаральд 566, 568
Хэмминг Ричард Уэсли 28

Чебышев Пафнутий Львович 26, 94, 110, 111, 113, 228, 298, 326, 470–473
Черчилль Уинстон 599
Честертон Гилберт Кит 525
Чистяков Владимир Павлович 13, 17

Шидак З. 333
Шильд А. 546
Ширяев Альберт Николаевич 27

Эггенбергер Ф. 39
Эйлер Леонард 132, 137
Эрланг Агнер Краруп 36, 52, 66, 67, 136, 599
Эсхил 163