

**Вероятностно-статистический анализ
хаотических процессов с помощью
смешанных гауссовских моделей.**

**Декомпозиция волатильности финансовых
индексов и турбулентной плазмы**

В. Ю. Королев

МГУ им. М. И. Ломоносова, факультет ВМиК;
Институт проблем информатики РАН.
E-mail: vkorolev@comtv.ru

11 февраля 2008 г.

Содержание

Введение	8
1. Теоретические основы негауссовых вероятностных моделей хаотических процессов	28
1.1. Случайные блуждания с дискретным временем	28
1.2. Однородные случайные блуждания с непрерывным временем	33
1.3. Неоднородные случайные блуждания, порождаемые обобщенными дважды стохастическими пуассоновскими процессами (обобщенными процессами Кокса)	37
1.3.1. Процессы Кокса: определения и примеры	37
1.3.2. Обобщенные процессы Кокса	41
1.3.3. Центральная предельная теорема для обобщенных процессов Кокса	43
1.3.4. Сходимость распределений обобщенных процессов Кокса к устойчивым законам	45
1.3.5. Оценки точности аппроксимации распределений обобщенных процессов Кокса масштабными смесями нормальных законов	47
1.3.6. Закон больших чисел для обобщенных процессов Кокса	53
2. Моделирование распределений приращений финансовых индексов смесями нормальных законов	55
2.1. Неоднородность операционного времени и нормальные смеси	55
2.2. Неоднородный дискретный стохастический хаос и процессы Кокса	59
2.3. Островеершинность масштабных смесей нормальных законов	72
2.4. Случай элементарных приращений с ненулевыми средними	74
2.5. Модели в рамках асимптотической схемы серий	77
2.6. От одномерных смешанных нормальных распределений к процессам Леви со случайным временем	82
2.6.1. Предварительные сведения. Процессы Леви. Пространство Скорохода	83

2.6.2. Функциональные предельные теоремы для обобщенных процессов Кокса	89
2.6.3. Примеры предельных процессов	94
3. Некоторые свойства смесей нормальных законов	115
3.1. Основные определения	116
3.2. Идентифицируемость смесей вероятностных распределений	120
3.3. Устойчивость нормальных смесей относительно смешивающего распределения	125
3.3.1. Прямая задача	126
3.3.2. Обратная задача	129
3.4. Самоподобие смешанных гауссовских процессов	139
3.5. Сужение класса смешивающих законов. Масштабные смеси нормальных законов как результат симметризации	142
3.5.1. Сверточная модель симметризации	142
3.5.2. Рандомизационная модель симметризации	150
3.6. Взаимосвязь поведения хвостов масштабных смесей нормальных законов и соответствующих смешивающих распределений	155
3.6.1. Экспоненциальное убывание хвоста смеси	156
3.6.2. Экспоненциально-степенное убывание хвоста смеси	160
3.6.3. Степенное убывание хвоста смеси	166
4. Волатильность	173
4.1. Понятие волатильности	173
4.2. Возможные подходы к определению и интерпретации волатильности	176
4.2.1. Теоретико-вероятностные характеристики изменчивости случайного процесса	176
4.2.2. Историческая (статистическая) волатильность	183
4.2.3. Неявная (подразумеваемая) волатильность	186
4.2.4. Вариационная волатильность	189
4.2.5. Различные интерпретации волатильности	189
4.2.6. Волатильность и информация. “Структурные” и “временные” компоненты волатильности	191
4.3. Подход, основанный на модели смеси распределений вероятностей	193
4.3.1. Декомпозиция волатильности	193
4.3.2. Разложение классической волатильности на динамическую и диффузионную компоненты	195
4.3.3. Преобразование волатильности при временном скейлинге	198
4.3.4. Диффузионный спектр и предполагаемый диффузионный спектр	201

5. Статистическое разделение смесей вероятностных распределений	205
5.1. Общая схема решения задачи	205
5.2. Задача разделения смесей вероятностных распределений и методы ее решения	207
5.2.1. Задача разделения смесей вероятностных распределений	207
5.2.2. Метод моментов	209
5.2.3. Метод максимального правдоподобия	209
5.3. ЕМ-алгоритм и его применение к задаче оценивания неизвестных параметров смесей распределений вероятностей	210
5.3.1. Предварительные сведения	212
5.3.2. Общее описание ЕМ-алгоритма	215
5.3.3. Монотонность ЕМ-алгоритма	216
5.3.4. Проксимальные алгоритмы (РР-алгоритмы) и их свойства	217
5.3.5. ЕМ-алгоритм как проксимальный алгоритм	217
5.3.6. Неподвижные точки ЕМ-алгоритма и локальные максимумы функции правдоподобия	220
5.3.7. Решение задачи разделения смесей вероятностных распределений с помощью ЕМ-алгоритма	221
5.3.8. Разделение конечных смесей нормальных распределений с помощью ЕМ-алгоритма	226
5.3.9. Выбор начальных приближений для ЕМ-алгоритма	230
5.3.10. Правила остановки ЕМ-алгоритма	230
5.4. Модификации ЕМ-алгоритма	232
5.4.1. Медианные модификации ЕМ-алгоритма	232
5.4.2. SEM-алгоритм	237
5.4.3. CEM-алгоритм	242
5.4.4. MCEM и SAEM-алгоритмы	243
5.5. ЕМ-алгоритм с большим числом компонент как средство построения непараметрических оценок плотности	246
5.6. Разделение смесей многомерных нормальных распределений с помощью ЕМ-алгоритма	247
5.7. Приближенное разделение конечных смесей с помощью метода фиксированных компонент для выбора начального приближения ЕМ-алгоритма	249
5.7.1. Основная идея метода фиксированных компонент	249
5.7.2. Разделение конечных смесей вероятностных распределений с фиксированными компонентами при помощи метода наименьших квадратов	252
5.7.3. Разделение конечных смесей вероятностных распределений с фиксированными компонентами при помощи метода наименьших модулей	256

5.7.4.	Разделение конечных смесей вероятностных распределений с фиксированными компонентами при помощи “усеченного” ЕМ-алгоритма	259
5.7.5.	Разделение конечных смесей вероятностных распределений с фиксированными компонентами при помощи байесовской классификации	260
5.8.	Выбор модели (определение типа и числа компонент смеси)	260
5.8.1.	Некоторые сведения из теории проверки сложных статистических гипотез	262
5.8.2.	Проверка значимости динамической составляющей волатильности с помощью критерия отношения правдоподобия	265
5.8.3.	“Последовательный” критерий отношения правдоподобия для определения числа компонент смеси	267
5.8.4.	Определение числа компонент смеси с помощью SEM-алгоритма	269
5.8.5.	Информационные критерии выбора модели (числа компонент и типа смеси)	271
5.9.	Примеры применения ЕМ-алгоритма в “статическом” режиме	278
6.	Скольльзящее (динамическое) разделение смесей вероятностных распределений. СРС-метод: описание и практическое применение	281
6.1.	Общая идея метода скольльзящего разделения смесей	281
6.2.	Оценивание предполагаемого диффузионного спектра с помощью СРС-метода	283
6.3.	СРС-метод с оцениванием значений компонент волатильности	285
6.3.1.	Критерий остановки	286
6.3.2.	Зависимость СРС-метода от выбора начальных приближений	286
6.3.3.	Визуализация результатов	292
6.3.4.	СРС-метод с ядерными оценками компонент волатильности	294
6.3.5.	О влиянии ширины окна на результат применения СРС-метода	298
6.3.6.	О влиянии масштаба (длины интервала времени между наблюдениями) на результат применения СРС-метода	300
6.3.7.	Промежуточные итоги	303
6.4.	СРС-метод с фиксированными значениями компонент волатильности	304
6.4.1.	Основная идея СРС-метода с фиксированными значениями компонент волатильности	304
6.4.2.	Дискретизация пространства компонент волатильности	305
6.4.3.	Сравнительная эффективность СРС-метода с фиксированными значениями компонент волатильности	308
6.4.4.	Сравнение версий СРС-метода с оцениваемыми и фиксированными значениями компонент волатильности	311

6.4.5. Эффект “насыщения” (эффект “достаточности” числа компонент) масштабных смесей, оцененных СРС-методом. Еще раз о том, что такое волатильность	312
6.5. СРС-метод с оцениванием диффузионных и динамической компонент волатильности	314
6.6. Исследование динамики корреляционной структуры хаотических стохастических процессов с помощью скользящего разделения смесей многомерных наблюдений	316
6.7. Примеры применения СРС-метода к анализу влияния информационных интервенций на состояние финансовых рынков	317
6.7.1. Взрывы в лондонском метро в июле 2005 года	317
6.7.2. Падение российского рынка в период с февраля по март 2007 года	320
6.7.3. События в США 11 сентября 2001 года	322
7. Применение СРС-метода к изучению турбулентности	324
7.1. Гидродинамическая и плазменная турбулентность	324
7.2. Феноменология турбулентности. Перемежаемость	326
7.3. Параметрические смешанные гауссовские модели турбулентности	330
7.4. “Статическая” подгонка смесей нормальных законов к эмпирическим распределениям, полученным в экспериментах с турбулентной плазмой	338
7.5. Результаты применения метода скользящего разделения смесей	342
Список литературы	346
Иллюстрации	375

Моей Королеве

Введение

“Во многих случаях я принужден лишь гадать или предполагать. Я нисколько не обманываюсь и отлично знаю, что не только не решаю вопроса во всей его полноте, но что остается поработать над ним в 1000 раз больше, чем я работал. Моя цель – возбудить к нему интерес, указав на великое значение его в будущем и на возможность его решения...”

В далеком будущем уже виднеются сквозь туман перспективы до такой степени обольстительные и важные, что о них едва ли теперь кто мечтает.”

К. Э. Циолковский. *“Ракета в космическое пространство”*.

Главную роль в данной работе будут играть специальные модели хаотических стохастических процессов, прежде всего, процессов, описывающих эволюцию финансовых индексов, и процессов, наблюдаемых в турбулентных потоках. Эти модели основаны на предельных теоремах для обобщенных дважды стохастических пуассоновских процессов (обобщенных процессов Кокса), приводящих к тому, что распределения приращений рассматриваемых процессов могут быть представлены в виде смеси нормальных законов. Перед тем как предоставить необходимую теоретическую базу для подобных моделей хаотических стохастических процессов, здесь следует дать краткое введение в проблематику математического моделирования хаотических стохастических процессов, историю математического моделирования эволюции финансовых индексов и функционирования финансовых рынков в целом, остановиться на особенностях анализа волатильности и сформулировать общепринятые предположения, делающие возможным решение упомянутых задач.

Все предлагаемые и описываемые в данной книге модели и методы могут быть использованы не только при анализе процессов, характеризующих эволюцию финансово-экономических показателей. Они также применимы в других областях, где рассматриваются процессы с хаотическим

и/или непредсказуемым поведением. Примером является физика турбулентной плазмы. Казалось бы, хаотические процессы на финансовых рынках принципиально отличаются от хаотических процессов плазменной турбулентности – их формирование оказывается сильно зависимым от того, знают ли наблюдатели (участники рынка), как развивается процесс, в то время как процессы, протекающие в токамаках или стеллараторах, будут развиваться так, как они должны были развиваться, вне зависимости от того, правильно или неправильно выписаны уравнения, описывающие соответствующие физические процессы. Однако при математическом моделировании финансовых рынков и турбулентных потоков заметны некоторые принципиальные аналогии. Процессы, протекающие на финансовых рынках и в турбулентных потоках, проявляют идентичные свойства: и тем и другим свойственны непредсказуемость, *перемежаемость* и в той или иной степени (локальная) *масштабная инвариантность* (или (мульти-) *фрактальность*). Под перемежаемостью обычно понимают нерегулярное чередование периодов времени, в течение каждого из которых вероятностно-статистические закономерности развития процесса сохраняются примерно неизменными. Масштабная инвариантность позволяет автоматически пересчитывать закономерности, наблюдаемые на конкретных временных масштабах (например, малых), на другие масштабы (например, большие) с помощью простого (степенного или логарифмически-линейного) преобразования.

В турбулентных потоках или в турбулентной плазме выявлены различные типы структур – вихри, солитоны, кавитоны и пр. Точно так же, информированность участников рынка о поведении биржевых цен создает свои типы структур в динамике финансовых рынков.

Несмотря на большое число выдающихся математических результатов в этих областях, ни в финансовой математике, ни в теории турбулентности пока нет всеобъемлющей математической теории, позволяющей с единых позиций описать все многообразие различных эффектов, наблюдаемых на финансовых рынках и в турбулентных потоках. Как следствие, и в финансовой математике, и в теории турбулентности имеется очень много феноменологических (в том числе теоретико-вероятностных) моделей, ориентированных на описание конкретных эффектов.

В данной книге сделана попытка предоставить дополнительные аргументы в пользу весьма удобного класса математических моделей стохастических хаотических процессов – так называемых подчиненных винеровских процессов, или процессов броуновского движения со случайным временем. Вследствие отмеченного выше отсутствия единой теории, объ-

ясняющей все эффекты, наблюдаемые на финансовых рынках и в турбулентных потоках, особую важность приобретают именно *внутренние*, аналитические механизмы формирования математических (прежде всего, вероятностных) моделей хаотических процессов. В данной книге в качестве аргументации в пользу указанных моделей типа процессов броуновского движения со случайным временем используется асимптотический подход, основанный на предельных теоремах теории вероятностей, в частности, на предельных теоремах для обобщенных дважды стохастических пуассоновских процессов (обобщенных процессов Кокса), которые в определенном смысле являются наилучшими математическими моделями неоднородных (и даже нестационарных) хаотических потоков. Такой подход дает возможность получить не только сами формальные вероятностные модели хаотических стохастических процессов, но и в некотором смысле дать разумное теоретическое объяснение их адекватности на основе минимальных предположений о внутренней структуре изучаемых характеристик, что чрезвычайно важно при решении задач анализа хаотических процессов в условиях стохастической неопределенности.

Как мы увидим ниже, в рамках указанных моделей взаимодействие частиц в плазме оказывается формально аналогичным процессу заключения сделок (контрактов купли-продажи) на современных финансовых рынках. Поэтому оказывается, что та же самая модель, которая разрабатывается в главе 2 для процесса эволюции финансовых индексов, может быть эффективно использована и при обработке данных, накопленных в экспериментах с турбулентной плазмой в токамаках и стеллараторах (см. главу 7).

Итак, для начала рассмотрим основные особенности используемого подхода к моделированию, анализу и интерпретации хаотических процессов, протекающих на финансовых рынках (и в турбулентной плазме).

Современные финансовые рынки представляют собой исключительно сложные системы, функционирование которых определяется очень большим (и практически не поддающимся учету) числом факторов и зависит от очень большого числа параметров. Эволюция биржевых цен, финансовых индексов (таких как, например, Dow–Jones, XETRA DAX, NIKKEI, NASDAQ и др.) и других величин, характеризующих объекты купли-продажи, подвержена влиянию таких факторов как общий экономический климат, политическая ситуация, микро- и макроэкономические показатели, рынок производных ценных бумаг (опционов, фьючерсов и др.) и, наконец, не в последнюю очередь, состояние самого рынка. Для простоты мы не будем делать различия между объектами купли-продажи на рын-

ке и будем говорить просто о *процессе* или *показателе*. То же относится и к другим возможным областям применения методов, разрабатываемых ниже. Под *процессом* или *показателем* мы будем подразумевать любой наблюдаемый и изучаемый процесс или временной ряд. Для основного наблюдаемого процесса мы будем использовать обозначение $P(t)$, $t \geq 0$.

Всюду ниже процесс (показатель) $P(t)$ будет интерпретироваться как *стохастический* процесс. Такая точка зрения основывается на предположении о том, что будущее развитие процесса нельзя предсказать абсолютно точно, в то время как механизмы, обуславливающие эволюцию показателя, чрезвычайно сложны. Другими словами, процесс $P(t)$ не будет считаться детерминированным. Такой подход, как правило, воспринимается как нечто само собой разумеющееся и подразумевает использование методов теории вероятностей и математической статистики для описания и анализа эволюции рассматриваемого процесса, что в дальнейшем не вызывает никакого дискомфорта. Однако, строго говоря, использование теоретико-вероятностных рассуждений при моделировании процессов эволюции финансовых индексов и особенно процессов плазменной турбулентности нуждается в особом соглашении, дополнительном комментарии о том, что при таком подходе можно получить, а чего нельзя, чем можно пренебречь, а чем нельзя.

В самом деле, наряду с теорией стохастических процессов (являющейся разделом теории вероятностей), в последние десятилетия быстро и эффективно развивается теория детерминированного хаоса. В рамках этой теории известно много примеров абсолютно детерминированных (нелинейных) динамических систем, эволюция которых при заданном начальном состоянии теоретически абсолютно точно известна (предсказуема), но которые в то же время обладают свойствами неустойчивости относительно начального состояния или “перемешивания”, проявляющимися в том, что малая погрешность в определении начального состояния приводит к экспоненциально возрастающему расхождению фазовых траекторий или “хаотическому размазыванию” траекторий по фазовому пространству. Асимптотическое поведение таких динамических систем характеризуется наличием так называемых *хаотических*, или *странных*, или *гиперболических* аттракторов фазовых траекторий. Подобные системы реализуют общепринятые представления о системах с хаотическим и/или непредсказуемым поведением. В то же время они наглядно проявляют парадоксальное свойство: *теоретически*, если начальное состояние известно абсолютно точно, то дальнейшая эволюция такой системы однозначно предопределена; однако *на практике* вычисленная траектория развития системы оказывается

произвольно далекой от реальной, так как начальное состояние, как правило, либо не известно абсолютно точно, либо не может быть абсолютно точно определено, поскольку, например, при компьютерном моделировании (скажем, численном решении соответствующих уравнений) информация представляется в дискретном виде, так что *всегда* существует реальный порог чувствительности (различимости).

В отличие от языка теории детерминированного хаоса в рамках исследования нелинейных динамических систем, язык теоретико-вероятностных и статистических рассуждений и моделей позволяет, избегая абсолютно точного описания поведения траектории сложной системы, выявить, тем не менее, некоторые закономерности в ее развитии и определить в некотором смысле интегральные характеристики и *их* эволюцию. Другими словами, язык стохастики используется для маскировки практической невозможности абсолютно точного описания поведения сложных хаотических динамических систем.

Однако для адекватности стохастических моделей необходимо, чтобы имелась хотя бы теоретическая возможность воспроизведения моделируемой ситуации в остающихся неизменными условиях как угодно много раз. Подход к моделированию и анализу хаотических процессов, описываемый ниже, сочетает это требование (на финансовых рынках взаимодействуют очень много участников, равно как в плазме взаимодействуют очень много частиц) с возможностью отслеживания и визуализации динамики изменений закономерностей в поведении рассматриваемой системы.

История рассматриваемой задачи тесно связана с развитием теории вероятностей и весьма примечательна. Одну из первых серьезных математических моделей для процесса биржевых цен предложил Л. Башелье в своем основополагающем труде “Теория спекуляции”, опубликованном в 1900 году (Bachelier, 1900). Отталкиваясь от предположений некоррелированности и стационарности приращений, он пришел к той же конструкции, на основе которой позже (в 1905 году) А. Эйнштейн предложил модель для процесса броуновского движения (Einstein, 1905), а именно, к одномерному уравнению диффузии. Как известно, дальнейшее развитие этой конструкции приводит к использованию винеровских процессов в качестве моделей для броуновского движения частицы и, стало быть, для процессов цен акций.

Приращения винеровского процесса имеют нормальное распределение. Винеровские процессы могут считаться в определенном смысле предельными моделями и тесно связаны с центральной предельной теоремой, применение которой к анализу распределений приращений биржевых цен сво-

дится к следующим рассуждениям.

Разбивая интервал времени $[t, t + \tau]$ на некоторое число непересекающихся подынтервалов, можно заметить, что приращение $P(t + \tau) - P(t)$ представляет собой сумму приращений процесса $P(t)$ на подынтервалах. В предположении, что приращения на подынтервалах независимы (а отмеченная выше некоррелированность приращений делает это предположение весьма разумным), в соответствии с центральной предельной теоремой распределение приращения $P(t + \tau) - P(t)$ приближается к нормальному при увеличении числа подынтервалов, если выполнено условие Линдеберга. Это условие означает, что ни одно из слагаемых не играет доминирующей роли в сумме. Если подынтервалы имеют одинаковую длину, то естественно считать, что распределения приращений на подынтервалах одинаковы. В этом случае условие Линдеберга сводится к требованию конечности дисперсий элементарных приращений.

Однако статистический анализ реальных данных показывает, что распределения приращений процессов биржевых цен на интервалах времени сравнительно небольшой длины (до 2–3 недель) отличны от нормальных. Первые работы, в которых отмечено это явление, появились еще в 1915 году. Результаты очень серьезного статистического анализа, подтверждающего отличие упомянутых распределений от нормального, были опубликованы М. Кендаллом в 1953 году (М. Kendall, 1953). Отмеченный феномен является всеобщим: не-нормальность распределений приращений проявляется на всех биржах независимо от объекта торговли. Переход к логарифмам, целесообразность которого описывается далее и который приводит к так называемому геометрическому броуновскому движению, не спасает ситуацию. Распределения приращений логарифмов биржевых цен также не нормальны. Поэтому винеровские процессы оказались отнюдь не бесспорными моделями динамики биржевых цен.

Отмеченная не-нормальность распределений приращений проявляется в том, что в действительности наблюдается заметно больше очень больших и очень маленьких по абсолютной величине значений приращений, нежели их должно быть в соответствии с нормальным распределением. Другими словами, наблюдаемые распределения приращений биржевых цен на интервалах времени умеренной длины являются более островершинными, нежели нормальные, имея заметно более тяжелые хвосты. Точно такое же свойство присуще эмпирическим распределениям вероятностей, возникающим при обработке данных, полученных в экспериментах с турбулентной плазмой (правда физики иногда вместо “хвостов” распределений говорят об их “крыльях”).

Наряду с островершинностью и наличием тяжелых хвостов у эмпирических распределений приращений рассматриваемых процессов статистический анализ иногда показывает наличие некоторой зависимости: по крайней мере последовательные приращения на интервалах времени очень малой длины зачастую оказываются коррелированными.

Предпринимались многие попытки объяснить неприменимость центральной предельной теоремы к анализу реальных хаотических процессов. В разделе 1.1 будут обсуждаться некоторые из них, в частности, попытки, основанные на идее Б. Мандельброта об использовании вместо нормального так называемых устойчивых законов (Mandelbrot, 1963), обусловленном предположением отсутствия второго момента (и, следовательно, дисперсии) у распределений приращений рассматриваемого процесса на элементарных временных интервалах. Но для отсутствия дисперсии необходимо, чтобы упомянутые элементарные приращения могли с положительной вероятностью принимать как угодно большие по абсолютной величине значения. В то же время в большинстве приложений нет реальных оснований отвергать предположение об ограниченности влияния каждого случайного фактора, воздействующего на регистрируемый процесс, и, как следствие, об ограниченности приращений регистрируемых процессов в пространстве.

В некоторых работах (см. обзоры, например, в (Ширяев, 1998) и (Hull, 2003)) отмечается, что другой возможной причиной упомянутых отклонений эмпирических распределений приращений биржевых цен от нормального является то, что на практике элементарные слагаемые (приращения) оказываются *неодинаково* распределенными, в частности, имеют разные дисперсии (другими словами, наблюдается эффект *гетероскедастичности*), что как раз и приводит к наличию островершинности и тяжелых хвостов у соответствующих распределений вероятностей. Во многих работах неоднородность приращений процесса биржевых цен описывается с помощью так называемой “скачкообразно-диффузионной” (jump-diffusion) модели, суть которой в том, что наблюдаемый процесс (с точностью до перехода к логарифмам) представляется в виде комбинации процесса броуновского движения и обобщенного (compound) пуассоновского процесса, скачки которого соответствуют “резким” и относительно большим изменениям цен вследствие поступления неких новостей. В терминах, принятых в теории процессов Леви, такие модели называются процессами Леви с конечной активностью (см. раздел 2.6.1). Самый простой вариант такой модели был предложен С. Прессом (Press, 1967). Затем эта модель усовершенствовалась многими авторами (см., например, (Merton, 1976), (Kou, 2002) и (Maheu and McCurdy, 2004) и списки литературы в этих рабо-

тах). Несмотря на то что подобные модели иногда демонстрируют высокую адекватность, будучи *параметрическими*, они все же не учитывают *все* многообразие причин неоднородности приращений финансовых индексов. К сожалению, основные предпосылки этих моделей довольно механистичны и не могут дать исчерпывающий ответ на вопрос, почему все устроено именно так, а не иначе.

Ниже мы систематически рассмотрим эффект гетероскедастичности и попытаемся изучить его внутренние причины. Кратко основное направление наших исследований можно описать следующим образом. Отличительной особенностью всех классических подходов является предположение *однородности* операционного времени и *стационарности* приращений основного процесса. Именно это предположение и будет той ключевой точкой, на которую будет направлено “острие атаки” описываемых ниже методов на рассматриваемую проблему.

При статистическом анализе временных рядов вне зависимости от природы данных неизбежно возникает проблема, связанная с тем, что реально наблюдается не вся реализация (траектория) *непрерывного* процесса, но лишь ее значения в некоторые, как правило, равноотстоящие моменты времени, образующие *дискретное* множество. При этом поведение процесса между отсчетами (моментами фиксации наблюдений) неизвестно. Разумная реконструкция этого поведения, построение соответствующей адекватной модели как раз и представляет собой основную задачу статистического исследования временных рядов.

Чтобы построить исходную элементарную модель типа случайного блуждания, из которой с помощью предельного перехода затем будет получена наша основная модель хаотического случайного процесса, мы будем использовать не схему равноотстоящих во времени наблюдений, типичную для временных рядов, но сосредоточим внимание на схеме, типичной для анализа тиковой информации, в которой основную роль играют моменты (тики, ticks) изменения значения индекса. Многие исследователи отмечали, что реальная интенсивность торгов на бирже заметно изменчива. Однако первым, кто систематически исследовал неоднородность операционного времени на биржах, был П. Кларк. В начале семидесятых годов прошлого века в работах (Clark, 1970), (Clark, 1973) на примере биржевых цен на хлопок он предложил моделировать эволюцию финансовых индексов (с точностью до перехода к логарифмам) не с помощью стандартных винеровских процессов $W(t)$, но с помощью так называемых *подчиненных* винеровских процессов со случайным временем $W(X(t))$. В таком контексте (случайный) процесс $X(t)$, называемый *субординатором*, играет роль

операционного времени и имеет неубывающие траектории, выходящие из нуля, приращения которых в определенном смысле пропорциональны активности (интенсивности) торгов.

По Кларку, применение классической центральной предельной теоремы к решению задачи моделирования биржевых цен приводит к неадекватным результатам, так как на практике нарушается ключевое модельное предположение неслучайности числа слагаемых в сумме. Кларк отмечает, что *“число индивидуальных эффектов, суммарное влияние которых определяет изменение цены в течение дня, изменчиво и фактически является случайной величиной”*. В следующих разделах мы продемонстрируем, что даже если все остальные условия центральной предельной теоремы выполнены, путем специального подбора считающего случайного процесса, описывающего число элементарных слагаемых в сумме, можно в итоге вместо нормального закона получить предельные распределения с произвольно тяжелыми хвостами, у которых, в частности, отсутствуют моменты произвольно малого порядка.

Предваряя формальное описание модели, кратко остановимся на общей идее, используемой при ее построении. Пусть P_i – значение индекса (цены акции) в i -м заключенном контракте, а $N(t)$ – неотрицательный случайный процесс с неубывающими траекториями, описывающий количество контрактов, заключенных до момента времени t . Тогда абсолютное и относительное приращение процесса $P(t)$ могут, соответственно, быть записаны в виде

$$P(t + \tau) - P(t) = \sum_{i=N(t)}^{N(t+\tau)} [P_i - P_{i-1}] \quad (0.0.1)$$

и

$$\log(P(t + \tau)) - \log(P(t)) = \sum_{i=N(t)}^{N(t+\tau)} [\log(P_i) - \log(P_{i-1})]. \quad (0.0.2)$$

Предположим, что слагаемые в суммах удовлетворяют предположениям центральной предельной теоремы (стохастическая независимость и справедливость условия Линдеберга, то есть в определенном смысле “равноправие”). Дополнительно предположим, что слагаемые (элементарные приращения процесса или величины скачков) стохастически независимы от $N(t)$. Если мы, более того, предположим, что $N(t)$ – не “равномерно” возрастающий процесс, но имеет участки с большими (быстро накапливающимися) и малыми (медленно накапливающимися) приращениями, то сможем, даже не вдаваясь глубоко в необходимые математические формализмы,

объяснить некоторые наблюдаемые эффекты с помощью моделей (0.0.1) и (0.0.2).

Для начала рассмотрим нарушение условия одинаковой распределенности приращений на интервалах времени равной длины: если мы рассмотрим приращения процессов (0.0.1) или (0.0.2) на интервалах времени равной длины, то заметим, что в силу аддитивности дисперсии (при условии независимости (и, добавим, несмещенности) слагаемых) дисперсии каждого “суммарного” приращения будут в определенном смысле пропорциональны соответствующему приращению процесса $N(t)$, и, если $N(t)$, как отмечено выше, не является “равномерно” возрастающим процессом, то мы вправе в результате ожидать несовпадения дисперсий приращений процессов (0.0.1) или (0.0.2) на интервалах времени равной длины.

Нарушение условия независимости приращений может быть объяснено аналогично: если последовательные интервалы времени попадают в область высокой торговой активности (то есть быстрого возрастания $N(t)$), внутри которой приращения процесса $N(t)$ не являются независимыми – большое приращение $N(t)$ за единицу времени, скорее всего, вызовет большее его приращение за следующую единицу времени, – то все приращения процессов (0.0.1) или (0.0.2) будут иметь большие дисперсии и не будут независимыми. Аналогично, если последовательные интервалы времени попадают в область низкой торговой активности (то есть медленного возрастания $N(t)$), внутри которой приращения процесса $N(t)$ не являются независимыми – малое приращение $N(t)$ за единицу времени, скорее всего, вызовет малое его приращение за следующую единицу времени, – то все приращения процессов (0.0.1) или (0.0.2) будут иметь малые дисперсии и не будут независимыми. Отметим, что эти рассуждения справедливы, даже если элементарные приращения (последовательные скачки процесса $P(t)$) стохастически независимы от интенсивности торгов!

Как будет показано ниже, предельные распределения приращений процессов вида $\sum_{i=1}^{N(t)} X_i$ (см. (0.0.1) и (0.0.2)) и, стало быть, конечномерные распределения предельных процессов типа $X(t) = W(T(t))$, при довольно широких и естественных условиях принадлежат к классу сдвиг-масштабных смесей нормальных законов. Пример подгонки таких распределений к экспериментальным данным приведен на рис. 1.

Как видно из графиков, приведенных на рис. 1, подогнанные четырех-компонентные смеси демонстрируют высокое согласие с реальными данными и, как мы увидим в главах 5 – 7, могут дать много информации о том, в каком состоянии находится финансовый рынок. В то же время, три примера, приведенных на рис. 1, показывают, что нет “универсальной”

четырёхкомпонентной смеси, и более того, не видно даже “шаблона” или общего правила, по которому смешиваются нормальные законы, обеспечивающие наилучшее согласие, хотя все три примера относятся к одному и тому же временному ряду.

В качестве базовой математической модели процесса $N(t)$, описывающего число информативных событий – скачков рассматриваемого показателя – до момента времени t мы будем рассматривать дважды стохастический пуассоновский процесс (процесс Кокса), то есть пуассоновский процесс со случайной интенсивностью.

Как известно (см., например, (Bening and Korolev, 2002)), наилучшей моделью *однородного* стохастического хаотического потока событий является пуассоновский процесс, характеризуемый тем обстоятельством, что интервалы времени между событиями потока стохастически независимы и имеют одинаковое показательное распределение. Напомним, что привлекательность пуассоновского процесса в качестве модели однородного дискретного хаоса обусловлена как минимум двумя обстоятельствами. Во-первых, показательное распределение обладает максимальной дифференциальной энтропией среди всех абсолютно непрерывных распределений вероятностей, сосредоточенных на всей положительной полуоси и имеющих конечное математическое ожидание, а энтропия, как известно, является очень удобной численной характеристикой неопределенности. Во-вторых, точки (события) пуассоновского потока равномерно распределены на оси времени в том смысле, что для любого конечного интервала времени $[t_1, t_2]$ условное совместное распределение точек пуассоновского потока, попавших в интервал $[t_1, t_2]$, при условии, что в этот интервал попало фиксированное число, скажем, n точек, совпадает с совместным распределением вариационного ряда, построенного по независимой однородной выборке объема n из равномерного на $[t_1, t_2]$ распределения. Равномерное же распределение обладает максимальной дифференциальной энтропией среди всех абсолютно непрерывных распределений вероятностей, сосредоточенных на конечных интервалах, и очень хорошо соответствует общепринятому представлению об абсолютно непредсказуемой ограниченной случайной величине.

Однако в реальных “хаотических” системах хаос практически никогда не бывает однородным в пространстве или времени. Неоднородный и даже стохастический характер интенсивности потока информативных событий может быть вызван многими причинами. Эти причины могут быть как внутренними по отношению к изучаемой сложной системе, так и внешними. Внутренние причины могут проявляться, например, как непероди-

ческие или периодические компоненты (тренды), вызванные изменениями локальных (во времени) тенденций, например, связанных с проявлениями эффекта толпы на биржах или эффекта перемежаемости в турбулентной среде. Внешние причины могут проявляться, например, как следствие изменения внешних параметров в турбулентной плазме тороидальных магнитных ловушек (токамаков и стеллараторов) или как следствие тех или иных политических событий, оказывающих информационное влияние на финансовые рынки. Неоднородность хаоса может быть вызвана некоторыми случайно возникающими (не поддающимися абсолютно надежному прогнозированию) причинами. Как известно, наиболее разумными стохастическими моделями неоднородных хаотических точечных процессов являются *дважды стохастические пуассоновские процессы*, иначе называемые *процессами Кокса* (см., например, (Bening and Korolev, 2002)).

С помощью аппарата предельных теорем для обобщенных дважды стохастических пуассоновских процессов (обобщенных процессов Кокса) в этой книге будет предпринята попытка перекинуть “теоретический” мостик между описанными выше элементарными моделями динамики финансовых индексов типа чисто скачкообразных процессов с конечной активностью (конечным числом скачков на любых интервалах времени конечной длины) и весьма популярными ныне моделями типа подчиненных процессов Леви типа $P(t) = W(X(t))$. Примерами таких моделей являются популярные ныне в финансовой математике обобщенные гиперболические процессы, нормальные \ обратные гауссовские процессы, дисперсионные гамма-процессы, двусторонние гамма-процессы. Предельный переход осуществляется при неограниченно возрастающей интенсивности потока информативных событий, в результате чего упомянутые модели будут процессами с бесконечной активностью. При этом, как уже отмечалось, исходя из соображений практической целесообразности, можно считать, что у допредельных процессов величины скачков имеют конечные дисперсии. Заметим, что это предположение в определенном смысле приводит к противоречию с принятым многими исследователями принципом масштабной инвариантности или скейлинга (самоподобия) результирующих моделей хаотических стохастических процессов эволюции финансовых индексов или турбулентности (см., например, (Mandelbrot, 1997)).

Схема подчинения оказывается очень удобной для учета наличия интервалов времени с разной интенсивностью потока информативных событий, приводящего к заметно варьирующейся скорости изменений значений основного рассматриваемого процесса. В рамках этой схемы скорость изменений значений основного процесса естественно описывается субординато-

ром – “внутренним” процессом $X(t)$ в представлении $P(t) = W(X(t))$ (или $P(t) = \exp\{W(X(t))\}$, если рассматриваются относительные приращения). Субординатор часто называют внутренним или операционным временем или “хронометром” (chronometer, clock). Ниже будет показано, что его изменения напрямую связаны с изменением волатильности рассматриваемого процесса.

Следует отметить, что еще одним весомым доводом в пользу использования моделей типа $P(t) = W(X(t))$ (или $P(t) = \exp\{W(X(t))\}$) является следующее обстоятельство. Основным математическим аппаратом, при помощи которого вычисляются различные функционалы от случайных процессов, описывающих эволюцию финансовых индексов, например, рациональную стоимость опционов, является стохастическое исчисление (стохастический анализ). Этот аппарат хорошо развит для довольно обширного класса случайных процессов – так называемых семимартингалов (см., например, (Ширяев, 1998), том 1). Хорошо известно, что любой семимартингал может быть представлен в виде подчиненного винеровского процесса. А именно, в работе (Монрое, 1978) доказана теорема, согласно которой для любого семимартингала $Z(t)$, траектории которого стартуют из нуля, существует процесс броуновского движения (винеровский процесс) $W(t)$ и случайный субординатор (положительный семимартингал с возрастающими траекториями, выходящими из нуля) $X(t)$ такие, что $Z(t) = W(X(t))$. В финансовой математике хорошо известен результат, согласно которому условие отсутствия арбитража необходимо влечет существование такой меры, относительно которой процесс дисконтированной цены акции является мартингалом (см., например, (Ширяев, 1998)). Следовательно, как отмечено, скажем, в работе (Geman and Ané, 1996), относительно реальной меры этот процесс должен быть семимартингалом. Это же свойство присуще и логарифму дисконтированной цены акции. Таким образом, условие отсутствия арбитража необходимо влечет возможность представления $P(t) = W(X(t))$ (или $P(t) = \exp\{W(X(t))\}$).

Элементарные модели типа обобщенных процессов Кокса оказываются весьма удобными, поскольку на примере таких моделей можно наглядно проследить тесную взаимосвязь интенсивности потока информативных событий – потока скачков обобщенного процесса Кокса – с изменчивостью (волатильностью) рассматриваемого показателя $P(t)$ и субординатором (“хронометром”) $X(t)$ в представлении $P(t) = W(X(t))$ (или $P(t) = \exp\{W(X(t))\}$). В частности, если $P(t)$ – процесс цены акции, то, как показано, например, в работе (Ané and Geman, 2000) на основе статистического анализа большого массива экспериментальных данных об

объемах торгов, количествах сделок и соответствующих изменениях цены акции, из всех разумных характеристик именно процесс накопленного количества сделок $X_{\#}(t)$ оказался наилучшим субординатором в том смысле, что распределение приращений процесса логарифма цены акций $\log P(X_{\#}^{-1}(t))$ после обратного преобразования времени оказывается практически не отличимым от нормального. (Заметим, что асимптотическое поведение процесса Кокса совпадает с поведением соответствующей накопленной интенсивности, см., например, (Bening and Korolev, 2002).)

Ниже именно скорость изменений (изменчивость или *волатильность*) рассматриваемого процесса будет основным объектом нашего интереса. Пока мы не будем придерживаться никакого строгого математического определения волатильности. Обычно волатильность ассоциируется с коэффициентом диффузии или его аналогом в негауссовских моделях. Понятие волатильности давно используется в финансовом анализе, однако, насколько известно автору, в работах, посвященных изучению плазменной турбулентности, это понятие не встречается.

По-видимому, дело здесь в том, что в тех формальных конструкциях, которые используются для описания процессов плазменной турбулентности, изменчивость присутствует в совершенно определенном виде – как коэффициент диффузии в соответствующих уравнениях. К сожалению, существующие методы статистического оценивания коэффициента диффузии по экспериментальным данным приводят к крайне неустойчивым оценкам, так что при практических расчетах значение коэффициента диффузии выписывается в значительной степени “от лукавого”, что приводит к результатам, весьма далеким от реальности. Отмеченная неустойчивость обусловлена не тем, что имеются проблемы с вычислением эмпирической дисперсии, но тем, что существующие модели слишком упрощенно и потому грубо интерпретируют изменчивость рассматриваемых стохастических хаотических процессов как некий постоянный скаляр – число, которое наделяется смыслом “коэффициента диффузии”. Как будет показано ниже, в реальных процессах изменчивость, во-первых, имеет существенно многомерную структуру и, во-вторых, изменчива сама. Поэтому, не сводя изменчивость к коэффициенту диффузии, и при рассмотрении процессов плазменной турбулентности для описания изменчивости в целях единообразия будет использоваться термин “волатильность”.

В финансовой математике упомянутая выше схема “подчинения” естественно приводит к концепции переменной и даже стохастической волатильности, изменение или стохастический характер которой описывается с помощью субординатора $X(t)$.

Известно много моделей процессов, описывающих эволюцию финансовых индексов, которые учитывают стохастический характер волатильности. Достаточно упомянуть основополагающие работы Р. Ф. Энгла (Engle, 1982), где была предложена ARCH- (AutoRegressive Conditional Heteroskedastic) модель, и Т. Боллерслева (Bollerslev, 1986), в которой была предложена GARCH- (Generalized AutoRegressive Conditional Heteroskedastic) модель. Также заслуживает быть упомянутой работа (Hull and White, 1987), посвященная оцениванию опционов на активы со стохастической волатильностью, в которой и цена основного актива, и ее волатильность описываются стохастическими дифференциальными уравнениями. К счастью, нет необходимости приводить здесь подробный обзор результатов, полученных в рамках подхода, основанного на стохастической волатильности, поскольку имеется обстоятельная книга (Ширяев, 1998), в которой этому и многим другим разделам финансовой математики уделено значительное внимание и в которой заинтересованный читатель сможет найти подробное описание упомянутых выше моделей.

Во всех существующих работах, в которых используется концепция стохастической волатильности, последняя трактуется как некий случайный процесс, каждая траектория которого представляет собой скалярную (то есть одномерную) функцию времени. В противоположность такой традиции, основной результат данной работы – описание и обоснование подхода, основанного на методе скользящего разделения смесей (*CPC-методе*), который дает возможность интерпретировать волатильность как многомерную (векторную) функцию времени с помощью специального и довольно естественного разложения волатильности на динамическую и диффузионные компоненты.

Ниже будут приведены некоторые *различные* формальные определения понятия волатильности (одного и того же процесса!), используемые в работах по финансовой математике. Такая ситуация напоминает старинную индийскую притчу о пяти слепых мудрецах, каждого из которых спросили, что такое слон. Первый из них сказал, что слон – это бревно: он ощупал ногу слона. Второй сказал, что слон – это веревка: он держал слона за хвост. Третий сказал, что слон – это труба: он держал слона за хобот. Четвертый мудрец сравнил слона с большим листом пальмы: он потрогал ухо слона. Наконец, пятый мудрец сказал, что слон – это большой шар: ему довелось коснуться брюха слона.

Чтобы постараться в максимально возможной степени не уподобляться этим мудрецам и дать более полный ответ на вопрос о том, что такое волатильность, помня, тем не менее, о том, что среди финансовых анали-

тиков “волатильность” – это ныне скорее обыденное понятие, нежели математический термин, необходимо заметить, что волатильность процесса (то есть его изменчивость) обусловлена как минимум двумя типами факторов. Первый тип факторов может быть условно назван *динамическим*. Влияние факторов такого типа проявляется в том, что процесс изменяется из-за наличия некоторого тренда или комбинации, взаимодействия трендов, отражающих интересы некоторых (нескольких) групп участников рынка (простейшим примером таких групп являются “быки”, играющие на повышение, и “медведи”, играющие на понижение). Упрощенная физическая интерпретация действия этой группы факторов может быть проиллюстрирована примером движения, скажем, щепки в горной реке, где имеются ярко выраженные течения или комбинации течений, перемещающие щепку. Следует отметить, что обычно при описании методов анализа эволюции финансовых индексов понятие тренда в некотором смысле противопоставляется понятию волатильности, и оба эти понятия изучаются раздельно. Ниже будет показано, что на самом деле между этими понятиями имеется глубокая и непростая взаимосвязь, которую, к счастью, можно довольно разумно описать, используя аппарат теории вероятностей и математической статистики.

Факторы второго типа могут быть названы *стохастическими* или *диффузионными*. Как мы уже отмечали, число участников финансовых рынков огромно, так что на практике просто нереально учесть влияние каждого из них в отдельности. Физическим примером действия факторов диффузионного типа может служить суммарное воздействие молекул среды на частицу, испытывающую броуновское (тепловое) движение.

Конечно, эта классификация факторов скорее условна, нежели строго определена. Однако, как будет ясно из дальнейшего, СРС-метод, описываемый в данной работе, спонтанно и довольно естественно выделяет факторы указанных типов и автоматически дает возможность проследить их взаимосвязь. В соответствии со сказанным выше, в рамках основной модели, используемой в данной работе, волатильность довольно естественно раскладывается на *динамическую* и *диффузионную* составляющие. В свою очередь, специальными математико-статистическими процедурами диффузионная составляющая раскладывается на несколько разных компонент, каждая из которых имеет свое собственное происхождение. Таким образом, в данной работе предлагается метод статистического разложения волатильности *скалярного* процесса в нетривиальную и существенно *многомерную* информативную картину.

Как будет показано на примере исследования конкретных временных

рядов, описывающих эволюцию конкретных финансовых индексов, динамическая и диффузионные составляющие вносят в итоговую волатильность примерно одинаковый вклад, причем в разные моменты времени доминировать могут разные компоненты.

С методологической точки зрения, предлагаемый подход в определенном смысле аналогичен дисперсионному анализу: в обоих методах именно поведение изменчивости служит основой для статистических выводов о наличии или отсутствии влияния определенных факторов и тем самым о структуре динамической системы, в которой эволюционирует изучаемый процесс.

Общая схема анализа хаотических стохастических процессов, предлагаемая в данной работе, такова.

- (i) В качестве “элементарных” математических моделей динамики финансовых индексов или хаотических процессов плазменной турбулентности на очень малых временных масштабах рассматриваются обобщенные дважды стохастические пуассоновские процессы (обобщенные процессы Кокса) со скачками, имеющими конечную дисперсию. В соответствии с принципом максимальной энтропии такие процессы являются наилучшими моделями неоднородных хаотических потоков.
- (ii) Асимптотический подход, основанный на предельных теоремах для обобщенных процессов Кокса как моделей неоднородных хаотических случайных блужданий естественно приводит к макромоделям стохастических хаотических процессов типа подчиненных винеровских процессов (процессов броуновского движения со случайными и зависящими от времени коэффициентами сноса и диффузии). Таким образом, аппроксимации для распределений (логарифмов) приращений процессов эволюции финансовых индексов или процессов плазменной турбулентности следует искать в виде общих сдвиг/масштабных смесей нормальных законов, в которых смешивающий закон определяется накопленной (интегральной) интенсивностью потоков соответствующих информативных событий (элементарных скачков, “тиков”).
- (iii) Проблема статистической реконструкции распределений приращений упомянутых процессов сводится к задаче статистического оценивания смешивающего распределения, которое является параметром этой задачи.

- (iv) В самой общей постановке задача статистического оценивания смешивающего распределения является некорректной, так как общие сдвиг/масштабные смеси нормальных законов не являются идентифицируемыми. Таким образом, в рамках общего принципа регуляризации некорректных задач, исходная задача заменяется задачей отыскания решения, наиболее близкого к истинному в классе конечных дискретных сдвиг/масштабных смесей нормальных законов. Эта “редуцированная” задача уже является корректной и имеет единственное решение, так как семейство конечных дискретных сдвиг/масштабных смесей нормальных законов идентифицируемо. При этом возникает необходимость исследования свойства устойчивости упомянутых смесей относительно смешивающего закона, чтобы, зная оценки устойчивости, вычислить погрешности, образующиеся при замене исходной задачи редуцированной. При упомянутой регуляризации происходит автоматическое выделение типичных или более или менее устойчивых структур в эволюции рассматриваемых сложных систем. Для решения задачи оценки параметров конечных сдвиг/масштабных смесей нормальных законов используется ЕМ-алгоритм.
- (v) Представление распределений (логарифмов) приращений процессов эволюции финансовых индексов или процессов плазменной турбулентности в виде конечных сдвиг/масштабных смесей нормальных законов естественно приводит к многомерной интерпретации волатильности рассматриваемого процесса и к возможности разложения волатильности на динамическую и диффузионные компоненты.
- (vi) Статистические закономерности поведения рассматриваемых процессов, формализованные в пунктах (i) – (ii), изменяются во времени, результатом чего является отсутствие универсального смешивающего закона. Таким образом, чтобы изучить динамику изменения статистических закономерностей в поведении исследуемого хаотического процесса, задача статистического разделения конечных смесей нормальных законов должна быть последовательно решена на интервалах времени, постоянно сдвигающихся в направлении “астрономического” времени. Тем самым параметры смесей (параметры сдвига (дрейфа), масштаба (диффузии) и веса компонент) оцениваются как функции времени. При этом естественно возникают задачи, связанные как с выбором подходящих методов оценивания параметров сдвига и масштаба, так и с выбором оптимальных параметров вычис-

лительных процедур, реализующих эти методы: начального приближения, ширины скользящего интервала (окна), правила остановки и др.

- (vii) Наконец, для адекватной интерпретации результатов и для идентификации феноменологически выделенных (статистически оцененных) компонент, то есть для адекватного сопоставления статистически оцененных компонент с реальными процессами или явлениями, необходимо из многих возможных моделей выбрать наиболее адекватную, например, проверить, является выделенная динамическая компонента волатильности статистически значимой или нет.

В сжатом виде основные элементы предлагаемого подхода описаны в статьях (Королев, 2007а), (Королев, 2007с).

Чтобы заключить введение, необходимо заметить, что, как следует из вышесказанного, СРС-метод вбирает в себя много статистических процедур. Несмотря на то что СРС-метод имеет очень серьезные теоретические основания, такие как предельные теоремы для обобщенных процессов Кокса, оценки устойчивости специальных представлений вероятностных распределений в виде смесей нормальных законов и результаты о сходимости ЕМ-алгоритма, все же некоторые элементы метода основаны скорее на здравом смысле, нежели на строгих математических результатах. Поэтому к СРС-методу надо относиться как к довольно гибкому средству анализа данных, нежели как к идеальному математическому инструменту. Применение этого метода к реальным данным в большей степени представляет собой творчество, нежели серийное производство. Однако эти слова ни в коей мере не должны испугать или разочаровать читателя. Напротив, как будет показано, при умелом применении СРС-метод обладает большой силой, реально работает и дает чрезвычайно информативные результаты.

Картина, получаемая в результате применения СРС-метода к анализу реальных хаотических стохастических процессов эволюции финансовых индексов и плазменной турбулентности соответствует следующей феноменологической модели рассматриваемых процессов. Состояние наблюдаемой динамической системы описывается несколькими процессами типа броуновского движения с переменными (случайными) сносом и диффузией. Состояние системы наблюдается в равноотстоящие моменты времени, причем в каждый из них наблюдению доступен лишь один из упомянутых процессов, то есть в каждый момент времени наблюдается значение только одного из этих процессов. При этом в каждый момент времени наблюдаемый процесс выбирается случайно, и соответствующие вероятности

также изменяются во времени. Эти броуновские движения могут исчезать и возникать, вообще говоря, в любые (случайные) моменты времени. При этом удивительно то, что как при моделировании эволюции финансовых индексов, так и при исследовании процессов плазменной турбулентности, на практике для адекватного описания развития соответствующих процессов с помощью указанного подхода в каждый момент времени оказывается достаточным не более пяти-шести процессов типа броуновского движения с переменной диффузией.

На финансовом рынке взаимодействует огромное число участников, у каждого из которых свои интересы и своя стратегия поведения. Наблюдаемое значение финансового индекса в каждый момент времени является интегральным (суммарным, усредненным) результатом взаимодействия интересов и стратегий участников рынка. Подобно тому, как Исаак Ньютон с помощью призмы разложил белый свет на составляющие его цвета радуги, с помощью СРС-метода наблюдаемый интегральный индекс, характеризующий состояние рынка, можно в определенном смысле разложить на составляющие его компоненты, отражающие текущие интересы и стратегии характерных групп участников рынка.

Автор выражает искреннюю признательность С. Я. Шоргину, поддержавшему идею издания данной книги, Ю. В. Жукову, Т. А. Маравиной и М. Рэю за помощь в проведении вычислений и подготовку рисунков для данной книги, Н. Н. Скворцовой за полезные обсуждения и возможность использовать данные, полученные в реальных экспериментах с турбулентной плазмой, Д. Е. Кашееву, С. Н. Кокшарову, К. Ф. Соколовой и И. Г. Шевцовой, любезно разрешившим включить в книгу фрагменты их диссертаций. Автор также благодарен В. В. Саенко, внимательно прочитавшему рукопись и указавшему на некоторые погрешности, А. И. Володину, оказавшему ощутимую библиографическую поддержку, а также Т. В. Захаровой за полезные советы по оформлению оригинал-макета.

По соображениям полиграфической целесообразности рисунки помещены не в основной текст, а вынесены в отдельный раздел в конце книги.

Довольно обширная (но ни в коей мере не претендующая на полноту) библиография, приведенная в конце книги, призвана помочь читателю, заинтересовавшемуся затронутыми в книге вопросами, ориентироваться в литературе по вероятностно-статистическому моделированию хаотических стохастических процессов и смежным областям.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проекты 05-01-00396, 05-01-00583, 05-01-00535, 05-07-90103.

1.

Теоретические основы негауссовых вероятностных моделей хаотических процессов

И свое формируя сознание
С каждым днем от момента рожденья,
Мы бредем по дороге познания,
А с познанием приходит сомненье.

Андрей Макаревич.

В этой главе будут рассмотрены общие математические конструкции и схемы, представляющие собой фундамент для моделей и методов, описываемых в этой книге.

1.1. Случайные блуждания с дискретным временем

Чтобы объяснить отклонения наблюдаемых распределений от нормального (гауссовского), сначала рассмотрим математические обоснования самих гауссовских моделей. В основе гауссовских вероятностных моделей многих физических процессов лежит понятие однородного случайного блуждания. Суть этого понятия такова. Рассмотрим частицу, которая в *равноотстоящие* моменты времени испытывает мгновенные перемещения на случайные

расстояния. Формально координата частицы после $n \geq 1$ перемещений может быть записана в виде

$$S_n = X_1 + \dots + X_n, \quad (1.1.1)$$

где X_i – i -е перемещение частицы. В предположении о том, что $X_1, X_2, \dots$ – независимые случайные величины с конечными дисперсиями, удовлетворяющие так называемому *условию Линдеберга*, суть которого, грубо говоря, в том, что все слагаемые в суммах (1.1.1) в определенном смысле равноправны (условие Линдеберга, например, запрещает ситуацию, когда при $n \rightarrow \infty$ итоговое перемещение является результатом лишь конечного числа скачков), из центральной предельной теоремы теории вероятностей вытекает, что при $n \gg 1$ случайная величина S_n имеет нормальное (гауссово) распределение, задаваемое плотностью

$$\phi_{a,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x-a)^2}{2\sigma^2} \right\}, \quad x \in \mathbb{R}.$$

Здесь $\sigma > 0$ и $a \in \mathbb{R}$ – параметры масштаба и сдвига, соответственно. Стандартные нормальная плотность и функция распределения (соответствующие параметрам $\sigma = 1$ и $a = 0$) по традиции будут обозначаться $\phi(x)$ и $\Phi(x)$,

$$\phi(x) = \phi_{0,1}(x), \quad \Phi(x) = \int_{-\infty}^x \phi(y) dy \quad x \in \mathbb{R}.$$

Если слагаемые $X_1, X_2, \dots$ имеют одинаковые распределения, то условие Линдеберга сводится к требованию конечности дисперсий слагаемых и становится необходимым и достаточным для выполнения центральной предельной теоремы.

Доказаны так называемые *функциональные предельные теоремы*, устанавливающие сходимость при $n \rightarrow \infty$ (и одновременном соответствующем изменении масштаба) траекторий случайного блуждания (1.1.1) к траекториям винеровского процесса, являющегося математической моделью процесса броуновского движения (см., например, (Прохоров, 1956), (Леви, 1972), (Биллингсли, 1977)).

Винеровский процесс является гауссовским, то есть все его приращения имеют нормальное распределение. Винеровские процессы могут считаться в определенном смысле асимптотическими моделями и тесно связаны с центральной предельной теоремой, применение которой к анализу распределений приращений реальных процессов с непрерывным временем

сводится к следующим рассуждениям. Деля интервал времени $[t, t + \tau]$ ($t \geq 0, \tau > 0$) на некоторое число подынтервалов равной длины, мы замечаем, что приращение $P(t + \tau) - P(t)$ рассматриваемого физического процесса может быть представлено в виде суммы приращений процесса $P(t)$ на подынтервалах. В предположении, что приращения на подынтервалах независимы, в соответствии с центральной предельной теоремой распределение приращения $P(t + \tau) - P(t)$ приближается к нормальному при увеличении числа подынтервалов, если выполнено условие Линдеберга. Как уже говорилось, это условие означает, что никакая конечная группа слагаемых не играет доминирующей роли в сумме. Так как подынтервалы имеют одинаковую длину, то естественно считать, что распределения приращений на подынтервалах одинаковы. В этом случае условие Линдеберга сводится к требованию конечности дисперсий элементарных приращений.

Однако, как уже было отмечено выше, распределения приращений регистрируемых процессов заметно отличны от нормальных. Поэтому винеровские процессы оказываются отнюдь не бесспорными математическими моделями рассматриваемых процессов. Здесь следует также отметить, что не-нормальность распределений приращений регистрируемых процессов также делает весьма проблематичной адекватность таких модных ныне вероятностных моделей как фрактальные броуновские движения, которые, хотя и учитывают возможную зависимость приращений процесса (обладают “памятью”), но остаются гауссовыми – все их конечномерные распределения нормальны.

Как уже было сказано, отмеченная не-нормальность распределений приращений проявляется в том, что в действительности наблюдается заметно больше очень больших и очень маленьких по абсолютной величине значений приращений, нежели их должно быть в соответствии с нормальным распределением. Другими словами, наблюдаемые распределения приращений реальных процессов на небольших интервалах времени являются более островершинными, нежели нормальные, имея заметно более тяжелые хвосты.

Важная попытка объяснить замеченную островершинность этих распределений была предпринята в 60-х годах XX столетия. Обычно эту попытку ассоциируют с исследованиями Б. Мандельброта (Mandelbrot, 1963), (Mandelbrot, 1967) и Э. Фамы (Fama, 1963), (Fama, 1965) в области моделирования процессов биржевых цен. Одна из основных идей в их рассуждениях может быть кратко изложена так: отклонение распределений приращений от нормального означает, что классическая центральная предельная теорема здесь не применима, так как нарушаются ее условия.

Если не выходить за рамки общей структурной модели, согласно которой итоговое приращение процесса $P(t)$ на каждом интервале является суммой независимых приращений этого же процесса на подынтервалах одинаковой длины, и предполагать, что элементарные приращения имеют одинаковые распределения, то единственным условием, которое должно нарушаться, остается требование конечности дисперсии элементарных приращений. Поэтому вместо классической центральной предельной теоремы в качестве основы для построения моделей распределений приращений биржевых цен Б. Мандельброт и Э. Фама предложили использовать предельные теоремы, описывающие сходимость сумм независимых слагаемых с бесконечными дисперсиями.

Как известно, если слагаемые одинаково распределены, то в таких теоремах в качестве предельных распределений выступают так называемые устойчивые законы. Напомним, что функция распределения $G(x)$ называется *устойчивой*, если для любых $a_1 \in \mathbb{R}$, $a_2 \in \mathbb{R}$, $b_1 > 0$, $b_2 > 0$ существуют $a \in \mathbb{R}$ и $b > 0$ такие, что

$$G(b_1x + a_1) * G(b_2x + a_2) = G(bx + a), \quad x \in \mathbb{R}$$

(здесь символ $*$ обозначает свертку). Характеристическая функция любого устойчивого распределения допускает представление в виде

$$g(s) = \exp \{ i\gamma s - c|s|^\alpha (1 + i\beta\omega(s, \alpha) \operatorname{sign} s) \}, \quad s \in \mathbb{R},$$

где $\gamma \in \mathbb{R}$, $c \geq 0$, $-1 \leq \beta \leq 1$, $0 < \alpha \leq 2$ – параметры, определяющие тип распределения, $\omega(s, \alpha) = \operatorname{tg}(\pi\alpha/2)$, если $\alpha \neq 1$, и $\omega(s, 1) = \frac{2}{\pi} \log |s|$. Параметр α , определяющий скорость убывания хвостов устойчивого закона, называется *характеристическим*.

Все невырожденные устойчивые распределения абсолютно непрерывны. В 1925 г. П. Леви доказал, что для того чтобы функция распределения $F(x)$ могла быть предельной при $n \rightarrow \infty$ для распределений сумм $S_n = (X_1 + \dots + X_n - a_n)/b_n$ независимых одинаково распределенных случайных величин при некотором выборе чисел $a_n \in \mathbb{R}$ и $b_n > 0$, необходимо и достаточно, чтобы она была устойчивой (см., например, (Хинчин, 1938) или (Гнеденко и Колмогоров, 1949)).

Устойчивые законы являются хорошо известными примерами распределений с тяжелыми хвостами. Если $G_\alpha(x)$ – устойчивая функция распределения с характеристическим показателем $\alpha \in (0, 2)$, то

$$G_\alpha(-x) + 1 - G_\alpha(x) \sim \frac{c}{x^\alpha}$$

при $x \rightarrow \infty$, где $c > 0$. Более того, если Z_α – случайная величина с функцией распределения $G_\alpha(x)$, $\alpha \in (0, 2)$, то $E|Z_\alpha|^\gamma < \infty$ для любого $\gamma < \alpha$, но моменты величины Z_α порядков, больше или равных α , не существуют. Таким образом, дисперсии всех устойчивых законов за исключением нормального не определены (иногда говорят, что дисперсии не-нормальных устойчивых законов “бесконечны”).

Исследования Б. Мандельброта и Э. Фамы вызвали взрыв интереса как к аналитическим свойствам устойчивых законов (см., например, монографии (Золотарев, 1983), (Samorodnitsky and Taqqu, 1994), (Uchaikin and Zolotarev, 1999)) так и к статистическим процедурам, связанным с устойчивыми семействами.

В функциональных предельных теоремах для однородных случайных блужданий с дискретным временем и скачками с бесконечной дисперсией в качестве предельных выступают так называемые устойчивые процессы Леви (процессы с независимыми и стационарными приращениями, имеющими устойчивые распределения, см., например, (Samorodnitsky and Taqqu, 1994)). Подобные модели обладают весьма привлекательным аналитическим свойством масштабной инвариантности, которое, по мнению некоторых исследователей, должно быть присуще описываемым этими моделями сложным системам (Mandelbrot, 1997).

Однако наряду с большим числом благоприятных аналитических качеств, устойчивые законы обладают одним неприятным свойством: за исключением четырех случаев (нормальное распределение, распределение Коши, плотность которого (с точностью до параметров сдвига и масштаба) имеет вид

$$p(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R},$$

распределение Леви, задаваемое плотностью

$$p(x) = \frac{1}{\sqrt{2\pi x^3}} \exp \left\{ -\frac{1}{2x} \right\}, \quad x > 0,$$

(распределение времени первого достижения процессом броуновского движения единичного уровня) и распределение, симметричное к распределению Леви), отсутствуют явные выражения устойчивых плотностей (которые всегда существуют) в терминах элементарных функций. Ненамного больше известно устойчивых плотностей, выражающихся через наиболее популярные специальные функции (лишь относительно недавно доказано, что все устойчивые плотности допускают представление в терминах так называемых функций Фокса или обобщенных J -функций Мейера, см.

(Schneider, 1986), (Uchaikin and Zolotarev, 1999)). Это обстоятельство затрудняет статистический анализ в рамках устойчивых моделей.

Невозможность применять классические статистические процедуры, основанные на использовании функции правдоподобия, не является единственным обстоятельством, заставляющим с некоторым сомнением относиться к целесообразности использования устойчивых законов для моделирования распределений приращений биржевых цен. Напомним, что основным предположением, обуславливающим их применение в подобных задачах, является отсутствие конечных дисперсий у приращений биржевых цен за интервалы времени (бесконечно) малой длины. Но, как уже отмечалось во введении, для отсутствия дисперсии необходимо, чтобы упомянутые элементарные приращения могли с положительной вероятностью принимать как угодно большие по абсолютной величине значения, что на практике невозможно, так как регистрируемые процессы ограничены в пространстве. К тому же, как было отмечено, например, в работах (Blattberg and Gonedes, 1974), (Akgiray and Booth, 1987), (Jansen and de Vries, 1991), (Lux, 1996), (Батанов, Колик и др., 2003), многие статистические тесты отвергают гипотезу о согласии распределений реальных данных с устойчивыми законами, поскольку, все же, слишком больших по абсолютной величине наблюдений в реальных данных меньше, чем должно быть при не-нормальных устойчивых распределениях с бесконечной дисперсией, хвосты которых спадают при $|x| \rightarrow \infty$ как $O(|x|^{-\alpha})$, $0 < \alpha < 2$. В частности, как показано в работах (Jansen and de Vries, 1991), (Lux, 1996), при статистической подгонке к финансовым данным распределений с хвостами, убывающими степенным образом, оцененное значение показателя α оказывается существенно больше двух.

1.2. Однородные случайные блуждания с непрерывным временем

Время шло и медленно и быстро.

“Медленно-то как идет время”, – думал Кошкин иной раз.

“Быстро-то как время бежит”, – думал он в другой раз.

Ю. И. Коваль. “Алый”.

Да и не все ли равно: быстро или медленно пойдет время?

Леонид Андреев. “Мельком”.

В поисках удовлетворительной элементарной стохастической модели, обобщающей (1.1.1), которая могла бы объяснить наблюдаемую ненормальность распределений приращений регистрируемых хаотических процессов, многие исследователи уделяют внимание однородным случайным блужданиям с непрерывным временем, в которых случайными являются не только перемещения частицы на очередном шаге, но и интервалы времени между последовательными (мгновенными) перемещениями (см., например, (Zaslavsky, 2002)).

Модель однородных случайных блужданий с непрерывным временем была введена Монтроллом и Вейссом в 1965 г. (Montroll and Weiss, 1965). С тех пор эта модель интенсивно изучалась и применялась, например, для описания турбулентных потоков (Shlesinger, West and Klafter, 1987), (Klafter, Blumen and Shlesinger, 1987), (Pedrosa et al., 1996), (Connor et al., 1999) (Batanov et al., 1998), переноса в неоднородной фрактальной среде (Scher and Lax, 1973), (Weiss and Rubin, 1983), (Bouchaud and Georges, 1990), (Metzler, Barkai and Klafter, 1999), (West and Deering, 1994), (Isichenko, 1992), динамики неустойчивых хаотических систем (Klafter, Zumofen and Shlesinger, 1993) недебаевской релаксации (Bendler and Shlesinger, 1985), (Shlesinger, 1984), проблем, связанных с потоками в пористых средах (Benson, Meerschaert, Schumer and Wheatcraft, 2001a), (Benson, Meerschaert, Schumer and Wheatcraft, 2001b), (Benson, Meerschaert and Wheatcraft, 2000), (Baeumer, Benson, Meerschaert and Wheatcraft 2001).

В работе (Kotulski, 1995) в модели однородных случайных блужданий с непрерывным временем впервые возникли новые негауссовские предельные (при $t \rightarrow \infty$) распределения. Специфическим свойством этих распределений является наличие тяжелых хвостов. Позднее (и независимо) эти же распределения возникли в работах (Saichev and Zaslavsky, 1997), (Korolev and Uchaikin, 1999) и (Королев и Учайкин, 2000).

Однородное случайное блуждание с непрерывным временем, которое также можно назвать *обобщенным (сotpound) процессом восстановления*, формально определяется следующим образом. Пусть $X_1, X_2, \dots$ – независимые одинаково распределенные случайные величины и пусть далее $T_1, T_2, \dots$ – неотрицательные одинаково распределенные случайные величины, независимые друг от друга и от последовательности $X_1, X_2, \dots$. Рассмотрим случайный процесс

$$S(t) = \sum_{i=1}^{N(t)} X_i, \quad t > 0,$$

где $N(t)$ – считающий процесс, определяемый соотношением

$$\sum_{i=1}^{N(t)} T_i \leq t < \sum_{i=1}^{N(t)+1} T_i, \quad t > 0,$$

(для определенности предполагается, что $\sum_{i=1}^0 = 0$). С физической точки зрения X_i означает величину i -го мгновенного скачка частицы, T_i имеет смысл случайного времени ожидания между $(i-1)$ -м и i -м скачками, а $S(t)$ имеет смысл координаты блуждающей частицы в момент времени t .

Наряду с условием $ET_1 < \infty$ мы также рассмотрим случай, когда распределение случайной величины T_i имеет тяжелый хвост (исходя из сказанного выше, мы всюду ниже будем предполагать, что $DX_1 = \sigma^2 < \infty$). Под “тяжелохвостостью” распределения времени ожидания T_1 будем понимать, что $P(T_1 > t) = O(t^{-\beta})$ при некотором $\beta \in (0, 1)$. Более формально, предположим, что распределение случайной величины T_1 принадлежит области нормального притяжения строго устойчивого закона $G(x; \beta)$, определяемого характеристической функцией (преобразованием Фурье–Стилтьеса)

$$\tilde{g}(s; \beta) = \exp \left\{ -|s|^\beta \exp \left(-i \frac{\pi\beta}{2} \operatorname{sign} s \right) \right\}, \quad s \in \mathbb{R},$$

где $0 < \beta < 1$. Этот факт будем обозначаться как $\mathcal{L}(T_1) \in DNA(G(x; \beta))$. Символ $\Rightarrow$ будет обозначать сходимость по распределению.

Сходимость одномерных распределений процесса $S(t)$ изучалась во многих работах. В частности, в работах (Kotulski, 1995), (Korolev and Uchaikin, 1999), (Королев и Учайкин, 2000), (Kolokoltsov, Korolev and Uchaikin, 2001) были сформулированы и доказаны различные версии результата, частным случаем которого является следующее утверждение.

ТЕОРЕМА 1.2.1. *Предположим, что распределение случайной величины X_1 удовлетворяет условиям $EX_1 = 0$ и $DX_1 = \sigma^2 < \infty$. Тогда:*

(i) *если $ET_1 < \infty$, то существует такая постоянная $c > 0$, что*

$$P\left(\frac{S(t)}{c\sqrt{t}} < x\right) \Rightarrow \Phi(x), \quad t \rightarrow \infty;$$

(ii) *если $\mathcal{L}(T_1) \in DNA(G(x; \beta))$, $0 < \beta < 1$, то существует такая постоянная $c > 0$, что*

$$P\left(\frac{S(t)}{ct^{\beta/2}} < x\right) \Rightarrow \int_0^\infty \Phi(xu^{\beta/2}) dG(u; \beta) \equiv Q(x; \beta), \quad t \rightarrow \infty. \quad (1.2.1)$$

При этом для плотности $q(x; \beta) = Q'(x; \beta)$ справедливо представление

$$q(x; \beta) = \frac{g(|x|^{-2/\beta}; \beta/2)}{\beta|x|^{1+2/\beta}}, \quad x \in \mathbb{R},$$

где $g(y; \beta/2)$, $y \geq 0$, – плотность одностороннего устойчивого распределения с характеристическим показателем $\beta/2$. Распределение $Q(x; \beta)$ обладает моментами всех порядков. Более того, для $n = 1, 2, \dots$ справедливы соотношения

$$\int_{-\infty}^{\infty} x^{2n} q(x; \beta) dx = \frac{4n! \Gamma(n + \frac{1}{2})}{\sqrt{\pi} \Gamma(n\beta + 1)}, \quad \int_{-\infty}^{\infty} x^{2n-1} q(x; \beta) dx = 0,$$

где $\Gamma(\cdot)$ – эйлерова гамма-функция.

Распределение $Q(x; \beta)$, стоящее в правой части (1.2.1), является представителем семейства так называемых *дробно-устойчивых* законов. Разнообразные аналитические свойства дробно-устойчивых законов (в том числе и распределения $Q(x; \beta)$) описаны в статье (Бенинг, Королев и др., 2003), также см. работу (Bening, Korolev et al., 2005). Обратим особое внимание на то, что распределение $Q(x; \beta)$ является масштабной смесью нормальных законов с нулевым средним. Это обстоятельство в дальнейшем будет играть важнейшую роль в контексте подхода, которому посвящена данная книга.

Переходя от рассмотрения конечномерных распределений процесса $S(t)$ к изучению слабой сходимости этого процесса в пространстве Скорохода $\mathcal{D}([0, \infty), \mathbb{R})$ (см. раздел 2.6.1, дальнейшие сведения можно найти, например, в (Биллингсли, 1977)), можно отметить, не останавливаясь на деталях, что в работе (Meerschaert and Scheffler, 2001) доказано, что при соответствующей нормировке процесс случайного блуждания с непрерывным временем, удовлетворяющий условию $\mathcal{L}(T_1) \in DNA(G(x; \beta))$ при $\beta < 1$, слабо сходится к суперпозиции $Z(t) = W(B(t))$, где $W(t)$ – винеровский процесс, $B(t)$ – самоподобный случайный процесс с показателем Херста $1/\beta$. Более того, процессы $W(t)$ и $B(t)$ независимы, процесс $B(t)$ имеет конечную дисперсию, пропорциональную $t^{2\beta}$, и математическое ожидание, пропорциональное t^β (откуда следует, что при $\beta < 1$ приращения $B(t)$ не стационарны), процесс $B(t)$ не является процессом с независимыми приращениями. Процесс $Z(t)$ является самоподобным, но не является процессом со стационарными приращениями.

В работе (Piryatinska, Saichev and Woyczynski, 2005) (по-видимому, независимо от статьи (Meerschaert and Scheffler, 2001), в которой процесс

$Z(t)$ получен с помощью предельного перехода) процесс $Z(t)$ назван *дробной аномальной субдиффузией* и введен (с точностью до постоянного коэффициента) как суперпозиция $W(M(t))$ стандартного винеровского процесса и независимого от него процесса *обратного случайного времени*

$$M(t) = \min\{\tau : T(\tau) \leq t\}, \quad t > 0,$$

где $T(\tau)$ – процесс Леви с неубывающими траекториями такой, что $T(0) = 0$ и $T(1)$ имеет одностороннее устойчивое распределение $G(x; \beta)$ с параметром $\beta \in (0, 1)$.

Однако статистический анализ экспериментальных данных показывает, что далеко не всегда наблюдаемые распределения приращений регистрируемых процессов удовлетворительно согласуются с моделью, описываемой правой частью (1.2.1) (см., например, (Королев и Скворцова – ред., 2003)). К тому же, негауссовы распределения возникают в модели однородного случайного блуждания с непрерывным временем, только если среднее время ожидания очередного скачка бесконечно, что возможно, только если время ожидания очередного скачка может с положительной вероятностью быть произвольно большим, что плохо согласуется с традиционными представлениями, так как регистрируемые процессы ограничены во времени.

Таким образом, учитывая сказанное в этом и предыдущем разделах, можно прийти к выводу, что *однородные случайные блуждания с дискретным или непрерывным временем, вообще говоря, не могут быть адекватными элементарными моделями рассматриваемых процессов эволюции финансовых индексов или плазменной турбулентности.*

1.3. Неоднородные случайные блуждания, порождаемые обобщенными дважды стохастическими пуассоновскими процессами (обобщенными процессами Кокса)

1.3.1. Процессы Кокса: определения и примеры

Естественным обобщением рассмотренной в предыдущем разделе модели является неоднородное случайное блуждание с непрерывным временем. От однородного блуждания оно отличается тем, что распределения случайных величин $T_1, T_2, \dots$ являются, вообще говоря, различными. Пред-

положение неоднородности (различия распределений времен между последовательными скачками) случайного блуждания хорошо согласуется с представлением о том, что интенсивность изменений координаты частицы, испытывающей броуновское движение в турбулентной среде, существенно непостоянна. Это непостоянство может быть вызвано многими причинами, проявляющимися, например, как непериодические или периодические компоненты (тренды), связанные с изменениями локальных (во времени) тенденций, например, панического характера на биржах или внешних параметров в турбулентной плазме тороидальных магнитных ловушек (токамаков и стеллараторов). Кроме того, участки нестационарности могут быть вызваны некоторыми случайно возникающими (не поддающимися абсолютно надежному прогнозированию) причинами.

Пусть, как и выше, $N(t)$ – число скачков случайно блуждающей частицы за период времени $[0, t]$, $t \geq 0$. Моменты скачков образуют хаотический точечный случайный процесс на оси времени. Однако этот хаотический случайный процесс не является однородным в силу указанных выше представлений. Как известно, наиболее разумными стохастическими моделями неоднородных хаотических точечных процессов являются *дважды стохастические пуассоновские процессы*, иначе называемые *процессами Кокса* (см., например, (Bening and Korolev, 2002)). Определение процессов Кокса следует предварить кратким экскурсом в описание свойств однородных и неоднородных пуассоновских процессов.

Как известно (см., например, (Bening and Korolev, 2002)), наилучшей моделью *однородного* стохастического хаотического потока событий является пуассоновский процесс, характеризуемый тем обстоятельством, что интервалы времени между событиями потока стохастически независимы и имеют одинаковое показательное распределение. Напомним, что привлекательность пуассоновского процесса в качестве модели однородного дискретного хаоса обусловлена как минимум двумя обстоятельствами. Во-первых, показательное распределение обладает максимальной дифференциальной энтропией среди всех абсолютно непрерывных распределений вероятностей, сосредоточенных на всей положительной полуоси и имеющих конечное математическое ожидание, а энтропия, как известно, является очень удобной численной характеристикой неопределенности. Во-вторых, точки (события) пуассоновского потока равномерно распределены на оси времени в том смысле, что для любого конечного интервала времени $[t_1, t_2]$, $t_1 < t_2$, условное совместное распределение точек пуассоновского потока, попавших в интервал $[t_1, t_2]$, при условии, что в этот интервал попало фиксированное число, скажем, n точек, совпадает с совместным распределени-

ем вариационного ряда, построенного по независимой однородной выборке объема n из равномерного на $[t_1, t_2]$ распределения. Равномерное же распределение обладает максимальной дифференциальной энтропией среди всех абсолютно непрерывных распределений вероятностей, сосредоточенных на конечных интервалах, и очень хорошо соответствует общепринятому представлению об абсолютно непредсказуемой ограниченной случайной величине.

Вспомним, что, если $N_\lambda(t)$ – однородный пуассоновский процесс с некоторой интенсивностью $\lambda > 0$, то параметр λ , называемый интенсивностью, имеет смысл среднего числа скачков процесса (количества событий наблюдаемого потока) в единицу времени. При этом

$$P(N_\lambda(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k = 0, 1, \dots \quad (1.3.1)$$

и

$$EN_\lambda(t) = DN_\lambda(t) = \lambda t, \quad t > 0,$$

так что для любых $s, t > 0$

$$\frac{EN_\lambda(t+s) - EN_\lambda(t)}{s} = \lambda = \text{const.}$$

Из соотношения (1.3.1) легко видеть, что для любого $k = 0, 1, \dots$

$$P(N_\lambda(t) = k) = P(N_1(\lambda t) = k), \quad t \geq 0, \quad (1.3.2)$$

то есть процессы $N_\lambda(t)$ и $N_1(\lambda t)$ *стохастически эквивалентны*.

Однако, как уже было отмечено, в реальной практике хаотические потоки не бывают однородными. Для наглядности, перед тем как определить дважды стохастический пуассоновский процесс, рассмотрим неоднородный пуассоновский поток событий $N^*(t)$, определяемый следующим образом. Пусть $\lambda(t)$ – некоторая положительная функция. Предположим, что функция $\lambda(t)$ интегрируема, и обозначим

$$\Lambda(t) = \int_0^t \lambda(\tau) d\tau, \quad t \geq 0.$$

Неоднородный пуассоновский процесс $N^*(t)$, $t \geq 0$, определятся как случайный процесс с независимыми приращениями, траектории которого стартуют из нуля ($N^*(0) = 0$) и

$$P(N^*(t) = k) = e^{-\Lambda(t)} \frac{(\Lambda(t))^k}{k!}, \quad k = 0, 1, \dots \quad (1.3.3)$$

Тогда

$$\mathbb{E}N^*(t) = \Lambda(t), \quad t \geq 0,$$

так что

$$\lim_{s \downarrow 0} \frac{\mathbb{E}N^*(t+s) - \mathbb{E}N^*(t)}{s} = \lim_{s \downarrow 0} \frac{\Lambda(t+s) - \Lambda(t)}{s} = \frac{d\Lambda(t)}{dt} = \lambda(t), \quad t \geq 0.$$

Другими словами, при таком определении функция $\lambda(t)$ играет роль *мгновенной интенсивности* процесса в точке t . При этом по аналогии с (1.3.2) из (1.3.3) вытекает, что

$$\mathbb{P}(N^*(t) = k) = \mathbb{P}(N_1(\Lambda(t)) = k), \quad k = 0, 1, \dots, \quad t \geq 0,$$

то есть процессы $N^*(t)$ и $N_1(\Lambda(t))$ стохастически эквивалентны.

Наконец, рассмотрим следующее естественное обобщение неоднородного пуассоновского процесса.

Пусть $N_1(t)$, $t \geq 0$, – однородный пуассоновский процесс с единичной интенсивностью, а $\Lambda(t)$, $t \geq 0$, – независимый от $N_1(t)$ случайный процесс, обладающий следующими свойствами: $\Lambda(0) = 0$, $\mathbb{P}(\Lambda(t) < \infty) = 1$ для любого $t > 0$, траектории $\Lambda(t)$ не убывают и непрерывны справа. Дважды стохастический пуассоновский процесс $N(t)$, называемый также процессом Кокса, определяется как суперпозиция $N_1(t)$ и $\Lambda(t)$:

$$N(t) = N_1(\Lambda(t)), \quad t \geq 0.$$

В этом случае будем говорить, что процесс Кокса $N(t)$ управляется процессом $\Lambda(t)$. В частности, если процесс $\Lambda(t)$ допускает представление

$$\Lambda(t) = \int_0^t \lambda(\tau) d\tau, \quad t \geq 0, \tag{1.3.4}$$

в котором $\lambda(t)$ – положительный случайный процесс с интегрируемыми траекториями, то $\lambda(t)$ можно интерпретировать как мгновенную стохастическую интенсивность процесса $N(t)$. Поэтому иногда процесс $\Lambda(t)$, управляющий процессом Кокса $N(t)$, будет называться *накопленной интенсивностью* процесса $N(t)$.

Несложно убедиться, что для процесса Кокса $N(t)$, управляемого процессом $\Lambda(t)$, справедливы соотношения

$$\mathbb{E}N(t) = \mathbb{E}\Lambda(t), \quad \mathbb{D}N(t) = \mathbb{E}\Lambda(t) + \mathbb{D}\Lambda(t).$$

Свойства процессов Кокса подробно описаны в книгах (Grandell, 1978) и (Bening and Korolev, 2002).

1.3.2. Обобщенные процессы Кокса

В основе конструкции базовых математических моделей неоднородных хаотических процессов, рассматриваемых в этой книге, лежит асимптотическая схема, основанная на принципе, который может быть наглядно проиллюстрирован на примере простейшей задачи из теории измерений (см., например, (Romanowski, 1979), (Новицкий и Зограф, 1991)). Погрешность измерения является результатом суммарного воздействия большого числа случайных факторов, ни один из которых не является доминирующим, и потому, согласно центральной предельной теореме, должна иметь нормальное распределение. Однако на *разные* измерения воздействует, вообще говоря, *разное* число случайных факторов, то есть число случайных факторов, определяющих погрешность, само является случайным фактором. Поэтому вместо классической центральной предельной теоремы здесь более уместно пользоваться предельными теоремами для сумм случайного числа независимых случайных величин.

Теория случайного суммирования довольно хорошо развита (см., например, монографии (Круглов и Королев, 1990), (Gnedenko and Korolev, 1996), (Bening and Korolev, 2002)). Не ставя перед собой цели привести результаты этой теории во всей полноте, сосредоточимся лишь на очень частном конкретном варианте постановки задачи, когда число слагаемых в суммах формируется в соответствии с дважды стохастическим пуассоновским процессом (процессом Кокса). Этот случай имеет чрезвычайно важное практическое значение.

Целью данного раздела является описание задачи моделирования неоднородных хаотических потоков событий с помощью обобщенных процессов Кокса. В следующих разделах будет продемонстрировано, что отклонения распределений наблюдаемых процессов от нормального могут быть успешно объяснены наличием существенной изменчивости интенсивности неоднородных хаотических потоков событий, описываемых процессами Кокса.

Пусть $X_1, X_2, \dots$ – одинаково распределенные случайные величины. Предположим, что при каждом $t \geq 0$ случайные величины $N(t), X_1, X_2, \dots$ независимы. Процесс

$$S(t) = \sum_{j=1}^{N(t)} X_j, \quad t \geq 0, \quad (1.3.5)$$

назовем обобщенным процессом Кокса (при этом для определенности считаем, что $\sum_{j=1}^0 = 0$). Процессы вида (1.3.5) играют чрезвычайно важ-

ную роль во многих прикладных задачах. Достаточно сказать, что при $\Lambda(t) \equiv \lambda t$ с $\lambda > 0$ процесс $S(t)$ превращается в классический обобщенный пуассоновский процесс, широко используемый при моделировании многих явлений в физике, теории надежности, финансовой и актуарной деятельности, биологии и т. д. Большое число разнообразных прикладных задач, приводящих к обобщенным пуассоновским процессам, описано в книгах (Gnedenko and Korolev, 1996) и (Bening and Korolev, 2002).

Приведем примеры некоторых задач, сводящихся к обобщенным процессам Кокса.

ПРИМЕР 1.3.1. Как уже говорилось, пуассоновский точечный процесс в пространстве описывает абсолютно хаотическое распределение точек. Рассмотрим следующее вполне естественное обобщение этого свойства применительно к задаче о суммарной массе тел. Пусть $\rho(x)$, $x \in \mathbb{R}^3$, – плотность (концентрация) каких-либо материальных частиц в пространстве. Предположим, что частицы – это материальные точки, распределенные в пространстве в соответствии с неоднородным трехмерным пуассоновским точечным процессом, так что если $N(B)$ – число точек в некотором (измеримом) множестве $B \subset \mathbb{R}^3$, то

$$P(N(B) = k) = \exp\{-\lambda(B)\} \frac{(\lambda(B))^k}{k!}, \quad k = 0, 1, 2, \dots,$$

где $\lambda(B) = \int_B \rho(x) dx$. Положим $\Lambda(t) = \int_{|x| \leq t} \rho(x) dx$. Пусть X_j – масса, сосредоточенная в j -ой точке. Тогда процесс $S(t)$, определяемый соотношением (1.3.5), описывает суммарную массу, содержащуюся в сфере радиуса t . Подобные объекты представляют традиционный интерес в космологии.

ПРИМЕР 1.3.2. Рассмотрим следующую модель одномерного броуновского движения – теплового движения некоторой частицы, испытывающей соударения с молекулами вещества, заполняющего среду, в которой движется частица. Пусть $N(t)$ – число соударений частицы с молекулами до момента t . Если бы среда и свойства частицы не изменялись во времени, то было бы естественно предполагать, что $N(t)$ является однородным пуассоновским процессом. Однако если допустить, что среда или свойства частицы изменяются во времени, то интенсивность столкновений перестает быть постоянной. Поэтому, обозначив перемещение частицы в результате j -го столкновения через X_j , можно прийти к выводу, что процесс $S(t)$, определяемый соотношением (1.3.5), описывает координату частицы в момент t , если она испытывает броуновское движение в неоднородной (случайной) среде.

Всюду в дальнейшем будем считать, что у случайных величин $\{X_j\}_{j \geq 1}$ имеется, по крайней мере, два первых момента. Обозначим $EX_1 = a$, $DX_1 = \sigma^2$, $0 < \sigma^2 < \infty$. Как будет показано, даже в таких предположениях предельные распределения обобщенных процессов Кокса могут иметь произвольно тяжелые хвосты. В данной главе для большей наглядности изложение будет сосредоточено на ситуации, в которой $a = 0$. Это сделано по нескольким причинам. Во первых, цель данной главы – описать базовые *принципы* построения адекватных стохастических моделей процессов плазменной турбулентности или эволюции финансовых индексов, не вдаваясь в излишние детали. Во-вторых, как будет показано ниже, на практике значения параметра a как правило оказываются очень близкими к нулю. Чтобы успокоить дотошного читателя, оговоримся, что общей ситуации ($a \neq 0$) ниже будет уделено должное внимание.

Несложно убедиться, что если $E\Lambda(t) < \infty$, то

$$ES(t) = aE\Lambda(t) \quad (1.3.6)$$

и, если дополнительно $E(\Lambda(t))^2 < \infty$, то

$$DS(t) = (a^2 + \sigma^2)E\Lambda(t) + a^2D\Lambda(t). \quad (1.3.7)$$

1.3.3. Центральная предельная теорема для обобщенных процессов Кокса

Время послушно затормозилось, секунды стали длинными-длинными, и в течение каждой можно было сделать очень много разных движений, нанести много ударов и видеть всех сразу.

А. и Б. Стругацкие. “Обитаемый остров”.

В этом разделе будут приведены необходимые и достаточные условия сходимости одномерных распределений обобщенных процессов Кокса со скачками, обладающими свойствами, указанными в разделе 1.3.1, без каких бы то ни было моментных условий на управляющий процесс. Мы продемонстрируем, что асимптотическое поведение процесса $S(t)$ полностью определяется асимптотическим поведением накопленной интенсивности $\Lambda(t)$. Более того, мы обратим внимание на то, что тяжелые (например, типа Парето) хвосты распределений, предельных для сумм (1.3.5), могут быть обусловлены не “плохим” поведением слагаемых (например, отсутствием у них моментов), а чрезмерно большим разбросом значений управляющего процесса $\Lambda(t)$. Паретовский характер хвостов распределений величин, связанных с турбулентной плазмой, отмечен, например, в

работах (Carreras, van Milligen, Pedrosa et al., 1998), (Скворцова, Сарксян и Харчев, 1999), (Carreras, van Milligen, Hidalgo et al., 1999), (Skvortsova, Batanov, Kolik et al., 2002).

Здесь будут приведены критерии асимптотической нормальности распределений обобщенных процессов Кокса. В качестве следствия в следующем разделе будет сформулирован критерий асимптотической устойчивости распределений обобщенных процессов Кокса.

Пусть $d(t) > 0$ – функция, неограниченно возрастающая при $t \rightarrow \infty$. Роль функции $d(t)$ – нормировка или масштабирование во времени (или, как будет показано, в пространстве, если “бесконечно большой” параметр t не является астрономическим временем).

ТЕОРЕМА 1.3.1. *Предположим, что $\Lambda(t) \rightarrow \infty$ по вероятности при $t \rightarrow \infty$. Для того чтобы одномерные распределения нормированного обобщенного процесса Кокса слабо сходились к распределению некоторой случайной величины Z ,*

$$\frac{S(t)}{\sigma\sqrt{d(t)}} \Rightarrow Z \quad (t \rightarrow \infty),$$

необходимо и достаточно, чтобы существовала неотрицательная случайная величина U такая, что

$$1^\circ \quad \mathbf{P}(Z < x) = \int_0^\infty \Phi(xy^{-1/2}) d\mathbf{P}(U < y) \quad (= \mathbf{E}\Phi(xU^{-1/2})), \quad x \in \mathbb{R};$$

$$2^\circ \quad \frac{\Lambda(t)}{d(t)} \Rightarrow U \quad (t \rightarrow \infty).$$

Д о к а з а т е л ь с т в о см. в (Королев, 1998) или (Bening and Korolev, 2002).

Обратим внимание на то, что условие 2° теоремы 1.3.1 может быть интерпретировано как требование статистической регулярности накопленной интенсивности: предел отношения $\Lambda(t)/d(t)$ при $t \rightarrow \infty$ может быть случайным, но он должен существовать. Еще одна интерпретация этого условия заключается в том, что при больших t распределение случайной величины $\Lambda(t)/d(t)$ практически не зависит от t .

При этом условие 1° означает, что, вообще говоря, предельным распределением обобщенного процесса Кокса является масштабная смесь нормальных законов, имеющая, как будет видно ниже, (вообще говоря, про-

извольно) более тяжелые хвосты, нежели само нормальное распределение. Идея использования масштабных смесей нормальных законов в качестве островершинных (и обладающих тяжелыми хвостами) альтернатив нормальному закону в задачах обработки измерений отнюдь не является новой. В частности, такую идею высказал еще в 1882 г. американский астроном С. Ньюкомб (Newcomb, 1882), см. также (Stigler, 1973). Обсуждение возможности применения таких смесей к описанию распределения скоростей в турбулентной среде можно найти в работе (Dutton, 1970), где в качестве физического обоснования этого подхода используется неоднородный, перемежающийся характер турбулентности. Другие ссылки на работы из разных областей, в которых обсуждается применимость масштабных смесей нормальных законов при обработке реальных данных, можно найти, например, в статье (Barndorff-Nielsen, 1977).

Из теоремы 1.3.1 и свойства идентифицируемости (см. раздел 2.2) семейства масштабных смесей нормальных законов немедленно вытекает следующее утверждение.

СЛЕДСТВИЕ 1.3.1. *В условиях теоремы 1.3.1*

$$P\left(\frac{S(t)}{\sigma\sqrt{d(t)}} < x\right) \Rightarrow \Phi(x) \quad (t \rightarrow \infty)$$

тогда и только тогда, когда

$$\frac{\Lambda(t)}{d(t)} \Rightarrow 1 \quad (t \rightarrow \infty).$$

Другими словами, предельное распределение обобщенного процесса Кокса может быть нормальным только в том случае, когда случайная величина $\Lambda(t)/d(t)$ асимптотически (при $t \rightarrow \infty$) неслучайна.

1.3.4. Сходимость распределений обобщенных процессов Кокса к устойчивым законам

В качестве еще одного следствия теоремы 1.3.1 здесь будет приведен критерий сходимости одномерных распределений обобщенных процессов Кокса с нулевым средним и *конечными дисперсиями* к устойчивым законам. Будет показано, что одномерные распределения обобщенных процессов Кокса с описанными выше свойствами асимптотически строго устойчивы тогда и только тогда, когда асимптотически строго устойчивы их управляющие процессы.

Пусть $G_{\alpha,\theta}(x)$ – строго устойчивая функция распределения с показателем α и параметром θ , которой соответствует характеристическая функция

$$\tilde{g}_{\alpha,\theta}(s) = \exp \left\{ -|s|^\alpha \exp \left\{ -i \frac{\pi\theta\alpha}{2} \text{sign}s \right\} \right\}, \quad s \in \mathbb{R},$$

где $0 < \alpha \leq 2$, $|\theta| \leq \theta_\alpha = \min(1, 2/\alpha - 1)$.

ТЕОРЕМА 1.3.2. *Предположим, что управляющий процесс $\Lambda(t)$ неограниченно возрастает по вероятности при $t \rightarrow \infty$. Для того чтобы*

$$\lim_{t \rightarrow \infty} \mathbf{P} \left(\frac{S(t)}{\sigma \sqrt{d(t)}} < x \right) = G_{\alpha,0}(x), \quad x \in \mathbb{R},$$

необходимо и достаточно, чтобы

$$\lim_{t \rightarrow \infty} \mathbf{P} \left(\frac{\Lambda(t)}{d(t)} < x \right) = G_{\alpha/2,1}(x), \quad x \in \mathbb{R}.$$

Д о к а з а т е л ь с т в о этого утверждения заключается в последовательном применении теоремы 1.3.1 и теоремы 3.3.1 из (Золотарев, 1983), согласно которой

$$G_{\alpha,0}(x) = \int_0^\infty \Phi(x/\sqrt{u}) dG_{\alpha/2,1}(u), \quad x \in \mathbb{R},$$

с использованием свойства идентифицируемости масштабных смесей нормальных законов (см. раздел 4.9) и абсолютной непрерывности устойчивых распределений.

Теперь рассмотрим ситуацию с дискретным временем $t = n = 1, 2, \dots$ и предположим, что управляющий процесс $\Lambda(n)$ имеет вид

$$\Lambda(n) = \sum_{i=1}^n Z_i,$$

где $\{Z_i\}$ – независимые одинаково распределенные случайные величины, $Z_i \geq 0$, $i \geq 1$. Такое представление возникает в ситуации, когда $\Lambda(t)$ – однородный процесс с независимыми приращениями, а обобщенный процесс Кокса наблюдается в равноотстоящие моменты времени, то есть Z_i – приращения управляющего процесса $\Lambda(t)$ за интервалы времени между наблюдениями. В соответствии с (1.3.5) полагаем

$$S(n) = \sum_{j=1}^{N_1(\Lambda(n))} X_j.$$

В этой ситуации с помощью теоремы 1.3.2 и теоремы 2 параграфа 35 из (Гнеденко и Колмогоров, 1949) можно легко получить следующее утверждение.

ТЕОРЕМА 1.3.3. *Одномерные распределения нормированного обобщенного процесса Кокса с дискретным временем $S(n)/\delta_n$ слабо сходятся к строго устойчивому распределению $G_{\alpha,0}$ при некотором выборе констант δ_n тогда и только тогда, когда для любого $k > 0$*

$$\lim_{x \rightarrow \infty} \frac{P(Z_1 \geq x)}{P(Z_1 \geq kx)} = k^{\alpha/2}.$$

Другими словами, часто наблюдаемые на практике тяжелые хвосты (устойчивых) распределений, предельных для обобщенных процессов Кокса при возрастающей интенсивности, могут возникать не только в той ситуации, где тяжелые хвосты присущи распределениям скачков. Как видно из теоремы 1.3.3, даже при произвольно легких хвостах распределений скачков тяжелые хвосты предельного закона могут возникать из-за того, что тяжелые (паретовские) хвосты имеются у распределений приращений управляющего процесса. Этот вывод подтверждается результатами экспериментов, см., например, (Batanov, Kolik, Petrov et al., 2003).

Доказательства приведенных выше результатов можно найти в (Королев, 1998) или (Bening and Korolev, 2002).

1.3.5. Оценки точности аппроксимации распределений обобщенных процессов Кокса масштабными смесями нормальных законов

В данном разделе будет показано, что при оценивании точности аппроксимации распределений обобщенных процессов Кокса масштабными смесями нормальных законов можно успешно использовать аналогичные оценки для пуассоновских случайных сумм.

Несложно видеть, что в силу стохастической независимости всех случайных величин и процессов, вовлеченных в определение обобщенного процесса Кокса, распределение смешанной пуассоновской случайной суммы $S(t)$ допускает представление

$$P(S(t) < x) = \int_0^\infty P\left(\sum_{j=1}^{N_1(\lambda)} X_j < x\right) dP(\Lambda(t) < \lambda), \quad x \in \mathbb{R}. \quad (1.3.8)$$

Всюду далее в этом разделе для простоты будет предполагаться, что $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = 1$. Пусть $\delta > 0$ – некоторое число. Обозначим $\beta_{2+\delta} = \mathbb{E}|X_1|^{2+\delta}$ и предположим, что

$$\beta_{2+\delta} < \infty \quad (1.3.9)$$

для некоторого $\delta \in (0, 1]$. Пусть $d(t)$ – некоторая положительная функция, определенная для $t > 0$ и неограниченно возрастающая при $t \rightarrow \infty$.

Цель данного раздела – привести некоторые оценки скорости сходимости в теореме 1.3.1.

Введем величину

$$C(\delta) = \frac{2^{1-\delta/2}\Gamma(\frac{2+\delta}{2})}{\pi(1+\delta)(2+\delta)},$$

где $\Gamma(\cdot)$ – эйлерова гамма-функция. Несложно видеть, что $C(\delta)$ как функция $\delta \in (0, 1]$ монотонно убывает, причем

$$0.0665 \approx \frac{1}{6\sqrt{2\pi}} = C(1) \leq C(\delta) \leq C(0+) = \frac{1}{\pi} \approx 0.3183$$

(см., например, (Королев и Шевцова, 2005а), (Королев и Шевцова, 2005b)). Для $\lambda > 0$ обозначим

$$\rho(\lambda) = \sup_x \left| \mathbb{P}\left(\frac{1}{\sqrt{\lambda}} \sum_{j=1}^{N_1(\lambda)} X_j < x\right) - \Phi(x) \right|.$$

ЛЕММА 1.3.1. *Предположим, что выполнено условие (1.3.9). Тогда*

$$\rho(\lambda) \leq C^*(\delta) \cdot \frac{\beta_{2+\delta}}{\lambda^{\delta/2}},$$

где

$$C^*(\delta) = \begin{cases} C(\delta), & \text{если } 0 < \delta < 1, \\ 0.7056, & \text{если } \delta = 1. \end{cases}$$

Доказательство этого результата приведено в работе (Шевцова, 2006). Также см. (Королев, Бенинг и Шоргин, 2007).

Если выполнено условие 2° теоремы 1.3.1, то имеет место сходимость распределений обобщенных процессов Кокса, причем предельное распределение имеет вид масштабной смеси нормальных законов, определенной

в пункте 1° теоремы 1.3.1. Обозначим

$$\rho_t = \sup_x \left| \mathbf{P} \left(\frac{S(t)}{\sqrt{d(t)}} < x \right) - \int_0^\infty \Phi \left(\frac{x}{\sqrt{\lambda}} \right) d\mathbf{P}(\Lambda < \lambda) \right|,$$

$$\Delta_t = \sup_x \left| \mathbf{P} \left(\frac{\Lambda(t)}{d(t)} < x \right) - \mathbf{P}(\Lambda < x) \right|.$$

Из представления (1.3.8) и леммы 1.3.1 вытекает следующее утверждение.

ТЕОРЕМА 1.3.4. *Предположим, что выполнено условие (1.3.9). Пусть Λ – неотрицательная случайная величина. Тогда для любого $t > 0$ справедлива оценка*

$$\rho_t \leq C^*(\delta) \beta_{2+\delta} \mathbf{E}[\Lambda(t)]^{-\delta/2} + \frac{1}{2} \Delta_t.$$

Д о к а з а т е л ь с т в о. приведено в статье (Гавриленко и Королев, 2006), также см. (Королев, Бенинг и Шоргин, 2007).

В качестве примера применения теоремы 1.3.4 рассмотрим ситуацию, когда при каждом $t > 0$ случайная величина $\Lambda(t)$ имеет гамма-распределение. Этот случай имеет большое значение для обоснования адекватности таких моделей динамики финансовых индексов как популярные ныне дисперсионные гамма-процессы Леви (VG-процессы) и двусторонние гамма-процессы Леви, о которых речь подробно пойдет в разделе 2.6.3.

Плотность гамма-распределения с параметром формы $r > 0$ и параметром масштаба $s > 0$, как известно, имеет вид

$$g_{r,s}(x) = \frac{s^r}{\Gamma(r)} e^{-sx} x^{r-1}, \quad x > 0.$$

Смешанное пуассоновское распределение с таким смешивающим гамма-распределением, тем самым, имеет характеристическую функцию

$$\begin{aligned} \psi(t) &= \int_0^\infty \exp\{y(e^{it} - 1)\} \frac{s^r}{\Gamma(r)} e^{-sy} y^{r-1} dy = \\ &= \frac{s^r}{\Gamma(r)} \int_0^\infty \exp \left\{ -sy \left(1 + \frac{1 - e^{it}}{s} \right) \right\} y^{r-1} dy = \left(1 + \frac{1 - e^{it}}{s} \right)^{-r}. \end{aligned}$$

Вводя перепараметризацию

$$s = \frac{p}{1-p} \quad \left(p = \frac{s}{1+s} \right), \quad p \in (0, 1), \quad (1.3.10)$$

окончательно получаем

$$\psi(t) = \left(\frac{p}{1 - (1-p)e^{it}} \right)^r, \quad t \in \mathbb{R},$$

что совпадает с характеристической функцией отрицательного биномиального распределения с параметрами $r > 0$ и $p \in (0, 1)$. Таким образом, в рассматриваемом случае при каждом $t > 0$ случайная величина $N(t)$ имеет отрицательное биномиальное распределение с параметрами $r > 0$ и $p \in (0, 1)$:

$$P(N(t) = n) = C_{r+n-1}^n p^r (1-p)^n, \quad n = 0, 1, 2, \dots \quad (1.3.11)$$

Здесь $r > 0$ и $p \in (0, 1)$ – параметры, и для нецелых r величина C_{r+n-1}^n определяется как

$$C_{r+n-1}^n = \frac{\Gamma(r+n)}{n! \cdot \Gamma(r)}.$$

В частности, при $r = 1$ соотношение (1.3.11) задает геометрическое распределение. Отрицательное биномиальное распределение с натуральным r допускает наглядную интерпретацию в терминах испытаний Бернулли. А именно, случайная величина, имеющая отрицательное биномиальное распределение с параметрами $r > 0$ и p , – это число испытаний Бернулли, проведенных до осуществления r -й по счету неудачи, если вероятность успеха в одном испытании равна $1 - p$.

Таким образом, отрицательное биномиальное распределение с параметрами $r > 0$ и $p \in (0, 1)$ представляет собой смешанное пуассоновское распределение, в котором параметр имеет гамма-распределение с параметром масштаба $s = p/(1-p)$ и параметром формы r . С учетом параметризации (1.3.10) о таком распределении мы будем говорить как об отрицательном биномиальном с параметрами $r > 0$ и s .

Функцию гамма-распределения с параметром масштаба s и параметром формы r обозначим $G_{r,s}(x)$,

$$G_{r,s}(x) = \int_0^x g_{r,s}(z) dz.$$

Несложно убедиться, что

$$G_{r,s}(x) \equiv G_{r,1}(sx). \quad (1.3.12)$$

Случайную величину с функцией распределения $G_{r,s}(x)$ обозначим $U(r, s)$. Хорошо известно, что

$$\mathbb{E}U(r, s) = \frac{r}{s}.$$

Будем считать, что параметр r фиксирован, и в качестве случайной величины $\Lambda(t)$ возьмем величину $U(r, s)$, считая, что $t = s^{-1}$:

$$\Lambda(t) = U(r, t^{-1}).$$

В качестве функции $d(t)$ возьмем

$$d(t) \equiv \mathbb{E}\Lambda(t) = \mathbb{E}U(r, t^{-1}).$$

Несложно видеть, что в терминах новой параметризации

$$\mathbb{E}U(r, t^{-1}) = rt.$$

Тогда с учетом (1.3.12)

$$\begin{aligned} \mathbb{P}\left(\frac{\Lambda(t)}{d(t)} < x\right) &= \mathbb{P}(U(r, t^{-1}) < xrt) = \mathbb{P}(U(r, 1) < xr) = \\ &= \mathbb{P}(U(r, r) < x) = G_{r,r}(x). \end{aligned}$$

Обратим внимание, что функция распределения, стоящая в правой части последнего соотношения, не зависит от t . Поэтому при указанном выборе функции $d(t)$ условие 2° теоремы 1.3.1 выполняется тривиальным образом, причем для всех $t > 0$

$$\Delta_t = 0.$$

Вычислим $\mathbb{E}[\Lambda(t)]^{-\delta/2}$, предполагая, что выполнено условие

$$r > \frac{\delta}{2}. \quad (1.3.13)$$

Имеем

$$\mathbb{E}[\Lambda(t)]^{-\delta/2} = \mathbb{E}[U(r, t^{-1})]^{-\delta/2} = \int_0^\infty \frac{e^{-x/t} x^{r-1-\delta/2}}{tr\Gamma(r)} dx = \frac{\Gamma(r - \frac{\delta}{2})}{t^{\delta/2}\Gamma(r)}.$$

Таким образом, фактически получено следующее утверждение, являющееся частным случаем теоремы 1.3.4.

СЛЕДСТВИЕ 1.3.2. Пусть случайная величина $\Lambda(t)$ имеет гамма-распределение с параметром формы $r > 0$ и параметром масштаба $s = 1/t$, где $t > 0$. Предположим, что выполнены условия (1.3.9) и (1.3.13). Тогда для каждого $t > 0$

$$\sup_x \left| \mathbf{P}(S(t) < x\sqrt{rt}) - \int_0^\infty \Phi\left(\frac{x}{\sqrt{y}}\right) dG_{r,r}(y) \right| \leq C(r, \delta) \cdot \frac{\beta_{2+\delta}}{t^{\delta/2}},$$

где

$$C(r, \delta) = \frac{C^*(\delta)\Gamma(r - \frac{\delta}{2})}{\Gamma(r)} = \begin{cases} \frac{\delta \mathbf{B}(r - \frac{\delta}{2}, \frac{\delta}{2})}{2^{\delta/2}\pi(1+\delta)(2+\delta)}, & \text{если } 0 < \delta < 1, \\ 0.7056 \frac{\Gamma(r - \frac{1}{2})}{\Gamma(r)}, & \text{если } \delta = 1, \end{cases}$$

$\mathbf{B}(\cdot, \cdot)$ – эйлерова бета-функция,

$$\mathbf{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad a > 0, b > 0.$$

Несложно видеть, что при $r = 1$ предельная функция распределения представляет собой распределение Лапласа (также называемое двойным показательным распределением), определяемое плотностью

$$\ell(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}, \quad x \in \mathbb{R}.$$

Функцию распределения Лапласа, соответствующую плотности $\ell(x)$, обозначим $L(x)$,

$$L(x) = \begin{cases} \frac{1}{2} e^{\sqrt{2}x}, & \text{если } x < 0, \\ 1 - \frac{1}{2} e^{-\sqrt{2}x}, & \text{если } x \geq 0. \end{cases}$$

Роль распределения Лапласа при моделировании характеристик плазменной турбулентности будет подробно обсуждаться в разделе 7.3.

СЛЕДСТВИЕ 1.3.3. Пусть случайная величина $\Lambda(t)$ имеет экспоненциальное распределение с параметром $s = 1/t$, где $t > 0$. Предположим, что $\mathbf{E}|X_1|^3 \equiv \beta^3 < \infty$. Тогда для каждого $t > 0$

$$\sup_x |\mathbf{P}(S(t) < x\sqrt{t}) - L(x)| \leq 1.2507 \cdot \frac{\beta^3}{\sqrt{t}}.$$

1.3.6. Закон больших чисел для обобщенных процессов Кокса

Результат, приведенный в данном подразделе, не будет напрямую использован при построении методов анализа волатильности стохастических процессов. Тем не менее для полноты изложения и лучшего понимания того, насколько изменяются основные статистические закономерности в поведении финансовых индексов при непостоянной интенсивности потока информативных событий, здесь будет приведена простейшая формулировка закона больших чисел для обобщенных процессов Кокса с ненулевым средним.

ТЕОРЕМА 1.3.5. *Предположим, что $\Lambda(t) \xrightarrow{P} \infty$ при $t \rightarrow \infty$ и $\mathbf{E}X_1 = a \neq 0$. Для того чтобы существовала случайная величина Z такая, что*

$$\frac{S(t)}{t} \Rightarrow Z \quad (t \rightarrow \infty),$$

необходимо и достаточно, чтобы существовала неотрицательная случайная величина U такая, что

$$\frac{\Lambda(t)}{t} \Rightarrow U \quad (t \rightarrow \infty).$$

Более того, при этом $Z \stackrel{d}{=} aU$.

Теорема 1.3.4 представляет собой усиление одного результата О. Лундберга (Lundberg, 1964). Ее доказательство можно найти, например, в книге (Bening and Korolev, 2002).

Таким образом, как видно из теоремы 1.3.4, в законе больших чисел для обобщенных процессов Кокса (как и для других более общих случайных сумм) предел “средних арифметических” может быть случайным. Это никак не противоречит традиционному представлению о сути законов больших чисел. Действительно, в классических законах больших чисел при увеличении числа слагаемых в средних арифметических информация о конкретном виде распределения отдельного слагаемого как бы затухает, сжимаясь в информацию о значении лишь одной характеристики распределения – математического ожидания. Именно этот эффект и описывается теоремой 1.3.4, в которой (случайный) предел не зависит от вида распределения слагаемого, а зависит лишь от его математического ожидания. При этом тип предельного распределения полностью определяется статистическими закономерностями поведения накопленной интенсивности.

На основе приведенных выше предельных теорем для *одномерных* распределений обобщенных процессов Кокса в работе (Кащеев, 2001) (также см. (Bening and Korolev, 2002), глава 10) доказаны некоторые функциональные предельные теоремы, согласно которым предельными процессами для обобщенных процессов Кокса, у которых управляющий процесс имеет независимые и стационарные приращения (см. теорему 1.3.3), являются подчиненные процессы Леви, у которых субординатор определяется процессом $\Lambda(t)$, тогда как “внешний” процесс определяется свойствами случайных величин $X_1, X_2, \dots$. В рассматриваемом здесь случае, когда дисперсии скачков предполагаются конечными, “внешний” процесс является винеровским (или процессом броуновского движения). Эти результаты, которые будут приведены и более подробно обсуждены ниже в разделе 2.6, можно считать своеобразным мостиком, связывающим микромоделли процессов плазменной турбулентности или эволюции финансовых индексов (на очень малых интервалах времени или, что то же самое, при малых временных масштабах) с популярными ныне макромоделями (крупномасштабными моделями) этих процессов вида броуновского движения со случайными сносом и диффузией (подобные модели описывают тепловое движение частицы в среде со случайной температурой и/или вязкостью). Таким образом, приведенные выше результаты дают еще одно теоретическое обоснование целесообразности моделирования распределений приращений (логарифмов) процессов плазменной турбулентности или эволюции финансовых индексов с помощью масштабных (или, в общем случае, сдвиг/масштабных) смесей нормальных законов. При этом, в отличие от случайных блужданий с непрерывным временем, случайные блуждания, порожденные обобщенными процессами Кокса, приводят к негауссовым предельным распределениям, хвосты которых могут убывать произвольно медленно или произвольно быстро (но не быстрее хвостов самого нормального закона).

2.

Моделирование распределений приращений финансовых индексов смесями нормальных законов

В данной главе схема неоднородных случайных блужданий с непрерывным временем, рассмотренная в предыдущем разделе, будет использована для моделирования эволюции финансовых индексов и, прежде всего, биржевых цен. Эта предметная область позволяет наглядно продемонстрировать основные принципы моделирования многих стохастических хаотических процессов, поскольку наблюдается явная аналогия между механизмами формирования процессов биржевых цен и других не менее важных хаотических процессов в других областях, например, траффика в сложных коммуникационных сетях (таких как Интернет), процессов, протекающих в турбулентной плазме. Здесь будет построена конкретная модель, являющаяся основой СРС-метода анализа компонент волатильности.

2.1. Неоднородность операционного времени и нормальные смеси

А время... текло – и быстро, и медленно, и рвалось, и перекачывалось по камням в отчаянной усталости.

В. Орлов. “Альтист Данилов”.

Многие исследователи обращали внимание на то, что реальная интен-

сивность биржевых торгов крайне изменчива, и предпринимали попытки объяснить островершинность распределений приращений биржевых цен, основываясь на этой изменчивости. По-видимому, первым, кто систематически исследовал неоднородность операционного времени на биржах, был П. Кларк (Clark, 1970), (Clark, 1973). Как уже отмечаось, вместо обычных винеровских процессов П. Кларк предложил в качестве моделей динамики биржевых цен использовать подчиненные винеровские процессы со случайным временем вида $W(X(t))$, где $W(t)$, $t \geq 0$ – винеровский процесс, а $X(t)$ – процесс с неубывающими траекториями, начинающимися в нуле.

По Кларку, при использовании центральной предельной теоремы в качестве базы при построении моделей для распределений приращений биржевых цен из-за неоднородности времени нарушается не конечность дисперсий, а другое структурное предположение, именно, предположение о том, что число слагаемых в сумме неслучайно.

На основе догадки Кларка сейчас будет показано, как построить более реалистичную и довольно простую модель процесса биржевых цен на интервалах времени малой длины (так сказать, элементарную модель на микроуровне), а затем перейти к моделям типа $W(X(t))$ на макроуровне с помощью нескольких версий центральной предельной теоремы для случайных сумм. Эти модели, по сути, являются удобными аппроксимациями реальных процессов.

Начнем с простейшей модели. Рассмотрим процесс цены акции на интервале времени малой длины. Основная посылка будет заключаться в том, что *формальное выражение цены акции существует только в заключенном на бирже контракте* (сделке). Эта посылка не предполагает существования некоего “объективно существующего” “основного” (“underlying”) или “истинного” процесса эволюции биржевой цены и прямо противоречит принимаемой во многих исследованиях концепции, согласно которой проблемы участников рынка обусловлены незнанием этого процесса, характеризующего “истинную” цену. Таким образом в соответствии с нашим основным предположением на интервалах времени между последовательными заключениями контрактов цена акции остается постоянной и равной той цене, которая указана в последнем заключенном до рассматриваемого момента времени контракте. Тогда изменение цены акции может быть описано маркированным точечным процессом $\{(T_j, P_j)\}_{j \geq 1}$, где T_j – момент заключения j -го контракта, а P_j – цена акции, указанная в j -ом контракте (контракты нумеруются в хронологическом порядке). Таким образом, если $P(\tau)$ – цена акции в момент τ , то

$$P(\tau) = P_j \text{ при } T_j \leq \tau < T_{j+1}.$$

Пусть $N(\tau)$ – число контрактов, заключенных к моменту τ (мы считаем, что отсчет начинается с $\tau = 0$). Как отмечено, например, в (Hull, 2003), ключевым моментом формирования цен акций и стратегии игры на бирже является то, что инвесторы, главным образом, заинтересованы в прибыли, выраженной в процентах от стоимости акций (то есть в относительной прибыли) и не зависящей от самой цены. Поэтому с учетом сказанного выше будем рассматривать следующее представление процесса $P(\tau)$:

$$\frac{P(\tau)}{P(0)} = \prod_{j=1}^{N(\tau)} \frac{P_j}{P_{j-1}}, \quad P_0 = P(0). \quad (2.1.1)$$

(Если бы нас интересовали *абсолютные* прибыли, то мультипликативное представление (2.1.1) следовало бы заменить его аддитивным аналогом

$$P(\tau) - P(0) = \sum_{j=1}^{N(\tau)} (P_j - P_{j-1}), \quad P_0 = P(0),$$

что просто привело бы к тому, что в нижеследующих выкладках надо было бы опустить знак логарифма.)

Обозначим

$$S(\tau) = \log P(\tau) - \log P(0), \quad X_j = \log P_j - \log P_{j-1}.$$

Тогда из (2.1.1) вытекает, что

$$S(\tau) = \sum_{j=1}^{N(\tau)} X_j \quad \text{и} \quad P(\tau) = P(0) \exp\{S(\tau)\}.$$

Предположим, что $X_1, X_2, \dots$ – независимые (и пока, вообще говоря, необязательно одинаково распределенные) случайные величины. Рассмотрим поведение распределения случайных величин $P(t)$ при неограниченном возрастании $N(t)$. Отметим, что это не означает, что обязательно $t \rightarrow \infty$. Интервал $[0, t]$ (пока) остается фиксированным. Для удобства построения аппроксимаций формально рассмотрим последовательность целочисленных неотрицательных случайных величин $N_1, N_2, \dots$ и предположим, что при каждом $k \geq 1$ случайные величины N_k и $X_1, X_2, \dots$ независимы.

Пусть $f_k(s)$ – характеристическая функция случайной величины $S(\tau)$, в которой $N(\tau)$ заменено на N_k :

$$f_k(s) = \mathbb{E} \exp \left\{ i s \sum_{j=1}^{N_k} X_j \right\}, \quad s \in \mathbb{R}.$$

Напомним, что семейство случайных величин $\{\xi_n\}_{n \geq 1}$ называется *слабо компактным*, если

$$\lim_{x \rightarrow \infty} \sup_n \mathbf{P}(|\xi_n| > x) = 0.$$

Справедлив следующий результат.

ТЕОРЕМА 2.1.1. *Пусть распределения слагаемых $X_1, X_2, \dots$ таковы, что для некоторых $b_k > 0$, $b_k \rightarrow \infty$ ($k \rightarrow \infty$) имеет место сходимость*

$$\mathbf{P}\left(\frac{1}{b_k} \sum_{j=1}^k X_j < x\right) \Rightarrow \Phi(x) \quad (k \rightarrow \infty). \quad (2.1.2)$$

Предположим, что при некоторых $d_k > 0$, $d_k \rightarrow \infty$ ($k \rightarrow \infty$) семейство случайных величин $\{b_{N_k}/d_k\}$ слабо компактно. Тогда при каждом $s \in \mathbb{R}$

$$\lim_{k \rightarrow \infty} \left| f_k\left(\frac{s}{d_k}\right) - \int_0^\infty \exp\left\{-\frac{x^2 s^2}{2}\right\} dA_k(x) \right| = 0,$$

где $A_k(x) = \mathbf{P}(b_{N_k} < x d_k)$.

Доказательство этого результата можно найти в (Gnedenko and Korolev, 1996).

Очевидно, что, если обозначить $U_k = b_{N_k}/d_k$, то

$$\int_0^\infty \exp\left\{-\frac{x^2 s^2}{2}\right\} dA_k(x) = \mathbf{E} \exp\{is(XU_k)\}, \quad s \in \mathbb{R},$$

где $\mathbf{P}(X < x) = \Phi(x)$; более того, X и U_k независимы при каждом $k \geq 1$. Таким образом, с помощью теоремы 2.1.1 мы приходим к следующему заключению: если число контрактов, заключенных в течение интервала времени $[0, t]$, довольно велико и последовательные изменения цены акции от контракта к контракту удовлетворяют условиям центральной предельной теоремы, то для некоторой неотрицательной случайной величины U справедливо приближенное равенство

$$\mathbf{P}(S(t) < x) \approx \mathbf{P}(XU < x) = \mathbf{E} \Phi\left(\frac{x}{U}\right), \quad x \in \mathbb{R}, \quad (2.1.3)$$

где $\mathbf{P}(X < x) = \Phi(x)$ и случайные величины X и U независимы. Другими словами, законы, аппроксимирующие распределения приращений логарифмов процессов цен акций следует искать среди масштабных смесей

нормальных законов. Отметим, что если случайные величины $X_1, X_2, \dots$ одинаково распределены, то (2.1.2) эквивалентно конечности дисперсии случайной величины X_1 . При этом $b_k = \sqrt{k \mathbb{D} X_1}$.

Теорема 2.1.1 описывает *сближение* распределений приращений логарифмов биржевых цен с масштабными смесями нормальных законов. Если воспользоваться утверждениями о *сходимости* распределений случайных сумм, например, теоремой 1.3.1, то можно заметить, что тип смешивающего распределения в правой части (2.1.3) определяется предельным поведением случайных величин b_{N_k}/d_k . В связи с этим теорема 1.3.1 позволяет ставить вопрос о реконструкции распределения U_k по распределению U , так как согласно этой теореме в предположении (2.1.2) условие $U_k \implies U$ не только достаточно, но и необходимо для возможности замены приближенного неравенства в (2.1.3) точным в пределе при $N_k \rightarrow \infty$ (по вероятности) при надлежащей нормировке случайной величины $S(t)$. Справедливость сказанного будет проиллюстрирована в следующем параграфе, тесно связанном с обобщенными процессами Кокса.

2.2. Неоднородный дискретный стохастический хаос и процессы Кокса

Хаос, однажды выбранный, хаос застывший – есть уже система. Сильнее было бы усовершенствовать решение так, чтобы хаос еще хаотически менялся.

А. И. Солженицын. “В круге первом”.

В этом разделе будут конкретизированы выводы предыдущего раздела. Опираясь на представление (2.1.1) для процесса цены акции, как и прежде, будем использовать обозначение $S(\tau) = \log P(\tau) - \log P(0)$, $X_j = \log P_j - \log P_{j-1}$. Напомним, что из представления (2.1.1) при этом вытекает, что

$$S(\tau) = \sum_{j=1}^{N(\tau)} X_j.$$

Попробуем предложить более или менее реалистичные модели для процесса $N(\tau)$ в этом представлении.

Имеются все основания считать, что моменты времени, в которые осуществляются последовательные сделки на биржах, распределены хаотически по оси времени, так что случайные величины T_j , $j \geq 1$, образуют “хаотический” точечный стохастический процесс.

Здесь необходимо сделать некоторое терминологическое и методологическое отступление. Среди физиков и специалистов по динамическим системам бытует мнение, что в определенном смысле понятие стохастичности является более сильным, нежели понятие хаотичности, другими словами, что стохастические системы являются более неопределенными, чем хаотические. На самом деле такая классификация является более чем условной и, вообще говоря, неверной. Как уже отмечалось во введении, язык теории стохастических процессов и язык теории детерминированного хаоса в равной степени используются для описания поведения непредсказуемых систем. Но ни в коем случае термины “стохастический” и “хаотический” не относятся к степеням проявления *одного и того же* качества или свойства – непредсказуемости или неопределенности. Во-первых, если стохастические свойства системы формализованы с помощью теории вероятностей, построенной на колмогоровских аксиомах, то математические модели, полученные в результате, будут адекватны, только если поведение системы может быть воспроизведено (или хотя бы предположительно может быть воспроизведено) неограниченно много раз в остающихся неизменными начальных условиях. Это обстоятельство, вообще говоря, создает определенные проблемы при теоретико-вероятностном моделировании *уникальных* хаотических динамических систем. В то же время, такие системы могут успешно изучаться с помощью методов теории детерминированного хаоса.

Напротив, поведение стохастических систем может в большей или меньшей степени быть хаотичным, если под “хаотичностью” поведения подразумевается его непредсказуемость. Используя рассуждения, связанные с понятием энтропии как меры неопределенности, можно убедиться, что при определенных условиях те или иные распределения вероятностей являются неопределенными (то есть непредсказуемыми и, стало быть, хаотическими) в большей или меньшей степени. Многочисленные примеры распределений, являющихся “наиболее неопределенными” при тех или иных условиях или ограничениях, приведены в книге (Karim, 1989).

Более того, детерминированные “хаотические” системы могут обладать различными “стохастическими” свойствами. Например, в книге (Анищенко и др., 2003) приведено много примеров вычислений вероятностных характеристик (в частности, эмпирических плотностей распределения вероятностей), описывающих стохастические свойства детерминированных хаотических систем. (Кстати, распределения вероятностей, вычисленные (статистически оцененные) в (Анищенко и др., 2003) для некоторых хаотических динамических систем, никак нельзя отнести к наиболее энтропийным и, стало быть, неопределенным, то есть “стохастически хаотическим”

распределениям).

Итак, наилучшей моделью однородного хаотического считающего процесса является, как можно было убедиться в предыдущем разделе, пуассоновский процесс. Однако, поскольку биржевое время неоднородно, что проявляется в стохастичности интенсивности торгов, разумно считать $N(\tau)$ процессом Кокса, управляемым процессом $\Lambda(\tau)$, производная которого по τ характеризует интенсивность торгов (напомним, что процесс Кокса, иначе называемый дважды стохастическим пуассоновским процессом, определяется как суперпозиция однородного пуассоновского процесса $N_1(t)$ с единичной интенсивностью и независимого от него процесса $\Lambda(t)$, $t \geq 0$, траектории которого стартуют из нуля, неограниченно монотонно не убывают и непрерывны справа). Действительно, моменты заключения контрактов на бирже хаотически рассредоточены на оси времени. Однако этот хаос не является однородным, так как интенсивность торгов изменяется в течение биржевого дня: сразу после начала торгов совершают свои сделки те трейдеры, которые получили заказ до открытия торгов, затем интенсивность несколько спадает. В течение дня наблюдается несколько систематических всплесков интенсивности торгов, имеющих, так сказать, “арбитражный” характер и связанных с поступлением информации о положении дел на открывающихся биржах, расположенных в других часовых поясах. Ближе к концу торгов интенсивность возрастает, так как совершают свои сделки те трейдеры, которые выжидали в течение дня. Наряду с “систематическими”, то есть наблюдаемыми день ото дня, всплесками интенсивности, конечно же, возможны и случайные всплески, связанные, например, с теми или иными политическими событиями. Поэтому процессы Кокса действительно являются разумными моделями для реальных потоков заключения контрактов на биржах.

Таким образом, будем считать, что введенная выше случайная величина $S(\tau)$ является значением обобщенного процесса Кокса в момент τ . Одними из основных характеристик, интересующих тех, кто играет на бирже, являются максимальная и минимальная цены за определенный интервал времени, например, биржевой день. Информация об этих характеристиках существенно определяет стратегию поведения того или иного игрока на следующем интервале времени. Обозначим

$$\underline{P}(t) = \min_{0 \leq \tau \leq t} P(\tau), \quad \overline{P}(t) = \max_{0 \leq \tau \leq t} P(\tau). \quad (2.2.1)$$

Будем считать, что $E\Lambda(\tau) = \lambda\tau$, $\lambda > 0$. В этом разделе будут построены *асимптотические* аппроксимации для распределений случайных

величин $\underline{P}(t)$ и $\overline{P}(t)$. Но по смыслу задачи разумно считать фиксированной длину t интервала времени, на котором рассматривается экстремумы процесса цен. Поэтому поведение случайных величин $\underline{P}(t)$ и $\overline{P}(t)$ будем изучать при неограниченно растущей средней интенсивности торгов, то есть при $\lambda \rightarrow \infty$. Такой подход вполне оправдан, поскольку реально на крупных биржах за интервалы времени $[0, t]$, представляющие разумный интерес (часы, дни), заключается огромное число контрактов. Таким образом, в задачу вводится еще один параметр λ и поэтому фактически $\Lambda(\tau) = \Lambda(\tau, \lambda)$, $P(\tau) = P(\tau, \lambda)$, $\underline{P}(t) = \underline{P}(t, \lambda)$, $\overline{P}(t) = \overline{P}(t, \lambda)$. Функция распределения максимума стандартного винеровского процесса на отрезке $[0, 1]$ будет обозначаться $G(x)$,

$$G(x) = 2\Phi(\max\{0, x\}) - 1, \quad x \in \mathbb{R}.$$

Предположим, что случайные величины $X_j = \log P_j - \log P_{j-1}$, $j \geq 1$, независимы и одинаково распределены с $\mathbf{E}X_j = 0$ и $0 < \sigma^2 = \mathbf{D}X_j < \infty$. Справедливо следующее утверждение.

ТЕОРЕМА 2.2.1. *Предположим, что $\Lambda(\tau, \lambda) \rightarrow \infty$ по вероятности при $\lambda \rightarrow \infty$. Тогда следующие условия эквивалентны:*

$$(i) \quad \mathbf{P} \left(\frac{\log P(t, \lambda) - \log P(0)}{\sigma \sqrt{t\lambda}} < x \right) \Rightarrow F(x) \quad (\lambda \rightarrow \infty);$$

$$(ii) \quad \mathbf{P} \left(\frac{\log \underline{P}(t, \lambda) - \log P(0)}{\sigma \sqrt{t\lambda}} < x \right) \Rightarrow \underline{F}(x) \quad (\lambda \rightarrow \infty);$$

$$(iii) \quad \mathbf{P} \left(\frac{\log \overline{P}(t, \lambda) - \log P(0)}{\sigma \sqrt{t\lambda}} < x \right) \Rightarrow \overline{F}(x) \quad (\lambda \rightarrow \infty);$$

(iv) *существует неотрицательная случайная величина U такая, что*

$$\frac{\Lambda(t, \lambda)}{t\lambda} \Rightarrow U \quad (\lambda \rightarrow \infty).$$

При этом

$$F(x) = \mathbf{E}\Phi(x/\sqrt{U}), \tag{2.2.2}$$

$$\overline{F}(x) = \mathbf{E}G(x/\sqrt{U}) = 2\mathbf{E}\Phi(\max\{0, x\}/\sqrt{U}) - 1,$$

$$\underline{F}(x) = 1 - \mathbf{E}G(-x/\sqrt{U}) = 2 \left[1 - \mathbf{E}\Phi(-\min\{0, x\}/\sqrt{U}) \right], \quad x \in \mathbb{R}.$$

Д о к а з а т е л ь с т в о см. в (Королев, 2000).

Таким образом, при больших λ справедливы следующие приближенные соотношения:

$$P(P(t, \lambda) < x) \approx E\Phi\left(\frac{\log x - \log P(0)}{\sigma\sqrt{t\lambda U}}\right), \quad x > 0, \quad (2.2.3)$$

$$\begin{aligned} P(\bar{P}(t, \lambda) < x) &\approx EG\left(\frac{\log x - \log P(0)}{\sigma\sqrt{t\lambda U}}\right) = \\ &= 2E\Phi\left(\frac{\max\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda U}}\right) - 1, \quad x > 0, \end{aligned} \quad (2.2.4)$$

$$\begin{aligned} P(\underline{P}(t, \lambda) < x) &\approx 1 - EG\left(-\frac{\log x - \log P(0)}{\sigma\sqrt{t\lambda U}}\right) = \\ &= 2\left[1 - E\Phi\left(-\frac{\min\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda U}}\right)\right], \quad x > 0. \end{aligned} \quad (2.2.5)$$

Обратим внимание, что аппроксимирующие законы в соотношениях (2.2.3) – (2.2.5) имеют существенно более тяжелые хвосты по сравнению с лог-нормальным или лог- G распределениями. Для иллюстрации рассмотрим некоторые примеры.

ПРИМЕР 2.2.1. Рассмотрим ситуацию, в которой управляющий процесс Λ имеет геометрическое распределение. Укажем конкретный вид аппроксимирующих распределений в правых частях (2.2.3) – (2.2.5) для этой ситуации. Данная модель оказывается весьма поучительной. Пусть $\Lambda(t, \lambda)$ имеет геометрическое распределение с параметром $(t\lambda)^{-1}$. Несложно видеть, что в этом случае предельная при $\lambda \rightarrow \infty$ случайная величина U имеет стандартное показательное распределение, а из (2.2.3) – (2.2.5) вытекает, что

$$P(P(t, \lambda) < x) \approx \begin{cases} 0, & \text{если } x < 0, \\ \frac{1}{2} \left(\frac{x}{P(0)}\right)^{\sqrt{2}(\sigma\sqrt{\lambda t})^{-1}}, & \text{если } 0 \leq x \leq P(0), \\ 1 - \frac{1}{2} \left(\frac{P(0)}{x}\right)^{\sqrt{2}(\sigma\sqrt{\lambda t})^{-1}}, & \text{если } x > P(0), \end{cases}$$

$$P(\bar{P}(t, \lambda) < x) \approx \begin{cases} 0, & \text{если } x < P(0), \\ 1 - \left(\frac{P(0)}{x}\right)^{(\sigma\sqrt{2\lambda t})^{-1}}, & \text{если } x \geq P(0). \end{cases}$$

Обратим внимание на паретовский (степенной) характер убывания хвостов обоих этих распределений. Каким бы малым ни было $\delta_0 > 0$, можно указать столь большое λ , что у обоих этих распределений будут отсутствовать моменты всех порядков $\delta > \delta_0$.

В некоторых случаях теорема 2.2.1 дает возможность довольно просто построить распределения случайных величин $\bar{P}(t, \lambda)$ и $\underline{P}(t, \lambda)$ по распределению случайной величины $P(t, \lambda)$. А именно, если аппроксимацию распределения логарифма случайной величины $P(t, \lambda)$ удастся представить в виде масштабной смеси нормальных законов, то распределения случайных величин $\bar{P}$ и $\underline{P}$ оказываются масштабными смесями, соответственно, функций распределения $G(x)$ и $1 - G(-x)$ с тем же самым смешивающим распределением (этот феномен является проявлением принципа инвариантности). Поясним сказанное примерами.

ПРИМЕР 2.2.2. Лог-устойчивый тип. Устойчивые законы весьма популярны в качестве моделей распределений логарифмов приращений биржевых цен. Предположим, что распределение логарифма случайной величины $P(t, \lambda)$ аппроксимируется устойчивым законом $G_\alpha(x)$ с характеристической функцией $g_\alpha(t) = \exp\{-|t|^\alpha\}$, $0 < \alpha \leq 2$. Тогда согласно теореме 3.3.1 в (Золотарев, 1983), справедливо представление

$$G_\alpha(x) = \int_0^\infty \Phi(x/\sqrt{y}) dG_{\alpha/2,1}(y), \quad (2.2.6)$$

где $G_{\alpha/2,1}(y)$ – функция распределения строго устойчивого закона с характеристическим показателем $\alpha/2$, сосредоточенного на положительной полуоси и определяемого характеристической функцией

$$g_{\alpha/2,1}(t) = \exp\left\{-|t|^{\alpha/2} \exp\left\{-\frac{i\pi\alpha}{4}\text{sign}t\right\}\right\}.$$

Представление (2.2.6) по форме совпадает с (2.2.2). Предполагая выполненными условия теоремы 2.2.1 (большая интенсивность торгов, конечность дисперсий логарифмов элементарных приращений и равенство нулю их ожиданий), с помощью соотношений (2.2.6) и (2.2.4) можно заключить, что в данном случае

$$\begin{aligned} P(\bar{P}(t, \lambda) < x) &\approx \int_0^\infty G\left(\frac{\log x - \log P(0)}{\sigma\sqrt{t\lambda y}}\right) dG_{\alpha/2,1}(y) = \\ &= 2 \int_0^\infty \Phi\left(\frac{\max\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda y}}\right) dG_{\alpha/2,1}(y) - 1 = \end{aligned}$$

$$= 2G_\alpha \left(\frac{\max\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda}} \right) - 1, \quad x > 0.$$

Аналогично,

$$P(\underline{P}(t, \lambda) < x) \approx 2 \left[1 - G_\alpha \left(-\frac{\min\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda}} \right) \right], \quad x > 0.$$

ПРИМЕР 2.2.3. Лог-симметричный NIG-тип. Согласно данным статистического анализа приращений цен на датских и немецких биржах, проведенного О. Барндорфф-Нильсеном и его коллегами (Barndorff-Nielsen, 1994), распределение логарифмов приращений биржевых цен хорошо аппроксимируется так называемым нормальным\обратным гауссовским (Normal\Inverse Gaussian) или сокращенно NIG-распределением, определяемым плотностью

$$\begin{aligned} p(x; \alpha, \beta, \delta, \mu) &= \\ &= A(\alpha, \beta, \delta, \mu) \left[Q\left(\frac{x - \mu}{\delta}\right) \right]^{-1} K_1\left(\delta \alpha Q\left(\frac{x - \mu}{\delta}\right)\right) e^{\beta x}, \quad x \in \mathbb{R}, \end{aligned}$$

где $A(\alpha, \beta, \delta, \mu) = \frac{\alpha}{\pi} \exp\{\delta\sqrt{\alpha^2 - \beta^2} - \beta\mu\}$, $Q(x) = \sqrt{1 + x^2}$, K_1 — модифицированная функция Бесселя третьего рода с индексом 1 (иногда это распределение называют гауссовским\обратным гауссовским и используют аббревиатуру GIG (Gaussian\Inverse Gaussian); здесь же используются иные термины и сокращения, чтобы избежать недоразумений, связанных с тем, что аббревиатура GIG, как правило, используется для так называемого обобщенного обратного гауссовского распределения). Соответствующую функцию распределения обозначим $B(x; \alpha, \beta, \delta, \mu)$. Известно (Barndorff-Nielsen and Blaesild, 1981), что

$$B(x; \alpha, \beta, \delta, \mu) = \mathbb{E}\Phi\left(\frac{x - \mu - \beta U}{\sqrt{U}}\right), \quad (2.2.7)$$

где U — случайная величина с обратным нормальным распределением, определяемым плотностью

$$\begin{aligned} q(u; \delta, \sqrt{\alpha^2 - \beta^2}) &= \\ &= \frac{\delta}{\sqrt{2\pi}u^3} \exp\left\{\delta\sqrt{\alpha^2 - \beta^2}\right\} \exp\left\{-\frac{1}{2}\left[(\alpha^2 - \beta^2)u + \frac{\delta^2}{u}\right]\right\}, \quad u > 0. \end{aligned} \quad (2.2.8)$$

Из (2.2.7) вытекает, что для $B(x; \alpha, 0, \delta, 0)$ справедливо представление (2.2.2). Поэтому, предполагая, что в условиях теоремы 2.2.1 распределение логарифма случайной величины $P(t, \lambda)$ аппроксимируется функцией

распределения $B(x; \alpha, 0, \delta, 0)$, из (2.2.7) и (2.2.3) – (2.2.5) можно заключить, что для $x > 0$

$$\begin{aligned} P(\bar{P}(t, \lambda) < x) &\approx 2B\left(\frac{\max\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda}}; \alpha, 0, \delta, 0\right) - 1; \\ P(\underline{P}(t, \lambda) < x) &\approx 2\left[1 - B\left(-\frac{\min\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda}}; \alpha, 0, \delta, 0\right)\right]. \end{aligned}$$

Ниже, в частности, в разделе 2.6 (см. пример 2.6.4) мы еще вернемся к данному примеру.

ПРИМЕР 2.2.4. Лог-симметричный гиперболический тип. В некоторых работах (см., например, (Eberlein and Keller, 1995)) высокое согласие реальных данных об эволюции финансовых индексов с так называемым гиперболическим распределением. Как и NIG-распределение, оно введено в практику О. Барндорфф-Нильсеном (Barndorff-Nielsen, 1977). Гиперболическое распределение имеет плотность

$$\begin{aligned} h(x; \alpha, \beta, \delta, \mu) &= \\ &= C(\alpha, \beta, \delta) \exp\left\{-\alpha\sqrt{\delta^2 + (x - \mu)^2} + \beta(x - \mu)\right\}, \quad x \in \mathbb{R}, \end{aligned}$$

где

$$C(\alpha, \beta, \delta) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})},$$

K_1 – модифицированная функция Бесселя третьего рода с индексом 1, $\alpha > 0$, $0 \leq |\beta| < \alpha$, $\mu \in \mathbb{R}$, $\delta \geq 0$. Соответствующую функцию распределения обозначим $H(x; \alpha, \beta, \delta, \mu)$. Известно (Barndorff-Nielsen, 1977), что

$$H(x; \alpha, \beta, \delta, \mu) = \mathbf{E}\Phi\left(\frac{x - \mu - \beta U}{\sqrt{U}}\right), \quad (2.2.9)$$

где U – случайная величина с обратным нормальным распределением, определяемым плотностью

$$\begin{aligned} q^*(u; \delta, \alpha^2 - \beta^2) &= \\ &= \frac{\sqrt{\alpha^2 - \beta^2}}{2\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} \exp\left\{-\frac{1}{2}\left((\alpha^2 - \beta^2)x + \frac{\delta^2}{x}\right)\right\}, \quad u > 0. \end{aligned} \quad (2.2.10)$$

Из (2.2.9) вытекает, что для $H(x; \alpha, 0, \delta, 0)$ справедливо представление (2.2.2). Поэтому, предполагая, что в условиях теоремы 2.2.1 распределение логарифма случайной величины $P(t, \lambda)$ аппроксимируется функцией

распределения $H(x; \alpha, 0, \delta, 0)$, из (2.2.9) и (2.2.3) – (2.2.5) можно заключить, что для $x > 0$

$$\begin{aligned} P(\bar{P}(t, \lambda) < x) &\approx 2H\left(\frac{\max\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda}}; \alpha, 0, \delta, 0\right) - 1; \\ P(\underline{P}(t, \lambda) < x) &\approx 2\left[1 - H\left(-\frac{\min\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda}}; \alpha, 0, \delta, 0\right)\right]. \end{aligned}$$

Мы еще вернемся к этому примеру в разделе 2.6 (см. примеры 2.6.2 и 2.6.3).

ПРИМЕР 2.2.5. В работе (Holdom, 1998) для моделирования финансовых данных используется идея “турбулентной” перемежаемости (см. раздел 7.1). При этом в качестве смешивающего используется распределение с плотностью

$$v(x; \theta) = \frac{2\theta^\theta}{\Gamma(\theta)} x^{2\theta-1} e^{-\theta x^2}, \quad x \geq 0.$$

В частности, при $\theta = \frac{1}{2}$ плотность $v(x; \theta) = v(x; \frac{1}{2})$ имеет вид

$$v(x; \frac{1}{2}) = \frac{\sqrt{2}}{\sqrt{\pi}} e^{-x^2/2}, \quad x \geq 0,$$

и совпадает с плотностью вероятностей $g(x) = G'(x)$ максимума стандартного винеровского процесса на отрезке $[0, 1]$. Соответствующее смешанное гауссовское распределение задается плотностью вида

$$\begin{aligned} h^*(x; \theta) &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{v(s; \theta)}{s} \exp\left\{-\frac{x^2}{2s^2}\right\} ds = \\ &= \frac{2^{(3-2\theta)/4} \theta^{(1+2\theta)/4}}{\sqrt{\pi} \Gamma(\theta)} |x|^{\theta-1/2} K_{\theta-1/2}(\sqrt{2\theta}|x|), \quad x \in \mathbb{R}, \end{aligned} \quad (2.2.11)$$

где $K_\nu(x)$ – модифицированная функция Бесселя второго рода с индексом ν . При $\theta = 1$ плотность $h^*(x; \theta)$ задает двойное показательное распределение (распределение Лапласа)

$$h^*(x; 1) = \frac{\sqrt{2}}{2} \exp\{-\sqrt{2}|x|\}, \quad x \in \mathbb{R}.$$

Функцию распределения, соответствующую плотности $h^*(x; \theta)$, обозначим $H^*(x; \theta)$:

$$H^*(x; \theta) = \int_{-\infty}^x h^*(y; \theta) dy.$$

Тогда из теоремы 2.2.1 вытекает, что

$$\begin{aligned} P(P(t, \lambda) < x) &\approx H^* \left(\frac{\max\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda}}; \theta \right); \\ P(\overline{P}(t, \lambda) < x) &\approx 2H^* \left(\frac{\max\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda}}; \theta \right) - 1; \\ P(\underline{P}(t, \lambda) < x) &\approx 2 \left[1 - H^* \left(-\frac{\min\{0, \log x - \log P(0)\}}{\sigma\sqrt{t\lambda}}; \theta \right) \right]. \end{aligned}$$

Семейство распределений типа (2.2.11) обладает свойством замкнутости относительно операции свертки. В работе (Holdom, 1998) с помощью распределения (2.2.11) удалось хорошо промоделировать следующий эффект, наблюдаемый как в турбулентных потоках, так и при анализе финансовых данных: хотя наблюдаемая автокорреляция самих изменений логарифмов биржевых цен пренебрежимо мала, автокорреляция некоторых функций от изменений логарифмов биржевых цен довольно заметна.

За дальнейшими сведениями о существующих моделях распределений приращений (логарифмов) биржевых цен заинтересованный читатель может обратиться к книге (Ширяев, 1998). Некоторые из них будут более подробно рассмотрены ниже в разделе 2.6.

С помощью теоремы 2.2.1 можно в определенном смысле уточнить утверждения, объясняющие наблюдаемую не-нормальность распределений логарифмов приращений биржевых цен (или “не-лог-нормальность” распределений самих этих приращений) на интервалах времени умеренной длины, которая проявляется в более тяжелых хвостах у реальных распределений. В то же время, у процессов биржевых цен отсутствует масштабная (по времени) инвариантность в том смысле, что распределения приращений процессов за интервалы времени большей длины намного ближе к нормальным, нежели распределения приращений процессов за интервалы меньшей длины. А именно, при увеличении длины интервала времени, приращения за который рассматриваются, наблюдается эффект сближения распределений логарифмов относительных приращений финансовых индексов, в частности, биржевых цен, с нормальным законом (так называемый эффект “унормализования”). При рассмотрении экстремумов процессов цен аналогом этого феномена должно быть сближение распределения максимума процесса цен с функцией распределения $G(x)$, а распределения минимума этого процесса – с функцией распределения $1 - G(-x)$.

Построенные выше модели эволюции биржевых цен довольно естественно объясняют этот эффект. В противоположность тем рассуждениям, которые привели к теореме 2.2.1, сейчас длина интервала времени t

не будет считаться фиксированной, а напротив, в предположении, что она измеряется в каких-либо *целых* единицах (например, биржевых днях), она будет устремляться к бесконечности при постоянной средней интенсивности торгов λ . Итак, пусть $t = n = 1, 2, \dots$. Рассмотрим поведение распределений случайных величин $\underline{P}(n) = \underline{P}(n, \lambda)$, и $\bar{P}(n) = \bar{P}(n, \lambda)$, определяемых соотношением (2.2.1) с заменой t на n , при $n \rightarrow \infty$. В этой ситуации разумно считать, что управляющий процесс $\Lambda(n, \lambda)$ имеет вид

$$\Lambda(n, \lambda) = \sum_{i=1}^n \Lambda_i,$$

где $\Lambda_1, \Lambda_2, \dots$ – независимые одинаково распределенные случайные величины с $E\Lambda_i = \lambda$, $i \geq 1$. Тогда в силу закона больших чисел

$$\frac{\Lambda(n, \lambda)}{n\lambda} = \frac{1}{n\lambda} \sum_{i=1}^n \Lambda_i \implies 1 \quad (n \rightarrow \infty),$$

то есть распределение случайной величины U в (2.2.2) вырождено в единице. Поэтому в силу теоремы 2.2.1 в этом случае

$$F(x) = \Phi(x), \quad \bar{F}(x) = G(x), \quad \underline{F}(x) = 1 - G(-x).$$

Базовая модель рассматриваемого процесса, основанная на его представлении в виде обобщенного процесса Кокса (с точностью до взятия логарифма) позволяет *напрямую связать интенсивность потока информативных событий* (“*тиков*”) с волатильностью. Действительно, как уже отмечалось во введении, накопленная интенсивность потока информативных событий – накопленная интенсивность $\Lambda(\tau)$ потока скачков обобщенного процесса Кокса – статистически эквивалентна субординатору (“хронометру”) $X(\tau)$ в представлении $S(\tau) = W(X(\tau))$ (или $P(\tau) = \exp\{W(X(\tau))\}$). А именно, как показано в работе (Ané and Geman, 2000) на основе статистического анализа большого массива экспериментальных данных об объемах торгов, количествах сделок и соответствующих изменениях цены акции, из всех разумных характеристик именно процесс накопленного количества сделок $X_{\#}(\tau)$ оказывается наилучшим субординатором в том смысле, что распределение процесса логарифма цены акций $S(\tau) = \log P(X_{\#}^{-1}(\tau))$ после обратного преобразования времени оказывается практически неотличимым от нормального. (Заметим, что асимптотическое поведение процесса Кокса совпадает с поведением соответствующей накопленной интенсивности $\Lambda(\tau)$, см., например, (Bening and Korolev, 2002).)

Далее, пусть для простоты процесс $\Lambda(\tau)$ допускает интегральное представление (1.3.4) с помощью случайного процесса $\lambda(\tau)$, интерпретируемого как мгновенная интенсивность потока информативных событий (см. раздел 1.3), а $EX_1 = 0$. Забегая вперед, договоримся под волатильностью случайного процесса понимать его дисперсию (или среднеквадратическое отклонение). Под условной *мгновенной* волатильностью процесса $S(\tau)$ относительно процесса $\lambda(\tau)$ в момент t договоримся понимать предел

$$\lim_{s \downarrow 0} \frac{1}{s} D(S(t+s) - S(t) | \lambda(\cdot)),$$

понимаемый, например, в смысле сходимости в среднем квадратичном. Тогда, используя соотношение (1.3.7), легко убедиться, что при некоторых дополнительных условиях регулярности для условной мгновенной волатильности процесса $S(\tau)$ в момент t справедливо соотношение

$$\lim_{s \downarrow 0} \frac{1}{s} D(S(t+s) - S(t) | \lambda(\cdot)) = \sigma^2 \lambda(t).$$

При этом процесс $\lambda(\tau)$ имеет смысл (условного) среднего количества контрактов, заключенных в течение интервала времени единичной длины. Именно это обстоятельство лежит в основе метода расчета волатильности, предложенного в статье (Lo, MacKinlay and Zhang, 2002), согласно которому в качестве приближенного значения волатильности используется логарифм *количества сделок в течение предыдущего часа*.

Как уже было отмечено, класс масштабных смесей нормальных законов $E\Phi(x/U)$ очень богат и содержит, помимо самого нормального распределения, например, распределения Коши, Стюдента, Лапласа (двойное или двустороннее экспоненциальное распределение). Нелишне отметить, что в виде масштабных смесей нормальных законов представимы и симметричные устойчивые распределения (см. (Золотарев, 1983), раздел 2.2.3). В связи с этим утверждения, приведенные в этом и предыдущих разделах, позволяют заключить, что лишь некоторые устойчивые законы могут представлять интерес в качестве моделей распределений приращений (логарифмов) биржевых цен, постольку-поскольку они могут быть представлены в виде смесей нормальных законов.

Приведенные выше теоремы можно рассматривать как теоретическое обоснование целесообразности использования смесей нормальных законов в качестве моделей распределений приращений (логарифмов) хаотических стохастических процессов. Применение нормальных смесей в таких целях отнюдь не является новой идеей. Достаточно вспомнить, что классические

ARCH- и GARCH-модели были предложены для анализа таких временных рядов, условные распределения которых являются гауссовскими, вследствие чего их безусловные распределения являются смесями нормальных законов. То же самое относится и к модели геометрического броуновского движения со стохастической волатильностью, задаваемой стохастическим дифференциальным уравнением (Hull and White, 1987).

Имеется большое число публикаций, посвященных изучению конкретных смесей со специальными смешивающими законами. Помимо упоминавшихся выше работ П. Кларка (Clark, 1970), (Clark, 1973), в которых автор использует логнормальное смешивающее распределение, заслуживают быть отмеченными работы (Praetz, 1972) и (Blattberg and Gonedes, 1974), где рассматривается распределение Стьюдента, (Brada and Van Tassel, 1966), (Grandier and Morgenstern, 1970), (Boness, Chen and Jatusipitak, 1974), (Oldfield, Rogalski and Jarrow, 1977), (Cox and Ross, 1976), (Merton, 1973), (Kon, 1984), (Holdom, 1998), где рассматриваются нормальные смеси с различными смешивающими распределениями. В разделе 2.6.3 мы подробно остановимся еще на нескольких примерах и убедимся, что практически все используемые сейчас разумные математические модели процессов эволюции финансовых индексов, демонстрирующие хорошее согласие с экспериментальными данными (гиперболические процессы, обобщенные гиперболические процессы, нормальные\обратные гауссовские процессы (NIG-процессы), дисперсионные гамма-процессы (VG-процессы), двусторонние гамма-процессы, процессы Мейкснера и др.) имеют одномерные распределения, представимые в виде сдвиг-масштабных смесей нормальных законов.

Однако, основываясь на теореме 2.2.1, можно высказать предположение, что, скорее всего, не существует одной универсальной случайной величины U , точнее, одного разумным образом параметризованного семейства таких случайных величин, которое удовлетворительно описывает поведение всех акций на всех биржах. Действительно, условие (iv) теоремы 2.2.1 напрямую связывает случайную величину U с изменением интенсивностей торгов или волатильностями, которые, разумеется, сильно зависят от объекта торговли, времени и места торгов. По-видимому, самое большее, на что можно надеяться, — это более или менее удачная подгонка “локально универсальной” параметрической модели, то есть для каждого конкретного объекта торговли, каждой конкретной биржи можно попробовать подобрать свою случайную величину U , то есть ее распределение, параметры которого все же заметно меняются во времени. Ситуация заметно усложняется при обобщении моделей на несимметричные случаи, которые

будут рассмотрены ниже. Скользящее (во времени) оценивание параметров таких моделей как гиперболические или NIG-модели, обеспечивающих хорошее “локальное” согласие с реальными данными, представляет собой весьма трудоемкую вычислительную задачу. Для более эффективного решения подобной проблемы и предназначен описываемый ниже СРС-метод.

2.3. Островершинность масштабных смесей нормальных законов

“Taking Three as the subject to reason about –
A convenient number to state...”

Льюис Кэрролл. “Охота на Снарка”.

Хотя описанию основных (в контексте данной работы) свойств смесей нормальных законов посвящена следующая глава, именно здесь уместно обсудить интересное и важное с точки зрения моделей, построенных выше, свойство масштабных смесей нормальных законов, а именно, их островершинность и “тяжелохвостость”.

Понятие островершинности распределения можно определять по-разному. К примеру, в прикладных исследованиях в качестве численной характеристики островершинности часто рассматривается коэффициент эксцесса $\varkappa(Z)$, который для случайной величины Z с $EZ^4 < \infty$ определяется как

$$\varkappa(Z) = E\left(\frac{Z - EZ}{\sqrt{DZ}}\right)^4.$$

Если $P(X < x) = \Phi(x)$, то $\varkappa(X) = 3$. Плотности с более острыми вершинами (и, соответственно, более тяжелыми хвостами), чем у нормальной плотности, имеют $\varkappa > 3$, а для плотностей с менее острой вершиной $\varkappa < 3$. Следующее утверждение устанавливает, что смеси $E\Phi(x/\sqrt{U})$ всегда являются более островершинными и, следовательно, имеют более тяжелые хвосты, нежели нормальный закон, если в качестве характеристики островершинности рассматривается коэффициент эксцесса.

ЛЕММА 2.3.1. *Пусть X и Y – независимые случайные величины с конечными четвертыми моментами. Предположим, что $EX = 0$ и $P(Y \geq 0) = 1$. Тогда*

$$\varkappa(XY) \geq \varkappa(X).$$

Более того,

$$\kappa(XY) = \kappa(X)$$

тогда и только тогда, когда

$$P(Y = \text{const}) = 1.$$

Доказательство ((Gnedenko and Korolev, 1996), (Bening and Korolev, 2002)). Из независимости X и Y следует, что

$$\begin{aligned} \kappa(XY) &= E\left(\frac{XY - EXY}{\sqrt{DXY}}\right)^4 = \frac{E(XY - EXY)^4}{(E(XY - EXY)^2)^2} = \\ &= \frac{E(XY - EXEY)^4}{(E(XY - EXEY)^2)^2} = \frac{EX^4EY^4}{(EX^2)^2(EY^2)^2} = \kappa(X) \cdot \frac{EY^4}{(EY^2)^2}. \end{aligned} \quad (2.3.1)$$

Но согласно неравенству Иенсена, $EY^4 \geq (EY^2)^2$. Поэтому правая часть соотношения (2.3.1) всегда не меньше, чем $\kappa(X)$. Более того, она равна $\kappa(X)$ тогда и только тогда, когда $EY^4 = (EY^2)^2$, что возможно лишь, только если $P(Y = \text{const}) = 1$. Лемма доказана.

Таким образом, если X – стандартная нормальная случайная величина, а U – неотрицательная случайная величина с $EU^2 < \infty$, независимая от X , то по теореме Фубини $P(X \cdot \sqrt{U} < x) = E\Phi(xU^{-1/2})$. Более того, $\kappa(X \cdot \sqrt{U}) \geq 3$ и $\kappa(X \cdot \sqrt{U}) = 3$, если и только если U вырождена, то есть неслучайна.

К слову сказать, равенство коэффициента эксцесса минимально возможному значению (равному трем) характеризует нормальный закон не только в классе масштабных смесей нормальных законов (см. вышеприведенную лемму), но и в классе всех безгранично делимых законов с конечным четвертым моментом (Круглов, 2003).

Понятие “распределения с тяжелыми хвостами”, естественно, тесно связанное с понятием островершинности, также можно определить по-разному. В этой книге мы не придерживаемся какого-либо одного формального определения распределения с тяжелыми хвостами, понимая под таковым распределение, хвосты которого имеют более низкую скорость убывания по сравнению с нормальным законом при неограниченном росте аргумента. В статье (Birnbau, 1948) предложено вместо “абсолютной” тяжести хвостов распределений вероятности рассматривать “относительную”, сравнивая вероятности больших отклонений. Следуя этому подходу и используя неравенство Иенсена, легко получить неравенства, напрямую связывающие хвосты масштабных смесей нормальных законов с хвостами

самого нормального распределения. Пусть, как и ранее, X – стандартная нормальная случайная величина, а U – неотрицательная случайная величина, независимая от X . Плотность случайной величины $Z = X \cdot \sqrt{U}$ (всегда существующую, симметричную и одновершинную) обозначим $p_Z(x)$. Легко видеть, что $P(Z > x) = 1 - E\Phi(x/\sqrt{U})$ для $x > 0$.

ЛЕММА 2.3.2. Для $x > 0$ справедливо неравенство

$$P(Z > x) \geq 1 - \Phi(\sqrt{2\pi x} p_Z(0)).$$

Если случайная величина U удовлетворяет условию нормировки

$$EU^{-1/2} = 1,$$

то

$$P(Z > x) \geq 1 - \Phi(x), \quad x > 0.$$

Из Леммы 2.3.2 вытекает, что, если $EU^{-1/2} = 1$, то для любого $x \geq 0$

$$P(|X \cdot \sqrt{U}| \geq x) \geq P(|X| \geq x) (= 2[1 - \Phi(x)]),$$

то есть масштабные смеси нормальных законов всегда имеют более тяжелые хвосты, нежели само нормальное распределение.

2.4. Случай элементарных приращений с ненулевыми средними

Иногда есть основания предполагать, что случайные слагаемые X_j в теоремах 2.1.1 и 2.2.1 не центрированы, то есть имеют ненулевое математическое ожидание. Так может быть, к примеру, когда рассматриваются участки стабильного роста или стабильного падения наблюдаемого процесса. Приведенные выше теоремы 2.1.1 и 2.2.1, в которых непременным условием является равенство нулю среднего значения приращений, в такой ситуации представляют собой довольно удобный инструмент исследования тренда исследуемого процесса.

Пусть $M(t), t \geq 0$, – тренд (общая тенденция) развития процесса. С формальной точки зрения под трендом традиционно понимается ожидаемое значение рассматриваемого процесса, то есть неслучайная функция. Однако в некоторых моделях сам тренд может считаться случайным, так как, вообще говоря, тенденцию развития процесса невозможно заранее предсказать во всех деталях. Таким образом, под трендом разумно понимать

условное математическое ожидание рассматриваемого процесса при условии, что известны некоторые дополнительные обстоятельства, под которыми в каждом отдельном случае подразумевается та или иная доступная информация, например, о развитии внешней или внутренней политической ситуации или макроэкономических условий.

Отклонение значений цены акции от $M(t)$ в каждом конкретном контракте обусловлено поведением конкретных участников сделки и потому может считаться “более случайным”, чем $M(t)$. Для простоты в дальнейшем будем считать $M(t)$ неслучайной функцией. Пусть $X_j^* = (\log P_j - \log P_{j-1})$ – приращение логарифма цены акции от $(j - 1)$ -го к j -му контракту и пусть T_j – момент заключения j -го контракта. Положим $\alpha_j^* = M(T_j) - M(T_{j-1})$. Тогда можно считать, что случайные величины $X_j = X_j^* - \alpha_j^*$ имеют нулевые математические ожидания и потому теоремы 2.1.1 и 2.2.1 дают возможность построить аппроксимации для распределений центрированных приращений логарифмов $Z(t) = (\log P(t) - M(t)) - (\log P(0) - M(0))$. Таким образом, получается простой критерий проверки правильности построения тренда: если наблюдаемое распределение случайной величины $Z(t)$ симметрично (и островершинно, см. раздел 2.5) то, скорее всего, тренд построен правильно. Точнее, если указанное распределение не является симметричным и островершинным, то тренд выбран неверно.

Отклонение распределения случайной величины $Z(t)$ от симметричного, в частности, может быть вызвано неправильным выбором (логарифмической) шкалы. Таким образом, теоремы 2.1.1 и 2.2.1 дают возможность проверять не только правильность подгонки тренда, но и правильность выбора шкалы по симметричности и островершинности распределений приращений (логарифмов) рассматриваемого процесса.

Выше везде предполагалось, что случайные величины $\{X_j = \log P_j - \log P_{j-1}\}_{j \geq 1}$ независимы. Однако, как следует из результатов работы (Королев, 1994), это предположение не играет решающей роли. Критичными являются условие независимости считающего процесса $N(\tau)$ от последовательности $X_1, X_2, \dots$ и условие асимптотической нормальности сумм неслучайного числа центрированных и нормированных слагаемых $X_1, X_2, \dots$. При этом нарушение условия независимости считающего процесса от последовательности элементарных приращений опять-таки не является абсолютно критическим для справедливости моделей, основанных на теоремах 2.1.1 и 2.2.1 и имеющих вид смесей нормальных законов. Можно показать, что при нарушении этого условия указанные теоремы останутся справедливыми в части достаточности. Но нарушение условия

асимптотической нормальности, например, замена его условием сходимости к другому закону, скажем, H , приведет к другим моделям, основанным на масштабных смесях распределения H . Например, обобщая подход Б. Мандельброта, при котором вместо сходимости к нормальному закону подразумевается сходимость к не-нормальному устойчивому распределению, в качестве итоговых моделей следовало бы рассматривать масштабные смеси устойчивых законов (см. раздел 1.2). Однако с физической точки зрения условие асимптотической нормальности должно выполняться практически всегда (к такому выводу можно прийти на основе термодинамических рассуждений). Так что можно считать, что условие асимптотической нормальности последовательности сумм неслучайного числа элементарных приращений является скорее правилом, нежели исключением.

В предыдущих разделах рассматривались модели распределений приращений хаотических стохастических процессов, полученные с помощью предельных теорем для специальных случайных сумм в схеме “нарастающих” сумм в предположении, что элементарные приращения рассматриваемого процесса (в частности, изменения процесса биржевой цены от предыдущего контракта к последующему) непосредственно или после центрирования правильно подобранным трендом имеют нулевые математические ожидания. Это предположение вкупе с общими рассуждениями, типичными для центральной предельной теоремы, непосредственно ведет к тому, что искомые модели распределений вероятностей имеют вид масштабных смесей нормальных законов с нулевым средним. Такие распределения островершинны, имеют тяжелые хвосты и симметричны. Однако на практике “правильно” определить тренд, вообще говоря, довольно трудно или невозможно. Поэтому, как правило, эмпирические распределения приходится строить без учета отклонения наблюдений от тренда. Иногда эти эмпирические распределения оказываются не только островершинными, но и асимметричными.

Выше рассматривались модели типа случайных сумм центрированных слагаемых, в которых сами суммы оказываются центрированными случайными величинами (случайными суммами центрирующих констант), то есть изучаемый процесс предполагается центрированным другим случайным процессом), что можно интерпретировать как желание “автоматически” учесть (вообще говоря, ненулевой) тренд. Переходя к рассмотрению моделей типа неслучайно центрированных случайных сумм, то есть к центрированию рассматриваемого процесса неслучайной функцией, с помощью более общих методов асимптотической теории случайного суммирования (см., например, (Королев, 1994), (Gnedenko and Korolev, 1996)) в пред-

положении о том, что элементарные приращения удовлетворяют условиям центральной предельной теоремы (а именно, имеют конечные дисперсии и в определенном смысле однородны или равноправны), можно естественно получить модели распределений приращений хаотических стохастических процессов, имеющие вид сдвиг/масштабных смесей нормальных законов $E\Phi((x - V)/U)$, где случайная величина U неотрицательна, причем случайные величины U и V не обязательно являются независимыми.

Такие смеси со специально связанными U и V также появляются в качестве предельных законов в асимптотической схеме серий, рассматриваемой в следующем разделе.

2.5. Модели в рамках асимптотической схемы серий

В рамках базовой схемы случайного блуждания

$$S(t) = X_1 + \dots + X_{N(t)}, \quad t \geq 0, \quad (2.5.1)$$

рассмотренной в предыдущих разделах, неизменным условием выступало предположение о том, что элементарные приращения (скачки) $X_1, X_2, \dots$ — это независимые и *одинаково распределенные* случайные величины. В асимптотической теории вероятностей схему (2.5.1) называют *схемой нарастающих сумм*. Как уже было показано, в рамках этой схемы с помощью соответствующих предельных теорем (при $t \rightarrow \infty$ или $N(t) \rightarrow \infty$ в каком-либо смысле) можно получить богатое семейство предельных распределений, которые можно использовать в качестве асимптотических аппроксимаций для распределений приращений реальных процессов. Особо подчеркнем, что в рамках схемы (2.5.1) при изменении (неограниченном росте) t или $N(t)$ распределения случайных слагаемых $X_1, X_2, \dots$ остаются неизменными.

В то же время в реальных процессах естественно возникают (иногда довольно продолжительные) разные периоды, на протяжении которых процесс имеет разные систематические смещения (негоризонтальный тренд) и/или разные степени изменчивости (волатильности), что, в частности, проявляется в том, что на разных участках приращения процесса имеют разные математические ожидания и/или дисперсии. В рамках одного периода стабильного изменения (роста или падения) по-прежнему можно предполагать совпадение распределений элементарных приращений процесса, но на протяжении длинных интервалов времени, содержа-

щих несколько подынтервалов различного стабильного изменения процесса, этого делать, вообще говоря, нельзя.

Таким образом, при построении асимптотических аппроксимаций естественно допускать возможность изменения распределений слагаемых $X_1, X_2, \dots$ при изменении (возрастании) t или $N(t)$. Это допущение можно реализовать в рамках так называемой *асимптотической схемы серий*, которая в определенном смысле является более общей, нежели схема нарастающих сумм.

Перейдем к формальному построению модели. Запишем основное соотношение в виде

$$\log P_n(t) - \log P_n(0) = \sum_{j=1}^{N_n} X_{n,j} \quad (2.5.2)$$

где $P_n(t)$ и $P_n(0)$ – значения рассматриваемого процесса биржевой цены, соответственно, в конце n -го интервала длины t и в его начале; $X_{n,1}, X_{n,2}, \dots$ – случайные величины, относительно которых будем предполагать, что при каждом конкретном n они стохастически независимы и одинаково распределены, $X_{n,1}, X_{n,2}, \dots$ имеют смысл значений элементарных приращений логарифма биржевой цены (от контракта к контракту) на n -ом интервале длины t ; $N_n = N_n(t)$ – целочисленные неотрицательные случайные величины, предполагаемые стохастически независимыми от $X_{n,1}, X_{n,2}, \dots$ при каждом $n \geq 1$ и играющие роль количества контрактов, заключенных на протяжении n -го интервала времени. Заметим, что параметр n допускает много интерпретаций, например, его можно считать просто “номером” интервала времени, на протяжении которого процесс имеет стабильное изменение. Можно также считать, что этот параметр не имеет никакого физического смысла и играет вспомогательную роль, позволяя строить более гибкие модели, приводящие к более гибким аппроксимациям для распределений приращений рассматриваемых процессов. Также можно считать, что параметр n зависит от времени t , и изучать распределения приращений процессов на интервалах времени увеличивающейся длины, например, эффект “унормализования” этого распределения (замечено, что чем больше длина интервала, приращения за который рассматриваются, тем ближе распределение (логарифма) приращения к нормальному; обычно нормальная аппроксимация хорошо работает уже при t , равном 3–4 неделям).

Итак, приступим к формализации наших рассуждений. Если k – натуральное число, то символом $S_{n,k}$ мы будем обозначать сумму

$$S_{n,k} = X_{n,1} + \dots + X_{n,k}.$$

В отношении слагаемых $X_{n,j}$ сделаем “хорошие” формальные предположения, заключающиеся в том, что они (i) предельно малы, то есть для любого $\epsilon > 0$

$$P(|X_{n,j}| > \epsilon) \longrightarrow 0$$

при $n \rightarrow \infty$ (так как при каждом n слагаемые имеют одно и то же распределение, то последнее соотношение справедливо при любом j), и (ii) удовлетворяют условиям центральной предельной теоремы в том смысле, что существуют натуральные k_n такие, что $k_n \rightarrow \infty$ при $n \rightarrow \infty$ и

$$P(S_{n,k_n} < x) \implies \Phi\left(\frac{x-a}{\sigma}\right) \quad (n \rightarrow \infty) \quad (2.5.3)$$

при некоторых $a \in \mathbb{R}$ и $0 < \sigma < \infty$ (включая случай $a = 0$ и $\sigma = 1$). Заметим, что условие (2.5.3) выполняется в следующей довольно общей ситуации. В соответствии со сказанным выше, предположим, что случайные величины $X_{n,j}$ имеют конечные дисперсии. Предположим, что величины $X_{n,j}$ могут быть представлены в виде

$$X_{n,j} = X_{n,j}^* + a_n,$$

где $a_n \in \mathbb{R}$ – число, характеризующее систематическое смещение (дрейф) элементарного приращения $X_{n,j}$ на n -ом интервале времени, а $X_{n,j}^*$ – случайная величина с $EX_{n,j}^* = 0$, $DX_{n,j}^* = \sigma_n^2 < \infty$, так что $EX_{n,1} = a_n$ и $DX_{n,1} = \sigma_n^2$. Предположим, что $a_n k_n \rightarrow a$ и $k_n \sigma_n^2 \rightarrow \sigma^2$ при $n \rightarrow \infty$. Тогда вследствие хорошо известного результата о необходимых и достаточных условиях сходимости к нормальному закону распределений сумм независимых случайных величин с конечными дисперсиями в схеме серий (см., например, (Гнеденко и Колмогоров, 1949)), можно заметить, что соотношение (2.5.3) имеет место тогда и только тогда, когда выполнено условие Линдеберга: для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} k_n E(X_{n,1}^*)^2 \mathbb{I}(|X_{n,1}^*| \geq \varepsilon) = 0,$$

(здесь $\mathbb{I}(A)$ – индикаторная функция множества (события) A), то есть квадратичные хвосты распределений элементарных приращений должны убывать достаточно быстро. Кстати, это условие очевидным образом противоречит гипотезе Мандельброта о неограниченности дисперсий элементарных приращений, но, как мы уже видели и еще убедимся в дальнейшем, также приводит к асимметричным предельным распределениям с тяжелыми хвостами.

Пусть $L_1(X, Y)$ – метрика в пространстве случайных величин, метризующая слабую сходимость. Примерами метрики L_1 могут быть сглаженное равномерное расстояние

$$L_1(X, Y) = \sup_{x \in \mathbb{R}} |(F_X * H)(x) - (F_Y * H)(x)|,$$

где $F_X(x) = P(X < x)$, $F_Y(x) = P(Y < x)$, H – произвольная абсолютно непрерывная функция распределения с ограниченной плотностью, $*$ – символ свертки, а также метрика Леви

$$L_1(X, Y) = \inf\{h > 0 : F_X(x-h) - h \leq F_Y(x) \leq F_X(x+h) + h \text{ для всех } x \in \mathbb{R}\}$$

(равная длине наибольшего квадрата, который можно вписать между графиками функций распределения $F_X(x) = P(X < x)$ и $F_Y(x) = P(Y < x)$). Теперь можно сформулировать следующую основную теорему.

ТЕОРЕМА 2.5.1. *Пусть случайные величины $\{X_{n,j}\}$ и N_n удовлетворяют сформулированным выше условиям. Предположим, что существует последовательность натуральных чисел $k_1, k_2, \dots$ такая, что выполнено условие (2.5.2) с некоторыми $a \in \mathbb{R}$ и $\sigma < \infty$. Тогда слабая сходимость распределений приращений логарифмов биржевых цен к некоторой случайной величине Z*

$$\log P_n(t) - \log P_n(0) - c_n \implies Z \quad (n \rightarrow \infty) \quad (2.5.4)$$

с некоторыми $c_n \in \mathbb{R}$ имеет место в том и только в том случае, когда существуют слабо компактная последовательность случайных величин $U_1, U_2, \dots$ и ограниченная числовая последовательность $v_1, v_2, \dots$ такие, что

$$E \exp\{isZ\} = E \exp\left\{is(U_n + v_n) - \frac{1}{2}s^2\sigma^2 U_n\right\} \quad (2.5.5)$$

для всех $n \geq 1$ и $s \in \mathbb{R}$, а также

$$L_1(N_n/k_n, U_n) \rightarrow 0 \quad (n \rightarrow \infty), \quad (2.5.6)$$

$$|v_n + c_n| \rightarrow 0 \quad (n \rightarrow \infty).$$

Это утверждение является непосредственным следствием теоремы 3 из (Korolev, 1995b) (см. также теорему 4.2.1 в книге (Gnedenko and Korolev, 1996)).

Предельная характеристическая функция (2.5.5) соответствует функции распределения

$$P(Z < x) = E\Phi\left(\frac{x - b - aU}{\sigma\sqrt{U}}\right), \quad (2.5.7)$$

где $b \in \mathbb{R}$, а случайная величина U неотрицательна. К сожалению, семейство сдвиг/масштабных смесей нормальных законов типа (2.5.7) не является идентифицируемым (см. (Teicher, 1961), подробности приведены в разделе 3.2). Другими словами, соотношению (2.5.7) могут удовлетворять разные случайные величины U с различными распределениями. Именно по этой причине в формулировке теоремы 2.5.1 присутствует условие (2.5.6), описывающее “слабое сближение” (то есть сближение в смысле расстояния, метризирующего слабую сходимость) последовательностей $\{N_n/k_n\}$ и $\{U_n\}$, а не более сильное и допускающее более наглядную интерпретацию условие $N_n/k_n \Rightarrow U$, которое является достаточным для (2.5.5) и, следовательно, для (2.5.3), но, к сожалению, не является необходимым.

Как уже отмечалось, сравнительно недавно О. Барндорфф-Нильсен и его коллеги установили (Barndorff-Nielsen, 1994), что наблюдается очень высокое согласие распределений приращений логарифмов цен акций на датских и немецких биржах с так называемым нормальным\\обратным гауссовым распределением (NIG-распределением; NIG=Normal\\Inverse Gaussian) (см. пример 2.2.3). Этой же теме посвящены публикации (Eberlein and Keller, 1995), (Prause, 1997), (Eberlein, Keller and Prause, 1998) и (Eberlein, 1999), связанные с анализом эволюции биржевых цен, основанном на аналогичных обобщенных гиперболических распределениях. При взгляде на весьма специфическое аналитическое выражение для плотности такого распределения (см. выражение (2.2.7)) у неподготовленного читателя может возникнуть недоумение, почему такое распределение обеспечивает столь высокое согласие с реальными данными, поскольку выражение (2.2.7) громоздко и трудно интерпретируемо. Однако с помощью теоремы 2.5.1 это недоумение весьма легко рассеять. Действительно, еще в 1977 было установлено, что NIG-распределение, равно как и другие обобщенные гиперболические законы, допускает представление в виде специальных сдвиг/масштабных смесей нормальных законов (Barndorff-Nielsen, 1977), также см. (Barndorff-Nielsen, 1978) и (Barndorff-Nielsen, Jensen and Sorensen 1982). А именно, NIG-распределение удовлетворяет соотношению (2.5.7), в котором случайная величина U имеет обратное нормальное распределение (2.2.8) (в примере 2.2.3 рассмотрен специальный случай $a = b = 0$). Именно в силу этого обстоятельства NIG-распределение получило свое название.

Обобщенным гиперболическим распределениям и обобщенным гиперболическим процессам Леви еще будет уделено подробное внимание в разделе 2.6.3 (примеры 2.6.2 – 2.6.4). Некоторые свойства моделей финансовой математики, основанных на NIG-распределениях, систематически описаны

в (Ширяев, 1995) и (Ширяев, 1998).

В связи с упоминавшимися работами О. Барндорфф-Нильсена и его коллег возникает вполне естественный вопрос о том, почему смешивающее распределение (2.2.8) обеспечивает столь высокое согласие с реальными данными. Первый и самый простой ответ заключается в том, что параметрическое семейство смешивающих обратных нормальных законов (2.2.8) достаточно богато, так что в нем всегда найдется представитель, удовлетворительно близкий к истинному смешивающему распределению. Другой ответ, возможно, удастся получить, если будет установлена аналитическая взаимосвязь поведения случайных величин N_n в теореме 2.5.1, основанной на модели (2.5.2), с рассуждениями О. Барндорфф-Нильсена, с помощью которых он при изучении распределений логарифмов частиц при дроблении пришел к нормальному\обратному нормальному (нормальному\обратному гауссовскому) закону (Barndorff-Nielsen, 1977) или с хорошо известным свойством обратного нормального закона как распределения времени достижения винеровским процессом некоторого уровня (детальное описание свойств обратного нормального закона можно найти в статье (Jørgensen, 1982) и книгах (Chhikara and Folks, 1989) и (Seshadri, 1999)).

Роль обобщенных гиперболических распределений в описании негауссовой диффузии, связанной с турбулентностью, описана в статьях (Barndorff-Nielsen, 1979), (Barndorff-Nielsen, P. Blæsild and Schmiegel, 2004). В разделе 7.1 эта проблема будет подвергнута более подробному обсуждению.

2.6. От одномерных смешанных нормальных распределений к процессам Леви со случайным временем

В данном разделе будет сделана попытка дать математическое обоснование целесообразности использования моделей типа так называемого геометрического броуновского движения со случайными сносом и диффузией в качестве макромоделей динамики финансовых индексов и, в первую очередь, биржевых цен. Подобные модели описывают броуновское движение в случайной среде, скажем, в жидкости со случайной температурой и/или вязкостью.

Вместе с тем, как уже можно было убедиться, на микроуровне (то есть

на очень малых интервалах времени) довольно адекватными моделями являются обобщенные процессы Кокса, естественным образом учитывающие неоднородность “биржевого времени”. Поскольку даже на очень малых интервалах времени число изменений биржевого курса велико, естественно возникают предпосылки к использованию асимптотических (предельных) аппроксимаций.

Для полноты описания предельного поведения обобщенных процессов Кокса именно как *случайных процессов* с учетом свойств их *траекторий* необходимо рассмотреть асимптотическое поведение обобщенных процессов Кокса как *случайных функций* в некоторых функциональных пространствах. В данном разделе будут рассмотрены функциональные предельные теоремы для обобщенных процессов Кокса.

2.6.1. Предварительные сведения. Процессы Леви. Пространство Скорохода

Не останавливаясь подробно на соответствующих определениях и предварительных результатах о функциональных предельных теоремах, отошлем заинтересованного читателя к книге (Биллингсли, 1977), где все эти сведения содержатся. Мы будем интенсивно использовать понятие процесса Леви. Под процессом Леви мы будем понимать случайный процесс $X(t)$, $t \geq 0$, со следующими свойствами:

- 1° $X(0) = 0$ с вероятностью единица.
- 2° $X(t)$ – процесс с независимыми приращениями, то есть для любых $N \geq 1$ и $t_0, t_1, \dots, t_N$ ($0 \leq t_0 \leq t_1 \leq \dots \leq t_N$) случайные величины $X(t_0), X(t_1) - X(t_0), \dots, X(t_N) - X(t_{N-1})$ независимы в совокупности.
- 3° $X(t)$ – однородный процесс, то есть для любых $s, t, h > 0$ случайные величины $X(t+h) - X(t)$ и $X(s+h) - X(s)$ имеют одинаковое распределение.
- 4° Процесс $X(t)$ стохастически непрерывен, то есть для любых $t \geq 0$ и $\epsilon > 0$

$$\lim_{s \rightarrow t} P(|X(t) - X(s)| > \epsilon) = 0.$$

- 5° Траектории процесса $X(t)$ непрерывны справа и имеют конечные пределы слева.

Пусть $X = X(t)$, $t \geq 0$, – стохастически непрерывный случайный процесс с независимыми однородными (по времени) приращениями, причем $P(X(0) = 0) = 1$. Тогда для фиксированного $t > 0$ и произвольного $n \geq 1$ имеем

$$\begin{aligned} X(t) &= X(0) + (X(t/n) - X(0)) + \dots + (X(t) - X(t - t/n)) \stackrel{d}{=} \\ &\stackrel{d}{=} X_{n,1} + X_{n,2} + \dots + X_{n,n}, \end{aligned}$$

где $\{X_{n,i}\}_{i=1}^n$ – независимые одинаково распределенные случайные величины, причем $X_{n,1} \stackrel{d}{=} X(t/n)$. Следовательно, *при любом* $t > 0$ случайная величина $X(t)$ имеет *безгранично делимое* распределение вероятностей.

Обозначим характеристическую функцию случайной величины $X(t)$ символом $\psi_t(s)$ ($\psi_t(s) = E e^{isX(t)}$, $s \in \mathbb{R}$). Тогда справедливо соотношение

$$\psi_t(s) = E e^{is(X_{n,1} + \dots + X_{n,n})} = [\psi_{t/n}(s)]^n \neq 0, \quad s \in \mathbb{R}.$$

Следовательно, $\psi_{t/n}(s) = [\psi_t(s)]^{1/n}$ – характеристическая функция случайной величины $X(t/n)$, и $\psi_{m/n}(s) = [\psi_1(s)]^{m/n}$.

Поскольку по предположению процесс $X = X(t)$, $t \geq 0$, стохастически непрерывен, для любой последовательности $\Delta_n \rightarrow 0$ выполнено соотношение $\psi_{\Delta_n}(s) \rightarrow 1$. Используя это соотношение с $\Delta_n \equiv t - \frac{[tn]}{n} \rightarrow 0$ ($n \rightarrow \infty$), для произвольного $s \in \mathbb{R}$ получим

$$\begin{aligned} \psi_t(s) &= \psi_{[tn]/n}(s) \cdot \psi_{t-[tn]/n}(s) = \\ &= [\psi_1(s)]^{[tn]/n} \cdot \psi_{t-[tn]/n}(s) \rightarrow [\psi_1(s)]^t \quad (n \rightarrow \infty). \end{aligned}$$

Функция $\psi_t(s)$ от n не зависит, следовательно, при любом $t > 0$

$$\psi_t(s) = [\psi_1(s)]^t, \quad s \in \mathbb{R}.$$

Таким образом, для любого однородного стохастически непрерывного процесса с независимыми приращениями (а значит, для любого процесса Леви) $X(t)$, $t \geq 0$, траектории которого стартуют из нуля, при любом $t > 0$ характеристическая функция случайной величины $X(t)$ безгранично делима и имеет вид

$$\psi_t(s) = [\psi_1(s)]^t = [E e^{isX(1)}]^t, \quad s \in \mathbb{R}.$$

Обратно, пусть Y – произвольная безгранично делимая случайная величина, тогда семейство безгранично делимых распределений с характеристическими функциями вида $[E e^{isY}]^t$ полностью определяет конечномерные

распределения однородного стохастически непрерывного процесса с независимыми приращениями $X(t)$, $t \geq 0$, траектории которого стартуют из нуля, причем $X(1) \stackrel{d}{=} Y$.

В силу указанного выше свойства характеристической функции $\psi_t(s)$ процесса Леви $X(t)$, $t \geq 0$, достаточно рассматривать ее в какой-либо фиксированной точке t , например, в точке $t = 1$. Обозначим $\psi(s) = \psi_1(s)$.

Поскольку характеристическая функция $\psi(s)$ безгранично делима, для нее справедливо *представление Леви–Хинчина*

$$\log \psi(s) = i\gamma s - \frac{\sigma^2 s^2}{2} + \int_{\mathbb{R} \setminus \{0\}} [e^{isx} - 1 - isx \mathbb{I}_{\{|x| < 1\}}(x)] L(dx),$$

где $\gamma \in \mathbb{R}$, $\sigma^2 \geq 0$, L – мера на $\mathbb{R} \setminus \{0\}$ такая, что

$$\int_{\mathbb{R} \setminus \{0\}} \min\{1, x^2\} L(dx) < \infty.$$

В представлении Леви–Хинчина меру L принято называть *спектральной мерой* или *мерой Леви*, параметры γ и σ^2 – параметрами сноса и диффузии, соответственно. Если спектральная мера Леви обладает плотностью относительно меры Лебега, то есть $L(dx) = \ell(x)dx$, то функция $\ell(x)$ называется *спектральной плотностью* (*плотностью Леви*) процесса $X(t)$. Физический смысл параметров γ , σ^2 и $\ell(x)$ раскрывается теоремой Леви–Ито о разложении процесса Леви, согласно которой процесс Леви $X(t)$, характеризуемый параметрами γ , σ^2 и L , допускает разложение в сумму трех независимых случайных процессов вида

$$X(t) = X^{(1)}(t) + X^{(2)}(t) + X^{(3)}(t), \quad t \geq 0,$$

где $X^{(1)}(t) = \gamma t + \sigma W_{0,1}(t)$, причем $W_{0,1}(t)$ – броуновское движение (стандартный винеровский процесс); характеристическая функция случайной величины $X^{(1)}(1)$ имеет вид

$$\log \psi^{(1)}(s) = i\gamma s - \frac{\sigma^2 s^2}{2}, \quad s \in \mathbb{R}.$$

Процесс $X^{(2)}(t)$ является обобщенным пуассоновским:

$$X^{(2)}(t) = \sum_{j=1}^{N(t)} Y_j,$$

где $N(t)$ – пуассоновский процесс с интенсивностью

$$\lambda = \int_{|x|>1} L(dx),$$

а $Y_1, Y_2, \dots$ – независимые случайные величины с одинаковым распределением $P(Y_1 \in [x, x + dx]) = \mathbb{I}_{\{|x|>1\}} L(dx)$. При этом процесс $N(t)$ и последовательность $Y_1, Y_2, \dots$ независимы. Другими словами, при $|x| > 1$ скачкообразный процесс $[X(t) - X(t-0)]\mathbb{I}_{\{X(t)-X(t-0) \in [x, x+dx]\}}$ является пуассоновским с интенсивностью $\ell(x)dx$, так что

$$X^{(2)}(t) = \sum_{\tau \leq t} [X(\tau) - X(\tau-0)]\mathbb{I}_{\{|X(\tau)-X(\tau-0)|>1\}}.$$

При этом характеристическая функция случайной величины $X^{(2)}(1)$ имеет вид

$$\log \psi^{(2)}(s) = \int (e^{isx} - 1)\mathbb{I}_{\{|x|>1\}}(x)L(dx), \quad s \in \mathbb{R}.$$

Процесс $X^{(3)}(t)$ – это процесс Леви, соответствующий характеристической функции

$$\log \psi^{(3)}(s) = \int_{\mathbb{R} \setminus \{0\}} (e^{isx} - 1 - isx)\mathbb{I}_{\{|x| \leq 1\}}(x)L(dx), \quad s \in \mathbb{R}.$$

Эта характеристическая функция является пределом характеристических функций суммы обобщенных пуассоновских процессов с малыми скачками, центрированных (компенсированных) своими математическими ожиданиями. В связи с этим выражение $isx\mathbb{I}_{\{|x|<1\}}$ в разложении Леви–Хинчина называют *компенсатором*. В силу указанного центрирования процесс $X^{(3)}(t)$ является мартингалом.

Из разложения Леви–Ито вытекает, что каждый процесс Леви является комбинацией процесса броуновского движения и суммы (возможно, бесконечного числа) обобщенных пуассоновских процессов. Это означает, что каждый процесс Леви может быть как угодно точно приближен скачкообразно-диффузионным процессом.

Среди чисто скачкообразных процессов Леви (то есть процессов Леви с $\sigma^2 = 0$ в представлении Леви–Хинчина) различают процессы с конечной

и бесконечной активностью. Говорят, что скачкообразный процесс Леви проявляет *конечную активность*, если

$$\lambda \equiv \int_{\mathbb{R} \setminus \{0\}} L(dx) < \infty.$$

Для процессов Леви с конечной активностью вероятность бесконечного числа скачков на любом интервале времени конечной длины равна нулю. Классическим примером процесса Леви с конечной активностью является скачкообразно-диффузионный (jump-diffusion) процесс Мертона (Merton, 1976), определяемый плотностью Леви

$$\ell(x) = \frac{\lambda}{\sqrt{2\pi}\delta} \exp \left\{ -\frac{(x-a)^2}{2\delta^2} \right\}, \quad x \in \mathbb{R}.$$

Процесс Мертона – это обычный обобщенный пуассоновский процесс с интенсивностью λ потока скачков и нормальным распределением величины скачков с параметрами a и δ^2 . Еще одним примером процесса Леви с конечной активностью является скачкообразно-диффузионный процесс с двойным экспоненциальным распределением величины скачков, предложенный в работах (Коу, 2002), (Kou and Wang, 2004). Этот процесс определяется спектральной плотностью Леви вида

$$\ell(x) = \frac{\lambda}{2\delta} \exp \left\{ -\frac{|x-a|}{\delta} \right\}, \quad x \in \mathbb{R}.$$

Различным вопросам применения скачкообразно-диффузионных процессов при решении задач финансовой и страховой математики посвящены работы (Cont and Tankov, 2004) и (Jang, 2007).

Процессы Леви с *бесконечной активностью* характеризуются условием

$$\int_{\mathbb{R} \setminus \{0\}} L(dx) = \infty.$$

Такие процессы на любом конечном интервале времени могут иметь бесконечно много скачков. Примерами процессов Леви с бесконечной активностью являются нормальный\обратный гауссовский (NIG) процесс Леви (Barndorff-Nielsen, 1998), обобщенный гиперболический процесс Леви (Eberlein, Keller and Prause, 1998), дисперсионный гамма-процесс Леви (Carr, Madan and Chang, 1998). Эти процессы будут подробно рассмотрены ниже в разделе 2.6.3.

Пусть $X = X(t)$, $t \geq 0$, – квадратично интегрируемый процесс Леви с характеристической функцией $\psi_t(s) = \mathbb{E} e^{isX(t)}$. Тогда

$$\begin{aligned}\psi_t(s) &= [\psi_1(s)]^t, & \frac{\partial \psi_t(s)}{\partial s} &= t[\psi_1(s)]^{t-1} \frac{\partial \psi_1(s)}{\partial s}, \\ \frac{\partial^2 \psi_t(s)}{\partial s^2} &= t \left[(t-1)[\psi_1(s)]^{t-2} \left(\frac{\partial \psi_1(s)}{\partial s} \right)^2 + [\psi_1(s)]^{t-1} \frac{\partial^2 \psi_1(s)}{\partial s^2} \right].\end{aligned}$$

Следовательно, используя хорошо известную связь между моментами случайных величин и производными соответствующих характеристических функций, можно получить соотношения

$$\mathbb{E}X(t) = t\mathbb{E}X(1), \quad \mathbb{E}X^2(t) = t[\mathbb{D}X(1) + t(\mathbb{E}X(1))^2],$$

откуда

$$\mathbb{D}X(t) = t\mathbb{D}X(1), \quad t \geq 0.$$

Свойства процессов Леви подробно описаны в книгах (Bertoin, 1996), (Sato, 1999). Книги (Barndorff-Nielsen, Mikosch and Resnick, 2001), (Schoutens, 2003) и обзорная статья (Geman, 2002) посвящены использованию процессов Леви в задачах моделирования динамики финансовых индексов.

Пусть $D = D[0, 1]$ – пространство вещественных функций, определенных на $[0, 1]$, непрерывных справа и имеющих конечные левосторонние пределы:

- (i) при $0 \leq t < 1$ предел $x(t+) = \lim_{s \downarrow t} x(s)$ существует и $x(t+) = x(t)$;
- (ii) при $0 < t \leq 1$ предел $x(t-) = \lim_{s \uparrow t} x(s)$ существует.

Пусть Δ – класс строго возрастающих непрерывных отображений отрезка $[0, 1]$ на себя. Пусть δ – неубывающая функция на $[0, 1]$, $\delta(0) = 0$, $\delta(1) = 1$. Положим

$$\|\delta\| = \sup_{s \neq t} \left| \log \frac{\delta(t) - \delta(s)}{t - s} \right|.$$

Если $\|\delta\| < \infty$, то функция δ непрерывна и строго возрастает и, следовательно, принадлежит классу Δ .

Определим расстояние $d_0(x, y)$ в множестве $D[0, 1]$ как нижнюю грань таких положительных чисел ϵ , для которых Δ содержит некоторую функцию δ такую, что

$$\|\delta\| \leq \epsilon$$

и

$$\sup_t |x(t) - y(\delta(t))| \leq \epsilon.$$

Можно показать, что пространство $D[0, 1]$ полно относительно метрики d_0 . Метрическое пространство $(D[0, 1], d_0)$ принято называть *пространством Скорохода*.

Задавшись произвольным $T > 0$ и проводя рассуждения, аналогичные представленным выше, можно определить метрическое пространство $(D[0, T], d_0)$, $T > 0$, также называемое *пространством Скорохода*.

Всюду в дальнейшем случайные процессы будут рассматриваться как случайные элементы со значениями в пространстве $\mathcal{D} = (D[0, T], d_0)$.

Поскольку несмотря на наличие “тяжелых хвостов” в большинстве приложений нет никаких оснований для того, чтобы отвергнуть предположение о конечности дисперсий элементарных скачков обобщенных процессов Кокса, здесь будет рассматриваться только случай конечных дисперсий. Материал этого раздела основан на результатах работы (Кащеев, 2001).

2.6.2. Функциональные предельные теоремы для обобщенных процессов Кокса

Для построения более гибких аппроксимаций, в отличие от предыдущих разделов, где в нашем распоряжении была *одна* последовательность $X_1, X_2, \dots$ и *один* процесс $N(t)$, здесь будет рассматриваться предельная схема серий, в которой присутствует дополнительный вспомогательный параметр n , вообще говоря, не имеющий никакого физического смысла.

Рассмотрим последовательность обобщенных процессов Кокса вида

$$Z_n(t) = \sum_{i=1}^{N_{n,1}(\Lambda_n(t))} X_{n,i}, \quad t \geq 0, \quad (2.6.1)$$

где при каждом $n = 1, 2, \dots$ $N_{n,1}(t)$, $t \geq 0$, – пуассоновский процесс с единичной интенсивностью; при каждом $n = 1, 2, \dots$ случайные величины $X_{n,1}, X_{n,2}, \dots$ одинаково распределены, причем для любого $t \geq 0$ величины $N_{n,1}(t)$, $X_{n,1}, X_{n,2}, \dots$ независимы; при каждом $n = 1, 2, \dots$ $\Lambda_n(t)$, $t \geq 0$, – это процесс Леви, независимый от процесса

$$X_n(t) = \sum_{i=1}^{N_{n,1}(t)} X_{n,i}, \quad t \geq 0, \quad (2.6.2)$$

и такой, что $\Lambda_n(0) = 0$, $P(\Lambda_n(t) < \infty) = 1$ для любого $t \geq 0$, почти все траектории процесса $\Lambda_n(t)$ не убывают и непрерывны справа, $E\Lambda_n^2(1) < \infty$.

Предположим, что $E X_{n,1}^2 < \infty$, $n = 1, 2, \dots$, и обозначим

$$a_n = E X_{n,1} \quad \sigma_n^2 = D X_{n,1}.$$

Из (2.6.1) и (2.6.2) легко видеть, что $Z_n(t) = X_n(\Lambda_n(t))$. Поскольку при каждом $n \geq 1$ $X_n(t)$ и $\Lambda_n(t)$ – независимые процессы Леви, то суперпозиция $Z_n(t) = X_n(\Lambda_n(t))$ также является процессом Леви.

Чтобы установить слабую сходимость случайных процессов $Z_n(t)$ в пространстве Скорохода $\mathcal{D} = (D[0, T], d_0)$, $T > 0$, сначала необходимо найти предельное распределение случайных величин $Z_n(t)$ при каждом $t > 0$.

Пусть $t = 1$. Обозначим $N_n = N_{n,1}(\Lambda_n(1))$. Предположим, что для некоторой неограниченно возрастающей последовательности $\{k_n\}_{n \geq 1}$ натуральных чисел при $n \rightarrow \infty$ выполнены условия

$$1^\circ \quad X_{n,1} + \dots + X_{n,k_n} \Longrightarrow Y;$$

$$2^\circ \quad \frac{N_n}{k_n} \Longrightarrow U,$$

где Y – безгранично делимая, а U – неотрицательная случайные величины. Тогда согласно теореме переноса Гнеденко–Фахима (Гнеденко и Фахим, 1969) имеет место сходимость

$$X_{n,1} + \dots + X_{n,N_n} \Longrightarrow Z \quad (n \rightarrow \infty), \quad (2.6.3)$$

где Z – случайная величина с характеристической функцией

$$E \exp\{isZ\} = \int_0^\infty h^u(s) dA(u), \quad s \in \mathbb{R},$$

а $h(s) = E \exp\{isY\}$, $A(u) = P(U < u)$.

Если при этом случайная величина U безгранично делима, то $Z \stackrel{d}{=} Y(U)$, где $Y(t)$ – это процесс Леви такой, что $Y(1) \stackrel{d}{=} Y$.

Обратим внимание, что условие 2° эквивалентно условию

$$\frac{\Lambda_n(1)}{k_n} \Longrightarrow U \quad (n \rightarrow \infty).$$

Заметим, что предельная случайная величина Z не обязана иметь симметричное распределение.

Поскольку случайная величина Z является безгранично делимой, может быть определен процесс Леви $Z(t), t \geq 0$, такой, что $Z(1) \stackrel{d}{=} Z$.

Далее, так как согласно (2.6.3)

$$Z_n(1) = \sum_{i=1}^{N_n} X_{n,i} \Longrightarrow Z(1),$$

а процессы $Z_n(t)$ и $Z(t)$ являются процессами Леви, то, используя свойство характеристических функций процессов Леви:

$$\mathbb{E} \exp\{isX(t)\} = [\mathbb{E} \exp\{isX(1)\}]^t, \quad s \in \mathbb{R}, \quad t \geq 0,$$

приходим к заключению о том, что для любого $t > 0$

$$Z_n(t) = \sum_{i=1}^{N_{n,1}(\Lambda_n(t))} X_{n,i} \Longrightarrow Z(t) \quad (n \rightarrow \infty).$$

Поскольку случайные процессы $Z_n(t)$, $0 \leq t \leq T$, и $Z(t)$, $0 \leq t \leq T$, являются процессами Леви, почти все их траектории принадлежат пространству Скорохода $\mathcal{D}$.

В дальнейшем основное внимание будет уделено изучению пространства Скорохода $(D[0, 1], d_0)$, то есть на случае $T = 1$. Общий случай может быть рассмотрен по аналогии.

Условия слабой сходимости обобщенного процесса Кокса $Z_n(t)$ к суперпозиции $Z(t) = Y(U(t))$ в пространстве $\mathcal{D}$ при $n \rightarrow \infty$ содержатся в следующем результате.

ТЕОРЕМА 2.6.1. Пусть справедливо хотя бы одно из двух условий: либо величины $\mathbb{E}|X_{n,1}| \cdot \mathbb{E}\Lambda_n(1)$ ограничены в совокупности, либо величины $\mathbb{E}(Z_n(1))^2 = (a_n^2 + \sigma_n^2) \mathbb{E}\Lambda_n(1) + a_n^2 \mathbb{E}\Lambda_n^2(1)$ ограничены в совокупности. Предположим, что для некоторой неограниченно возрастающей последовательности натуральных чисел $\{k_n\}_{n \geq 1}$ при $n \rightarrow \infty$ справедливы соотношения

$$(i) \quad X_{n,1} + \dots + X_{n,k_n} \Longrightarrow Y;$$

$$(ii) \quad k_n^{-1} \Lambda_n(1) \Longrightarrow U \geq 0,$$

где Y и U — не являющиеся вырожденными в нуле безгранично делимые случайные величины такие, что $\mathbb{E}Y^2 < \infty$ и $\mathbb{E}U^2 < \infty$. Тогда обобщенные

процессы Кокса $Z_n(t)$ слабо сходятся в пространстве $\mathcal{D} = (D[0, T], d_0)$ к процессу Леви $Z(t)$ такому, что распределение случайной величины $Z(1)$ определяется характеристической функцией

$$\mathbb{E} e^{isZ(1)} = \mathbb{E} [\mathbb{E} e^{isY}]^U, \quad s \in \mathbb{R}.$$

В качестве следствий теоремы 2.6.1 приведем функциональную центральную предельную теорему и функциональный закон больших чисел для нецентрированных обобщенных процессов Кокса. Следующая теорема представляет собой вариант функциональной центральной предельной теоремы.

ТЕОРЕМА 2.6.2. Пусть для некоторой неограниченно возрастающей последовательности натуральных чисел $\{k_n\}_{n \geq 1}$ выполнены условия:

(а) существуют числа $a \in \mathbb{R}$ и $\sigma \in (0, \infty)$ такие, что

$$k_n \mathbb{E} X_{n,1} \longrightarrow a, \quad k_n \mathbb{D} X_{n,1} \longrightarrow \sigma^2 \quad (n \rightarrow \infty);$$

(б) (условие Линдеберга) для любого $\epsilon > 0$

$$\lim_{n \rightarrow \infty} k_n \mathbb{E} [(X_{n,1} - a_n)^2 \mathbb{I}(|X_{n,1} - a_n| \geq \epsilon)] = 0,$$

где $\mathbb{I}(A)$ – индикатор события A ;

(с) $k_n^{-1} \Lambda_n(1) \Longrightarrow U \quad (n \rightarrow \infty)$,

где U – некоторая безгранично делимая случайная величина такая, что $\mathbb{P}(U = 0) < 1$, $\mathbb{P}(U \geq 0) = 1$, $\mathbb{E} U^2 < \infty$;

(д) $\sup_n k_n^{-2} \mathbb{E} (\Lambda_n(1))^2 < \infty$.

Тогда обобщенные процессы Кокса $Z_n(t) = X_n(\Lambda_n(t))$ слабо сходятся в $\mathcal{D}$ к процессу Леви $X(t)$ такому, что

$$X(1) \stackrel{d}{=} \sigma \sqrt{U} \cdot N(0, 1) + aU,$$

где $N(0, 1)$ – случайная величина со стандартным нормальным распределением, независимая от U .

Теперь сформулируем вариант функционального закона больших чисел для обобщенных процессов Кокса.

ТЕОРЕМА 2.6.3. Пусть для некоторой неограниченно возрастающей последовательности натуральных чисел $\{k_n\}_{n \geq 1}$ выполнены следующие условия:

(е) существует действительное число $b \neq 0$ такое, что

$$k_n \mathbf{E} X_{n,1} \longrightarrow b, \quad k_n \mathbf{D} X_{n,1} \longrightarrow 0 \quad (n \rightarrow \infty);$$

(f) существует безгранично делимая случайная величина V такая, что $\mathbf{P}(V = 0) < 1$, $\mathbf{P}(V \geq 0) = 1$, $\mathbf{E} V^2 < \infty$ и

$$k_n^{-1} \Lambda_n(1) \Longrightarrow V \quad (n \rightarrow \infty);$$

(g) справедливо хотя бы одно из двух предположений:

(g1) величины $\mathbf{E}|X_{n,1}| \cdot \mathbf{E} \Lambda_n(1)$ ограничены в совокупности;

(g2) величины $k_n^{-2} \mathbf{E} \Lambda_n^2(1)$ ограничены в совокупности.

Тогда обобщенные процессы Кокса $Z_n(t) = X_n(\Lambda_n(t))$ слабо сходятся в $\mathcal{D}$ к процессу Леви $X(t)$ такому, что $X(1) \stackrel{d}{=} bV$.

Заметим, что условия (с) и (d) теоремы 2.6.2 и условия (f) и (g) (точнее, (g2)) теоремы 2.6.3 очевидным образом выполнены для процессов Кокса–Леви с управляющими процессами вида $\Lambda_n(t) = k_n \Lambda(t)$, где $\Lambda(t)$ – процесс Леви, а k_n – надлежащим образом подобранные числа. Здесь управляющий процесс $\Lambda(t)$, в дополнение к обычным требованиям

(i) $\Lambda(0) = 0$;

(ii) $\mathbf{P}(\Lambda(t) < \infty) = 1$ для любого $t > 0$;

(iii) почти все траектории процесса $\Lambda(t)$ не убывают и непрерывны справа,

удовлетворяет условию

(iv) $\mathbf{E} \Lambda^2(1) < \infty$.

Доказательства теорем 2.6.1 – 2.6.3 приведены в работе (Кащеев, 2001); также см. главу 10 в (Bening and Korolev, 2002).

Функциональные предельные теоремы данного раздела представляют собой теоретический мостик, который соединяет модели эволюции биржевых цен на микроуровне (то есть при малых временных интервалах), описанные в предыдущем разделе, и популярные в настоящее время макромодели типа геометрического броуновского движения со случайными сносом и диффузией, и тем самым дают дополнительное обоснование моделям распределений биржевых цен вида сдвиг-масштабных смесей нормальных законов.

2.6.3. Примеры предельных процессов

Каждая попытка научно обосновать ответы, возникающие в ситуации сложной неопределенности, является... тоталитарной пародией на считающийся разумным процесс познания.

А. Ф. М. Смит ((Smith, 1984), русский перевод см. (Бернштейн, 2006), с. 151)

В этом разделе будут приведены примеры некоторых процессов, появляющихся в качестве пределов в функциональных теоремах предыдущего параграфа. Будет показано, что практически все наиболее популярные модели динамики финансовых индексов (обобщенные гиперболические процессы Леви, нормальные\обратные гауссовские процессы Леви, дисперсионные гамма-процессы (VG-процессы), двусторонние гамма-процессы Леви, процессы Мейкснера) являются подчиненными винеровскими процессами и, стало быть, имеют распределения типа сдвиг-масштабных смесей нормальных законов.

Броуновское движение

ПРИМЕР 2.6.1. Прежде всего отметим, что если в теореме 2.6.2 случайная величина U является вырожденной в некоторой точке $\gamma > 0$, то в качестве предельного процесса будет выступать процесс Леви $X = X(t)$ такой, что

$$X(1) \stackrel{d}{=} \sigma' N(0, 1) + a',$$

где $\sigma' > 0$, $a' \in \mathbb{R}$. Здесь и далее обозначение $N(a, b)$ используется для случайной величины, имеющей нормальное распределение со средним a и дисперсией b . Таким образом, очевидно, что процесс $S(t) = S(0) \exp \{X(t)\}$ есть обычное геометрическое броуновское движение.

Обобщенные гиперболические процессы Леви

ПРИМЕР 2.6.2. Для начала напомним, что функция Бесселя третьего рода порядка λ (иногда говорят “с индексом λ ”) может быть записана в интегральной форме как

$$K_\lambda(z) = \frac{1}{2} \int_0^\infty y^{\lambda-1} \exp \left\{ -\frac{z}{2} \left(y + \frac{1}{y} \right) \right\} dy.$$

Из этого представления вытекает, что для любых $\gamma > 0$ и $\delta > 0$ функция

$$p_{GIG}(x; \lambda, \delta, \gamma) = \left(\frac{\gamma}{\delta} \right)^\lambda \frac{1}{2K_\lambda(\delta\gamma)} x^{\lambda-1} \exp \left\{ -\frac{1}{2} \left(\frac{\delta^2}{x} + \gamma^2 x \right) \right\}, \quad x > 0, \quad (2.6.4)$$

где

$$\begin{aligned}\delta &\geq 0, \gamma > 0, \text{ если } \lambda > 0; \\ \delta &> 0, \gamma > 0, \text{ если } \lambda = 0; \\ \delta &> 0, \gamma \geq 0, \text{ если } \lambda < 0,\end{aligned}$$

представляет собой плотность распределения вероятностей положительной случайной величины. Распределение, определяемое плотностью (2.6.4), называется *обобщенным обратным гауссовским* (или GIG-распределением). Случайную величину с плотностью вероятностей (2.6.4) обозначим $GIG(\lambda, \delta, \gamma)$.

С учетом того факта, что $K_{-\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2}}x^{-1/2}e^{-x}$, можно заметить, что при $\lambda = -\frac{1}{2}$ GIG-распределение сводится к обычному обратному гауссовскому (IG) распределению с плотностью

$$p_{IG}(x; \delta, \gamma) = \frac{\delta}{\sqrt{2\pi}} \cdot x^{-\frac{3}{2}} \cdot \exp\left\{-\frac{\gamma^2}{2x}\left(x - \frac{\delta}{\gamma}\right)^2\right\}, \quad x > 0. \quad (2.6.5)$$

Случайную величину с плотностью (2.6.5) обозначим $IG(\delta, \gamma)$. Хорошо известно, что $IG(\delta, \gamma)$ можно проинтерпретировать как момент первого достижения уровня δ процессом броуновского движения с единичным коэффициентом диффузии и коэффициентом сноса, равным γ .

Характеристическая функция обобщенного обратного гауссовского распределения равна

$$\begin{aligned}\psi_{GIG}(s; \lambda, \delta, \gamma) &= \int_0^\infty e^{itx} p_{GIG}(x; \lambda, \delta, \gamma) dx = \\ &= \left(\frac{\gamma}{\sqrt{\gamma^2 - 2it}}\right)^\lambda \cdot \frac{K_\lambda(\delta\sqrt{\gamma^2 - 2it})}{K_\lambda(\delta\gamma)}, \quad s \in \mathbb{R}.\end{aligned}$$

Отсюда, учитывая известное соотношение

$$K'_\lambda(z) = -\frac{\lambda}{z}K_\lambda(z) - K_{\lambda-1}(z),$$

для целых $r \geq 0$ получаем

$$\mathbb{E}[GIG(\lambda, \delta, \gamma)]^r = \left(\frac{\delta}{\gamma}\right)^r \cdot \frac{K_{\lambda+r}(\delta\gamma)}{K_\lambda(\delta\gamma)}.$$

В частности,

$$\mathbb{E}GIG(\lambda, \delta, \gamma) = \frac{\delta}{\gamma} \cdot \frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)}, \quad (2.6.6)$$

$$DGIG(\lambda, \delta, \gamma) = \left(\frac{\delta}{\gamma}\right)^2 \cdot \left[\frac{K_{\lambda+2}(\delta\gamma)}{K_{\lambda}(\delta\gamma)} - \left(\frac{K_{\lambda+1}(\delta\gamma)}{K_{\lambda}(\delta\gamma)}\right)^2 \right]. \quad (2.6.7)$$

Кстати, устремляя $\delta \downarrow 0$ и используя тот факт, что $K_{\lambda}(z) \sim \frac{1}{2}\Gamma(\lambda)(z/\lambda)^{-\lambda}$ при $z \rightarrow \infty$, легко убедиться в том, что

$$\lim_{\delta \downarrow 0} \psi_{GIG}(s; \lambda, \delta, \gamma) = \left(1 - \frac{2it}{\gamma^2}\right)^{-\lambda}, \quad s \in \mathbb{R},$$

причем правая часть последнего соотношения совпадает с характеристической функцией гамма-распределения с параметром формы λ и параметром масштаба $a = \gamma^2/2$, определяемого плотностью вероятностей

$$p_{\Gamma}(x) = \frac{a^{\lambda}}{\Gamma(\lambda)} \cdot x^{\lambda-1} e^{-ax}, \quad x > 0.$$

Аналогично, при $\lambda < 0$ и $\delta > 0$, устремляя $\delta \downarrow 0$, в пределе получается плотность *обратного гамма-распределения* с параметрами λ и $\delta^2/2$:

$$\lim_{\delta \downarrow 0} P_{GIG}(x; \lambda, \delta, \gamma) = \left(\frac{2}{\delta^2}\right)^{\lambda} \frac{x^{\lambda-1}}{\Gamma(|\lambda|)} \exp\left\{-\frac{\delta^2}{2x}\right\}, \quad x > 0.$$

Подробнее о других свойствах обратного гауссовского распределения и обобщенного обратного гауссовского распределения см. в (Jørgensen, 1982), (Chhikara and Folks, 1989) и (Seshadri, 1999).

В 1977 г. О. Барндорфф-Нильсен ввел класс *обобщенных гиперболических распределений* как класс сдвиг-масштабных смесей нормальных законов вида

$$F_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = E\Phi\left(\frac{x - \mu - \beta GIG(\lambda, \delta, \sqrt{\alpha^2 - \beta^2})}{GIG(\lambda, \delta, \sqrt{\alpha^2 - \beta^2})}\right), \quad x \in \mathbb{R}, \quad (2.6.8)$$

где

$$\begin{aligned} \mu \in \mathbb{R}; \quad & \delta \geq 0, \quad |\beta| < \alpha, \quad \text{если } \lambda > 0; \\ & \delta > 0, \quad |\beta| < \alpha, \quad \text{если } \lambda = 0; \\ & \delta > 0, \quad |\beta| \leq \alpha, \quad \text{если } \lambda < 0. \end{aligned}$$

Случайную величину с функцией распределения (2.6.8) обозначим $GH(\lambda, \alpha, \beta, \delta, \mu)$. Можно убедиться, что величина $GH(\lambda, \alpha, \beta, \delta, \mu)$ обладает плотностью вероятностей вида

$$p_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = \frac{(\alpha^2 - \beta^2)^{\lambda/2}}{\sqrt{2\pi} \alpha^{\lambda - \frac{1}{2}} \delta^{\lambda} K_{\lambda}(\delta \sqrt{\alpha^2 - \beta^2})} (\delta^2 + (x - \mu)^2)^{(\lambda - \frac{1}{2})/2} \times$$

$$\times K_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2 + (x-\mu)^2}) \cdot \exp(\beta(x-\mu)), \quad x \in \mathbb{R}. \quad (2.6.9)$$

Обобщенное гиперболическое распределение имеет “полутяжелые” хвосты в том смысле, что его плотность удовлетворяет асимптотическому соотношению

$$p_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) \sim |x|^{\lambda-1} \exp\{(\pm\alpha + \beta)x\}$$

с точностью до постоянного множителя при $x \rightarrow \pm\infty$.

Характеристическая функция случайной величины $GH(\lambda, \alpha, \beta, \delta, \mu)$ имеет вид

$$\psi_{GH}(s; \lambda, \alpha, \beta, \delta, \mu) = e^{i\mu s} \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + is)^2} \right)^{\lambda/2} \cdot \frac{K_\lambda(\delta\sqrt{\alpha^2 - (\beta + is)^2})}{K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})}, \quad s \in \mathbb{R}.$$

С учетом (2.6.8), (2.6.6) и (2.6.7) отсюда вытекает, что

$$EGH(\lambda, \alpha, \beta, \delta, \mu) = \mu + \beta EGIG(\lambda, \delta, \gamma) = \mu + \frac{\delta\beta}{\gamma} \cdot \frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)},$$

$$\begin{aligned} DGH(\lambda, \alpha, \beta, \delta, \mu) &= EGIG(\lambda, \delta, \gamma) + \beta^2 DGIG(\lambda, \delta, \gamma) = \\ &= \frac{\delta}{\gamma} \cdot \frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)} + \left(\frac{\beta\delta}{\gamma} \right)^2 \cdot \left[\frac{K_{\lambda+2}(\delta\gamma)}{K_\lambda(\delta\gamma)} - \left(\frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)} \right)^2 \right]. \end{aligned}$$

Отметим, что различные значения параметра $\lambda \in \mathbb{R}$ определяют различные подклассы обобщенных гиперболических распределений.

Из определения случайной величины $GH(\lambda, \alpha, \beta, \delta, \mu)$ (см. (2.6.8)) вытекает, что

$$GH(\lambda, \alpha, \beta, \delta, \mu) \stackrel{d}{=} X(GIG(\lambda, \delta, \sqrt{\alpha^2 - \beta^2})) + \mu, \quad (2.6.10)$$

где $X(t)$, $t \geq 0$, – процесс Леви такой, что для любого $t > 0$ $X(t) \stackrel{d}{=} \beta t + N(0, t)$.

Соотношение (2.6.10) может быть переписано в виде

$$\begin{aligned} GH(\lambda, \alpha, \beta, \delta, \mu) &\stackrel{d}{=} \sqrt{GIG(\lambda, \delta, \sqrt{\alpha^2 - \beta^2})} \cdot N(0, 1) + \\ &+ \beta \cdot GIG(\lambda, \delta, \sqrt{\alpha^2 - \beta^2}) + \mu. \end{aligned} \quad (2.6.11)$$

Поскольку GIG -распределение является безгранично делимым, то соотношение (2.6.10) гарантирует *безграничную делимость* обобщенных гиперболических распределений. Можно определить *обобщенный гиперболический процесс Леви* $Y_{GH}(t)$ такой, что $Y_{GH}(1) \stackrel{d}{=} GH(\lambda, \alpha, \beta, \delta, \mu)$.

Из сказанного выше (и из того, что $E(GIG)^2 < \infty$) следует, что обобщенный гиперболический процесс Леви может появляться в качестве предела в теореме 2.6.2 со случайной величиной $U \stackrel{d}{=} GIG$. Допредельный процесс Леви $\Lambda_n(t)$ может быть взят, например, в виде $\Lambda_n(t) = k_n \Lambda(t)$, где $\Lambda(1) \stackrel{d}{=} GIG$.

Отметим, что в работах (Eberlein and Keller, 1995), (Prause, 1997), (Eberlein, Keller and Prause, 1998), (Eberlein and Prause, 1998) установлено хорошее согласие обобщенного гиперболического распределения, *гиперболического* ($\lambda = 1$) и *нормального обратного гауссовского* ($\lambda = -1/2$) распределений, входящих в семейство обобщенных гиперболических распределений вероятностей, с данными о ценах на датских и немецких биржах, финансовыми индексами: NYSE, DAX, обменными курсами валют и так далее. С точки зрения поведения “хвостов”, эти распределения занимают как бы промежуточное положение между устойчивыми распределениями с индексом $\alpha < 2$ и гауссовскими распределениями $\alpha = 2$: их “хвосты” убывают быстрее, чем у устойчивых распределений ($\alpha < 2$), но медленнее гауссовских.

Класс обобщенных гиперболических распределений весьма широк. В дополнение к уже упомянутым представителям этого класса, в частности, заметим, что распределение случайной величины $GH(-\frac{\nu}{2}, 0, 0, \sqrt{\nu}, 0)$ совпадает с распределением Стьюдента с ν степенями свободы.

Пяти параметров семейства обобщенных гиперболических распределений более чем достаточно для успешной (в некоторых случаях “неприлично хорошей”) подгонки таких законов к эмпирическим данным; подклассы гиперболических и нормальных\обратных гауссовских распределений вероятностей в большинстве случаев обеспечивают практически такое же качество аппроксимации. Опишем данные классы более подробно.

Гиперболические процессы Леви

ПРИМЕР 2.6.3. Если в соотношениях (2.6.8) и (2.6.9) положить $\lambda = 1$, то получается подкласс *гиперболических* (H) распределений, плотность которых устроена гораздо проще:

$$p_H(x) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} \cdot \exp \left\{ -\alpha\sqrt{\delta^2 + (x - \mu)^2} + \beta(x - \mu) \right\}.$$

Происхождение термина *гиперболический* связано с тем, что график логарифма $\ln p_H(x)$ является *гиперболой* (в то время как у гауссовской

плотности график логарифма представляет собой *параболу*). Именно данный подкласс распределений был первым востребован для анализа финансовых индексов (Eberlein and Keller, 1995).

Используя H -распределенную случайную величину, можно определить *гиперболический процесс Леви*.

Нормальные\обратные гауссовские процессы Леви

ПРИМЕР 2.6.4. Другим подклассом обобщенных гиперболических законов, обеспечивающим хорошее согласие с финансовыми индексами, является подкласс *нормальных\обратных гауссовских* (NIG) распределений, для которого $\lambda = -1/2$. Плотность NIG -распределения вероятностей имеет вид

$$p_{NIG}(x) = \frac{\alpha\delta}{\pi} \cdot \frac{\exp\{\delta\sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)\}}{\sqrt{\delta^2 + (x - \mu)^2}} \cdot K_1\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right).$$

NIG -распределение имеет “полутяжелые” хвосты в том смысле, что его плотность удовлетворяет асимптотическому соотношению

$$p_{NIG}(x) \sim \frac{\exp\{(\pm\alpha + \beta)x\}}{|x|^{3/2}}$$

с точностью до постоянного множителя при $x \rightarrow \pm\infty$.

Отметим, что при $\lambda = -1/2$ плотность смешивающего GIG -распределения (2.6.9) принимает вид

$$p_{IG}(x) = \left(\frac{\delta^2}{2\pi}\right)^{\frac{1}{2}} \cdot x^{-\frac{3}{2}} \cdot \exp\left\{-\frac{\gamma^2}{2x}\left(x - \frac{\delta}{\gamma}\right)^2\right\}, \quad x > 0,$$

то есть в таком случае мы получаем плотность *обратного гауссовского* (IG) распределения вероятностей.

Характеристическая функция NIG -распределения имеет вид

$$\psi_{NIG}(s) = \mathbf{E} e^{isNIG} = e^{i\mu s} \cdot \exp\left\{\delta\left[\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + is)^2}\right]\right\}.$$

Отсюда видно, что если $Y_1, \dots, Y_n$ – независимые NIG -распределенные случайные величины с *одинаковыми* α и β , то их сумма $Y = Y_1 + \dots + Y_n$ также имеет NIG -распределение с теми же α и β и с $\mu = \mu_1 + \dots + \mu_n$, $\delta = \delta_1 + \dots + \delta_n$. Таким образом, NIG -распределение замкнуто (в указанном смысле) относительно свертки.

Отметим, что никакой другой подкласс обобщенных гиперболических распределений ($\lambda \neq -1/2$) не обладает указанным привлекательным свойством (см., например, (Eberlein, 1999)). В том же источнике отмечается, что, хотя приращения обобщенного гиперболического процесса Леви на интервалах времени единичной длины и имеют (по построению) требуемое GH -распределение, однако, за исключением NIG -процессов Леви, приращение процесса на интервале, отличном от единичного, не принадлежит к тому же подклассу.

Нетрудно показать, что характеристическая функция $IG(\delta, \gamma)$ -распределения имеет вид

$$\mathbb{E} e^{isIG} = \exp \left\{ \delta \left(\gamma - \sqrt{\gamma^2 - 2is} \right) \right\}.$$

Отсюда для любого $c > 0$

$$c \cdot IG(\delta, \gamma) \stackrel{d}{=} IG\left(\sqrt{c}\delta, \frac{\gamma}{\sqrt{c}}\right).$$

Тогда для получения в пределе NIG -процесса Леви в качестве допредельного управляющего процесса может быть выбран процесс Леви $\Lambda_n(t)$, $t \geq 0$, такой, что для любого $t > 0$

$$\Lambda_n(t) \stackrel{d}{=} IG\left(\sqrt{k_n}\delta t, \frac{\sqrt{\alpha^2 - \beta^2}}{\sqrt{k_n}}\right).$$

NIG -процесс Леви $Z_{NIG}(t)$ определяется его спектральной плотностью Леви

$$\ell_{NIG}(x) = \frac{\delta\alpha}{\pi} \cdot \frac{e^{\beta x} K_1(\alpha|x|)}{|x|},$$

где $K_1(\cdot)$ – модифицированная функция Бесселя третьего рода с индексом 1. При этом для моментов случайной величины $Z_{NIG}(1) = NIG$ справедливы соотношения

$$\mathbb{E} Z_{NIG}(1) = \delta\beta\sqrt{\alpha^2 - \beta^2}, \quad \mathbb{D} Z_{NIG}(1) = \alpha^2\delta(\alpha^2 - \beta^2)^{-3/2}.$$

Коэффициенты асимметрии и эксцесса (островершинности) случайной величины $Z_{NIG}(1) = NIG$, соответственно, равны

$$\begin{aligned} \mathbb{E} \left(\frac{Z_{NIG}(1) - \mathbb{E} Z_{NIG}(1)}{\sqrt{\mathbb{D} Z_{NIG}(1)}} \right)^3 &= \frac{3\beta}{\alpha\sqrt{\delta}(\alpha^2 - \beta^2)^{1/4}}, \\ \mathbb{E} \left(\frac{Z_{NIG}(1) - \mathbb{E} Z_{NIG}(1)}{\sqrt{\mathbb{D} Z_{NIG}(1)}} \right)^4 &= 3 \left(1 + \frac{\alpha^2 + 4\beta^2}{\delta\alpha^2\sqrt{\alpha^2 - \beta^2}} \right). \end{aligned}$$

“Физический” смысл параметров NIG -модели можно проследить в рамках следующей конструкции (Barndorff-Nielsen, 1998). Пусть $W_{0,1}(t)$ – процесс броуновского движения с единичным коэффициентом диффузии и нулевым коэффициентом сноса (то есть $W_{0,1}(t)$ – стандартный винеровский процесс). Определим процесс $W_{\nu,1}(t) = \nu t + W_{0,1}(t)$ ($W_{\nu,1}(t)$ – процесс броуновского движения с единичным коэффициентом диффузии и коэффициентом сноса, равным ν). Пусть $T_\nu(t)$ – время, требующееся процессу $W_{\nu,1}(t)$, чтобы достичь уровня $t > 0$. При этом для преобразования Лапласа–Стилтьеса случайной величины $T_\nu(t)$ справедлива формула

$$\mathbf{E} \exp\{-sT_\nu(t)\} = \exp\left\{-t(\sqrt{2s + \nu^2} - \nu)\right\}.$$

Пусть, далее, $\widehat{W}_{0,1}(t)$ – еще один стандартный винеровский процесс, независимый от $W_{0,1}(t)$. Определим процесс $\widehat{W}_{\theta,\sigma}(t) = \theta t + \sigma \widehat{W}_{0,1}(t)$ ($\widehat{W}_{\theta,\sigma}(t)$ – процесс броуновского движения с коэффициентом диффузии σ и коэффициентом сноса θ). Определим нормальный \\\ обратный гауссовский процесс Леви $Z_{NIG}^\circ(t)$ как

$$Z_{NIG}^\circ(t) = \widehat{W}_{\theta,\sigma}(T_\nu(t)) = \theta T_\nu(t) + \sigma \widehat{W}_{0,1}(T_\nu(t)), \quad t \geq 0.$$

Несложно видеть, что при каждом $t > 0$

$$Z_{NIG}^\circ(t) \stackrel{d}{=} \theta \cdot T_\nu(t) + \sqrt{T_\nu(t)} \cdot N(0, \sigma^2),$$

причем случайные величины $T_\nu(t)$ и $N(0, \sigma^2)$ независимы. Тогда характеристическая функция случайной величины $Z_{NIG}^\circ(t)$ имеет вид

$$\mathbf{E} \exp\{isZ_{NIG}^\circ(t)\} = \exp\left\{-t\left[|\nu| + \sqrt{\nu^2 + 2i\theta s + \sigma^2 s^2}\right]\right\}, \quad s \in \mathbb{R}.$$

При этом для спектральной плотности Леви случайной величины $Z_{NIG}^\circ(1)$ справедливо представление

$$\ell_{Z_{NIG}^\circ(1)}(x) = \sqrt{\frac{2}{\pi}} \sigma \left(\frac{\nu^2}{\sigma^2} + \frac{\theta^2}{\sigma^4} \right) \exp\left\{\frac{\theta x}{\sigma^2}\right\} \frac{K_1(x)}{|x|},$$

в то время как плотность случайной величины $Z_{NIG}^\circ(t)$ равна

$$\begin{aligned} p_{Z_{NIG}^\circ(t)}(x) &= \frac{t\sqrt{\sigma^2\nu^2 + \theta^2}}{\sigma\pi} \cdot \frac{\exp\{\theta x\sigma^{-2} + \sigma^2\nu\}}{\sqrt{t^2\sigma^2 + x^2}} \times \\ &\times K_1\left(\sigma^{-2}\sqrt{(\sigma^2\nu^2 + \theta^2)(t^2\sigma^2 + x^2)}\right), \quad x \in \mathbb{R}, \end{aligned}$$

где $K_1(\cdot)$ – модифицированная функция Бесселя третьего рода порядка 1. При этом параметры α , β и δ связаны с параметрами ν , θ и σ соотношениями

$$\delta = \sigma, \quad \alpha^2 = \frac{\sigma^2\nu^2 + \theta^2}{\sigma^4}, \quad \beta = \frac{\theta}{\sigma^2}.$$

Дисперсионные гамма-процессы (VG-процессы)

ПРИМЕР 2.6.5. Если в соотношении (2.6.5) положить $\lambda > 0$, $\delta = 0$ и $\gamma > 0$, то получается плотность

$$p_{\Gamma}(x) = \left(\frac{\gamma^2}{2}\right)^{\lambda} \cdot \frac{1}{\Gamma(\lambda)} \cdot x^{\lambda-1} \cdot \exp\left\{-\frac{1}{2}\gamma^2 x\right\}, \quad x > 0.$$

Это плотность *гамма-распределения* с параметрами $\alpha_0^* = \lambda$, $\beta^* = \gamma^2/2$ ($\Gamma(\alpha_0^*, \beta^*)$ -распределения). Тогда, если в теореме 2.6.2 в качестве U взять случайную величину, имеющую $\Gamma(\alpha_0^*, \beta^*)$ -распределение, то предельным процессом будет являться процесс Леви $Z(t)$, $t \geq 0$, такой, что

$$Z(1) \stackrel{d}{=} \sigma \cdot \sqrt{\Gamma(\alpha_0^*, \beta^*)} \cdot N(0, 1) + a \cdot \Gamma(\alpha_0^*, \beta^*) + \mu. \quad (2.6.12)$$

В силу известного свойства гамма-распределений допредельный процесс Леви $\Lambda_n(t)$ может быть взят таким, что $\Lambda_n(1) \stackrel{d}{=} \Gamma(\alpha_0^*, \beta^*/k_n)$. Легко убедиться, что при этом случайная величина $N_n(t) = N_{n,1}(\Lambda_n(t))$ будет иметь отрицательное биномиальное распределение.

Полагая без ограничения общности $\sigma = 1$, из (2.6.11) и (2.6.12) получаем, что

$$Z(1) \stackrel{d}{=} GH(\lambda, \alpha, \beta, 0, \mu),$$

где

$$\lambda = \alpha_0^*, \quad \beta = a, \quad \frac{\gamma^2}{2} = \frac{\alpha^2 - \beta^2}{2} = \beta^*, \quad \mu \in \mathbb{R}.$$

Таким образом, случайная величина $Z(1)$ имеет обобщенное гиперболическое распределение $GH(\lambda = \alpha_0^*, \alpha = \sqrt{2\beta^* + a^2}, \beta = a, 0, \mu)$.

Рассматриваемая модель является обобщением предложенного в работах (Madan and Seneta, 1990), (Carr, Madan and Chang, 1998) так называемого *дисперсионного гамма-процесса* (Variance Gamma (VG) process). Этот процесс определяется следующим образом. Пусть $W_{0,1}(t)$ – процесс броуновского движения с единичным коэффициентом диффузии и нулевым коэффициентом сноса (то есть $W_{0,1}(t)$ – стандартный винеровский процесс). Определим процесс $W_{\nu,1}(t) = \nu t + W_{0,1}(t)$ ($W_{\nu,1}(t)$ – процесс броуновского движения с единичным коэффициентом диффузии и коэффициентом сноса, равным ν). Пусть $\Gamma_{\nu}(t)$ – гамма-процесс Леви такой, что $E\Gamma_{\nu}(1) = 1$ и $D\Gamma_{\nu}(1) = \nu$. Тогда плотность распределения случайной величины $\Gamma_{\nu}(t)$ имеет вид

$$p_{\Gamma_{\nu}(t)}(x) = \frac{x^{t/\nu-1} e^{-x/\nu}}{\nu^{t/\nu} \Gamma(t/\nu)}, \quad x > 0.$$

Здесь $\Gamma(\cdot)$ – эйлерова гамма-функция. Соответственно, преобразование Лапласа–Стилтьеса случайной величины $\Gamma_\nu(t)$ имеет вид

$$f_{\Gamma_\nu(t)}(s) = \mathbb{E} \exp\{-s\Gamma_\nu(t)\} = \frac{1}{\nu^{t/\nu}\Gamma(t/\nu)} \int_0^\infty \frac{x^{t/\nu-1} e^{-x(s+1/\nu)} dx}{\nu^{t/\nu}\Gamma(t/\nu)} = \frac{1}{(1+s\nu)^{t/\nu}},$$

$s \geq 0$. Пусть, далее, $\widehat{W}_{0,1}(t)$ – еще один стандартный винеровский процесс, независимый от $W_{0,1}(t)$. Определим процесс $\widehat{W}_{\theta,\sigma}(t) = \theta t + \sigma \widehat{W}_{0,1}(t)$ ($\widehat{W}_{\theta,\sigma}(t)$ – процесс броуновского движения с коэффициентом диффузии σ и коэффициентом сноса θ). Определим дисперсионный гамма-процесс как

$$Z_{VG}(t) = \widehat{W}_{\theta,\sigma}(\Gamma_\nu(t)) = \theta \Gamma_\nu(t) + \sigma \widehat{W}_{0,1}(\Gamma_\nu(t)), \quad t \geq 0.$$

Несложно видеть, что при каждом $t > 0$

$$Z_{VG}(t) \stackrel{d}{=} \theta \cdot \Gamma_\nu(t) + \sqrt{\Gamma_\nu(t)} \cdot N(0, \sigma^2),$$

причем случайные величины $\Gamma_\nu(t)$ и $N(0, \sigma^2)$ независимы. Плотность распределения случайной величины $Z_{VG}(t)$ имеет вид

$$\begin{aligned} p_{VG}(x) &= \frac{1}{\sqrt{2\pi\nu^{t/\nu}\Gamma(t/\nu)}} \int_0^\infty z^{t/\nu-1/2} \exp\left\{-\frac{(x-\theta z)^2}{2\sigma^2 z} - \frac{z}{\nu}\right\} dz = \\ &= \frac{\sqrt{2}e^{\theta x/\sigma^2}}{\pi\sigma\nu^{t/\nu}\Gamma(t/\nu)} \left(\frac{|x|}{\sqrt{\theta^2 + 2\sigma^2/\nu}}\right)^{t/\nu-1/2} K_{t/\nu-1/2}\left(\frac{|x|}{\sigma^2} \sqrt{\theta^2 + 2\sigma^2/\nu}\right), \quad x \in \mathbb{R}, \end{aligned} \quad (2.6.13)$$

где $K_\gamma(\cdot)$ – модифицированная функция Бесселя третьего рода с индексом γ . Характеристическая функция случайной величины $Z_{VG}(t)$ имеет более простой вид:

$$\begin{aligned} \mathbb{E} \exp\{isZ_{VG}(t)\} &= \mathbb{E} \mathbb{E}[\exp\{isZ_{VG}(t)\} | \Gamma_\nu(t)] = \\ &= \mathbb{E} \mathbb{E}[\exp\{is[\theta\Gamma_\nu(t) + \sigma W_{0,1}(\Gamma_\nu(t))]\} | \Gamma_\nu(t)] = \\ &= \mathbb{E} \exp\{is\theta\Gamma_\nu(t) - \tfrac{1}{2}s^2\sigma^2\Gamma_\nu(t)\} = \mathbb{E} \exp\{i(s\theta + \tfrac{1}{2}s^2\sigma^2)\Gamma_\nu(t)\} = \\ &= f_{\Gamma_\nu(t)}(-i[s\theta + \tfrac{1}{2}s^2\sigma^2]) = \frac{1}{(1 - is\theta\nu + \tfrac{1}{2}\sigma^2\nu s^2)^{t/\nu}}, \quad s \in \mathbb{R}. \end{aligned}$$

При этом для спектральной плотности Леви случайной величины $Z_{VG}(1)$ справедливо представление (Carr, Geman, Madan and Yor, 2002)

$$\ell_{VG}(x) = \frac{\exp\{xR(x; \theta, \nu, \sigma)\}}{\nu|x|},$$

где

$$R(x; \theta, \nu, \sigma) = \frac{2}{-\operatorname{sign} x \sqrt{\theta^2 \nu^2 + 2\sigma^2 \nu} - \theta \nu}.$$

Для моментов случайной величины $Z_{VG}(1)$ справедливы соотношения

$$\mathbb{E}Z_{VG}(1) = \theta, \quad \mathbb{D}Z_{VG}(1) = \sigma^2 + \nu\theta,$$

коэффициенты асимметрии и эксцесса (островершинности), соответственно, равны

$$\begin{aligned} \mathbb{E}\left(\frac{Z_{VG}(1) - \mathbb{E}Z_{VG}(1)}{\sqrt{\mathbb{D}Z_{VG}(1)}}\right)^3 &= \frac{\theta\nu(3\sigma^2 + 2\nu\theta^2)}{(\sigma^2 + \nu\theta^2)^{3/2}}, \\ \mathbb{E}\left(\frac{Z_{VG}(1) - \mathbb{E}Z_{VG}(1)}{\sqrt{\mathbb{D}Z_{VG}(1)}}\right)^4 &= 3\left[1 + 2\nu - \frac{\nu\sigma^4}{(\sigma^2 + \nu\theta^2)^2}\right]. \end{aligned}$$

Возвращаясь к более общему процессу Леви $Z(t)$, определяемому соотношением (2.6.12), заметим, что, если $T_{\alpha_0^*, \beta^*}(t)$, $t \geq 0$, есть процесс Леви такой, что

$$T_{\alpha_0^*, \beta^*}(1) \stackrel{d}{=} \Gamma(\alpha_0^*, \beta^*),$$

то

$$T_{\alpha_0^*, \beta^*}(t) \stackrel{d}{=} \Gamma(\alpha_0^* t, \beta^*), \quad t \geq 0.$$

Пусть $W_{0,1}(t)$ – стандартный винеровский процесс, независимый от процесса $T_{\alpha_0^*, \beta^*}(t)$. Тогда процесс Леви $Z(t)$, определяемый соотношением (2.6.12), может быть представлен в виде

$$Z(t) = \sigma W_{0,1}(T_{\alpha_0^*, \beta^*}(t)) + a T_{\alpha_0^*, \beta^*}(t), \quad t \geq 0.$$

Будем говорить, что случайная величина Y имеет *симметризованное гамма-распределение* с параметрами α_1, β_1 ($SG(\alpha_1, \beta_1)$ -распределение), если $Y \stackrel{d}{=} X_1 - X_2$, где X_1 и X_2 независимые одинаково распределенные случайные величины, причем случайная величина X_1 имеет $\Gamma(\alpha_1, \beta_1)$ -распределение (см. пример 3.5.1).

Заметим, что если в (2.6.12) $a = 0$ и $\sigma = 1$, то случайная величина $Z(1) - \mu$ имеет $SG(\tilde{\alpha}, \tilde{\beta})$ -распределение, где $\tilde{\alpha} = \alpha_0^*$, $\tilde{\beta} = \sqrt{2\beta^*}$.

Двусторонние гамма-процессы

ПРИМЕР 2.6.6. Анализируя характеристические функции, можно убедиться, что введенная в предыдущем примере случайная величина

$$Z(1) - \mu \stackrel{d}{=} \sqrt{\Gamma(\alpha_0^*, \beta^*)} \cdot N(0, 1) + a \cdot \Gamma(\alpha_0^*, \beta^*)$$

может быть представлена как разность двух независимых гамма-распределенных случайных величин с одинаковым параметром формы α , но разным масштабным параметром β :

$$Z(1) - \mu \stackrel{d}{=} \Gamma(\alpha, \beta_1) - \Gamma(\alpha, \beta_2).$$

Действительно, аналогично тому, как была вычислена характеристическая функция дисперсионного гамма-процесса, можно убедиться, что

$$\mathbb{E} e^{is(Z(1)-\mu)} = \frac{(\beta^*)^{\alpha_0^*}}{\left(\frac{s^2}{2} + \beta^* - isa\right)^{\alpha_0^*}}.$$

Но с другой стороны,

$$\mathbb{E} e^{is(\Gamma(\alpha, \beta_1) - \Gamma(\alpha, \beta_2))} = \frac{\beta_1^\alpha}{(\beta_1 - is)^\alpha} \cdot \frac{\beta_2^\alpha}{(\beta_2 + is)^\alpha} = \left(\frac{\frac{1}{2} \cdot \beta_1 \cdot \beta_2}{\frac{s^2}{2} + \frac{\beta_1 \cdot \beta_2}{2} - is \cdot \frac{\beta_2 - \beta_1}{2}} \right)^\alpha. \quad (2.6.14)$$

Теперь достаточно положить

$$\beta^* = \frac{\beta_1 \beta_2}{2}, \quad \alpha_0^* = \alpha, \quad a = \frac{\beta_2 - \beta_1}{2}.$$

Как показано в работах (Carr, Madan and Chang, 1998), (Carr and Madan, 1999), (Кащеев, 2001), (Fu et al., 2007), обсуждаемый обобщенный дисперсионный гамма-процесс Леви является довольно адекватной моделью динамики финансовых индексов. Поэтому имеет смысл рассмотреть его более подробно. Будем использовать представление случайной величины $Z(1)$ в виде (2.6.12) и соответствующую параметризацию.

Полагая без ограничения общности $\sigma = 1$, для произвольного $t > 0$ получаем, что функция распределения случайной величины $Z(t)$ имеет вид

$$\begin{aligned} F_{Z(t)}(x) &= \mathbb{P}(\sqrt{\Gamma(\alpha_0^* t, \beta^*)} \cdot N(0, 1) + a \cdot \Gamma(\alpha_0^* t, \beta^*) + \mu t < x) = \\ &= \mathbb{P}(\sqrt{\Gamma(\alpha_0^* t, 1)} \cdot N(0, 1) + \frac{a}{\sqrt{\beta^*}} \cdot \Gamma(\alpha_0^* t, 1) < (x - \mu t) \cdot \sqrt{\beta^*}) = \\ &= F_{\tilde{Z}(t)}((x - \mu t) \cdot \sqrt{\beta^*}), \end{aligned}$$

где

$$\tilde{Z}(t) \stackrel{d}{=} \sqrt{\Gamma(\alpha_0^* t, 1)} \cdot N(0, 1) + \frac{a}{\sqrt{\beta^*}} \cdot \Gamma(\alpha_0^* t, 1). \quad (2.6.15)$$

Являясь сдвиг-масштабной смесью гауссовских законов, распределение случайной величины $\tilde{Z}(t)$ имеет вид

$$F_{\tilde{Z}(t)}(u) = \mathbb{E} \Phi \left(\frac{u - \frac{a}{\sqrt{\beta^*}} \Gamma(\alpha_0^* t, 1)}{\sqrt{\Gamma(\alpha_0^* t, 1)}} \right).$$

Обозначая $\delta_1 = a/\sqrt{\beta^*}$, получаем

$$F_{\tilde{Z}(t)}(u) = \frac{1}{\Gamma(\alpha_0^* t)} \int_0^\infty \Phi \left(\frac{u - \delta_1 y}{\sqrt{y}} \right) y^{\alpha_0^* t - 1} e^{-y} dy.$$

Чтобы найти плотность случайной величины $\tilde{Z}(t)$, заметим, что $\tilde{Z}(t)$ имеет $GH(\lambda = \alpha_0^* t, \alpha = \sqrt{2 + \delta_1^2}, \beta = \delta_1, \delta = 0, \mu = 0)$ -распределение (см. (2.6.12), (2.6.15)).

Так как для модифицированной функции Бесселя третьего рода порядка ν справедливы соотношения

$$\begin{aligned} K_{-\nu}(x) &= K_\nu(x), \\ K_\nu(x) &\sim \Gamma(\nu) \cdot 2^{\nu-1} \cdot x^{-\nu} \quad (x \downarrow 0, \nu > 0), \\ K_0(x) &\sim -\ln x \quad (x \downarrow 0), \end{aligned} \tag{2.6.16}$$

(см. с. 196 в (Абрамовиц и Стиган, 1979)), то при $\lambda > 0, \delta \downarrow 0$ (см. (2.6.4))

$$\delta^\lambda K_\lambda(\delta \sqrt{\alpha^2 - \beta^2}) \sim \delta^\lambda \Gamma(\lambda) 2^{\lambda-1} \delta^{-\lambda} (\sqrt{\alpha^2 - \beta^2})^{-\lambda} = \frac{\Gamma(\lambda) 2^{\lambda-1}}{(\sqrt{\alpha^2 - \beta^2})^\lambda}.$$

Таким образом, обозначая $\alpha_t^* = \alpha_0^* t$, из (2.6.4) получаем

$$p_{\tilde{Z}(t)}(u) = \sqrt{\frac{2}{\pi}} \frac{e^{\delta_1 u}}{\Gamma(\alpha_t^*)} \left(\frac{|u|}{\sqrt{2 + \delta_1^2}} \right)^{\alpha_t^* - \frac{1}{2}} \cdot K_{\alpha_t^* - \frac{1}{2}} \left(|u| \sqrt{2 + \delta_1^2} \right), \quad u \neq 0, \tag{2.6.17}$$

(что согласуется с (2.6.9)). Если $\alpha_t^* > 1/2$, то значение плотности $p_{\tilde{Z}(t)}(u)$ при $u = 0$ может быть доопределено по непрерывности с учетом (2.6.16):

$$p_{\tilde{Z}(t)}(0) = \frac{1}{\sqrt{2\pi}} \cdot \left(\frac{2}{2 + \delta_1^2} \right)^{\alpha_t^* - \frac{1}{2}} \cdot \frac{\Gamma(\alpha_t^* - \frac{1}{2})}{\Gamma(\alpha_t^*)}.$$

Из (2.6.17) и (2.6.16) следует, что при $\alpha_t^* \leq 1/2$ плотность $p_{\tilde{Z}(t)}(u)$ в точке $u = 0$ не ограничена.

Так как $Z(t) \stackrel{d}{=} \tilde{Z}(t)/\sqrt{\beta^*} + \mu t$, то

$$p_{Z(t)}(x) = p_{\tilde{Z}(t)}((x - \mu t)\sqrt{\beta^*}) \cdot \sqrt{\beta^*}.$$

Одним из методов, используемым для оценки параметров распределения случайной величины $Z(1)$, является *метод моментов*. Опуская несколько страниц утомительных вычислений, приведем формулы для среднего и центральных моментов случайной величины $Z(1)$:

$$\begin{aligned} \mathbb{E} Z(1) &= a \frac{\alpha_0^*}{\beta^*} + \mu, \quad \mathbb{D} Z(1) = \frac{\alpha_0^*}{\beta^*} \left(\frac{a^2}{\beta^*} + 1 \right), \\ \mathbb{E} (Z(1) - \mathbb{E} Z(1))^3 &= \frac{\alpha_0^*}{(\beta^*)^{3/2}} \left[2 \left(\frac{a}{\sqrt{\beta^*}} \right)^3 + 3 \left(\frac{a}{\sqrt{\beta^*}} \right) \right], \\ \mathbb{E} (Z(1) - \mathbb{E} Z(1))^4 &= \\ &= \frac{3 \alpha_0^*}{(\beta^*)^2} \left(\left(\frac{a}{\sqrt{\beta^*}} \right)^4 (2 + \alpha_0^*) + 2 \left(\frac{a}{\sqrt{\beta^*}} \right)^2 (2 + \alpha_0^*) + 1 + \alpha_0^* \right). \end{aligned} \quad (2.6.18)$$

Формулы для начальных моментов здесь не приводятся, поскольку они громоздки и неудобны для оценки параметров по методу моментов.

Соотношения для центральных моментов случайной величины $Z(t)$ легко получаются из (2.6.18) заменой α_0^* на $\alpha_0^* t$ и μ на μt .

Используя (2.6.18), можно получить следующие выражения для *коэффициентов асимметрии и эксцесса* случайной величины $Z(1)$:

1) коэффициент асимметрии

$$\frac{\mathbb{E} (Z(1) - \mathbb{E} Z(1))^3}{(\mathbb{D} Z(1))^{3/2}} = \frac{2 \delta_1^3 + 3 \delta_1}{\sqrt{\alpha_0^*} (\delta_1^2 + 1)^{3/2}}; \quad (2.6.19)$$

2) коэффициент эксцесса

$$\frac{\mathbb{E} (Z(1) - \mathbb{E} Z(1))^4}{(\mathbb{D} Z(1))^2} = \frac{6}{\alpha_0^*} \left(1 - \frac{1}{2 (\delta_1^2 + 1)^2} \right) + 3 > 3, \quad (2.6.20)$$

где $\delta_1 = a/\sqrt{\beta^*}$.

Соотношения (2.6.12), (2.6.19) показывают, что распределение случайной величины $Z(t)$ симметрично только в том случае, если $\delta_1 = 0$ ($a = (\beta_2 - \beta_1)/2 = 0$, см. (2.6.14)); *скошенность* в ту или иную сторону определяется знаком параметра a .

Соотношение (2.6.20) в явном виде демонстрирует действие общего результата леммы 2.3.1. Поскольку гауссовские законы имеют нулевой коэффициент асимметрии и коэффициент эксцесса, равный трем, то тот факт, что, как видно из (2.6.19) и (2.6.20), при $t \rightarrow \infty$ коэффициенты асимметрии и эксцесса случайной величины $Z(t)$ стремятся, соответственно, к нулю и трем, указывает на возможность “унормализования” распределения $Z(t)$ с ростом t .

Рассмотрим случайную величину

$$\bar{Z}(t) = \sqrt{\Gamma(\alpha_0^* t, \beta^*)} \cdot N(0, 1) + a \cdot \Gamma(\alpha_0^* t, \beta^*).$$

Введем случайную величину $\bar{Z}_0(t)$, равную

$$\bar{Z}_0(t) = \frac{\bar{Z}(t) - \mathbb{E} \bar{Z}(t)}{\sqrt{\mathbb{D} \bar{Z}(t)}} = \frac{\beta^* \bar{Z}(t)}{\sqrt{\alpha_0^* (a^2 + \beta^*) t}} - a \sqrt{\frac{\alpha_0^* t}{a^2 + \beta^*}}.$$

Используя стандартную технику, легко показать, что $\bar{Z}_0(t) \Rightarrow N(0, 1)$ ($t \rightarrow \infty$). Таким образом, можно прийти к выводу, что случайная величина $\bar{Z}(t)$ имеет асимптотически нормальное распределение со средним $a \frac{\alpha_0^*}{\beta^*} t$ и дисперсией $\frac{\alpha_0^*}{\beta^*} \left(\frac{a^2}{\beta^*} + 1 \right) t$. Как было отмечено ранее, распределения доходностей финансовых активов также обнаруживают тенденцию к “унормализованию” с увеличением периода времени, за который рассчитываются доходности.

Аналогично, рассмотрим разность $\Gamma(\alpha_1, \beta) - \Gamma(\alpha_2, \beta)$ двух независимых гамма-распределенных случайных величин с одинаковыми масштабными параметрами β , но разными параметрами формы α . Не ограничивая общности, предположим, что $\alpha_1 > \alpha_2$. Ее характеристическая функция имеет вид

$$\begin{aligned} \mathbb{E} \exp \{ i s (\Gamma(\alpha_1, \beta) - \Gamma(\alpha_2, \beta)) \} &= \left(\frac{\beta}{\beta - i s} \right)^{\alpha_1} \left(\frac{\beta}{\beta + i s} \right)^{\alpha_2} = \\ &= \left(\frac{\beta^2}{\beta^2 + s^2} \right)^{\alpha_2} \left(\frac{\beta}{\beta - i s} \right)^{\alpha_1 - \alpha_2}, \quad s \in \mathbb{R}. \end{aligned} \quad (2.6.21)$$

Первый сомножитель в правой части (2.6.21) представляет собой характеристическую функцию симметризованного гамма-распределения $SG(\alpha_2, \beta)$. В разделе 3.5 (см. пример 3.5.1) можно будет убедиться, что такая характеристическая функция соответствует произведению стандартной нормальной случайной величины $N(0, 1)$ и корня квадратного из независимой от нее гамма-распределенной случайной величины $\Gamma(\alpha_2, \frac{1}{2}\beta^2)$ с параметром формы α_2 и параметром масштаба $\frac{1}{2}\beta^2$. Очевидно, что второй

сомножитель в правой части (2.6.21) представляет собой характеристическую функцию гамма-распределенной случайной величины $\Gamma(\alpha_1 - \alpha_2, \beta)$ с параметром формы $\alpha_1 - \alpha_2$ и параметром масштаба β . Таким образом,

$$\Gamma(\alpha_1, \beta) - \Gamma(\alpha_2, \beta) \stackrel{d}{=} \sqrt{\Gamma(\alpha_2, \frac{1}{2}\beta^2)} \cdot N(0, 1) + \Gamma(\alpha_1 - \alpha_2, \beta),$$

где все случайные величины в правой части независимы. Поэтому, если, как и ранее, $T_{\alpha, \beta}(t)$, $t \geq 0$, является процессом Леви таким, что

$$T_{\alpha, \beta}(1) \stackrel{d}{=} \Gamma(\alpha, \beta),$$

то есть

$$T_{\alpha, \beta}(t) \stackrel{d}{=} \Gamma(\alpha t, \beta), \quad t \geq 0,$$

и $W_{0,1}(t)$ – стандартный винеровский процесс, то процесс Леви $Z(t)$, такой, что $Z(1) \stackrel{d}{=} \Gamma(\alpha_1, \beta) - \Gamma(\alpha_2, \beta)$, может быть представлен в виде

$$Z(t) = W_{0,1}(T_{\alpha_2, \frac{1}{2}\beta^2}(t)) + T_{\alpha_1 - \alpha_2, \beta}(t), \quad t \geq 0,$$

причем все процессы в правой части независимы. Другими словами, процесс Леви, равный разности двух гамма-процессов Леви с одинаковыми масштабными параметрами β , но разными параметрами формы α , представляется в виде подчиненного винеровского процесса со случайным сносом. При этом одномерные распределения такого процесса являются сдвиг-масштабными смесями нормального закона:

$$\begin{aligned} P(Z(t) < x) &= \\ &= \frac{\beta^{\alpha_1 - \alpha_2}}{2^{\alpha_2} \Gamma(\alpha_2) \Gamma(\alpha_1 - \alpha_2)} \int_0^\infty \int_0^\infty \Phi\left(\frac{x - v}{\sqrt{u}}\right) u^{\alpha_2 - 1} v^{\alpha_1 - \alpha_2} \exp\left\{-\frac{\beta^2 u}{2} - \beta v\right\} du dv, \end{aligned}$$

$x \in \mathbb{R}$.

Наконец, заметим, что рассматривая два подходящих независимых допредельных процесса, удовлетворяющих условиям теоремы 2.6.3, возможно получить более общую модель, связанную с гамма-распределениями, а именно, процесс Леви $Z(t)$, $t \geq 0$, такой, что $Z(1) \stackrel{d}{=} \Gamma(\alpha_1, \beta_1) - \Gamma(\alpha_2, \beta_2)$, где слагаемые независимы и все параметры различны. Как показано в работе (Кащеев, 2001) (также см. (Bening and Korolev, 2002), глава 10), к такому предельному процессу можно придти и в рамках *единого* допредельного процесса с помощью функциональных предельных теорем для неслучайно централированных обобщенных процессов Кокса. Свойства таких общих

двусторонних гамма-процессов (bilateral gamma-processes) Леви и их применения в финансовой математике рассмотрены в работах (Кащеев, 2001), (Kühler and Tappe, 2007a), (Kühler and Tappe, 2007b). В частности, в работе (Kühler and Tappe, 2007a) показано, что спектральная плотность двусторонних гамма-процессов Леви имеет вид

$$\ell_{BG}(x) = \frac{\alpha_1}{x} e^{-\beta_1 x} \mathbb{I}_{(0,\infty)}(x) + \frac{\alpha_2}{|x|} e^{-\beta_2 |x|} \mathbb{I}_{(-\infty,0)}(x). \quad (2.6.22)$$

Не ограничивая общность, предположим, что $\alpha_1 > \alpha_2$. Тогда характеристическая функция случайной величины $Z(1)$ примет вид

$$\begin{aligned} \psi_{Z(1)}(s) &= \left(\frac{\beta_1}{\beta_1 - is} \right)^{\alpha_1} \left(\frac{\beta_2}{\beta_2 + is} \right)^{\alpha_2} = \\ &= \left(\frac{\beta_1}{\beta_1 - is} \right)^{\alpha_2} \left(\frac{\beta_2}{\beta_2 - is} \right)^{\alpha_2} \left(\frac{\beta_1}{\beta_1 - is} \right)^{\alpha_1 - \alpha_2}, \quad s \in \mathbb{R}, \end{aligned}$$

откуда с учетом (2.6.14) вытекает, что

$$Z(1) \stackrel{d}{=} \sqrt{\Gamma(\alpha_2, \frac{1}{2}\beta_1\beta_2)} \cdot N(0, 1) + \frac{1}{2}(\beta_2 - \beta_1)\Gamma(\alpha_2, \frac{1}{2}\beta_1\beta_2) + \Gamma(\alpha_1 - \alpha_2, \beta_1),$$

где случайные величины $N(0, 1)$, $\Gamma(\alpha_2, \frac{1}{2}\beta_1\beta_2)$ и $\Gamma(\alpha_1 - \alpha_2, \beta_1)$ независимы. Это означает, что в терминах гамма-процессов Леви $T_{\alpha,\beta}(t)$, введенных ранее, двусторонний гамма-процесс Леви $Z(t)$ может быть записан в виде подчиненного винеровского процесса со случайным дрейфом (сносом):

$$Z(t) = W_{0,1}(T_{\alpha_2, \frac{1}{2}\beta_1\beta_2}(t)) + \frac{1}{2}(\beta_2 - \beta_1)T_{\alpha_2, \frac{1}{2}\beta_1\beta_2}(t) + T_{\alpha_1 - \alpha_2, \beta_1}(t), \quad t \geq 0,$$

где винеровский процесс $W_{0,1}(t)$ и гамма-процессы Леви $T_{\alpha_2, \frac{1}{2}\beta_1\beta_2}(t)$ и $T_{\alpha_1 - \alpha_2, \beta_1}(t)$ независимы. Таким образом, распределение случайной величины $Z(1)$ является сдвиг-масштабной смесью нормальных законов

$$\begin{aligned} P(Z(t) < x) &= \\ &= \left(\frac{\beta_1\beta_2}{2} \right)^{\alpha_2} \frac{\beta_1^{\alpha_1 - \alpha_2}}{\Gamma(\alpha_2)\Gamma(\alpha_1 - \alpha_2)} \int_0^\infty \int_0^\infty \Phi\left(\frac{x - \frac{1}{2}(\beta_2 - \beta_1)u - v}{\sqrt{u}} \right) \times \\ &\quad \times u^{\alpha_2 - 1} v^{\alpha_1 - \alpha_2 - 1} \exp\left\{ -\frac{\beta_1\beta_2}{2}u - \beta_1 v \right\} du dv, \quad x \in \mathbb{R}. \end{aligned}$$

Для моментов случайной величины $Z(1)$ справедливы выражения

$$EZ(1) = \frac{\alpha_1}{\beta_1} - \frac{\alpha_2}{\beta_2}, \quad DZ(1) = \frac{\alpha_1}{\beta_1^2} + \frac{\alpha_2}{\beta_2^2},$$

причем коэффициенты асимметрии и эксцесса, соответственно, имеют вид

$$\begin{aligned} \mathbb{E} \left(\frac{Z(1) - \mathbb{E}Z(1)}{\sqrt{\mathbb{D}Z(1)}} \right)^3 &= 2 \left(\frac{\alpha_1}{\beta_1^3} - \frac{\alpha_2}{\beta_2^3} \right) \left(\frac{\alpha_1}{\beta_1^2} + \frac{\alpha_2}{\beta_2^2} \right)^{-3/2}, \\ \mathbb{E} \left(\frac{Z(1) - \mathbb{E}Z(1)}{\sqrt{\mathbb{D}Z(1)}} \right)^4 &= 3 + 6 \left(\frac{\alpha_1}{\beta_1^4} + \frac{\alpha_2}{\beta_2^4} \right) \left(\frac{\alpha_1}{\beta_1^2} + \frac{\alpha_2}{\beta_2^2} \right)^{-2}. \end{aligned}$$

Из представления (2.6.22) вытекает, что двусторонние гамма-процессы Леви представляют собой частный случай так называемых CGMY-процессов, являющихся весьма популярными сейчас моделями динамики финансовых индексов. Название CGMY-процессов представляет собой аббревиатуру фамилий авторов статьи (Carr, Geman, Madan and Yor, 2002), в которой они были предложены как обобщение VG-процессов. CGMY-процессы Леви определяются спектральной плотностью

$$\ell_{CGMY}(x) = \frac{C}{x^{1+Y}} e^{-Mx} \mathbb{I}_{(0,\infty)}(x) + \frac{C}{|x|^{1+Y}} e^{-G|x|} \mathbb{I}_{(-\infty,0)}(x),$$

где C , G , M и Y – положительные параметры.

Процессы Мейкснера

ПРИМЕР 2.6.7. Распределение Мейкснера (Meixner distribution) $M(a, b, d, m)$ возникает в теории ортогональных полиномов. Оно определяется плотностью вероятностей

$$p_M(x; a, b, d, m) = \frac{(2 \cos(b/2))^{2d}}{2a\pi\Gamma(2d)} \exp \left\{ \frac{b(x-m)}{a} \right\} \left| \Gamma \left(d + \frac{i(x-m)}{a} \right) \right|^2, \quad x \in \mathbb{R},$$

где $a > 0$, $-\pi < b < \pi$, $d > 0$ и $m \in \mathbb{R}$. Пусть Z – случайная величина с распределением Мейкснера. У этой случайной величины существуют моменты всех порядков. При этом

$$\mathbb{E}Z = m + ad \operatorname{tg}(b/2), \quad \mathbb{D}Z = \frac{a^2 d}{\cos^2(b/2)},$$

$$\mathbb{E} \left(\frac{Z - \mathbb{E}Z}{\sqrt{\mathbb{D}Z}} \right)^4 = 3 + \frac{3 - 2 \cos^2(b/2)}{d}.$$

Распределение Мейкснера имеет “полутяжелые” хвосты в том смысле, что

$$p_M(x; a, b, d, m) \sim C_- |x|^{2d-1} \exp \left\{ - \frac{(\pi - b)|x|}{a} \right\}, \quad x \rightarrow -\infty,$$

$$p_M(x; a, b, d, m) \sim C_+ x^{2d-1} \exp \left\{ -\frac{(\pi + b)x}{a} \right\}, \quad x \rightarrow +\infty,$$

для некоторых положительных констант C_- и C_+ .

Характеристическая функция распределения Мейкснера имеет вид

$$\psi_M(s, a, b, d, m) = e^{ims} \left[\frac{\cos(b/2)}{\cosh(\frac{1}{2}(as - ib))} \right]^{2d}, \quad s \in \mathbb{R}.$$

Очевидно, эта характеристическая функция является безгранично делимой. Поэтому можно определить процесс Леви $Z(t)$ такой, что $Z(1) \stackrel{d}{=} Z$. Такой процесс называется *процессом Мейкснера*. При этом для любого $t > 0$ распределение случайной величины $Z(t)$ является распределением Мейкснера $M(a, b, dt, mt)$. Процессы Мейкснера введены в работе (Schoutens and Teugels, 1998). В статье (Grigelionis, 1999) указано, что процессы Мейкснера могут быть использованы в качестве модели динамики финансовых индексов.

В разложении Леви–Хинчина характеристической функции $\psi_M(s; a, b, d, m)$ отсутствует диффузионная (броуновская) компонента, что означает, что процесс Мейкснера – чисто скачкообразный. При этом величина и интенсивность скачков процесса Мейкснера определяются его спектральной плотностью Леви

$$\ell_M(x) = \frac{\exp\{bx/a\}d}{x \sinh(\pi x/a)}.$$

Свойства процессов Мейкснера и возможности их приложения к анализу финансовых данных обсуждаются в работах (Grigelionis, 1999), (Schoutens, 2000a), (Schoutens, 2000b), (Schoutens, 2003). В контексте данной книги одним из важнейших свойств процессов Мейкснера является возможность их представления в виде подчиненного процесса броуновского движения. А именно, в работе (Biane, Pitman and Yor, 2001) введен случайный процесс

$$B(t) = \frac{2}{\pi^2} \sum_{n=1}^{\infty} \frac{T_{1,1}^{(n)}(t)}{(n - \frac{1}{2})^2}, \quad t \geq 0,$$

где $\{T_{1,1}^{(n)}(t)\}_{n \geq 1}$ – последовательность независимых гамма-процессов Леви таких, что

$$T_{1,1}^{(n)}(t) \stackrel{d}{=} \Gamma(t, 1), \quad t > 0; \quad n = 1, 2, \dots$$

Как показано в упомянутой работе, при любом $t > 0$ преобразование Лапласа–Стилтьеса случайной величины $B(t)$ имеет вид

$$\mathbb{E} \exp\{-sB(t)\} = (\cosh \sqrt{2s})^{-t}, \quad s \geq 0.$$

Теперь, если $W_{0,1}(t)$ – стандартный винеровский процесс, независимый от процесса $B(t)$, то подчиненный процесс $W_{0,1}(B(t))$ является процессом Леви, определяемым характеристической функцией

$$\mathbb{E} \exp\{isW_{0,1}(B(t))\} = \left(\frac{1}{\cosh s}\right)^t, \quad s \in \mathbb{R},$$

то есть является процессом Мейкснера $Z(t)$ с распределением $M(2, 0, t, 0)$. Описанная конструкция приводит к процессам Мейкснера с симметричными распределениями, являющимися масштабными смесями нормальных законов.

Как отмечено в работе (Grigelionis, 2000), распределения Мейкснера являются специальными представителями семейства так называемых *обобщенных z -распределений*. Обобщенное z -распределение $GZ(a, b_1, b_2, d, m)$ определяется характеристической функцией

$$\psi_{GZ}(s; a, b_1, b_2, d, m) = e^{ims} \left[\frac{B(b_1 + \frac{ias}{2\pi}, b_2 - \frac{ias}{2\pi})}{B(b_1, b_2)} \right]^{2d}, \quad s \in \mathbb{R},$$

где a, b_1, b_2 и d положительны, $m \in \mathbb{R}$, а $B(\cdot, \cdot)$ – эйлерова бета-функция. При этом в силу замкнутости класса сдвиг-масштабных смесей нормальных законов относительно операции свертки (см. раздел 3.4) из результатов статьи (Barndorff-Nielsen, Kent and Sørensen, 1982), в которой показано, что любое z -распределение с характеристической функцией

$$\psi_Z(s; a, b_1, b_2, m) = e^{ims} \cdot \frac{B(b_1 + \frac{ias}{2\pi}, b_2 - \frac{ias}{2\pi})}{B(b_1, b_2)}, \quad s \in \mathbb{R},$$

может быть представлено в виде сдвиг-масштабной смеси нормальных законов, вытекает, что любое обобщенное z -распределение может быть представлено в виде сдвиг-масштабной смеси нормальных законов. Таким образом, одномерные распределения любого процесса Мейкснера имеют вид сдвиг-масштабных смесей нормальных законов.

Заметим, что классы обобщенных гиперболических распределений и обобщенных z -распределений не имеют общих представителей.

Другую информацию о возможностях применения процессов Леви со случайным временем в качестве моделей при решении задач финансовой

математики можно найти, например, в работах (Carr, Geman, Madan and Yor, 2002), (Carr, Geman, Madan and Yor, 2003), (Carr and Wu, 2004), (Schoutens, 2003), (Fu et al., 2007). В частности, в работе (Carr, Geman, Madan and Yor, 2003) в качестве моделей динамики финансовых индексов рассматриваются “дважды стохастические” процессы Леви типа NIG -, VG - или $CGMY$ -процессов *со случайным временем*. Процессы Мейкснера со случайным временем (“со стохастической волатильностью”) рассмотрены в книге (Schoutens, 2003).

В заключение данного раздела отметим, что описанная выше асимптотическая схема, в рамках которой осуществляется переход от обобщенных процессов Кокса как элементарных моделей на микромасштабах к процессам Леви со случайным временем как моделям изучаемых процессов на макромасштабах, является не более чем удобной иллюстрацией общего подхода, реализующего переход от обобщенных процессов Кокса к гораздо более общим и необязательно стационарным моделям типа броуновского движения со случайным временем. В таких моделях внешний процесс (броуновское движение) характеризует эволюцию рассматриваемой системы при условии ее консервативности (замкнутости). При этом истоки конкретного вида внешнего процесса – процесса броуновского движения – лежат во втором начале термодинамики, согласно которому энтропия в консервативных системах не убывает. Как известно, нормальное распределение приращений броуновского движения обладает максимальной энтропией среди всех распределений с неограниченным носителем и конечным вторым моментом. Случайное время – субординатор – характеризует влияние тех факторов, которые можно считать внешними по отношению к той системе, эволюция которой описывается броуновским движением. В реальных сложных системах внешние факторы очень сложно (практически невозможно) исчерпывающим образом моделировать и предсказывать. К примеру, по отношению к финансовым рынкам внешними факторами являются все политические новости, появление и значение которых, очевидно, нельзя предсказать с абсолютной точностью. Поэтому модели типа процессов Леви, в той или иной степени обладающие свойством стационарности приращений, нельзя считать всеобъемлющими моделями хаотических стохастических процессов, протекающих в сложных системах.

3.

Некоторые свойства смесей нормальных законов

Класс чисто масштабных смесей нормальных законов с нулевыми средними, которые возникают в качестве предельных законов в предельных теоремах, приведенных в предыдущих разделах, очень богат и содержит много типов распределений с существенно различным поведением хвостов. К примеру, помимо самого нормального распределения с очень быстро (как $|x|^{-1}e^{-x^2}$) убывающими (при $|x| \rightarrow \infty$) хвостами, этот класс содержит распределение Лапласа (также называемое двойным экспоненциальным распределением), хвосты которого убывают при $|x| \rightarrow \infty$ как $e^{-|x|}$, распределение Стюдента с хвостами, убывающими как $|x|^{-\gamma}$ с произвольным $\gamma > 0$, в частности, распределение Коши, соответствующее случаю $\gamma = 1$, симметричные строго устойчивые законы с таким же степенным убыванием хвостов с $\gamma \in (0, 2)$, симметризованные гамма-распределения, хвосты которых ведут себя как $|x|^a e^{-b|x|}$ с $-1 < a < \infty$, $b > 0$, симметризованное распределение Вейбулла–Гнеденко с хвостами, убывающими как $e^{-|x|^d}$ при $0 < d < 1$, а также много других типов, см., например, (Keilson and Steutel, 1974), (Королев, 2004).

В дальнейшем будут рассматриваться также другие специальные смеси нормальных законов, которые не обязательно будут чисто масштабными смесями, но также более общими сдвиг/масштабными смесями нормальных законов. В этом разделе будут описаны некоторые основные свойства нормальных смесей. Другие их свойства будут описаны ниже при построении СРС-метода.

3.1. Основные определения

Рубин хорошо знал, что понятие “народ” есть понятие вымышленное, есть неправомерное обобщение, что всякий народ делится на классы, и даже классы меняются со временем.

А. И. Солженицын. “В круге первом”.

В дальнейшем изложении будут интенсивно использоваться некоторые специальные свойства нормальных смесей. Чтобы систематически исследовать эти свойства, сначала надо напомнить строгое определение смеси вероятностных распределений.

Рассмотрим функцию $F(x, \mathbf{y})$, определенную на множестве $\mathbb{R} \times \mathbb{Y}$. Для простоты будем предполагать, что $\mathbb{Y}$ – это некоторое подмножество m -мерного евклидова пространства, $\mathbb{Y} \subseteq \mathbb{R}^m$ при некотором $m \geq 1$, причем множество $\mathbb{Y}$ снабжено борелевской σ -алгеброй Σ . Более того, предположим, что при каждом фиксированном $\mathbf{y}$ функция $F(x, \mathbf{y})$ является функцией распределения по x , а при каждом фиксированном x функция $F(x, \mathbf{y})$ измерима по $\mathbf{y}$, то есть для любых $x \in \mathbb{R}$ и $c \in \mathbb{R}$ выполнено условие $\{\mathbf{y} : F(x, \mathbf{y}) < c\} \in \Sigma$. Пусть $\mathbf{Q}$ – вероятностная мера, определенная на измеримом пространстве $(\mathbb{Y}, \Sigma)$. Функция распределения

$$H(x) = \int_{\mathbb{Y}} F(x, \mathbf{y}) \mathbf{Q}(d\mathbf{y}), \quad x \in \mathbb{R}, \quad (3.1.1)$$

называется *смесью* функции распределения $F(x, \mathbf{y})$ по $\mathbf{y}$ относительно $\mathbf{Q}$. Распределение $F(x, \mathbf{y})$ называется *смешиваемым*, в то время как мера $\mathbf{Q}$ задает *смешивающее* распределение. Если $\mathbf{Y}$ – m -мерная тождественная случайная величина (то есть $\mathbf{Y}(y) \equiv y, y \in \mathbb{Y}$), определенная на вероятностном пространстве $(\mathbb{Y}, \Sigma, \mathbf{Q})$, то функция распределения $H(x)$ может быть записана в виде

$$H(x) = \mathbb{E}F(x, \mathbf{Y}), \quad x \in \mathbb{R}.$$

Если $f(x, \mathbf{y})$ – плотность распределения, соответствующая функции распределения $F(x, \mathbf{y})$,

$$f(x, \mathbf{y}) = \frac{d}{dx} F(x, \mathbf{y}),$$

то смеси $H(x)$ соответствует плотность

$$h(x) = \mathbb{E}f(x, \mathbf{Y}) = \int_{\mathbb{Y}} f(x, \mathbf{y}) \mathbf{Q}(d\mathbf{y}), \quad x \in \mathbb{R}.$$

Если случайный вектор $\mathbf{Y}$ имеет дискретное распределение и принимает значения $\mathbf{y}_1, \mathbf{y}_2, \dots$ с вероятностями $p_1, p_2, \dots$, соответственно, то получается смесь вида

$$H(x) = \mathbf{E}F(x, \mathbf{Y}) = \sum_{j \geq 1} p_j F(x, \mathbf{y}_j), \quad x \in \mathbb{R},$$

называемая *дискретной*. В таком случае функции распределения $F(x, \mathbf{y}_j)$ называются *компонентами* смеси $H(x)$, а числа p_j называются *весами* соответствующих компонент, $j \geq 1$. Если в дискретной смеси число ненулевых весов конечно (то есть случайный вектор $\mathbf{Y}$ принимает конечное число значений), то дискретная смесь называется *конечной*.

Если в таком случае функции распределения $F(x, \mathbf{y})$ соответствует плотность $f(x, \mathbf{y})$, то дискретной смеси $H(x)$ соответствует плотность

$$h(x) = \mathbf{E}f(x, \mathbf{Y}) = \sum_{j \geq 1} p_j f(x, \mathbf{y}_j), \quad x \in \mathbb{R}. \quad (3.1.3)$$

Особо подчеркнем, что при разных значениях параметра $\mathbf{y}$ функции распределения $F(x, \mathbf{y})$ могут относиться к *разным* типам распределения. Например, если смесь дискретна, то есть мера $\mathbf{Q}$ приписывает вероятности p_j точкам $\mathbf{y}_j$, $j = 1, 2, \dots$, причем $F(x, \mathbf{y}_j) = F_j(x)$, где $F_j(x)$ – функции распределения, возможно, соответствующие разным типам при разных j , то

$$H(x) = \mathbf{E}F(x, \mathbf{Y}) = \sum_{j \geq 1} p_j F_j(x), \quad x \in \mathbb{R}.$$

Более того, если в таком случае компоненты F_j абсолютно непрерывны и имеют плотности f_j , то смесь $H(x)$ также будет абсолютно непрерывной с плотностью

$$h(x) = \mathbf{E}f(x, \mathbf{Y}) = \sum_{j \geq 1} p_j f_j(x), \quad x \in \mathbb{R}.$$

В дальнейшем особую роль будут играть сдвиг/масштабные смеси. Формально они определяются следующим образом. Пусть в определении, сформулированном выше, $m = 2$. Предположим, что вектор $\mathbf{y}$ имеет вид

$$\mathbf{y} = (u, v),$$

где $u > 0$ и $v \in \mathbb{R}$, так что функция распределения $F(x, \mathbf{y})$ допускает представление

$$F(x, \mathbf{y}) = F\left(\frac{x - v}{u}\right), \quad x \in \mathbb{R}.$$

Тогда $\mathbb{Y}$ – это положительная полуплоскость, то есть $\mathbb{Y} = \mathbb{R}^+ \times \mathbb{R}$, и функция распределения

$$H(x) = \int_{\mathbb{Y}} F\left(\frac{x-v}{u}\right) \mathbf{Q}(du, dv), \quad x \in \mathbb{R}, \quad (3.1.2)$$

называется *сдвиг/масштабной смесью* функции распределения F относительно $\mathbf{Q}$. Здесь u – это параметр масштаба, а v – параметр сдвига (положения). Если функция распределения F имеет плотность f , то функции распределения $F((x-v)/u)$ соответствует плотность

$$f(x, \mathbf{y}) = \frac{1}{u} f\left(\frac{x-v}{u}\right),$$

так что смеси (3.1.2) соответствует плотность

$$h(x) = \int_{\mathbb{Y}} \frac{1}{u} f\left(\frac{x-v}{u}\right) \mathbf{Q}(du, dv), \quad x \in \mathbb{R}.$$

Если X , U и V – случайные величины, заданные на одном и том же достаточно богатом вероятностном пространстве, так что при каждом фиксированном значении (u, v) пары случайных величин (U, V) случайная величина X имеет функцию распределения $F((x-v)/u)$, то смесь (3.1.2) может быть записана в виде

$$H(x) = \mathbf{E} F\left(\frac{x-V}{U}\right), \quad x \in \mathbb{R}.$$

Более того, в таком случае из теоремы Фубини вытекает, что функция распределения $H(x)$ соответствует случайной величине $X \cdot U + V$, где случайная величина X и случайный вектор $\mathbf{Y} = (U, V)$ стохастически независимы. Кстати, легко убедиться, что в таком контексте чисто сдвиговая смесь $H(x) = \mathbf{E} F(x - V)$ является не чем иным как функцией распределения суммы двух независимых случайных величин X и V , то есть сверткой их функций распределения. В то же время, чисто масштабная смесь $H(x) = \mathbf{E} F(x/U)$ является не чем иным как функцией распределения произведения двух независимых случайных величин X и U .

Чтобы определить дискретную сдвиг/масштабную смесь функции распределения $F(x)$, положим $\mathbf{y}_j = (\sigma_j, a_j)$, где $a_j \in \mathbb{R}$, $\sigma_j > 0$, $j = 1, \dots, k$, и в качестве специального случая приведенного выше определения дискретной смеси получим

$$H(x) = \sum_{j=1}^k p_j F\left(\frac{x-a_j}{\sigma_j}\right), \quad x \in \mathbb{R}. \quad (3.1.4)$$

Если при этом $a_j = 0$, $j = 1, \dots, k$, то мы получаем чисто масштабную конечную смесь

$$H(x) = \sum_{j=1}^k p_j F\left(\frac{x}{\sigma_j}\right), \quad x \in \mathbb{R}.$$

Если функция распределения F абсолютно непрерывна и имеет плотность $f = F'$, то смеси (3.1.4) функций распределения соответствует смесь плотностей

$$h(x) = \sum_{j=1}^k \frac{p_j}{\sigma_j} f\left(\frac{x - a_j}{\sigma_j}\right), \quad x \in \mathbb{R}.$$

Физический смысл понятия смеси вероятностных распределений может быть проиллюстрирован на примере дискретной смеси. Рассмотрим некоторую популяцию, которая не является однородной и, в свою очередь, состоит из некоторого числа, скажем, k субпопуляций. Предположим, что наблюдаемый признак (или наблюдаемая характеристика) внутри j -й субпопуляции распределен в соответствии с функцией распределения $F_j(x) \equiv F(x, \mathbf{y}_j)$, которую можно интерпретировать как условную вероятность того, что значение наблюдаемого признака у случайно выбранного индивидуума будет меньше, чем x , при условии, что случайно выбранный индивидуум является представителем j -й субпопуляции. Пусть вероятность того, что при случайном выборе индивидуума из всей (генеральной) популяции будет выбран представитель именно j -й субпопуляции, равна p_j ($p_j \geq 0$, $p_1 + \dots + p_k = 1$). Тогда по формуле полной вероятности безусловная вероятность того, что значение наблюдаемого признака у индивидуума, случайно выбранного из всей генеральной популяции, будет меньше, чем x , окажется равной

$$H(x) = \sum_{j=1}^k p_j F_j(x).$$

В определенном смысле операция смешивания вероятностных распределений обеспечивает возможность формально интерпретировать популяции, реально являющиеся неоднородными, как однородные.

Очень часто нельзя непосредственно определить тип субпопуляции, к которой принадлежит очередное наблюдение. Вследствие этого вся (генеральная) популяция вынужденно считается однородной, хотя на самом деле она таковой не является и содержит индивидуумов, принадлежащих к существенно различным типам. Именно такая ситуация типична для

анализа финансовых временных рядов или процессов плазменной турбулентности. Поэтому чрезвычайно важно иметь возможность осуществить операцию, в некотором смысле обратную операции смешивания, а именно, операцию разделения (расщепления) смесей. Статистические процедуры, реализующие эту операцию, описываются ниже. Эти процедуры в значительной степени зависят от свойства идентифицируемости смесей вероятностных распределений.

3.2. Идентифицируемость смесей вероятностных распределений

Понятие идентифицируемой смеси интенсивно используется в прикладных задачах, связанных с декомпозицией (разделением, разложением, расщеплением) совокупностей (популяций). В качестве примеров можно упомянуть задачи классификации, распознавания образов или идентификации вероятностных распределений. Библиография по этим вопросам обширна: см., например, обзоры (Исаенко и Урбах, 1976), (Круглов, 1991) или книги (Titterton, Smith and Makov, 1987), (Айвазян, Бухштабер, Енюков и Мешалкин, 1989), (Prakasa Rao, 1992) и списки литературы в указанных источниках.

Напомним определение идентифицируемых семейств смесей распределений вероятностей. Оно было предложено в работе (Teicher, 1961).

Пусть функция $F(x, \mathbf{y})$ определена на множестве $\mathbb{R} \times \mathbb{Y}$. Для простоты предположим, что $\mathbb{Y} \subseteq \mathbb{R}^m$ при некотором $m \geq 1$ и множество $\mathbb{Y}$ снабжено борелевской σ -алгеброй Σ . Как и ранее, будем предполагать, что функция $F(x, \mathbf{y})$ измерима по $\mathbf{y}$ при каждом фиксированном x и является функцией распределения как функция аргумента x при каждом фиксированном $\mathbf{y}$. Пусть $\mathcal{Q}$ – семейство случайных величин, принимающих значения во множестве $\mathbb{Y}$. Обозначим

$$\mathcal{H} = \{H_Q(x) = EF(x, Q), x \in \mathbb{R} : Q \in \mathcal{Q}\}. \quad (3.2.1)$$

Семейство $\mathcal{H}$, определяемое ядром F и множеством $\mathcal{Q}$, называется *идентифицируемым*, если из равенства

$$EF(x, Q_1) = EF(x, Q_2), \quad x \in \mathbb{R},$$

с $Q_1 \in \mathcal{Q}$, $Q_2 \in \mathcal{Q}$ вытекает, что $Q_1 \stackrel{d}{=} Q_2$ (здесь и далее символ $\stackrel{d}{=}$ обозначает равенство по распределению, то есть совпадение распределений).

К примеру, идентифицируемыми являются конечные смеси нормальных распределений, показательных распределений, пуассоновских распределений и распределений Коши. Однако легко привести очень простые примеры неидентифицируемых семейств. В частности, в качестве ядра рассмотрим равномерное распределение и связанные с ним смеси

$$\frac{1}{3} \cdot 3 \cdot \mathbb{I}_{[0, \frac{1}{3})}(x) + \frac{2}{3} \cdot \frac{3}{2} \cdot \mathbb{I}_{[\frac{1}{3}, 1)}(x) = \frac{1}{2} \cdot 2 \cdot \mathbb{I}_{[0, \frac{1}{2})}(x) + \frac{1}{2} \cdot 2 \cdot \mathbb{I}_{[\frac{1}{2}, 1)}(x).$$

Здесь

$$\mathbb{I}_{[a, b)}(x) = \begin{cases} 1, & \text{если } x \in [a, b), \\ 0, & \text{если } x \notin [a, b). \end{cases}$$

В данной работе главным образом рассматриваются сдвиг/масштабные смеси, в которых $\mathbf{y} = (u, v)$, $F(x, \mathbf{y}) = F((x - v)/u)$. Поэтому все, что на самом деле нам нужно, – это результаты об идентифицируемости семейств сдвиг/масштабных смесей одномерных распределений.

Если X , U и V – случайные величины, определенные на одном и том же достаточно богатом вероятностном пространстве так, что (i) случайная величина X стохастически независима от пары (U, V) и (ii) для любых фиксированных значений u случайной величины U и v случайной величины V случайная величина X имеет функцию распределения $F((x - v)/u)$, то приведенное выше определение идентифицируемости применительно к сдвиг/масштабным смесям сводится к следующему. Сдвиг/масштабная смесь $H(x) = \mathbb{E}F((x - V)/U)$, порожденная ядром F (сдвиг/масштабная смесь функции распределения F) идентифицируема, если из соотношения

$$X_1 \cdot U_1 + V_1 \stackrel{d}{=} X_2 \cdot U_2 + V_2,$$

где (X_i, U_i, V_i) , $i = 1, 2$, – тройки случайных величин, обладающих точно такими же свойствами, что присущи описанным выше случайным величинам X, U, V , вытекает, что

$$(U_1, V_1) \stackrel{d}{=} (U_2, V_2).$$

Сужение определения идентифицируемости на класс конечных сдвиг/масштабных смесей сводится к следующему.

Семейство смесей

$$\mathcal{H} = \left\{ \sum_{j=1}^k p_j F\left(\frac{x - a_j}{\sigma_j}\right) : \right.$$

$$k \geq 1; p_j \geq 0, p_1 + \dots + p_k = 1; a_j \in \mathbb{R}, \sigma_j > 0, j = \overline{1, k} \Big\}$$

порожденное ядром F , идентифицируемо, если из равенства

$$\sum_{j=1}^k p_j F\left(\frac{x - a_j}{\sigma_j}\right) = \sum_{i=1}^m q_i F\left(\frac{x - b_i}{\delta_i}\right)$$

вытекает, что

- (i) $k = m$;
- (ii) для каждого индекса $j \in \{1, \dots, k\}$ существует индекс $i \in \{1, \dots, k\}$ такой, что

$$p_j = q_i, \quad a_j = b_i, \quad \sigma_j = \delta_i.$$

Теперь перейдем к рассмотрению условий идентифицируемости. Сначала рассмотрим условия идентифицируемости смесей, в которых смешивание производится либо по параметру сдвига, либо по параметру масштаба. Другими словами, сначала рассмотрим семейства однопараметрических смесей. Условия идентифицируемости таких семейств хорошо известны. Напомним некоторые из них.

Семейство функций распределения $\{F(x, y) : y > 0\}$ называется *аддитивно замкнутым*, если для любых $y_1 > 0, y_2 > 0$ справедливо соотношение

$$F(x, y_1) * F(x, y_2) \equiv F(x, y_1 + y_2). \quad (3.2.2)$$

Здесь символ $*$ обозначает свертку распределений: если F_1 и F_2 – функции распределения, то

$$F_1(x) * F_2(x) = \int_{-\infty}^{\infty} F_1(x - y) dF_2(y) = \int_{-\infty}^{\infty} F_2(x - y) dF_1(y).$$

Иногда свойство (3.2.2) семейств распределений вероятностей называется *воспроизводимостью* по параметру y .

Семейство нормальных законов с нулевым математическим ожиданием $\mathcal{N}_0 = \{\Phi(x/\sqrt{s}), s > 0\}$ является очевидным примером аддитивно замкнутого семейства (относительно дисперсии s).

Следующие результаты принадлежат Г. Тейчеру (Teicher, 1961).

ТЕОРЕМА 3.2.1. Семейство смесей (3.2.1) функций распределения $F(x, \cdot)$ из аддитивно замкнутого семейства является идентифицируемым.

Отсюда немедленно вытекает, что семейство $\mathcal{N}_0$ масштабных смесей нормальных законов с нулевым средним идентифицируемо.

Смеси, порождаемые ядрами из аддитивно замкнутых семейств, конечно же, не исчерпывают все примеры идентифицируемых смесей. Рассмотрим масштабные смеси распределений, сосредоточенных на неотрицательной полупрямой.

ТЕОРЕМА 3.2.2. Пусть $F(x, y) = F_0(xy)$, $y \geq 0$, где F_0 – функция распределения такая, что $F_0(0) = 0$. Предположим, что преобразование Фурье функции $w(y) = F_0(e^y)$, $y \geq 0$, нигде не обращается в нуль. Пусть $\mathcal{Q}^+$ – множество всех случайных величин Q таких, что $P(Q > 0) = 1$. Тогда семейство смесей

$$\mathcal{H} = \{H_Q(x) = EF(x, Q), x \geq 0 : Q \in \mathcal{Q}^+\}$$

идентифицируемо.

Аналогичное свойство присуще некоторым семействам сдвиговых смесей распределений вероятностей.

ТЕОРЕМА 3.2.3. Пусть $\mathcal{Q}$ – множество всех случайных величин. Семейство сдвиговых смесей

$$\mathcal{H} = \{H_Q(x) = EF(x - Q), x \in \mathbb{R} : Q \in \mathcal{Q}\},$$

идентифицируемо, если характеристическая функция, соответствующая функции распределения $F(x)$, нигде не обращается в нуль.

Теорема 3.2.1 гарантирует, что наряду со смесями нормальных законов с нулевым средним, идентифицируемыми являются семейства масштабных смесей любых строго устойчивых законов.

Теорема 3.2.3 гарантирует идентифицируемость семейств сдвиговых смесей любых безгранично делимых (в том числе устойчивых) законов.

Здесь упомянуты только те идентифицируемые семейства, которые так или иначе рассматриваются в данной работе в рамках описания СРС-метода. Многочисленные примеры других идентифицируемых семейств можно найти в работах (Medgyessy, 1961), (Teicher, 1961), (Teicher, 1963), (Yakowitz and Spragins, 1968).

Некоторые критерии (то есть необходимые и достаточные условия идентифицируемости семейств смесей доказаны в работе (Tallis, 1969), также см. обзор (Круглов, 1991). Итоги исследований в области идентифицируемости семейств смесей вероятностных распределений подведены

в обстоятельной книге (Prakasa Rao, 1992), где подробно обсуждаются как условия идентифицируемости, так и соответствующие примеры.

К сожалению, примеры ядер, порождающих идентифицируемые семейства сдвиг/масштабных смесей в общей ситуации (то есть без каких-либо дополнительных условий на смешивающие распределения) не известны. Такие примеры известны лишь для дискретных смешивающих законов. В частности, справедливо следующее утверждение, доказанное Г. Тейчером (Teicher, 1963).

ТЕОРЕМА 3.2.4. *Семейство конечных сдвиг/масштабных смесей нормальных законов идентифицируемо.*

Покажем, что семейство сдвиг/масштабных смесей нормальных законов при произвольном (двумерном) смешивающем законе не является идентифицируемым. Пусть X_1 , X_2 , U_1 и U_2 – независимые случайные величины такие, что X_1 и X_2 имеют одно и то же стандартное нормальное распределение, $P(U_1 = 1) = 1$, а случайная величина U_2 не вырождена, то есть ни при каком u не имеет места соотношение $P(U_2 = u) = 1$. Более того, предположим, что $P(U_2 > 0) = 1$. Тогда распределение случайной величины $Z = X_1 U_2 + X_2$ может быть записано как в виде

$$P(Z < x) = \int_0^{\infty} \int_{-\infty}^{\infty} \Phi\left(\frac{x-v}{u}\right) dP(X_2 < v) dP(U_2 < u), \quad (3.2.3)$$

так и в виде

$$P(Z < x) = \int_0^{\infty} \int_{-\infty}^{\infty} \Phi\left(\frac{x-v}{u}\right) dP(X_1 U_2 < v) dP(U_1 < u). \quad (3.2.4)$$

Легко видеть, что смешивающие двумерные распределения в представлениях (3.2.3) и (3.2.4) различны. Таким образом, можно заключить, что в общем случае задача разделения сдвиг/масштабных смесей, то есть задача реконструкции двумерного смешивающего распределения по распределению смеси является некорректной, так как она может иметь несколько различных решений. Это справедливо даже для случая сдвиг/масштабных смесей нормальных законов. Идея, лежащая в основе описываемого СРС-метода, заключается в замене общей некорректно поставленной задачи аналогичной очень близкой корректной (см. теорему 3.2.4) задачей разделения конечных сдвиг/масштабных смесей нормальных законов. Этот подход, в свою очередь, основан на том обстоятельстве, что любое распределение вероятностей может быть как угодно точно приближено конечным дискретным распределением.

В терминах примера с популяцией, состоящей из нескольких (возможно, бесконечно большого числа) субпопуляций, идея данного подхода может быть пояснена следующим образом. Область значений наблюдаемых характеристик (возможно, континуальная) разбивается на счетное (возможно, конечное) множество групп так, что каждая группа соответствует близким (в определенном смысле) субпопуляциям. Другими словами, все субпопуляции разбиваются на счетное (возможно, конечное) число классов (кластеров). Таким образом, все субпопуляции, попавшие в один класс (кластер) считаются неразличимыми. Другими словами, выделяется счетное (возможно, конечное) множество *типических* субпопуляций. К примеру, все люди, населяющие нашу планету, различны. Более того, численность населения Земли такова, что эта общая популяция может считаться континуальной (каких бы двух людей с характеризующими их признаками мы ни взяли, практически всегда найдется третий человек с промежуточными признаками). Однако, как известно, все люди подразделяются на конечное число рас.

Идея замены исходной некорректной задачи корректной задачей разделения конечных смесей будет работать на практике, только если решение “суженной” корректной задачи будет близким к множеству решений исходной общей задачи. Другими словами, для того чтобы такой подход привел к разумным результатам, требуется, чтобы семейство сдвиг/масштабных смесей обладало свойством устойчивости относительно смешивающего распределения. Это свойство применительно к смесям нормальных законов рассматривается в следующем разделе.

3.3. Устойчивость нормальных смесей относительно смешивающего распределения

В этом разделе будет показано, что семейство масштабных смесей нормальных законов обладает свойством устойчивости относительно смешивающего распределения. Несмотря на то что задача разделения общих масштабных смесей нормальных законов имеет единственное решение (см. теорему 3.2.1), на практике невозможно ее решить с абсолютной точностью, поскольку сама исходная смесь не известна. Даны лишь результаты наблюдений, о которых известно, что они представляют собой реализации случайных величин, которые распределены в соответствии с этой неизвестной смесью. На практике задача статистического разделения смесей совпадает с задачей статистической реконструкции самой смеси. Таким

образом, свойство устойчивости смесей относительно смешивающего распределения становится решающим для реалистичности и адекватности результатов статистического анализа, получаемых с помощью СРС-метода.

3.3.1. Прямая задача

Пусть X_i и Y_i – независимые случайные величины и пусть, более того, $P(Y_i \geq 0) = 1$, $i = 1, 2$. Функции распределения случайных величин X_i и Y_i будут, соответственно, обозначаться $F_i(x)$ и $G_i(x)$. Равномерное расстояние между случайными величинами X и Y с соответствующими функциями распределения $F_X(x)$ и $F_Y(x)$ будет обозначаться $\rho(X, Y)$,

$$\rho(X, Y) = \sup_x |F_X(x) - F_Y(x)|.$$

Наша ближайшая цель – оценить величину

$$\Delta = \rho\left(\frac{X_1}{Y_1}, \frac{X_2}{Y_2}\right)$$

через $\rho(X_1, X_2)$ и $\rho(Y_1, Y_2)$. Обозначим

$$\Delta(x) \equiv \left| P\left(\frac{X_1}{Y_1} < x\right) - P\left(\frac{X_2}{Y_2} < x\right) \right|.$$

Легко видеть, что для $-\infty < x < +\infty$ справедливо соотношение

$$\begin{aligned} P\left(\frac{X}{Y} < x\right) &= P(X < 0)P(Y = 0) + P(X < xY; Y > 0) = \\ &= F_X(0)[F_Y(0+) - F_Y(0)] + \int_{0+}^{\infty} F_X(xy) dF_Y(y). \end{aligned}$$

Поэтому

$$\begin{aligned} \Delta(x) &\leq F_1(0)[G_1(0+) - G_1(0)] + F_2(0)[G_2(0+) - G_2(0)] + \\ &+ \left| \int_{0+}^{\infty} F_1(xy) dG_1(y) - \int_{0+}^{\infty} F_2(xy) dG_2(y) \right| \equiv Q_1 + Q_2. \end{aligned}$$

Очевидно,

$$Q_1 = P(X_1 < 0)P(Y_1 = 0) + P(X_2 < 0)P(Y_2 = 0). \quad (3.3.1)$$

Оценим величину Q_2 следующим образом:

$$Q_2 \leq \left| \int_{0+}^{\infty} [F_1(xy) - F_2(xy)] dG_1(y) \right| + \left| \int_{0+}^{\infty} F_2(xy) d[G_1(y) - G_2(y)] \right| \equiv Q_{21} + Q_{22}.$$

Имеем

$$Q_{21} \leq \rho(X_1, X_2) \int_{0+}^{\infty} dG_1(y) = \rho(X_1, X_2) P(Y_1 > 0). \quad (3.3.2)$$

Если $x = 0$, то $\Delta(x) = 0$. Поэтому в дальнейшем будем предполагать, что $x \neq 0$. Рассмотрим Q_{22} . Интегрируя по частям, получаем

$$\begin{aligned} Q_{22} &= \left| F_2(xy) [G_1(y) - G_2(y)] \Big|_{y=0+}^{\infty} - \int_{0+}^{\infty} [G_1(y) - G_2(y)] d_y F_2(xy) \right| \leq \\ &\leq F_2(xy) [G_1(y) - G_2(y)] \Big|_{y=0+}^{\infty} + \left| \int_{0+}^{\infty} [G_1(y) - G_2(y)] d_y F_2(xy) \right| \equiv Q_{221} + Q_{222}. \end{aligned}$$

Очевидно,

$$Q_{221} \leq F_2(0+) |G_1(0+) - G_2(0+)| = P(X_2 \leq 0) |P(Y_1 = 0) - P(Y_2 = 0)|. \quad (3.3.3)$$

В то же время

$$Q_{222} \leq \rho(Y_1, Y_2) \left| \int_{0+}^{\infty} d_y F_2(xy) \right| = \rho(Y_1, Y_2) I_2(x), \quad (3.3.4)$$

где

$$I_2(x) = \begin{cases} P(X_2 > 0), & \text{если } x > 0, \\ P(X_2 < 0), & \text{если } x < 0. \end{cases} \quad (3.3.5)$$

Объединяя (3.3.1), (3.3.2), (3.3.3) и (3.3.4), мы приходим к следующему утверждению.

ЛЕММА 3.3.1. *При указанных выше условиях, налагаемых на случайные величины X_1 , X_2 , Y_1 и Y_2 справедлива оценка*

$$\begin{aligned} \Delta(x) &\leq P(X_1 < 0) P(Y_1 = 0) + P(X_2 < 0) P(Y_2 = 0) + \rho(X_1, X_2) P(Y_1 > 0) + \\ &+ P(X_2 \leq 0) |P(Y_1 = 0) - P(Y_2 = 0)| + \rho(Y_1, Y_2) I_2(x), \end{aligned}$$

где функция $I_2(x)$ определена в (3.3.5).

Поскольку в приведенных рассуждениях случайные величины X_1 и X_2 играют абсолютно равносильные роли, из леммы 3.3.1 можно получить следующее утверждение.

СЛЕДСТВИЕ 3.3.1. *Если в дополнение к условиям леммы 3.3.1 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$, то*

$$\Delta \leq \rho(X_1, X_2) + \rho(Y_1, Y_2) \min_{i=1,2} \max \{P(X_i < 0), P(X_i > 0)\}.$$

СЛЕДСТВИЕ 3.3.2. *Если в дополнение к условиям леммы 3.3.1 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$ и по крайней мере одна из случайных величин X_i , $i = 1, 2$, удовлетворяет условию $P(X_i < 0) = P(X_i > 0)$, то*

$$\Delta \leq \rho(X_1, X_2) + \frac{1}{2}\rho(Y_1, Y_2).$$

И конечно же, простейшим следствием из леммы 3.3.1 является

СЛЕДСТВИЕ 3.3.3. *Если в дополнение к условиям леммы 3.3.1 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$, то*

$$\Delta \leq \rho(X_1, X_2) + \rho(Y_1, Y_2).$$

Так как в случае $P(Y_1 = 0) = P(Y_2 = 0) = 0$ справедливы соотношения

$$\begin{aligned} \rho(Y_1^{-1}, Y_2^{-1}) &= \sup_x \left| P\left(\frac{1}{Y_1} < x\right) - P\left(\frac{1}{Y_2} < x\right) \right| = \\ &= \sup_x \left| P\left(Y_1 > \frac{1}{x}\right) - P\left(Y_2 > \frac{1}{x}\right) \right| = \\ &= \sup_x \left| \left[1 - P\left(Y_1 \leq \frac{1}{x}\right)\right] - \left[1 - P\left(Y_2 \leq \frac{1}{x}\right)\right] \right| = \\ &= \sup_z \left| P(Y_1 < z) - P(Y_2 < z) \right| = \rho(Y_1, Y_2), \end{aligned}$$

то из леммы 3.3.1 также можно получить следующие оценки равномерного расстояния между произведениями независимых случайных величин.

СЛЕДСТВИЕ 3.3.4. *Если в дополнение к условиям леммы 3.3.1 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$, то*

$$\rho(X_1 Y_1, X_2 Y_2) \leq \rho(X_1, X_2) + \rho(Y_1, Y_2) \min_{i=1,2} \max \{P(X_i < 0), P(X_i > 0)\}.$$

СЛЕДСТВИЕ 3.3.5. Если в дополнение к условиям леммы 3.3.1 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$ и, по крайней мере, одна из случайных величин X_i , $i = 1, 2$, удовлетворяет условию $P(X_i < 0) = P(X_i > 0)$, то

$$\rho(X_1 Y_1, X_2 Y_2) \leq \rho(X_1, X_2) + \frac{1}{2} \rho(Y_1, Y_2).$$

СЛЕДСТВИЕ 3.3.6. Если в дополнение к условиям леммы 3.3.1 выполнено условие $P(Y_1 = 0) = P(Y_2 = 0) = 0$, то

$$\rho(X_1 Y_1, X_2 Y_2) \leq \rho(X_1, X_2) + \rho(Y_1, Y_2).$$

Лемма 3.3.1 и следствия 3.3.1 – 3.3.6 содержат простые оценки близости масштабных смесей в терминах расстояний между соответствующими смешиваемыми и смешивающими распределениями.

Основная идея регуляризации общей некорректно поставленной задачи разделения смеси (на практике не известной точно) нормальных законов основана на замене неизвестного “истинного” смешивающего распределения его конечной дискретной аппроксимацией.

Рассмотрим некоторые вопросы, связанные с оцениванием потери точности, сопутствующей такой подмене, на примере масштабных смесей нормальных законов.

Пусть Y_1 и Y_2 – некоторые неотрицательные случайные величины. Из следствия 3.3.5 вытекает, что в условиях леммы 3.3.1 справедлива оценка

$$\sup_{x \in \mathbb{R}} |E\Phi(x \cdot \sqrt{Y_1}) - E\Phi(x \cdot \sqrt{Y_2})| \leq \frac{1}{2} \sup_{y > 0} |P(Y_1 < y) - P(Y_2 < y)|.$$

Таким образом, если известна оценка равномерного расстояния между “истинным” смешивающим распределением и его конечной дискретной аппроксимацией, то оценка равномерного расстояния между соответствующими “истинной” и приближенной смесью оказывается вдвое меньшей.

3.3.2. Обратная задача

На самом деле намного интереснее найти решение задачи, которая является в некотором смысле обратной к только что рассмотренной: по заданному расстоянию (например, равномерному) между смесями (которое может быть оценено статистически, например, с использованием модификации методов оценивания квантильных мер риска (VaR), основанных на статистике Колмогорова–Смирнова) найти оценку расстояния между

“истинным” и статистически подогнанным смешивающим распределениями. Один из возможных подходов к решению этой задачи таков. Расстояние между смесями может быть охарактеризовано расстоянием между соответствующими характеристическими функциями. Характеристическая функция $f_j(t)$, соответствующая смешанной функции распределения $E\Phi(x \cdot \sqrt{Y_j})$, имеет вид

$$f_j(t) = E \exp\{-\frac{1}{2}t^2 Y_j\}, \quad t \in \mathbb{R},$$

$j = 1, 2$. В правой части этого соотношения стоит преобразование Лапласа–Стилтьеса $\psi_j(s) = E e^{-sY_j}$ случайной величины Y_j в точке $s = \frac{1}{2}t^2$, $j = 1, 2$. Как известно, преобразование Лапласа–Стилтьеса однозначно определяет распределение неотрицательной случайной величины. Таким образом, расстояние (невязка) между смесями связано с “расстоянием” между смешивающими распределениями соотношением

$$|f_1(t) - f_2(t)| = |\psi_1(\frac{1}{2}t^2) - \psi_2(\frac{1}{2}t^2)|, \quad t \in \mathbb{R}.$$

Более того, по формуле обращения преобразования Лапласа–Стилтьеса (см., например, (Феллер, 1984), том 2) имеем

$$P(Y_j < x) = \lim_{t \rightarrow \infty} \sum_{n \leq \sqrt{2tx}} \frac{(-1)^n (2t)^{n/2}}{n!} f_j^{(n)}(\sqrt{2t}), \quad j = 1, 2.$$

Поэтому

$$|P(Y_1 < x) - P(Y_2 < x)| \leq \lim_{t \rightarrow \infty} \sum_{n \leq \sqrt{2tx}} \frac{(2t)^{n/2}}{n!} |[f_1(\sqrt{2t}) - f_2(\sqrt{2t})]^{(n)}|,$$

где, как и ранее, $f_j(t)$ – характеристическая функция, соответствующая смешанной функции распределения $E\Phi(x \cdot \sqrt{Y_j})$, $j = 1, 2$.

Как видно из последнего соотношения, для получения приемлемой оценки равномерного расстояния между смешивающими распределениями при таком подходе нужна существенно более подробная информация о расстоянии между смесями, нежели значение равномерной метрики. Вообще говоря, равномерное расстояние – “неправильная” метрика для решения обсуждаемой “обратной” задачи во всей ее полноте. Действительно, рассмотрим следующий пример.

Пусть $\epsilon > 0$ – произвольно малое положительное число, U_1 и U_2 – две “случайные” величины, имеющие вырожденные распределения, соответственно, в точках 1 и $1 + \epsilon'$, где $\epsilon' = \epsilon\sqrt{2\pi e}$. Очевидно, что при этом по

формуле Лагранжа

$$\begin{aligned} \left| \mathbb{E}\Phi\left(\frac{x}{U_1}\right) - \mathbb{E}\Phi\left(\frac{x}{U_2}\right) \right| &= \left| \Phi(x) - \Phi\left(\frac{x}{1+\epsilon'}\right) \right| = \\ &= \left| \frac{x\epsilon'}{1+\epsilon'} \phi\left(x\left[\delta + (1-\delta)\frac{1}{1+\epsilon'}\right]\right) \right| \end{aligned}$$

для некоторого $\delta \in [0, 1]$, где $\phi(x)$ – стандартная нормальная плотность. Легко видеть, что

$$\delta + (1-\delta)\frac{1}{1+\epsilon'} = \frac{\delta\epsilon' + 1}{1+\epsilon'} \geq 1.$$

Поэтому

$$\left| \mathbb{E}\Phi\left(\frac{x}{U_1}\right) - \mathbb{E}\Phi\left(\frac{x}{U_2}\right) \right| \leq \epsilon' \sup_x |x| \phi(x) = \frac{\epsilon'}{\sqrt{2\pi e}} = \epsilon,$$

в то время как

$$\rho(U_1, U_2) = 1.$$

По-видимому, для решения “обратной” задачи в общем случае следует использовать метрики, метризирующие слабую сходимость, например, метрику Леви или сглаженное равномерное расстояние (см. раздел 2.4).

Напомним определение метрики Леви. Пусть X и Y – две случайные величины с функциями распределения, соответственно, $F(x)$ и $G(x)$, $x \in \mathbb{R}$. Расстояние (метрика) Леви $L(X, Y)$ между случайными величинами X и Y определяется как

$$L(X, Y) = \inf\{h : G(x-h) - h \leq F(x) \leq G(x+h) + h \text{ для всех } x \in \mathbb{R}\}.$$

Расстояние Леви допускает наглядную геометрическую интерпретацию: оно равно длине стороны наибольшего квадрата, который можно вписать между графиками функций распределения $F(x)$ и $G(x)$ так, чтобы его стороны были параллельны координатным осям. Так как в определении метрики Леви задействованы не сами случайные величины X и Y , а только их функции распределения $F(x)$ и $G(x)$, то выражения $L(X, Y)$ и $L(F, G)$ не будут различаться:

$$L(X, Y) \equiv L(F, G).$$

Для равномерного расстояния между функциями распределения $F(x)$ и $G(x)$, как и ранее, будем использовать обозначение $\rho(X, Y) \equiv \rho(F, G)$,

$$\rho(X, Y) \equiv \rho(F, G) = \sup_x |F(x) - G(x)|.$$

По крайней мере, если рассматривается задача оценки устойчивости идентифицируемости *конечных* смесей нормальных законов, то в терминах метрики Леви можно получить вполне наглядные результаты.

Проиллюстрируем сказанное на примере обратной задачи оценивания устойчивости нормального закона по отношению к так называемому “загрязнению” (контаминации). А именно, рассмотрим схему Дж. Тьюки, в рамках которой вместо “чистого” нормального распределения $\Phi(x)$ имеется масштабная смесь нормальных законов

$$F_{p,\sigma}(x) = p\Phi(x) + (1-p)\Phi(\sigma x), \quad x \in \mathbb{R},$$

где $p \in (0, 1]$, $\sigma \in (0, 1]$ (Tukey, 1960). Если выборка представляет собой набор независимых реализаций случайной величины с функцией распределения $F_{p,\sigma}(x)$, то компонента $\Phi(\sigma x)$ смеси $F_{p,\sigma}(x)$ соответствует “загрязняющим” (нетипичным) наблюдениям, искажающим статистические выводы, полученные в предположении “чистой” нормальности распределения выборки.

Несложно видеть, что смесь $F_{p,\sigma}(x)$ – это функция распределения произведения стандартно нормально распределенной случайной величины Z и независимой от нее случайной величины $W_{p,\sigma}$, принимающей значения 1 и σ^{-1} с вероятностями p и $(1-p)$, соответственно. Обозначим $U_{p,\sigma} = W_{p,\sigma}^{-1}$.

Обратная задача оценивания устойчивости нормального закона по отношению к “загрязнению” может быть сформулирована следующим образом. Пусть $\epsilon > 0$ – произвольно малое число. Предположим, что

$$L(F_{p,\sigma}, \Phi) \leq \epsilon. \quad (3.3.6)$$

Что можно сказать о том, насколько близкими к единице в таком случае должны быть параметры p и σ ? Другими словами, насколько близким к вырожденному в единице должно быть распределение случайной величины $U_{p,\sigma}$?

Следует отметить, что “прямая” задача оценивания устойчивости нормального закона по отношению к “загрязнению” в схеме Тьюки является частным случаем задачи оценивания расстояния между масштабной смесью и порождающим ее смешиваемым законом. Пусть $U \geq 0$ – случайная величина. Пусть $F(x)$ – некоторая функция распределения. Задача оценивания расстояния между масштабной смесью и порождающим ее смешиваемым законом заключается в оценивании расстояния между функциями распределения $F(x)$ и $EF(U^{-1}x)$ в терминах расстояния между распределением случайной величины U и распределением, вырожденным в единице. Из результатов на эту тему для $F = \Phi$ можно выделить следующие.

ТЕОРЕМА 3.3.1. Пусть $E \max\{U, U^{-1}\} < \infty$. Тогда

$$\rho(\Phi(x), E\Phi(xU^{-1})) \leq 0.242 E(\max\{U, U^{-1}\} - 1).$$

Д о к а з а т е л ь с т в о. По формуле Лагранжа для некоторого $\delta = \delta(U) \in [0, 1]$ имеем

$$\begin{aligned} \Phi(x) - E\Phi(xU^{-1}) &= E[\Phi(x) - \Phi(xU^{-1})] = E[x(1 - U^{-1})\phi(x(\delta + (1 - \delta)U^{-1}))] \leq \\ &\leq E[|1 - U^{-1}| \sup_{x \geq 0} x\phi(x(\delta + (1 - \delta)U^{-1}))] = \\ &= \frac{1}{\sqrt{2\pi e}} E[|1 - U^{-1}| U(U\delta + 1 - \delta)^{-1}] \leq \frac{1}{\sqrt{2\pi e}} E(\max\{U, U^{-1}\} - 1), \end{aligned}$$

что и требовалось доказать.

ТЕОРЕМА 3.3.2. Пусть $EU^{-2} < \infty$. Тогда

$$\rho(\Phi(x), E\Phi(xU^{-1})) \leq E|U^{-2} - 1|.$$

Д о к а з а т е л ь с т в о см. в работе (Shaked, 1981).

То обстоятельство, что расстояние между представителями семейства масштабных смесей нормальных законов можно естественно охарактеризовать с помощью моментов смешивающих величин, по-видимому, впервые отмечено в статье (Keilson and Steutel, 1974). Пусть X – случайная величина со стандартной нормальной функцией распределения Φ . Тогда функция распределения $E\Phi(U^{-1}x)$ соответствует случайной величине $Z = X \cdot U$, где сомножители независимы. В работе (Keilson and Steutel, 1974), в частности, замечено, что близость представителя подмножества указанного семейства с $EU^4 < \infty$ к “чистому” нормальному закону удобно характеризовать с помощью величины

$$\gamma(Z) = \kappa(Z) - 3,$$

где

$$\kappa(Z) = E\left(\frac{Z - EZ}{\sqrt{DZ}}\right)^4$$

– коэффициент эксцесса случайной величины Z , $\kappa(X) = 3$ (см. раздел 2.3). Справедлив следующий результат.

ТЕОРЕМА 3.3.3. Пусть $EU^2 = 1$ и $EU^4 < \infty$. Тогда

$$\rho(\Phi(x), E\Phi(xU^{-1})) \leq 0.648 \gamma(XU).$$

Д о к а з а т е л ь с т в о см. в работе (Heyde and Leslie, 1976). Константа уточнена в статье (Hall, 1979).

ЗАМЕЧАНИЕ 3.3.1. Легко убедиться, что из условия $EU^2 = 1$ вытекают соотношения

$$\gamma(XU) = 3E(U^4 - 1) = 3E(U^2 - 1)^2 = 3DU^2.$$

Поэтому в условиях теоремы 3.3.3 справедливы оценки

$$\rho(\Phi(x), E\Phi(xU^{-1})) \leq 1.944 E(U^4 - 1) = 1.944 E(U^2 - 1)^2 = 1.944 DU^2.$$

Для масштабных смесей, в которых случайные величины U имеют моменты меньших порядков, справедлив следующий результат. Напомним, что символом $L(\cdot, \cdot)$ обозначается метрика Леви.

ТЕОРЕМА 3.3.4. Пусть $EU^\delta < \infty$ для некоторого $\delta > 0$. Тогда

$$L(\Phi(x), E\Phi(xU^{-1})) \leq \left(2 + \frac{2}{\delta\sqrt{2\pi e}}\right) L(U^\delta, 1).$$

Д о к а з а т е л ь с т в о см. в работе (Hall, 1979).

Вернемся к обратной задаче оценивания устойчивости нормального закона по отношению к “загрязнению”. Справедлив следующий результат.

ТЕОРЕМА 3.3.5. Пусть $p \in (0, 1]$, $\sigma \in (0, 1]$ и для некоторого $\epsilon > 0$ выполнено неравенство (3.3.6). Тогда существует абсолютная константа $C > 0$ такая, что

$$L(U_{p,\sigma}, 1) \leq C \sqrt{\epsilon}.$$

При этом в качестве C можно взять

$$C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405.$$

Д о к а з а т е л ь с т в о. Справедливо неравенство

$$\rho(\Phi, F_{p,\sigma}) \leq \left(1 + \frac{1}{\sqrt{2\pi}}\right) L(\Phi, F_{p,\sigma})$$

(см., например, (Золотарев, 1986), с. 107). Поэтому из неравенства (3.3.6) вытекает, что

$$\rho(\Phi, F_{p,\sigma}) \leq \left(1 + \frac{1}{\sqrt{2\pi}}\right) \cdot \epsilon.$$

Несложно видеть, что

$$\rho(\Phi, F_{p,\sigma}) = (1-p) \sup_x |\Phi(\sigma x) - \Phi(x)| \geq (1-p) |\Phi(\sigma) - \Phi(1)|.$$

Поэтому, если выполнено неравенство (3.3.6), то

$$(1-p) |\Phi(\sigma) - \Phi(1)| \leq \left(1 + \frac{1}{\sqrt{2\pi}}\right) \cdot \epsilon. \quad (3.3.7)$$

Вместе с тем по формуле Лагранжа в силу того, что $0 < \sigma \leq 1$, имеем

$$|\Phi(\sigma) - \Phi(1)| = (1-\sigma)\phi(\sigma') \geq (1-\sigma)\phi(1) = \frac{1-\sigma}{\sqrt{2\pi e}}, \quad (3.3.8)$$

где $\phi(x)$ – стандартная нормальная плотность, $\sigma' \in [\sigma, 1]$. Из (3.3.7) и (3.3.8) получаем, что, если выполнено неравенство (3.3.6), то

$$(1-p)(1-\sigma) \leq \sqrt{2\pi e} \cdot \left(1 + \frac{1}{\sqrt{2\pi}}\right) \cdot \epsilon = \sqrt{e}(1 + \sqrt{2\pi}) \cdot \epsilon < 5.7815 \epsilon. \quad (3.3.9)$$

Таким образом, если “загрязненная” смесь $F_{p,\sigma}(x)$ близка к “чистому” нормальному закону в смысле метрики Леви или равномерного расстояния, то параметры p и σ *должны* быть близки к единице в том смысле, что должно быть выполнено соотношение (3.3.9).

Но с другой стороны, используя геометрическую интерпретацию расстояния Леви, несложно убедиться, что

$$L(U_{p,\sigma}, 1) = \min\{1-p, 1-\sigma\}.$$

Следовательно, из (3.3.9) окончательно получаем, что

$$\begin{aligned} L^2(U_{p,\sigma}, 1) &= [\min\{1-p, 1-\sigma\}]^2 \leq (1-p)(1-\sigma) \leq \\ &\leq \sqrt{e}(1 + \sqrt{2\pi}) \cdot \epsilon < 5.7815 \epsilon. \end{aligned}$$

Теорема доказана.

Попутно была получена следующая характеристика нормального распределения в классе смесей типа $F_{p,\sigma}$. Пусть $\epsilon \rightarrow 0$. Пусть одновременно параметры $p = p(\epsilon)$ и $\sigma = \sigma(\epsilon)$ изменяются так, что неравенство (3.3.6) остается справедливым. При этом из теоремы (3.3.9) вытекает, что это возможно, только если выполнено условие

$$L(U_{p,\sigma}, 1) \longrightarrow 0.$$

Но это означает, что смесь $F_{p,\sigma}$ может быть “чистым” нормальным законом тогда *и только тогда*, когда $p = 1$ или $\sigma = 1$.

Аналогичную задачу можно рассмотреть и для сдвиговых смесей нормальных законов вида

$$G_{p,a}(x) = p\Phi(x) + (1-p)\Phi(x-a), \quad a \in \mathbb{R}, \quad x \in \mathbb{R}.$$

Несложно заметить, что смесь $G_{p,a}(x)$ – это функция распределения суммы стандартно нормально распределенной случайной величины Z и независимой от нее случайной величины $V_{p,a}$, принимающей значения 0 и a с вероятностями p и $(1-p)$, соответственно.

Пусть $|a| \leq 1$. Так как

$$\begin{aligned} \rho(G_{p,a}, \Phi) &= (1-p) \sup_x |\Phi(x-a) - \Phi(x)| \geq (1-p) \cdot |\Phi(a) - \Phi(0)| = \\ &= (1-p)|a|\phi(a') \geq (1-p)|a|\phi(1) \geq \frac{(1-p)|a|}{\sqrt{2\pi e}}, \end{aligned}$$

где $a' \in [0, a]$, то, используя ту же самую логику доказательства, что и в теореме 3.3.5, можно доказать следующее утверждение.

ТЕОРЕМА 3.3.6. Пусть $p \in (0, 1]$, $|a| \leq 1$ и для некоторого $\epsilon > 0$ выполнено неравенство

$$L(G_{p,a}, \Phi) \leq \epsilon.$$

Тогда существует абсолютная константа $C > 0$ такая, что

$$L(V_{p,a}, 0) \leq C \sqrt{\epsilon}.$$

При этом в качестве C можно взять $C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405$.

При этом смесь $G_{p,a}$ может быть “чистым” нормальным законом тогда *и только тогда*, когда $p = 1$ или $a = 0$.

Теоремы 3.3.5 и 3.3.6 допускают простое обобщение. Рассмотрим две смеси:

$$F_{p,\sigma}(x) = p\Phi(x) + (1-p)\Phi(\sigma x)$$

и

$$F_{q,\sigma}(x) = q\Phi(x) + (1-q)\Phi(\sigma x),$$

где $p \in [0, 1]$, $q \in [0, 1]$, $\sigma \in (0, 1]$. Так как

$$\rho(F_{p,\sigma}, F_{q,\sigma}) = |p - q| \sup_x |\Phi(x) - \Phi(\sigma x)| \geq$$

$$\geq |p - q| \cdot |\Phi(\sigma) - \Phi(1)| \geq \frac{|p - q| \cdot |1 - \sigma|}{\sqrt{2\pi e}},$$

то, используя ту же самую логику доказательства, что и в теореме 3.3.5, можно доказать, что, если

$$L(F_{p,\sigma}, F_{q,\sigma}) \leq \epsilon, \quad (3.3.9)$$

то

$$|p - q| \cdot |1 - \sigma| \leq \sqrt{e}(1 + \sqrt{2\pi}) \cdot \epsilon,$$

и справедливо следующее утверждение.

ТЕОРЕМА 3.3.7. Пусть $p \in [0, 1]$, $q \in [0, 1]$, $\sigma \in (0, 1]$ и для некоторого $\epsilon > 0$ выполнено неравенство (3.3.9). Тогда существует абсолютная константа $C > 0$ такая, что

$$L(U_{p,\sigma}, U_{q,\sigma}) \leq C \sqrt{\epsilon}.$$

При этом в качестве C можно взять $C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405$.

Аналогично, если рассмотреть смеси

$$G_{p,a}(x) = p\Phi(x) + (1 - p)\Phi(x - a)$$

и

$$G_{q,a}(x) = q\Phi(x) + (1 - q)\Phi(x - a),$$

где $p \in [0, 1]$, $q \in [0, 1]$, $|a| \leq 1$, то, используя ту же самую логику доказательства, что и в теореме 3.3.6, можно доказать, что, если

$$L(G_{p,a}, G_{q,a}) \leq \epsilon, \quad (3.3.10)$$

то

$$|a| \cdot |p - q| \leq \sqrt{e}(1 + \sqrt{2\pi}) \cdot \epsilon,$$

и справедливо следующее утверждение.

ТЕОРЕМА 3.3.8. Пусть $p \in [0, 1]$, $q \in [0, 1]$, $|a| \leq 1$ и для некоторого $\epsilon > 0$ выполнено неравенство (3.3.10). Тогда существует абсолютная константа $C > 0$ такая, что

$$L(V_{p,a}, V_{q,a}) \leq C \sqrt{\epsilon}.$$

При этом в качестве C можно взять $C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405$.

Пусть теперь $\sigma_1 \in (0, 1]$ и $\sigma_2 \in (0, 1]$. Так как

$$\begin{aligned} \rho(F_{p,\sigma_1}, F_{p,\sigma_2}) &= (1-p) \sup_x |\Phi(\sigma_1 x) - \Phi(\sigma_2 x)| \geq \\ &\geq (1-p) \cdot \left| \Phi\left(\frac{\sigma_1}{\sigma_2}\right) - \Phi(1) \right| \geq \frac{1-p}{\sqrt{2\pi e}} \cdot \left| 1 - \frac{\sigma_1}{\sigma_2} \right|, \end{aligned}$$

то, используя ту же самую логику доказательства, что и в теореме 3.3.5, можно доказать, что, если

$$L(F_{p,\sigma_1}, F_{p,\sigma_2}) \leq \epsilon, \quad (3.3.11)$$

то

$$(1-p) \cdot \max \left\{ \left| 1 - \frac{\sigma_1}{\sigma_2} \right|, \left| 1 - \frac{\sigma_2}{\sigma_1} \right| \right\} \leq \sqrt{e}(1 + \sqrt{2\pi}) \cdot \epsilon,$$

и справедливо следующее утверждение.

ТЕОРЕМА 3.3.9. Пусть $p \in [0, 1]$, $\sigma_1 \in (0, 1]$ и $\sigma_2 \in (0, 1]$ и для некоторого $\epsilon > 0$ выполнено неравенство (3.3.11). Тогда существует абсолютная константа $C > 0$ такая, что

$$L(U_{p,\sigma_1}, U_{p,\sigma_2}) \leq C \sqrt{\epsilon}.$$

При этом в качестве C можно взять $C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405$.

Пусть теперь $|a_1| \leq 1$ и $|a_2| \leq 1$. Так как

$$\begin{aligned} \rho(G_{p,a_1}, G_{p,a_2}) &= (1-p) \sup_x |\Phi(x - a_1) - \Phi(x - a_2)| \geq \\ &\geq (1-p) \cdot |\Phi(a_1 - a_2) - \Phi(0)| \geq \frac{(1-p)|a_1 - a_2|}{\sqrt{2\pi e}}, \end{aligned}$$

то, используя ту же самую логику доказательства, что и в теореме 3.3.6, можно доказать, что, если

$$L(G_{p,a_1}, G_{p,a_2}) \leq \epsilon, \quad (3.3.12)$$

то

$$(1-p)|a_1 - a_2| \leq \sqrt{e}(1 + \sqrt{2\pi}) \cdot \epsilon,$$

и справедливо следующее утверждение.

ТЕОРЕМА 3.3.10. Пусть $p \in [0, 1]$, $a_1 \leq 1$, $a_2 \leq 1$ и для некоторого $\epsilon > 0$ выполнено неравенство (3.3.12). Тогда существует абсолютная константа $C > 0$ такая, что

$$L(V_{p,a_1}, V_{p,a_2}) \leq C \sqrt{\epsilon}.$$

При этом в качестве C можно взять $C = [\sqrt{e}(1 + \sqrt{2\pi})]^{1/2} < 2.405$.

Указанные обобщения играют очень важную роль при обосновании возможности приближенного разделения смесей с помощью метода фиксированных компонент (см. раздел 5.7).

3.4. Самоподобие смешанных гауссовских процессов

Проблематика этого раздела тесно связана с задачами, которым посвящен раздел 4.3.3. Здесь будут рассмотрены условия, при которых стохастический процесс с независимыми приращениями, имеющими смешанное гауссовское распределение, обладает свойством самоподобия, иногда называемым автомодельностью.

Вообще говоря, свойство самоподобия (автомодельность) означает, что распределение приращений процесса на больших интервалах времени сходно (или совпадает) с распределением приращений того же процесса на малых интервалах времени при соответствующем изменении масштаба.

Сразу же заметим, что, если стохастический процесс имеет независимые приращения на интервалах единичной длины, причем эти приращения имеют вид масштабных смесей гауссовских законов, то такой процесс обладает свойством смешанного гауссовского самоподобия в том смысле, что все его приращения на интервалах времени с длинами, кратными длине единичного интервала (то есть интервалах времени, длины которых равны натуральным числам), снова будут масштабными смесями гауссовских законов.

Формально, предположим, что $X_1, \dots, X_n$ – случайные величины со стандартным нормальным распределением, $U_1, \dots, U_n$ – неотрицательные (необязательно одинаково распределенные) случайные величины. Более того, предположим, что случайные величины $X_1, \dots, X_n, U_1, \dots, U_n$ независимы в совокупности. Тогда распределение каждого произведения $X_j U_j$ будет иметь вид масштабной смеси нормальных законов. Теперь рассмотрим сумму

$$X_1 U_1 + \dots + X_n U_n.$$

Поскольку слагаемые в ней независимы, характеристическая функция

этой суммы при каждом $s \in \mathbb{R}$ имеет вид

$$\begin{aligned} \mathbb{E} \exp\{is(X_1U_1 + \dots + X_nU_n)\} &= \prod_{j=1}^n \mathbb{E} \exp\{is(X_jU_j)\} = \prod_{j=1}^n \mathbb{E} \exp\{-\frac{s^2}{2}U_j^2\} = \\ &= \mathbb{E} \prod_{j=1}^n \exp\{-\frac{s^2}{2}U_j^2\} = \mathbb{E} \exp\left\{-\frac{s^2}{2} \sum_{j=1}^n U_j^2\right\} = \mathbb{E} \exp\left\{isX\sqrt{U_1^2 + \dots + U_n^2}\right\}, \end{aligned} \quad (3.4.1)$$

где X – случайная величина со стандартным нормальным распределением, независимая от случайных величин $U_1, \dots, U_n$. В правой части этого соотношения стоит характеристическая функция, соответствующая масштабной смеси нормальных законов вида

$$\mathbb{E} \Phi\left(\frac{x}{\sqrt{U_1^2 + \dots + U_n^2}}\right), \quad x \in \mathbb{R}.$$

Обратим внимание, что до сих пор не предполагалось, что случайные величины $U_1, \dots, U_n$ одинаково распределены. Это означает, что для того чтобы случайный процесс с независимыми приращениями обладал свойством смешанного гауссовского самоподобия, совсем не обязательно, чтобы он был однородным (то есть с одинаково распределенными приращениями на интервалах времени одинаковой длины).

Теперь рассмотрим условия, обеспечивающие самоподобие смешанных гауссовских процессов в более традиционном смысле. Под самоподобием (в широком смысле) случайного процесса будем понимать совпадение распределений его приращений на любых интервалах времени ненулевой длины после соответствующего изменения масштаба.

Более строго, пусть $a(t)$ – положительная функция, $t \geq 0$. Случайный процесс $\xi(t)$, $t \geq 0$, называется *самоподобным* (в широком смысле) с масштабирующей функцией $a(t)$, если для любых $t > 0$ справедливо равенство

$$\xi(t) \stackrel{d}{=} a(t)\xi(1). \quad (3.4.2)$$

Теперь предположим, что процесс $X(t)$ является смешанным гауссовским, то есть для любого $t > 0$ справедливо соотношение

$$\xi(t) \stackrel{d}{=} X_t \cdot U(t),$$

где X_t – случайная величина со стандартным нормальным распределением, $U(t)$ – неотрицательная случайная величина, причем случайные величины X_t и $U(t)$ независимы.

ЛЕММА 3.4.1. Пусть $\xi(t)$ – смешанный гауссовский процесс. Процесс $\xi(t)$ является самоподобным в широком смысле с масштабирующей функцией $a(t)$ тогда и только тогда, когда смешивающий процесс $U(t)$ является самоподобным в широком смысле с той же самой масштабирующей функцией $a(t)$.

ДОКАЗАТЕЛЬСТВО. Поскольку характеристическая функция случайной величины $\xi(t)/a(t)$ имеет вид

$$\mathbb{E} \exp \left\{ i s \frac{\xi(t)}{a(t)} \right\} = \mathbb{E} \exp \left\{ - \frac{s^2 U^2(t)}{2 a^2(t)} \right\}, \quad s \in \mathbb{R},$$

условие (3.4.2) может быть записано в форме

$$\mathbb{E} \exp \left\{ - \frac{s^2 U^2(t)}{2 a^2(t)} \right\} = \mathbb{E} \exp \left\{ - \frac{s^2 U^2(1)}{2} \right\}, \quad s \in \mathbb{R}. \quad (3.4.3)$$

Теперь, если преобразование Лапласа–Стилтьеса неотрицательной случайной величины Z обозначить символом $\psi_Z(v)$,

$$\psi_Z(v) = \mathbb{E} e^{-vZ}, \quad v \geq 0,$$

то соотношение (3.4.3) означает не что иное, как

$$\psi_{U^2(t)/a^2(t)}(v) = \psi_{U^2(1)}(v), \quad v \geq 0.$$

Следовательно, поскольку $U(t) \geq 0$ и $a(t) \geq 0$, $t \geq 0$, из последнего равенства вытекает, что

$$U(t) \stackrel{d}{=} a(t) \cdot U(1).$$

Лемма доказана.

Например, если $U(t) = \sqrt{Y_\alpha(t)}$, где $Y_\alpha(t)$ – устойчивый процесс Леви с характеристическим показателем α (так как $U(t)$ неотрицателен, то $0 < \alpha < 1$, см., например, (Золотарев, 1983)), то, как известно,

$$Y_\alpha(t) \stackrel{d}{=} c t^{1/\alpha} Y_\alpha(1), \quad t \geq 0,$$

где $c > 0$ – некоторая константа, и, следовательно,

$$U(t) \stackrel{d}{=} c t^{1/2\alpha} U(1), \quad t \geq 0.$$

Тогда процесс $\xi(t)$ является самоподобным в широком смысле с масштабирующей функцией

$$a(t) = c \cdot t^{1/2\alpha}.$$

Более того, по “теореме умножения” (теорема 3.3.1 в (Золотарев, 1983)) в таком случае $\xi(t)$ также оказывается устойчивым процессом Леви, но с характеристическим показателем 2α .

3.5. Сужение класса смешивающих законов. Масштабные смеси нормальных законов как результат симметризации

Выше было показано, что использование смесей нормальных законов (в простейшей ситуации – масштабных) в качестве распределений приращений логарифмов биржевых цен имеет вполне разумное обоснование. В качестве, так сказать, побочного вывода из вышесказанного следует, что лишь некоторые устойчивые законы заслуживают внимания в связи с задачей описания распределения приращений логарифмов биржевых цен (и многих других реальных процессов), постольку-поскольку они представимы в виде смесей нормальных законов.

Другой вывод из предыдущих разделов заключается в том, что задача статистического анализа распределения логарифмов приращений биржевых цен сводится к статистическому определению смешивающего распределения (разделению смеси), которое является неизвестным параметром рассматриваемой статистической задачи. Без каких-либо дополнительных предположений класс смешивающих законов (параметрическое множество) совпадает с множеством всех распределений, сосредоточенных на неотрицательной полуоси. Выбор нужного закона при этом представляет собой чрезвычайно трудоемкую статистическую задачу. Поэтому вполне естественно стремиться сузить параметрическое множество, то есть семейство допустимых смешивающих законов за счет каких-либо дополнительных соображений. Один из возможных подходов к решению этой задачи и предлагается в данном разделе. Будут рассмотрены два простейших варианта задачи. Данный раздел основан на материале работы (Королев, 2000).

3.5.1. Сверточная модель симметризации

Статистический анализ распределения приращений логарифмов биржевых цен производится на основе выборки, содержащей как положительные, так и отрицательные наблюдения. Разобьем основную выборку на две подвыборки, содержащие, соответственно, только положительные (“ups”) и только неположительные (“downs”) наблюдения. Подгоняя распределения к каждой подвыборке по отдельности, можно прийти к заключению, что случайная величина P , описывающая итоговое приращение логариф-

ма приращения биржевой цены, представима в виде

$$P = U - D, \quad (3.5.1)$$

где U – случайная величина, описывающая положительную часть приращения, а D – случайная величина, описывающая абсолютную величину неположительной части приращения. Предположим, что в представлении (3.5.1) случайные величины U и D независимы. Это предположение согласуется с выводами статистического анализа реальных данных. Более того, соображения симметрии приводят нас к предположению о том, что распределения случайных величин U и D должны, по крайней мере, принадлежать к одному типу. Для простоты мы будем предполагать, что функции распределения случайных величин U и D совпадают и равны, скажем, $F(x)$. Характеристическую функцию, соответствующую функции распределения $F(x)$, обозначим $f(t)$. Тогда $E \exp\{isP\} = |f(s)|^2$, $s \in \mathbb{R}$. Последняя характеристическая функция вещественна, следовательно, распределение, ей соответствующее, является симметричным в том смысле, что для любого $x > 0$ справедливо соотношение $P(P < -x) = P(P > x)$.

Распределение, соответствующее характеристической функции $|f(s)|^2$, будем называть *сверточной симметризацией* распределения $F(x)$.

Таким образом, упомянутая выше задача сужения класса допустимых смесей нормальных законов сводится к следующей. В дальнейшем распределения вероятностей и соответствующие функции распределения, которые для определенности считаются непрерывными слева, будут отождествляться.

ЗАДАЧА 1. *Описать класс $\mathfrak{P}$ всех распределений F , сосредоточенных на неотрицательной полуоси и таких, что их сверточная симметризация представима в виде масштабной смеси нормальных законов.*

Решение задачи 1 дается следующей теоремой.

ТЕОРЕМА 3.5.1. *Распределение F принадлежит классу $\mathfrak{P}$ тогда и только тогда, когда*

- (i) $F(0) = 0$;
- (ii) *соответствующая ему характеристическая функция f удовлетворяет следующему условию: функция $|f(\sqrt{t})|^2$, $t \geq 0$, является вполне монотонной, то есть она бесконечно дифференцируема и при каждом $n \geq 1$*

$$(-1)^n \frac{d^n}{dt^n} |f(\sqrt{t})|^2 \geq 0, \quad t \geq 0.$$

Д о к а з а т е л ь с т в о (в целом аналогичное приводимому ниже доказательству теоремы 3.5.3) см. в работе (Korolev, 2004).

Заметим, что условие (ii) теоремы 3.5.1 представляет собой критерий представимости сверточной симметризации *произвольного* распределения F (не обязательно сосредоточенного на полуоси) с характеристической функцией f в виде масштабной смеси нормальных законов.

Класс $\mathfrak{P}$ не является пустым, что демонстрируют следующие примеры.

ПРИМЕР 3.5.1. Пусть $\Gamma_{\alpha,\lambda}$ – функция гамма-распределения с некоторыми параметром формы α и параметром масштаба λ , задаваемого плотностью вероятностей

$$p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x \geq 0,$$

где $\Gamma(\alpha)$ – эйлерова гамма-функция,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx.$$

Тогда $\Gamma_{\alpha,\lambda} \in \mathfrak{P}$. Действительно, функции распределения $\Gamma_{\alpha,\lambda}$ соответствует характеристическая функция

$$\gamma_{\alpha,\lambda}(t) = \left(\frac{\lambda}{\lambda - it} \right)^\alpha, \quad t \in \mathbb{R}.$$

Легко видеть, что, если W и W_1 – независимые случайные величины с одним и тем же гамма-распределением с некоторыми параметрами масштаба $\lambda > 0$ и формы $\alpha > 0$, то

$$\mathbb{E} e^{it(W-W_1)} = |\gamma_{\alpha,\lambda}(t)|^2 = \frac{1}{\left(1 - \frac{it}{\lambda}\right)^\alpha} \cdot \frac{1}{\left(1 + \frac{it}{\lambda}\right)^\alpha} = \left(\frac{\lambda^2}{\lambda^2 + t^2} \right)^\alpha, \quad (3.5.2)$$

то есть для $s \geq 0$

$$|\gamma_{\alpha,\lambda}(\sqrt{s})|^2 = \left(\frac{\lambda^2}{\lambda^2 + s} \right)^\alpha. \quad (3.5.3)$$

В правой части (3.5.3) стоит преобразование Лапласа-Стилтьеса гамма-распределения с параметром формы α и параметром масштаба λ^2 . По теореме С. Н. Бернштейна (см., например, (Феллер, 1984), том 2, глава XIII, § 4, теорема 1) эта функция вполне монотонна, и, следовательно, согласно Теореме 3.5.1, $\Gamma_{\alpha,\lambda} \in \mathfrak{P}$.

Несложно видеть, что распределению $\Gamma_{\alpha,\lambda} \in \mathfrak{P}$ соответствует смешивающая случайная величина $Y = \sqrt{Z}$, где Z имеет функцию распределения $\Gamma_{\alpha,\lambda^2/2}$. Действительно, убедимся, что $P \stackrel{d}{=} X \cdot W_2^{1/2}$, где случайная величина X имеет стандартное нормальное распределение и независима от случайной величины W_2 , имеющей гамма-распределение с параметром масштаба $\frac{1}{2}\lambda^2$ и параметром формы α . В таком случае по теореме Фубини

$$\begin{aligned} \mathbb{E} \exp \{itX\sqrt{W_2}\} &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-x(\frac{1}{2}t^2 + \frac{1}{2}\lambda^2)} x^{\alpha-1} dx = \\ &= \frac{\lambda^{2\alpha}}{2^\alpha \Gamma(\alpha) (\frac{1}{2}t^2 + \frac{1}{2}\lambda^2)^\alpha} \int_0^\infty e^{-y} y^{\alpha-1} dy = \left(\frac{\lambda^2}{\lambda^2 + t^2} \right)^\alpha. \end{aligned} \quad (3.5.4)$$

Правые части (3.5.2) и (3.5.4) совпадают, что и требовалось доказать.

Плотность симметризованного гамма-распределения имеет вид (2.6.9) с $\theta = 0$, $\sigma^2 = 1$.

Из этого примера, в частности, вытекает, что экспоненциальное распределение принадлежит к классу $\mathfrak{P}$, причем сверточной симметризацией экспоненциального распределения является распределение Лапласа, определяемое плотностью $\ell(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$, и функцией распределения

$$L(x) = \begin{cases} \frac{1}{2}e^x, & x \leq 0, \\ 1 - \frac{1}{2}e^{-x}, & x > 0. \end{cases}$$

При этом смешивающей случайной величиной является $Y = \sqrt{Z}$, где Z опять-таки имеет экспоненциальное распределение.

ПРИМЕР 3.5.2. Пусть $G_{\alpha,b,c}$ – функция распределения устойчивого закона, сосредоточенного на положительной полуоси, и которому соответствует характеристическая функция

$$g_{\alpha,b,c}(t) = \exp \left\{ ibt - c|t|^\alpha \left[1 - i \frac{t}{|t|} \tan \frac{\pi\alpha}{2} \right] \right\}, \quad t \in \mathbb{R}.$$

Здесь $0 < \alpha < 1$, $b \in \mathbb{R}$, $c > 0$. Тогда $G_{\alpha,b,c} \in \mathfrak{P}$.

Действительно, для $s \geq 0$ справедливо представление

$$|g_{\alpha,b,c}(\sqrt{s})|^2 = \exp \{ -2cs^{\alpha/2} \}. \quad (3.5.5)$$

По теореме 1 из (Феллер, 1984), Гл. XIII, с. 504–505) функция, стоящая в правой части (3.5.5), является преобразованием Лапласа-Стилтьеса сосредоточенного на положительной полуоси устойчивого закона с показателем $\alpha/2$. По уже упоминавшейся теореме С. Н. Бернштейна эта функция вполне монотонна, и следовательно, согласно теореме 3.5.1, $G_{\alpha,b,c} \in \mathfrak{P}$.

Найдем смешивающее распределение, соответствующее функции распределения $G_{\alpha,b,c}$. С этой целью заметим, что

$$|g_{\alpha,b,c}(t)|^2 = \exp \{-2c|t|^\alpha\}. \quad (3.5.6)$$

Не ограничивая общности, полагаем $2c = 1$ (это предположение фактически сводится к изменению масштаба в $(2c)^{1/\alpha}$ раз). Тогда в правой части (3.5.6) будет стоять характеристическая функция стандартного симметричного строго устойчивого закона с показателем $\alpha < 1$. По теореме 3.3.1 из (Золотарев, 1983), соответствующая такой характеристической функции функция распределения G_α может быть представлена в виде

$$G_\alpha(x) = \int_0^\infty \Phi(x/\sqrt{y}) dP(Y_{\alpha/2} < y), \quad x \in \mathbb{R},$$

где $\Phi(x)$ – стандартная нормальная функция распределения, а $Y_{\alpha/2}$ – положительная строго устойчивая случайная величина с показателем $\alpha/2$.

В частности, к классу $\mathfrak{P}$ принадлежит распределение Леви (устойчивое распределение с параметром $\alpha = 1/2$) с плотностью

$$p(x) = \frac{1}{\sqrt{2\pi x^3}} \exp \left\{ -\frac{1}{2x} \right\}, \quad x > 0.$$

Ему соответствует смешивающее сосредоточенное на положительной полуоси строго устойчивое распределение с характеристическим показателем $\alpha = 1/4$.

Практика показывает, что во многих случаях статистический анализ данных о приращениях биржевых цен (или экспериментальных данных, связанных с измерениями параметров турбулентной плазмы) удобно производить, разбивая общий массив данных (выборку) на два подмассива (две подвыборки), один из которых содержит только положительные, а другой – только неположительные данные, подгоняя распределение (которое в соответствии с моделью, рассматриваемой в данном разделе, должно выбираться из класса $\mathfrak{P}$) к каждой из подвыборок и в качестве распределения приращения используя симметризацию подогнанного распределения.

Преимущества данного подхода заключаются как минимум в двух обстоятельствах. Во-первых, статистическую подгонку распределения к выборке неотрицательных наблюдений производить во многих отношениях проще, нежели к выборке наблюдений обоих знаков. Во-вторых, для того чтобы определить характер поведения хвостов смеси нормальных законов, вообще говоря, необходимо проделать некоторую дополнительную работу, например, попытаться вычислить или аналитически оценить соответствующие (кратные) интегралы или использовать аппарат тауберовых или абелевых теорем, в то время как характер поведения хвоста распределения, сосредоточенного на полуоси, как правило, очевиден из его аналитического представления, а хвосты односторонних распределений ведут себя точно так же, как и хвосты их симметризаций. В справедливости последнего утверждения убеждает следующая лемма.

Пусть X и X' – независимые одинаково распределенные случайные величины. Сверточную симметризацию случайной величины X обозначим $X^{(s)}$. При этом $X^{(s)} = X - X'$.

ЛЕММА 3.5.1. *Для любого $x > 0$ справедливы неравенства*

$$\begin{aligned} \frac{1}{2}P(|X - \text{med}X| \geq x) &\leq P(|X^{(s)}| \geq x) \leq \\ &\leq 2 \inf_{b \in \mathbb{R}} P(|X - b| \geq \frac{x}{2}) [2 - P(|X - b| \geq \frac{x}{2})]. \end{aligned}$$

Д о к а з а т е л ь с т в о см. с. 259 в (Лозв, 1962).

Иногда говорят, что случайная величина X имеет распределение с *тяжелыми хвостами*, если для некоторых $C > 0$ и $\gamma > 0$

$$P(|X| \geq x) \sim \frac{C}{x^\gamma}$$

при $x \rightarrow \infty$. Тогда из леммы 3.5.1 вытекает, что “тяжесть” хвоста распределения с тяжелыми хвостами совпадает с аналогичной характеристикой его симметризации. А именно, справедливо

СЛЕДСТВИЕ 3.5.1. *Если случайная величина X имеет распределение с тяжелыми хвостами, характеризуемым параметром $\gamma > 0$, то*

$$\frac{1}{2} \leq \liminf_{x \rightarrow \infty} \frac{P(|X^{(s)}| \geq x)}{P(|X| \geq x)} \leq \limsup_{x \rightarrow \infty} \frac{P(|X^{(s)}| \geq x)}{P(|X| \geq x)} \leq 2^{\gamma+1}.$$

Довольно часто оказывается, что хвосты законов, подогнанных к подвыборкам одного знака, убывают (при $x \rightarrow \infty$) вейбулловским образом,

то есть как $O(\exp\{-x^\gamma\})$ с некоторым $\gamma \in (0, 1)$. Два следующих примера иллюстрируют возможность выбора соответствующего распределения из класса $\mathfrak{P}$.

Оба эти примера основаны на следующем замечательном утверждении, доказанном в работе (Багиров, 1988).

ТЕОРЕМА 3.5.2. Пусть $\mathfrak{A}$ – класс случайных величин, функции распределения которых представимы в виде масштабных смесей нормальных функций распределения с нулевым средним. Если V – некоторая случайная величина из класса $\mathfrak{A}$ и n – произвольное натуральное число, то распределение случайной величины V^{2n} принадлежит классу $\mathfrak{P}$.

ПРИМЕР 3.5.3. Пусть X – случайная величина, плотность распределения которой имеет вид

$$p(x) = \begin{cases} \frac{1}{k\sqrt{2\pi}} x^{-(2k-1)/2k} \exp\{-x^{1/k}\}, & x > 0, \\ 0, & x < 0 \end{cases}$$

при некотором натуральном k . Тогда $P(X < x) \in \mathfrak{P}$. Чтобы убедиться в этом, воспользуемся теоремой 3.5.2. В качестве V возьмем случайную величину со стандартным нормальным распределением. Очевидно, $V \in \mathfrak{A}$. Несложно проверить, что при этом рассматриваемая в данном примере плотность $p(x)$ соответствует случайной величине $X = V^{2k}$.

ПРИМЕР 3.5.4. Пусть W_γ – распределение Вейбулла с параметром γ , имеющее плотность

$$w_\gamma(x) = \begin{cases} \gamma x^{\gamma-1} \exp\{-x^\gamma\}, & x > 0, \\ 0, & x < 0, \end{cases}$$

причем $\gamma = (2k)^{-1}$ при некотором $k = 1, 2, \dots$. Тогда $W_\gamma \in \mathfrak{P}$. Действительно, выше (в частности, см. пример 3.5.1) было показано, что классу $\mathfrak{A}$ принадлежит случайная величина Λ , имеющая распределение Лапласа с плотностью $\ell(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$, и функцией распределения

$$L(x) = \begin{cases} \frac{1}{2}e^x, & x \leq 0, \\ 1 - \frac{1}{2}e^{-x}, & x > 0. \end{cases}$$

Поэтому, если доказать, что случайная величина Λ^{2k} имеет плотность $w_\gamma(x)$ с $\gamma = (2k)^{-1}$, то принадлежность распределения W_γ классу $\mathfrak{P}$ будет следовать из теоремы 3.5.2. Но для указанных γ при $x > 0$, очевидно,

имеем

$$P(\Lambda^{1/\gamma} < x) = P(|\Lambda| < x^\gamma) = 2L(x^\gamma) - 1 = 1 - \exp\{-x^\gamma\}$$

и

$$\frac{d}{dx}(1 - \exp\{-x^\gamma\}) = w_\gamma(x),$$

и тем самым принадлежность распределения Вейбулла с указанным параметром классу $\mathfrak{P}$ установлена.

Теорема 3.5.2 обеспечивает также присутствие в классе $\mathfrak{P}$ распределений (а стало быть, и соответствующих случайных величин в классе $\mathfrak{A}$), имеющих хвосты, убывающие степенным образом с произвольным показателем. В этом убеждает следующий пример.

ПРИМЕР 3.5.5. Пусть $F_{1,m}$ – распределение Снедекора–Фишера с $(1, m)$ степенями свободы ($m > 0$), задаваемое плотностью

$$p_{1,m}(x) = \frac{\Gamma((m+1)/2)m^{m/2}}{\sqrt{\pi}\Gamma(m/2)} \cdot \frac{1}{\sqrt{x}(m+x)^{(m+1)/2}}, \quad x > 0.$$

Тогда $F_{1,m} \in \mathfrak{P}$. Действительно, распределение Снедекора–Фишера с (n, m) степенями свободы ($n > 0, m > 0$), хорошо известное в математической статистике, соответствует случайной величине

$$Z_{n,m} = \frac{m\eta_n}{n\eta_m},$$

где η_n и η_m – независимые случайные величины, имеющие распределение хи-квадрат, соответственно, с n и m степенями свободы (при этом n и m не обязаны быть целыми). Отсюда видно, что распределение Снедекора–Фишера $F_{1,m}$ с $(1, m)$ степенями свободы соответствует квадрату случайной величины τ_m , имеющей распределение Стьюдента с m степенями свободы. Но, как хорошо известно, $\tau_m \in \mathfrak{A}$. Теперь требуемое утверждение непосредственно вытекает из теоремы 3.5.2.

Необходимо отметить, что в соответствии с введенной выше терминологией масштабная смесь нормальных законов, являющаяся сверточной симметризацией распределения Снедекора–Фишера $F_{1,m}$ с $(1, m)$ степенями свободы, по лемме 3.5.1 имеет тяжелые хвосты, убывающие при $x \rightarrow \infty$ как $O(x^{-m/2})$.

Можно сформулировать задачу, являющуюся в некотором смысле обратной к задаче 1.

ЗАДАЧА 2. *Описать класс $\mathfrak{M}$ всех распределений H , сосредоточенных на неотрицательной полуоси, обладающих следующим свойством: для распределения H существует распределение F , сосредоточенное на неотрицательной полуоси, сверточная симметризация которого совпадает с функцией распределения*

$$\int_0^{\infty} \Phi(x/y) dH(y).$$

Эта задача не является тривиальной, поскольку класс $\mathfrak{M}$ не совпадает с классом всех распределений, сосредоточенных на неотрицательной полуоси. Действительно, пусть H – вырожденная функция распределения с единственным единичным скачком в некоторой точке $a > 0$. Тогда масштабная смесь нормальных законов становится нормальным распределением, а основное уравнение принимает вид $X \stackrel{d}{=} U/a - U'/a$. Это уравнение относительно распределения случайной величины U по теореме Леви–Крамера о разложимости нормального закона лишь на нормальные компоненты имеет единственное решение: распределение случайной величины U само должно быть нормальным, но оно имеет точки роста на отрицательной полуоси, и, стало быть, вырожденное распределение не принадлежит к $\mathfrak{M}$.

Из этих рассуждений следует интересный вывод: в рамках сделанных выше предположений *нормальный закон неприемлем в качестве модели распределения приращений логарифмов биржевых цен.*

3.5.2. Рандомизационная модель симметризации

Рассмотрим еще один подход к симметризации распределений. Для наглядности вновь будем говорить о случайной величине P , равной приращению некоторого финансового индекса за рассматриваемый промежуток времени. Пусть X – абсолютная величина этого приращения. Если финансовый рынок находится в стационарном режиме, то заранее нельзя сделать абсолютно никакого прогноза о том, в какую сторону сдвинется рассматриваемый финансовый индекс, то есть о том, каким окажется знак случайной величины P . Формально это выражается в том, что приращение с вероятностью $\frac{1}{2}$ может быть положительным и с такой же вероятностью – отрицательным. Другими словами, каким бы ни было $x > 0$,

$$P(P < x) = \frac{1}{2}P(X < x) + \frac{1}{2}P(-X < x).$$

Абстрагируясь от нашей конкретной прикладной задачи, рассмотрим теперь произвольную случайную величину X и обозначим ее функцию распределения $F(x)$. Назовем *рандомизационной симметризацией* случайной величины X случайную величину $\tilde{X}$ такую, что

$$\tilde{X} = \begin{cases} X & \text{с вероятностью } \frac{1}{2}, \\ -X & \text{с вероятностью } \frac{1}{2}. \end{cases}$$

Очевидно, что

$$\begin{aligned} \mathbf{P}(\tilde{X} < x) &= \frac{1}{2}\mathbf{P}(X < x) + \frac{1}{2}\mathbf{P}(-X < x) = \\ &= \frac{1}{2}[F(x) + 1 - F(-x + 0)] = \frac{1}{2} + \frac{1}{2}[F(x) - F(-x + 0)]. \end{aligned}$$

Если при этом случайная величина X абсолютно непрерывна, то, обозначив плотности случайных величин X и $\tilde{X}$, соответственно, $p(x)$ и $q(x)$, из последнего равенства получим соотношение

$$q(x) = \frac{1}{2}[p(x) - p(-x)].$$

Более того, если $\mathbf{P}(X \geq 0) = 1$, то

$$\mathbf{P}(\tilde{X} < x) = \begin{cases} \frac{1}{2}[1 + F(x)], & \text{если } x > 0, \\ \frac{1}{2}[1 - F(-x + 0)], & \text{если } x \leq 0, \end{cases}$$

и

$$q(x) = \frac{1}{2}p(|x|).$$

По аналогии функцию распределения $Q(x) = \mathbf{P}(\tilde{X} < x)$ будем называть *рандомизационной симметризацией* функции распределения $F(x)$.

Рандомизационную симметризацию $\tilde{X}$ случайной величины X удобно интерпретировать следующим образом. Пусть Z – независимая от X случайная величина такая, что $\mathbf{P}(Z = 1) = \mathbf{P}(Z = -1) = \frac{1}{2}$. Тогда

$$\tilde{X} = X \cdot Z.$$

Если $f(s)$ – характеристическая функция случайной величины X , то

$$\begin{aligned} \mathbf{E} \exp \{is\tilde{X}\} &= \frac{1}{2}f(s) + \frac{1}{2}f(-s) = \\ &= \frac{1}{2}[\mathbf{E} \cos sX + i\mathbf{E} \sin sX + \mathbf{E} \cos sX - i\mathbf{E} \sin sX] = \mathbf{E} \cos sX = \operatorname{Re} f(s). \end{aligned}$$

Принимая во внимание приведенные выше доводы в пользу того, что распределения приращений (логарифмов) финансовых индексов являются смесями нормальных законов, приходим к следующей задаче.

ЗАДАЧА 3. *Описать класс $\mathfrak{Q}$ всех распределений F , сосредоточенных на неотрицательной полуоси и таких, что их рандомизационная симметризация представима в виде масштабной смеси нормальных законов.*

Решение задачи 3 дается следующей теоремой.

ТЕОРЕМА 3.5.3. *Распределение F принадлежит классу $\mathfrak{Q}$ тогда и только тогда, когда*

$$(iii) \quad F(0) = 0;$$

(iv) *соответствующая ему характеристическая функция $f(s)$ удовлетворяет следующему условию: функция $\text{Ref}(\sqrt{s})$, $s \geq 0$, является вполне монотонной, то есть она бесконечно дифференцируема и при каждом $n \geq 1$*

$$(-1)^n \frac{d^n}{dt^n} \text{Ref}(\sqrt{s}) \geq 0, \quad t \geq 0.$$

Д о к а з а т е л ь с т в о. В терминах случайных величин (отождествляя случайные величины и их функции распределения, а, по сути, рассматривая классы эквивалентности случайных величин, относя к одному классу эквивалентности все случайные величины с одной и той же функцией распределения) можно дать следующее определение класса $\mathfrak{Q}$: этот класс содержит те и только те случайные величины X , для каждой из которых найдутся независимая от X случайная величина Z такая, что $P(Z = 1) = 1 - P(Z = -1) = \frac{1}{2}$, и пара независимых случайных величин W и Y , такая, что W имеет стандартное нормальное распределение, $P(Y \geq 0) = 1$ и

$$X \cdot Z \stackrel{d}{=} W \cdot Y. \quad (3.5.7)$$

Переписав (3.5.7) в терминах характеристических функций, получим

$$\begin{aligned} \text{Ref}(s) &= E \exp\{isWY\} = \int_0^\infty \exp\left\{-\frac{s^2 y^2}{2}\right\} dP(Y < y) = \\ &= E \exp\left\{-s^2 \left(\frac{Y^2}{2}\right)\right\}, \quad s \in \mathbb{R}. \end{aligned} \quad (3.5.8)$$

Обозначим $V = Y^2/2$, $u = s^2$. Тогда, очевидно, $u \geq 0$ и $s = \pm\sqrt{u}$. В левой части соотношения (3.5.8) стоит четная функция, поэтому всегда

$$\operatorname{Ref}(-\sqrt{u}) = \operatorname{Ref}(\sqrt{u}).$$

Таким образом, соотношение (3.5.8) эквивалентно следующему:

$$\operatorname{Ref}(\sqrt{u}) = \mathbb{E} \exp\{-uV\}, \quad u \geq 0. \quad (3.5.9)$$

В правой части (3.5.9) стоит преобразование Лапласа–Стилтьеса некоторой неотрицательной случайной величины V . Поэтому можно заключить, что $F \in \mathfrak{Q}$, если и только если соответствующая функции распределения F характеристическая функция f удовлетворяет условию: функция $\operatorname{Ref}(\sqrt{u})$ является преобразованием Лапласа–Стилтьеса. Но, согласно теореме С. Н. Бернштейна (см., например, раздел XIII.4 в книге (Феллер, 1984)), для того чтобы некоторая функция была преобразованием Лапласа–Стилтьеса, необходимо и достаточно, чтобы (а) она была вполне монотонной и (б) в нуле она была равна единице. В нашем случае условие (б) выполнено автоматически, поскольку f – характеристическая функция. Условие (iii) при этом обеспечивает концентрацию F на неотрицательной полуоси. Теорема доказана.

Заметим, что условие (iv) теоремы 3.5.3 представляет собой критерий представимости рандомизационной симметризации *произвольного* распределения F (не обязательно сосредоточенного на полуоси) с характеристической функцией f в виде масштабной смеси нормальных законов.

Класс $\mathfrak{Q}$ не пуст. Например, он содержит экспоненциальное распределение. Действительно, характеристическая функция экспоненциального распределения с параметром λ , как известно, равна

$$f(s) = \frac{1}{1 - \frac{is}{\lambda}} = \frac{1 + \frac{is}{\lambda}}{\left(1 - \frac{is}{\lambda}\right)\left(1 + \frac{is}{\lambda}\right)} = \frac{1 + \frac{is}{\lambda}}{1 + \frac{s^2}{\lambda^2}}.$$

Стало быть,

$$\operatorname{Ref}(s) = \frac{\lambda^2}{\lambda^2 + s^2}, \quad s \in \mathbb{R},$$

так что

$$\operatorname{Ref}(\sqrt{s}) = \frac{\lambda^2}{\lambda^2 + s}, \quad s \geq 0.$$

Но последняя функция есть не что иное как преобразование Лапласа–Стилтьеса показательного распределения с параметром λ^2 и, стало быть,

вполне монотонна. По теореме 3.5.3 рандомизационная симметризация экспоненциального распределения представима в виде масштабной смеси нормальных законов. Более того, так как в данном случае $\operatorname{Ref}(s) = |f(s)|^2$, то рандомизационная симметризация показательного распределения совпадает с его сверточной симметризацией, которая, как уже было показано, представляет собой распределение Лапласа.

Строго говоря, рандомизационная симметризация не приводит к сужению класса масштабных смесей нормальных законов. В самом деле, можно заметить, что все масштабные смеси нормальных законов абсолютно непрерывны. Воспользовавшись уже приведенным соотношением $q(x) = \frac{1}{2}p(|x|)$, связывающим плотности $p(x)$ случайной величины $X \geq 0$ и $q(x)$ ее рандомизационной симметризации $\tilde{X}$, каждой масштабной смеси нормальных плотностей $q(x)$ (которая, очевидно, симметрична) можно поставить в соответствие ее “половинку” $p(x)$, отличную от нуля на неотрицательной полуоси. Ясно, что при этом $p(x)$ соответствует распределению из класса $\mathfrak{Q}$. Это означает, что соответствие между классом масштабных смесей нормальных законов и классом $\mathfrak{Q}$ взаимно однозначно.

Из соотношения $q(x) = \frac{1}{2}p(|x|)$ вытекает, что классу $\mathfrak{Q}$ принадлежат только унимодальные законы, имеющие моду в нуле.

Из этого замечания, в свою очередь, вытекает, что классы $\mathfrak{Q}$ и $\mathfrak{P}$ не совпадают. Действительно, в отличие от первой модели симметризации, рандомизационные симметризации гамма-распределений с параметром формы, превосходящим единицу, не могут быть представлены в виде масштабных смесей нормальных законов, так как рандомизационные симметризации указанных распределений не унимодальны.

В отличие от первой модели симметризации, в рамках второй модели, очевидно, нормальное распределение допустимо как распределение приращений финансовых индексов. Действительно, несложно видеть, что если $G(x) = (2\Phi(\max\{x, 0\}) - 1)$ – функция распределения максимума стандартного винеровского процесса на единичном отрезке, то рандомизационная симметризация функции распределения $G(x)$ совпадает с $\Phi(x)$.

3.6. Взаимосвязь поведения хвостов масштабных смесей нормальных законов и соответствующих смешивающих распределений

Как было показано в предыдущих разделах, класс масштабных смесей нормальных законов содержит законы, хвосты которых могут убывать довольно медленно. В данном разделе будет рассмотрена взаимосвязь поведения хвостов масштабных смесей нормальных законов с поведением хвостов соответствующих смешивающих распределений. При этом внимание будет сосредоточено на наиболее интересных типах убывания хвостов: степенном (типа распределения Стюдента), экспоненциальном (типа симметризованного гамма-распределения) и экспоненциально-степенном (типа симметризованного распределения Вейбулла). Материал данного раздела базируется на работах (Barndorff-Nielsen, Jensen and Sørensen, 1982), (Антонов и Кокшаров, 2006) и (Кокшаров, 2007). В первой из упомянутых работ используется аппарат тауберовых теорем, вполне естественный для рассматриваемых задач, поскольку, как было показано в разделе 3.5, характеристические функции масштабных смесей нормальных законов с точностью до степенного преобразования аргумента представляют собой преобразования Лапласа–Стилтьеса смешивающих законов. Однако здесь, следуя работам (Антонов и Кокшаров, 2006) и (Кокшаров, 2007), будут использоваться прямые методы анализа хвостов, поскольку такой подход приводит к более общим результатам.

В данной главе рассматриваются смеси нормального закона по масштабному параметру и изучается зависимость между скоростью убывания “хвоста” смешивающего распределения и скоростью убывания “хвоста” полученной смеси. Такая задача представляет интерес с точки зрения оценивания вероятностей критических значений, то есть “хвостов” распределения.

Будем рассматривать случайные величины $Y = X\sqrt{U}$, где случайная величина X стандартно нормально распределена, а случайная величина U – неотрицательна ($P(U \geq 0) = 1$), причем случайные величины X и U независимы. Обозначим функцию распределения случайной величины Y через $F(x) \equiv P(Y < x)$, а функцию распределения случайной величины U – через $G(x) \equiv P(U < x)$. Как обычно, функцию распределения стандартного нормального закона будем обозначать $\Phi(x)$. Функция распределения

$F(x)$ представима в виде

$$F(x) = \int_0^{\infty} \Phi\left(\frac{x}{\sqrt{y}}\right) dG(y) = \int_0^{\infty} \frac{1}{\sqrt{2\pi y}} \int_{-\infty}^x e^{-\frac{u^2}{2y}} du dG(y), \quad (3.6.1)$$

при этом ее характеристическая функция имеет вид

$$l(t) = \int_0^{\infty} \exp\left\{-\frac{t^2 y}{2}\right\} dG(y). \quad (3.6.2)$$

3.6.1. Экспоненциальное убывание хвоста смеси

Пусть $q > 0$ и $\mathcal{H}$ – класс положительных функций $h(x)$, неубывающих на $[q, \infty)$ и таких, что при $\lambda > 0$

$$\lim_{x \rightarrow \infty} \frac{h(\lambda x)}{h(x)} = \phi(\lambda),$$

где $\phi(\lambda)$ – конечная функция, положительная на некотором множестве положительной меры Лебега. Из леммы 1 и теоремы гл. VIII §8 в (Феллер, 1984) вытекает, что все функции класса $\mathcal{H}$ являются правильно меняющимися (см., например, (Феллер, 1984), гл. VIII, и (Сенета, 1985)), то есть для любой функции $h(x) \in \mathcal{H}$ справедливо представление

$$h(x) = x^\rho L(x), \quad (3.6.3)$$

где $\rho > 0$ – конечное число, а $L(x)$ удовлетворяет условию

$$\lim_{x \rightarrow \infty} \frac{L(\lambda x)}{L(x)} = 1,$$

то есть является медленно меняющейся на бесконечности. Из леммы 2 гл. VIII §8 в (Феллер, 1984) следует, что функция $L(x)$ не может расти быстрее, чем x в любой положительной степени.

ТЕОРЕМА 3.6.1 Пусть функция распределения $F(x)$ определена формулой (3.6.1), $h(x) \in \mathcal{H}$, $0 \leq \gamma < \infty$. Для справедливости соотношения

$$\liminf_{x \rightarrow \infty} \frac{-\ln[1 - F(x)]}{xh(\ln x)} = \frac{1}{\gamma} \quad (3.6.4)$$

необходимо и достаточно, чтобы

$$\liminf_{y \rightarrow \infty} \frac{-\ln [1 - G(y)]}{yh^2(\ln y)} = \frac{1}{2\gamma^2}. \quad (3.6.5)$$

Д о к а з а т е л ь с т в о. Для произвольного $z > 0$ справедливы неравенства

$$\begin{aligned} 1 - F(x) &\geq \int_z^\infty \left[1 - \Phi\left(\frac{x}{\sqrt{y}}\right)\right] dG(y) \geq \int_z^\infty \left[1 - \Phi\left(\frac{x}{\sqrt{z}}\right)\right] dG(y) \geq \\ &\geq [1 - G(z)] \cdot \left[1 - \Phi\left(\frac{x}{\sqrt{z}}\right)\right]. \end{aligned} \quad (3.6.6)$$

Тогда для произвольного $\alpha > 1$

$$\begin{aligned} 1 - F(x) &\geq [1 - G(z)] \cdot \left[1 - \Phi\left(\frac{x}{\sqrt{z}}\right)\right] \geq [1 - G(z)] \cdot \left[\frac{1}{\sqrt{2\pi}z} \int_x^{\alpha x} \exp\left\{-\frac{u^2}{2z}\right\} du\right] \geq \\ &\geq [1 - G(z)] \frac{x(\alpha - 1)}{\sqrt{2\pi}z} \exp\left\{-\frac{\alpha^2 x^2}{2z}\right\}. \end{aligned} \quad (3.6.7)$$

1. Докажем, что из справедливости (3.6.4) следует

$$\liminf_{z \rightarrow \infty} \frac{-\ln(1 - G(z))}{zh^2(\ln z)} \geq \frac{1}{2\gamma^2}.$$

Допустим, что соотношение (3.6.3) справедливо при $0 \leq \gamma < \infty$. Тогда для произвольного $\beta > \gamma$ при достаточно больших $x > x(\beta)$ справедливо неравенство

$$1 - F(x) \leq \exp\left\{-\frac{xh(\ln x)}{\beta}\right\}. \quad (3.6.8)$$

Возьмем произвольное $\alpha \in (1, \sqrt{2})$ и положим

$$z = \frac{x\beta}{h(\ln x)}.$$

При достаточно больших $x > x(\alpha)$ имеем:

$$1 - F(x) \geq \exp\left\{-\frac{\alpha^2 x h(\ln x)}{2\beta}\right\} \left[1 - G\left(\frac{x\beta}{h(\ln x)}\right)\right]. \quad (3.6.9)$$

Из неравенств (3.6.8) и (3.6.9) вытекает, что

$$1 - G(z) \leq \exp \left\{ - \frac{(2 - \alpha^2)xh(\ln x)}{2\beta} \right\} \leq \exp \left\{ - \frac{(2 - \alpha^2)zh^2(\ln z - \ln \beta)}{2\beta^2} \right\},$$

где $x > x(\alpha, \beta)$. Так как α можно взять сколь угодно близким к единице, а β – близким к γ , то

$$\liminf_{z \rightarrow \infty} \frac{-\ln[1 - G(z)]}{zh^2(\ln z)} \geq \frac{1}{2\gamma^2}. \quad (3.6.10)$$

2. Докажем, что из справедливости (3.6.5) следует, что

$$\liminf_{x \rightarrow \infty} \frac{-\ln[1 - F(x)]}{xh(\ln x)} \geq \frac{1}{\gamma}.$$

Пусть (3.6.5) справедливо при $0 \leq \gamma < \infty$. Тогда для произвольного $\beta > \gamma$ при достаточно больших $y > y(\beta)$ имеем:

$$1 - G(y) \leq \exp \left\{ - \frac{yh^2(\ln y)}{2\beta^2} \right\}. \quad (3.6.11)$$

Из неравенства Чебышева следует, что

$$\frac{1}{\sqrt{2\pi y}} \int_x^\infty \exp \left\{ - \frac{u^2}{2y} \right\} du \leq \sqrt{n} \exp \left\{ - \frac{x^2}{2y} \cdot \left(\frac{n-1}{n} \right) \right\}, \quad x > 0. \quad (3.6.12)$$

Рассмотрим функцию

$$g(k, x) = \frac{\nu^k \beta x}{2h(\ln x)},$$

где $k = 0, 1, \dots, n+1$, $\nu = 4^{1/n}$. Тогда в силу (3.6.3) и (3.6.11)

$$1 - G(g(k, x)) \leq \exp \left\{ - \frac{\nu^k \beta x}{4\beta^2 h(\ln x)} h^2 \left(\ln \left(\frac{\nu^k \beta x}{2h(\ln x)} \right) \right) \right\}.$$

Функция $h(x)$ является правильно меняющейся на бесконечности, поэтому для $0 < \delta < 1$ при достаточно больших $x > x(n)$ справедливо неравенство

$$h^2 \left(\ln \left(\frac{\nu^k \beta x}{2h(\ln x)} \right) \right) h^{-1}(\ln x) \geq (1 - \delta) h(\ln x).$$

Возьмем $\delta = 1 - \frac{1}{\nu}$. Тогда

$$1 - G(g(k, x)) \leq \exp \left\{ - \frac{\nu^{k-1} x h(\ln x)}{4\beta} \right\}, \quad (3.6.13)$$

где $k = 0, 1, 2, \dots, n+1$, $x > x(n)$. Из неравенства (3.6.12) следует, что

$$\begin{aligned} 1 - F(x) &= \int_0^\infty \frac{1}{\sqrt{2\pi y}} \int_x^\infty \exp \left\{ - \frac{u^2}{2y} \right\} du dG(y) \leq \\ &\leq \sqrt{n} \int_0^{g(0,x)} \exp \left\{ - \frac{x^2}{2y} \left(\frac{n-1}{n} \right) \right\} dG(y) + \sqrt{n} \sum_{k=0}^n \int_{g(k,x)}^{g(k+1,x)} \exp \left\{ - \frac{x^2}{2y} \left(\frac{n-1}{n} \right) \right\} dG(y) + \\ &\quad + \sqrt{n} \int_{g(n+1,x)}^\infty \exp \left\{ - \frac{x^2}{2y} \left(\frac{n-1}{n} \right) \right\} dG(y) \leq \\ &\leq \sqrt{n} \exp \left\{ - \frac{x h(\ln x)(n-1)}{n\beta} \right\} [G(g(0, x)) - G(0)] + \\ &\quad + \sqrt{n} \sum_{k=0}^n \exp \left\{ - \frac{x h(\ln x)(n-1)}{n\nu^{k+1}\beta} \right\} [G(g(k+1, x)) - G(g(k, x))] + \\ &\quad + \sqrt{n} [1 - G(g(n+1, x))]. \end{aligned} \quad (3.6.14)$$

Заметим, что

$$\frac{1}{\nu^{k+1}} \left(\frac{n-1}{n} \right) + \frac{\nu^{k-1}}{4} \geq \frac{1}{\nu} \sqrt{\frac{n-1}{n}} = \frac{1}{4^{1/n}} \sqrt{\frac{n-1}{n}}.$$

Тогда в силу (3.6.13) и (3.6.14) справедлива оценка

$$1 - F(x) \leq \sqrt{n} (n+3) \exp \left\{ - \frac{x h(\ln x)}{4^{1/n} \sqrt{\frac{n}{n-1}} \beta} \right\}, \quad x > x(n), \quad (3.6.15)$$

в которой n можно взять сколь угодно большим, а β сколь угодно близким к γ . Поэтому из (3.6.15) вытекает, что

$$\liminf_{x \rightarrow \infty} \frac{-\ln [1 - F(x)]}{x h(\ln x)} \geq \frac{1}{\gamma}. \quad (3.6.16)$$

3. Докажем, что в соотношениях (3.6.10) и (3.6.16) знак неравенства можно заменить на знак равенства. Допустим, что в неравенстве (3.6.10)

(неравенстве (3.6.16)) левая часть строго больше правой. Тогда, повторив рассуждения части 2 (рассуждения части 1), докажем, что в соотношении (3.6.16) (в соотношении (3.6.10)) выполнено строгое неравенство, но это противоречит утверждению (3.6.4) (утверждению (3.6.5)).

Теорема доказана.

Из неравенства (3.6.9) вытекает

СЛЕДСТВИЕ 3.6.1. В условиях теоремы 3.6.1 справедливо неравенство

$$\limsup_{x \rightarrow \infty} \frac{-\ln [1 - F(x)]}{xh(\ln x)} \leq \sqrt{2 \limsup_{y \rightarrow \infty} \frac{-\ln [1 - G(y)]}{yh^2(\ln y)}}.$$

СЛЕДСТВИЕ 3.6.2. Если выполнены условия теоремы 3.6.1 и существует предел

$$A = \lim_{x \rightarrow \infty} \frac{-\ln [1 - F(x)]}{xh(\ln x)}, \quad (3.6.17)$$

то справедливо равенство

$$A = \sqrt{2 \liminf_{y \rightarrow \infty} \frac{-\ln [1 - G(y)]}{yh^2(\ln y)}}.$$

СЛЕДСТВИЕ 3.6.3. Если выполнены условия теоремы 3.6.1 и существует предел

$$B = \lim_{y \rightarrow \infty} \frac{-\ln [1 - G(y)]}{yh^2(\ln y)},$$

то существует предел A (см. (3.6.17)), и справедливо соотношение

$$A = \sqrt{2B}.$$

3.6.2. Экспоненциально-степенное убывание хвоста смеси

Приведенные выше результаты можно легко перенести на более общий случай. Обозначим через $\mathcal{L}$ класс положительных медленно меняющихся функций

$$\mathcal{L} = \left\{ L : \mathbb{R} \rightarrow \mathbb{R}_+ \mid \lim_{x \rightarrow \infty} \frac{L(\lambda x)}{L(x)} = 1, \forall \lambda > 0 \right\},$$

где $\mathbb{R}_+$ – множество положительных действительных чисел.

ТЕОРЕМА 3.6.2. Пусть функция распределения $F(x)$ определена формулой (3.6.1), $L(x) \in \mathcal{L}$, $0 < \rho < 2$, $0 < \gamma < \infty$. Для справедливости соотношения

$$\liminf_{x \rightarrow \infty} \frac{-\ln [1 - F(x)]}{x^\rho L(x)} = \frac{1}{\gamma} \quad (3.6.18)$$

необходимо и достаточно, чтобы

$$\liminf_{y \rightarrow \infty} \frac{-\ln [1 - G(y)]}{y^{\frac{\rho}{2-\rho}} [L(y^{\frac{1}{2-\rho}})]^{\frac{2}{2-\rho}}} = \frac{1}{2\gamma^{\frac{2}{2-\rho}}}. \quad (3.6.19)$$

Доказательство. Воспользуемся неравенством (3.6.7). Пусть $0 < \beta < \infty$ – некоторое число. Положим

$$z = \frac{x^{2-\rho}\beta}{L(x)}.$$

Тогда при достаточно больших $x > x(\rho, L)$ имеем:

$$\frac{x(\alpha - 1)}{\sqrt{2\pi z}} = \frac{x^{\rho/2} \sqrt{L(x)}(\alpha - 1)}{\sqrt{2\pi\beta}} \geq 1.$$

Следовательно,

$$\begin{aligned} 1 - F(x) &\geq \exp \left\{ -\frac{\alpha^2 x^\rho L(x)}{2\beta} \right\} \left[1 - G\left(\frac{x^{2-\rho}\beta}{L(x)}\right) \right] \frac{x^{\rho/2} \sqrt{L(x)}(\alpha - 1)}{\sqrt{2\pi\beta}} \geq \\ &\geq \exp \left\{ -\frac{\alpha^2 x^\rho L(x)}{2\beta} \right\} \left[1 - G\left(\frac{x^{2-\rho}\beta}{L(x)}\right) \right]. \end{aligned} \quad (3.6.20)$$

1. Докажем, что из справедливости (3.6.18) следует

$$\liminf_{z \rightarrow \infty} \frac{-\ln [1 - G(z)]}{z^{\frac{\rho}{2-\rho}} (L(z^{\frac{1}{2-\rho}}))^{\frac{2}{2-\rho}}} \geq \frac{1}{2\gamma^{\frac{2}{2-\rho}}}.$$

Допустим, что соотношение (3.6.18) справедливо при $0 < \gamma < \infty$. Тогда для любого $\beta > \gamma$ и достаточно больших $x > x_1(\rho, \beta, L)$

$$1 - F(x) \leq \exp \left\{ -\frac{x^\rho L(x)}{\beta} \right\}. \quad (3.6.21)$$

Отсюда, используя неравенства (3.6.20) и (3.6.21), получаем

$$1 - G(z) \leq \exp \left\{ -\frac{(2 - \alpha^2)x^\rho L(x)}{2\beta} \right\} \leq \exp \left\{ -\frac{(2 - \alpha^2)z^{\frac{\rho}{2-\rho}} L^{\frac{2}{2-\rho}}(x)}{2\beta^{\frac{2}{2-\rho}}} \right\},$$

где $x > \max \{x(\rho, L), x_1(\rho, \beta, L)\}$. Выразив x через z , получим неравенство

$$1 - G(z) \leq \exp \left\{ - \frac{(2 - \alpha^2) z^{\frac{\rho}{2-\rho}} (L(z^{\frac{1}{2-\rho}}))^{\frac{2}{2-\rho}}}{2\beta^{\frac{2}{2-\rho}}} \left(\frac{L(z^{\frac{1}{2-\rho}} \beta^{-\frac{1}{2-\rho}} L^{\frac{1}{2-\rho}}(x))}{L(z^{\frac{1}{2-\rho}})} \right)^{\frac{2}{2-\rho}} \right\}. \quad (3.6.22)$$

Так как $L(x) \in \mathcal{L}$, то для любого $\varepsilon \in (0, 1)$ существует такое $x_2 = x_2(\rho, \beta, L)$, что при $x > x_2$

$$\frac{L(z^{\frac{1}{2-\rho}} \beta^{-\frac{1}{2-\rho}} L^{\frac{1}{2-\rho}}(x))}{L(z^{\frac{1}{2-\rho}})} \geq (1 - \varepsilon)^{\frac{2-\rho}{2}}. \quad (3.6.23)$$

Выберем α так, чтобы $\alpha \leq \sqrt{1 + \varepsilon}$. Для таких α выполняется соотношение

$$2 - \alpha^2 \geq 1 - \varepsilon. \quad (3.6.24)$$

При указанных ε , α и $x > \max \{x(\rho, L), x_1(\rho, \beta, L), x_2(\rho, \beta, L)\}$ из (3.6.21), (3.6.22) и (3.6.23) имеем:

$$1 - G(z) \leq \exp \left\{ - \frac{z^{\frac{\rho}{2-\rho}} L^{\frac{2}{2-\rho}}(z^{\frac{1}{2-\rho}})}{2\beta^{\frac{2}{2-\rho}}} (1 - \varepsilon)^2 \right\},$$

то есть

$$\frac{-\ln [1 - G(z)]}{z^{\frac{\rho}{2-\rho}} L^{\frac{2}{2-\rho}}(z^{\frac{1}{2-\rho}})} \geq \frac{(1 - \varepsilon)^2}{2\beta^{\frac{2}{2-\rho}}}.$$

Так как β можно выбрать произвольно близким к γ , а ε можно выбрать произвольно близким к нулю, то

$$\liminf_{z \rightarrow \infty} \frac{-\ln [1 - G(z)]}{z^{\frac{\rho}{2-\rho}} L^{\frac{2}{2-\rho}}(z^{\frac{1}{2-\rho}})} \geq \frac{1}{2\gamma^{\frac{2}{2-\rho}}}.$$

Первая часть доказательства завершена.

2. Докажем, что из справедливости (3.6.19) следует

$$\liminf_{x \rightarrow \infty} \frac{-\ln [1 - F(x)]}{x^\rho L(x)} \geq \frac{1}{\gamma}.$$

Пусть соотношение (3.6.19) справедливо при $0 < \gamma < \infty$. Тогда для произвольного $\beta > \gamma$ и достаточно больших $y > y(\beta, L)$ справедливо неравенство

$$1 - G(y) \leq \exp \left\{ - \frac{y^{\frac{\rho}{2-\rho}} L^{\frac{2}{2-\rho}}(y^{\frac{1}{2-\rho}})}{2\beta^{\frac{2}{2-\rho}}} \right\}. \quad (3.6.25)$$

Кроме того, для любых $n > 1$ и $x > 0$ справедлива оценка

$$\begin{aligned} 1 - \Phi\left(\frac{x}{\sqrt{y}}\right) &= \frac{1}{\sqrt{2\pi y}} \int_x^\infty \exp\left\{-\frac{u^2}{2y}\right\} du \leq \\ &\leq \frac{\sqrt{n}}{\sqrt{2\pi y n}} \int_x^\infty \exp\left\{-\frac{u^2}{2yn}\right\} \exp\left\{-\frac{u^2}{2y} \cdot \frac{n-1}{n}\right\} du \leq \\ &\leq \frac{\sqrt{n}}{2} \exp\left\{-\frac{x^2}{2y} \cdot \left(\frac{n-1}{n}\right)\right\}. \end{aligned} \quad (3.6.26)$$

Рассмотрим функцию $g(k, x) = \frac{1}{2} \nu^{k \frac{2-\rho}{\rho}} \beta x^{2-\rho} / L(x)$, $k = 0, 1, \dots, n$, $\nu = 2^{\frac{2}{2-\rho}n}$. В силу (3.6.25) при достаточно больших $x > x_1(\rho, \beta, L, n)$

$$\begin{aligned} 1 - G(g(k, x)) &\leq \\ &\leq \exp\left\{-\frac{\nu^k x^\rho L(x)}{2^{\frac{2}{2-\rho}} \beta} \left[L\left(\frac{\nu^{k \frac{1}{\rho}} \beta^{\frac{1}{2-\rho}} x}{2^{\frac{1}{2-\rho}} L^{\frac{1}{2-\rho}}(x)}\right) / L(x)\right]^{\frac{2}{2-\rho}}\right\}, \quad k = 1, 2, \dots, n. \end{aligned} \quad (3.6.27)$$

Так как $L \in \mathcal{L}$, то для любого $\varepsilon \in (0, 1)$ при $x > x_2(\rho, \beta, L, n)$

$$\left[L\left(\frac{\nu^{k \frac{1}{\rho}} \beta^{\frac{1}{2-\rho}} x}{2^{\frac{1}{2-\rho}} L^{\frac{1}{2-\rho}}(x)}\right) / L(x)\right]^{\frac{2}{2-\rho}} \geq 1 - \varepsilon, \quad k = 1, 2, \dots, n. \quad (3.6.28)$$

Таким образом, при $x > \max\{x_1(\rho, \beta, L, n), x_2(\rho, \beta, L, n)\}$ для указанного ε из (3.6.27) и (3.6.28) вытекает, что

$$1 - G(g(k, x)) \leq \exp\left\{-\frac{\nu^k x^\rho L(x)}{2^{\frac{2}{2-\rho}} \beta} (1 - \varepsilon)\right\}, \quad k = 1, 2, \dots, n. \quad (3.6.29)$$

Из неравенства (3.6.26) следует, что

$$\begin{aligned} 1 - F(x) &= \int_0^\infty \left[1 - \Phi\left(\frac{x}{\sqrt{y}}\right)\right] dG(y) \leq \frac{\sqrt{n}}{2} \int_0^{g(0, x)} \exp\left\{-\frac{x^2}{2y} \cdot \frac{n-1}{n}\right\} dG(y) + \\ &+ \frac{\sqrt{n}}{2} \sum_{k=0}^{n-1} \int_{g(k, x)}^{g(k+1, x)} \exp\left\{-\frac{x^2}{2y} \cdot \frac{n-1}{n}\right\} dG(y) + \frac{\sqrt{n}}{2} \int_{g(n, x)}^\infty \exp\left\{-\frac{x^2}{2y} \cdot \frac{n-1}{n}\right\} dG(y) \leq \\ &\leq \frac{\sqrt{n}}{2} \exp\left\{-\frac{x^\rho L(x)(n-1)}{n\beta}\right\} [G(g(0, x)) - G(0)] + \end{aligned}$$

$$\begin{aligned}
& + \frac{\sqrt{n}}{2} \sum_{k=0}^{n-1} \left(\exp \left\{ - \frac{x^\rho L(x)(n-1)}{n\nu^{(k+1)\frac{2-\rho}{\rho}}\beta} (1-\varepsilon) \right\} [G(g(k+1, x)) - G(g(k, x))] + \right. \\
& \left. + \frac{\sqrt{n}}{2} [1 - G(g(n, x))] \right). \quad (3.6.30)
\end{aligned}$$

Заметим, что

$$\begin{aligned}
& \frac{1}{\nu^{(k+1)\frac{2-\rho}{\rho}}} \binom{n-1}{n} + \frac{\nu^k}{2^{\frac{2}{2-\rho}}} \geq \frac{1}{\nu^{k\frac{1-\rho}{\rho} + \frac{2-\rho}{2\rho}}} \left(\frac{n-1}{n} \cdot \frac{1}{\nu^{\frac{k}{\rho} + \frac{2-\rho}{2\rho}}} + \frac{\nu^{\frac{k}{\rho} + \frac{2-\rho}{2\rho}}}{2^{\frac{2}{2-\rho}}} \right) \geq \\
& \geq \frac{2}{\nu^{k\frac{1-\rho}{\rho} + \frac{2-\rho}{2\rho}}} \sqrt{\frac{n-1}{n} \cdot \frac{1}{2^{\frac{2}{2-\rho}}}} \geq \frac{2^{1-\frac{1}{2-\rho}-\frac{1-\rho}{\rho}}}{\nu^{\frac{2-\rho}{2\rho}}} \sqrt{\frac{n-1}{n}} \geq \\
& \geq \frac{2^{\frac{2(\rho-1)^2}{(\rho-2)\rho}}}{\nu^{\frac{2-\rho}{2\rho}}} \sqrt{\frac{n-1}{n}} \geq \frac{1}{2^{\frac{2-\rho}{2n\rho}}} \sqrt{\frac{n-1}{n}}.
\end{aligned}$$

Тем самым k -ый элемент суммы в правой части (3.6.30) (обозначим его A_k) может быть оценен сверху следующим образом:

$$A_k \leq \exp \left\{ - \frac{x^\rho L(x)}{2^{\frac{2-\rho}{2n\rho}} \sqrt{\frac{n}{n-1}} \beta} (1-\varepsilon) \right\} \leq \exp \left\{ - \frac{x^\rho L(x)}{2^{\frac{2-\rho}{2n\rho}} \frac{n}{n-1} \beta} (1-\varepsilon) \right\}. \quad (3.6.31)$$

Кроме того,

$$\exp \left\{ - \frac{x^\rho L(x)(n-1)}{n\beta} \right\} \leq \exp \left\{ - \frac{x^\rho L(x)}{2^{\frac{2-\rho}{2n\rho}} \frac{n}{n-1} \beta} (1-\varepsilon) \right\}, \quad (3.6.32)$$

$$1 - G(g(n, x)) \leq \exp \left\{ - \frac{x^\rho L(x)}{\beta} (1-\varepsilon) \right\} \leq \exp \left\{ - \frac{x^\rho L(x)}{2^{\frac{2-\rho}{2n\rho}} \frac{n}{n-1} \beta} (1-\varepsilon) \right\}. \quad (3.6.33)$$

В силу соотношений (3.6.30) – (3.6.33) справедлива оценка

$$1 - F(x) \leq \frac{\sqrt{n}}{2} (n+2) \exp \left\{ - \frac{x^\rho L(x)}{2^{\frac{2-\rho}{2n\rho}} \frac{n}{n-1} \beta} (1-\varepsilon) \right\}, \quad (3.6.34)$$

где $x > \max \{x_1(\rho, \beta, L, n), x_2(\rho, \beta, L, n)\}$. Выберем N_0 так, чтобы при $n \geq N_0$

$$2^{\frac{2-\rho}{2n\rho}} \cdot \frac{n}{n-1} \leq \frac{1}{1-\varepsilon}. \quad (3.6.35)$$

Тогда

$$1 - F(x) \leq \frac{\sqrt{N_0}}{2} (N_0 + 2) \exp \left\{ - \frac{x^\rho L(x)}{\beta} (1 - \varepsilon)^2 \right\}. \quad (3.6.36)$$

Так как β можно выбрать сколь угодно близким к γ , то

$$\liminf_{x \rightarrow \infty} \frac{-\ln [1 - F(x)]}{x^\rho L(x)} \geq \frac{1}{\gamma}. \quad (3.6.37)$$

Вторая часть доказательства завершена.

3. Докажем, что в соотношениях (3.6.24) и (3.6.37) знак неравенства можно заменить на знак равенства. Допустим, что в неравенстве (3.6.24) (неравенстве (3.6.37)) левая часть строго больше правой. Тогда, повторив рассуждения пункта 2 (рассуждения пункта 1), докажем, что в соотношении (3.6.37) (в соотношении (3.6.24)) стоит строгое неравенство. Но это противоречит утверждению (3.6.18) (утверждению (3.6.19)).

Тем самым теорема 3.6.2 полностью доказана.

СЛЕДСТВИЕ 3.6.4. *В условиях теоремы 3.6.2 справедливо неравенство*

$$\limsup_{x \rightarrow \infty} \frac{-\ln [1 - F(x)]}{x^\rho L(x)} \leq \left(2 \limsup_{y \rightarrow \infty} \frac{-\ln [1 - G(y)]}{y^{\frac{\rho}{2-\rho}} L^{\frac{2}{2-\rho}} (y^{\frac{1}{2-\rho}})} \right)^{\frac{2-\rho}{2}}.$$

Это неравенство является следствием неравенства (3.6.20).

СЛЕДСТВИЕ 3.6.5. *Если выполнены условия теоремы 3.6.2 и существует предел*

$$A = \lim_{x \rightarrow \infty} \frac{-\ln [1 - F(x)]}{x^\rho L(x)}, \quad (3.6.38)$$

то справедливо равенство

$$A = \left(2 \liminf_{y \rightarrow \infty} \frac{-\ln [1 - G(y)]}{y^{\frac{\rho}{2-\rho}} L^{\frac{2}{2-\rho}} (y^{\frac{1}{2-\rho}})} \right)^{\frac{2-\rho}{2}}.$$

СЛЕДСТВИЕ 3.6.6. *Если выполнены условия теоремы 3.6.2 и существует предел*

$$B = \lim_{y \rightarrow \infty} \frac{-\ln [1 - G(y)]}{y^{\frac{\rho}{2-\rho}} L^{\frac{2}{2-\rho}} (y^{\frac{1}{2-\rho}})},$$

то существует предел A (см. (3.6.38)) и справедливо соотношение

$$A = (2B)^{\frac{2-\rho}{2}}.$$

ПРИМЕР 3.6.1 Возьмем в качестве смешивающего распределения показательное, имеющее плотность

$$g(x) = \begin{cases} 0, & x \leq 0, \\ e^{-x}, & x > 0. \end{cases} \quad (3.6.39)$$

Тогда характеристическая функция полученной смеси имеет вид

$$\varphi(t) = \frac{1}{1 + \frac{t^2}{2}},$$

то есть она совпадает с характеристической функцией двустороннего показательного распределения (распределения Лапласа), имеющего плотность

$$p(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}, \quad -\infty < x < \infty.$$

Таким образом, в обозначениях теоремы 3.6.2 имеем:

$$1 - G(x) = e^{-x}, \quad x > 0, \quad (3.6.40)$$

$$1 - F(x) = \frac{1}{2} e^{-x\sqrt{2}}, \quad x > 0. \quad (3.6.41)$$

Предположим, что известно только соотношение $1 - G(x) \sim e^{-x}$ ($x \rightarrow \infty$) для смешивающего распределения. Тогда, применяя теорему 3.6.2 с $\rho = 1$, $L(x) \equiv 1$ и $\gamma = \frac{1}{\sqrt{2}}$, получим, что

$$1 - F(x) \sim \frac{1}{2} e^{-x\sqrt{2}}, \quad x \rightarrow \infty,$$

что отлично согласуется с (3.6.41).

3.6.3. Степенное убывание хвоста смеси

В предыдущих разделах был рассмотрен случай экспоненциального убывания “хвостов” смешивающего распределения. Оказывается, аналогичные рассуждения можно провести и для случая степенного убывания “хвостов”.

ЛЕММА 3.6.1. Пусть функция распределения $F(x)$ определена формулой (3.6.1), $\rho > 0$, $C \geq 0$ и

$$\limsup_{x \rightarrow \infty} [1 - F(x)]x^\rho = C, \quad (3.6.42)$$

тогда

$$\limsup_{y \rightarrow \infty} [1 - G(y)]y^{\frac{\rho}{2}} \leq 2C. \quad (3.6.43)$$

Д о к а з а т е л ь с т в о. Для любого $z > 0$ справедливо неравенство (3.6.6), откуда вытекает, что

$$1 - G(z) \leq \frac{1 - F(x)}{1 - \Phi(x/\sqrt{z})}. \quad (3.6.44)$$

Положим $z = x^\alpha$, где $\alpha > 0$. Очевидно, что, если $\alpha < 2$, то правая часть неравенства (3.6.44) экспоненциально возрастает и ведет себя при $x \rightarrow \infty$ как функция

$$Cx^{-\rho}e^{c_1x^{1-\alpha/2}},$$

где c_1 – некоторая константа. Можно получить лучшую (в смысле порядка) оценку для правой части при $\alpha > 2$. Положим $z = x^{2+\beta}$, где $\beta > 0$. Для любого $\varepsilon > 0$ существует такое x_0 , что для всех $x > x_0$ выполняется неравенство

$$\frac{1}{1 - \Phi(x^{-\frac{\beta}{2}})} \leq 2(1 + \varepsilon). \quad (3.6.45)$$

Далее, в силу (3.6.42) для указанного ε существует такое x_1 , что для всех $x > x_1$ выполняется неравенство

$$1 - F(x) \leq \frac{C(1 + \varepsilon)}{x^\rho}, \quad (3.6.46)$$

тогда для выбранного $\varepsilon > 0$ и всех $x > \max\{x_0, x_1\}$

$$1 - G(z) \leq \frac{2C}{x^\rho}(1 + \varepsilon)^2 \leq \frac{2C}{z^{\frac{\rho}{2+\beta}}}(1 + \varepsilon)^2. \quad (3.6.47)$$

В силу произвольности $\beta > 0$, для любого фиксированного $x > \max\{x_0, x_1\}$ справедливо неравенство

$$1 - G(z) \leq \frac{2C}{z^{\frac{\rho}{2}}}(1 + \varepsilon)^2, \quad (3.6.48)$$

доказывающее лемму.

ЛЕММА 3.6.2. Пусть функция распределения $F(x)$ определена формулой (3.6.1), $\rho > 0$, $C \geq 0$ и

$$\limsup_{y \rightarrow \infty} [1 - G(y)] y^\rho = C, \quad (3.6.49)$$

тогда

$$\limsup_{x \rightarrow \infty} [1 - F(x)] x^{2\rho} \leq \frac{C}{2}. \quad (3.6.50)$$

Д о к а з а т е л ь с т в о. Для любых $n \geq 2$ и $x > 0$ справедливо соотношение (3.6.27), откуда вытекает неравенство

$$1 - \Phi(x) \leq e^{-\frac{3}{8}x^2}. \quad (3.6.51)$$

Положим $z = x^\alpha$, $0 < \alpha < 2$. Тогда

$$\begin{aligned} 1 - F(x) &= \int_0^\infty \left[1 - \Phi\left(\frac{x}{\sqrt{y}}\right)\right] dG(y) = \\ &= \int_0^z \left[1 - \Phi\left(\frac{x}{\sqrt{y}}\right)\right] dG(y) + \int_z^\infty \left[1 - \Phi\left(\frac{x}{\sqrt{y}}\right)\right] dG(y) \leq \\ &\leq 1 - \Phi\left(\frac{x}{\sqrt{z}}\right) + \frac{1}{2} [1 - G(z)] \leq e^{-\frac{3}{8}x^{2-\alpha}} + \frac{1}{2} [1 - G(x^\alpha)]. \end{aligned} \quad (3.6.52)$$

В силу (3.6.48) для произвольного $\varepsilon > 0$ существует такое x_0 , что для всех $x > x_0$ выполняется неравенство

$$1 - G(y) \leq \frac{C(1 + \varepsilon)}{y^\rho}.$$

Функция $g(x) = e^{-x^\nu}$, $\nu > 0$, при больших x убывает быстрее любой отрицательной степени x , то есть

$$\lim_{x \rightarrow \infty} \frac{e^{-x^\nu}}{x^{-\delta}} = 0$$

для любого $\delta > 0$. В силу того, что $1 - G(x)$ убывает при $x \rightarrow \infty$ степенным образом, первое слагаемое в правой части (3.6.52) при больших x будет бесконечно малым по сравнению со вторым. Следовательно, при больших x

$$1 - F(x) \leq \frac{C}{2x^{\rho\alpha}} (1 + \varepsilon).$$

В силу произвольности $0 < \alpha < 2$ для любого фиксированного $x > x_0$ выполнено неравенство

$$1 - F(x) \leq \frac{C}{2x^{2\rho}} (1 + \varepsilon).$$

Лемма доказана.

ТЕОРЕМА 3.6.3. Пусть функция распределения $F(x)$ определена формулой (3.6.1), а $\rho > 0$ и $C \geq 0$ – произвольные числа. Для того чтобы

$$\limsup_{x \rightarrow \infty} [1 - F(x)]x^\rho = C, \quad (3.6.53)$$

необходимо и достаточно, чтобы

$$\limsup_{y \rightarrow \infty} [1 - G(y)]y^{\frac{\rho}{2}} = 2C. \quad (3.6.54)$$

Доказательство. Необходимость. Предположим, что выполнено (3.6.53). Тогда из леммы 3.6.1 следует, что

$$\lim_{y \rightarrow \infty} [1 - G(y)]y^{\frac{\rho}{2}} \leq 2C.$$

Предположим, что

$$\lim_{y \rightarrow \infty} [1 - G(y)]y^{\frac{\rho}{2}} < 2C.$$

Тогда существует $\delta > 0$ такое, что

$$\lim_{y \rightarrow \infty} [1 - G(y)]y^{\frac{\rho}{2}} \leq 2(C - \delta).$$

Повторяя доказательство леммы 3.6.2, получаем, что

$$\lim_{x \rightarrow \infty} [1 - F(x)]x^\rho \leq C - \delta.$$

Следовательно,

$$\lim_{x \rightarrow \infty} [1 - F(x)]x^\rho < C,$$

что противоречит (3.6.53). Таким образом,

$$\lim_{y \rightarrow \infty} [1 - G(y)]y^{\frac{\rho}{2}} = 2C.$$

Достаточность. Пусть имеет место (3.6.54). Тогда из леммы 3.6.2 следует, что

$$\lim_{x \rightarrow \infty} [1 - F(x)]x^\rho \leq C.$$

Аналогично доказательству теоремы в части необходимости, предположим, что

$$\lim_{x \rightarrow \infty} [1 - F(x)]x^\rho < C.$$

Тогда существует $\delta > 0$ такое, что

$$\lim_{x \rightarrow \infty} [1 - F(x)]x^\rho \leq C - \delta.$$

Повторяя доказательство леммы 3.6.1, получаем

$$\lim_{y \rightarrow \infty} [1 - G(y)]y^{\frac{\rho}{2}} \leq 2(C - \delta),$$

то есть

$$\lim_{y \rightarrow \infty} [1 - G(y)]y^{\frac{\rho}{2}} < 2C,$$

что противоречит (3.6.54). Таким образом, доказано, что

$$\lim_{x \rightarrow \infty} [1 - F(x)]x^\rho = C.$$

Теорема доказана.

Приведенные выше утверждения можно обобщить с минимальными изменениями на класс функций $\mathcal{H}$, неубывающих при $x \geq 1$, таких, что для всех $h(x) \in \mathcal{H}$ и произвольных фиксированных $\gamma > 0$, $\sigma > 0$ выполнено соотношение

$$\lim_{x \rightarrow \infty} \frac{h(x^\gamma)}{e^{x^\sigma}} = 0. \quad (3.6.55)$$

Для функций из класса $\mathcal{H}$ справедливы следующие утверждения.

ЛЕММА 3.6.3. Пусть $h(x) \in \mathcal{H}$, функция распределения $F(x)$ определена формулой (3.6.1), $\rho > 0$, $C \geq 0$ и

$$\limsup_{x \rightarrow \infty} [1 - F(x)]h(x) = C. \quad (3.6.56)$$

Тогда

$$\limsup_{y \rightarrow \infty} [1 - G(y)]h(\sqrt{y}) \leq 2C. \quad (3.6.57)$$

Возьмем $z = x^\alpha$, $\alpha > 0$. Очевидно, что в силу неравенства (3.6.44) и того, что функция $h(x)$ удовлетворяет соотношению (3.6.55), для получения оптимальной (в смысле порядка) оценки для $1 - G(y)$ при $y \rightarrow \infty$ достаточно рассмотреть случай $\alpha > 2$. Пусть $z = x^{2+\beta}$, $\beta > 0$. Тогда для

произвольного $\varepsilon > 0$ существует такое x_0 , что для всех $x > x_0$ выполняется неравенство

$$\frac{1}{1 - \Phi(x^{-\beta/2})} \leq 2(1 + \varepsilon). \quad (3.6.58)$$

Далее, в силу (3.6.56) для выбранного ε существует такое x_1 , что при всех $x > x_1$

$$1 - F(x) \leq \frac{C(1 + \varepsilon)}{h(x)}. \quad (3.6.59)$$

Для указанного $\varepsilon > 0$ и всех $x > \max\{x_0, x_1\}$ из (3.6.44) вытекает неравенство

$$1 - G(z) \leq \frac{2C}{h(x)}(1 + \varepsilon)^2 \leq \frac{2C}{h(z^{1/(2+\beta)})}(1 + \varepsilon)^2. \quad (3.6.60)$$

В силу произвольности $\beta > 0$ для любого фиксированного $x > \max\{x_0, x_1\}$

$$1 - G(z) \leq \frac{2C}{h(\sqrt{z})}(1 + \varepsilon)^2, \quad (3.6.61)$$

что и доказывает лемму.

Аналогично лемме 3.6.2 можно доказать следующее утверждение.

ЛЕММА 3.6.3. Пусть $h(x) \in \mathcal{H}$, функция распределения $F(x)$ определена формулой (3.6.1), $\rho > 0$, $C \geq 0$ и

$$\limsup_{y \rightarrow \infty} [1 - G(y)]h(y) = C. \quad (3.6.62)$$

Тогда

$$\limsup_{x \rightarrow \infty} [1 - F(x)]h(x^2) \leq \frac{C}{2}. \quad (3.6.63)$$

Аналогично теореме 3.6.3 можно доказать следующее утверждение.

ТЕОРЕМА 3.6.4. Пусть $h(x) \in \mathcal{H}$, функция распределения $F(x)$ определена формулой (3.6.1) и $C \geq 0$. Для того чтобы

$$\limsup_{x \rightarrow \infty} [1 - F(x)]h(x) = C, \quad (3.6.64)$$

необходимо и достаточно, чтобы

$$\limsup_{y \rightarrow \infty} [1 - G(y)]h(\sqrt{y}) = 2C. \quad (3.6.65)$$

ПРИМЕР 3.6.2. Рассмотрим в качестве смешивающей функцию распределения Парето

$$G(x) = 1 - \frac{1}{x^{k-1/2}}, \quad x \geq 1, \quad k \in N. \quad (3.6.66)$$

Тогда для $1 - F(x)$ справедливо представление

$$\begin{aligned} 1 - F(x) &= \int_1^\infty \left[1 - \Phi\left(\frac{x}{\sqrt{y}}\right) \right] dG(y) = \frac{2k-1}{2\sqrt{2\pi}} \int_x^\infty \int_1^\infty \frac{e^{-\frac{u^2}{2y}}}{y^{k+1}} dy du = \\ &= \frac{2k-1}{2\sqrt{2\pi}} \int_x^\infty \int_0^1 e^{-\frac{u^2 t}{2}} t^{k-1} dt du. \end{aligned}$$

Введем обозначение

$$S(k, a) = \int_0^1 e^{-at} t^{k-1} dt, \quad k \in N, \quad a \neq 0.$$

С помощью простых вычислений несложно убедиться, что

$$S(k, a) = -\frac{e^{-a}}{a} \sum_{i=0}^{k-2} \frac{(k-1)!}{a^i (k-1-i)!} + \frac{(k-1)!}{a^k} (1 - e^{-a})$$

(для определенности полагаем $\sum_{i=0}^{-1} = 0$). Тогда

$$1 - F(x) = \frac{2k-1}{2\sqrt{2\pi}} \int_x^\infty S\left(k, \frac{u^2}{2}\right) du.$$

Так как функция e^{-x} убывает быстрее любой степени x , то

$$\lim_{x \rightarrow \infty} [1 - F(x)] x^{2k-1} = C.$$

При этом, применяя теорему 3.6.3 с учетом (3.6.66), получаем

$$\lim_{x \rightarrow \infty} [1 - F(x)] x^{2k-1} = \frac{1}{2}.$$

4.

Волатильность

4.1. Понятие волатильности

“Волатильность – это отвратительное, беспорядочное понятие, которое трудно понять и еще труднее обсуждать. В индустрии, полной словесной чепухи, волатильность является предметом нелепых и тщетных разговоров...”

С другой стороны, волатильность – это, пожалуй, наиболее важный объект в опционном бизнесе. Она почти всегда “в центре событий”. Она захватывает и манит. Ее сложность хороша для нас. Если бы она была проста и предсказуема, опционный бизнес, скорее всего, стал бы тривиальным и лишенным многих возможностей.”

Мартин О’Коннелл. “*The Business of Options*”.

Ни одно понятие финансовой математики в частности и теории финансов вообще не обладает столь противоречивыми толкованиями, как *волатильность*. Этот термин возникает в многочисленных контекстах и всегда означает нечто, связанное с *неопределенностью*, *изменчивостью* или *риском*. Однако единого универсального определения волатильности, справедливого во всех обстоятельствах, не существует. Разные источники по-разному трактуют волатильность. Для сравнения рассмотрим определения, приведенные в некоторых источниках.

Джон Халл (Hull, 2003) определяет волатильность как *меру неопределенности дохода, реализуемого на данном активе* (a measure of the uncertainty of the return realized on an asset).

На латыни слово *volatilis* означает *крылатый, летающий, быстрый, быстротечный, мимолетный, временный, преходящий, эфемерный*.

Англо-русский словарь В. К. Мюллера (Мюллер, 1967) дает следующие переводы слова “volatility”: 1) (хим.) *летучесть*; 2) *изменчивость, непостоянство*.

Словарь “Random House Webster’s concise dictionary” (Random House, New York, 1993) приводит такие интерпретации термина “volatile”: 1) *evaporating rapidly*; 2) *tending or threatening to erupt in violence; explosive*; 3) *changeable; unstable*, что дословно переводится на русский язык как 1) *быстро испаряющийся*; 2) *стремящийся или грозящий излиться или извергнуться со стремительностью или неистовством; взрывоопасный*; 3) *изменчивый; неустойчивый*. Эти интерпретации близки к оригинальным латинским корням слова “volatility”.

В англо-русском словаре математических терминов под редакцией П. С. Александрова, Л. Н. Большева и др. (Александров – ред., 1994) слово “volatility” трактуется как физический термин и переводится как *летучесть*. Интересно, что в этом словаре слово “volatile” связывается с “математическими машинами”, причем словосочетание “volatile memory” переводится как *оперативная память*.

Между тем, ни в русско-английском словаре математических терминов Ловатера (Boas, 1990), ни в англо-русском и русско-английском словаре по теории вероятностей, математической статистике и комбинаторике К. А. Боровкова (Боровков, 1994) слов “volatility” и “волатильность” вообще нет. Отсюда можно сделать очевидный вывод, что термин “волатильность” традиционно не считается математическим.

Тем не менее в финансовой математике в рамках некоторых специальных моделей эволюции финансовых индексов, таких как, скажем, геометрическое броуновское движение, волатильность отождествляется с *коэффициентом диффузии*, в свою очередь, отождествляемым со стандартным отклонением приращений соответствующих случайных процессов на временных интервалах единичной длины. В некоторых случаях этот коэффициент даже предполагается бесконечным, несмотря на то что в реальном мире бесконечно большие перемещения не осуществимы. Ниже мы остановимся на такой интерпретации волатильности подробнее. Здесь же в качестве примера приведем ссылку на Wikipedia.org (июнь 2005), где дается более или менее “математическое” объяснение термина “волатильность”: *Волатильность – это стандартное отклонение изменения стоимости финансового инструмента в рамках заданного временного горизонта*. Здесь отмечается еще одна особенность: волатильность также оказывается зависящей от временного масштаба.

Помимо приведенных выше толкований волатильности “вообще”, в тео-

рии финансов существуют как минимум три конкретных понятия волатильности: *историческая* волатильность, *неявная* или *подразумеваемая* (implied) волатильность и *вариационная* волатильность.

Существует несколько методов оценивания *исторической волатильности*. Некоторые из них будут рассмотрены в данном разделе. Само название явно указывает на то, что все эти методы используют исторические, то есть статистические данные, представляющие собой временной ряд последовательно наблюдавшихся значений рассматриваемого процесса. Как правило, эти методы используются для описания текущего или прошлого состояния рынка, но также могут быть использованы для прогнозирования.

Неявная (подразумеваемая) волатильность в корне отличается от исторической: она выражает то, что рынок ожидает от будущего. Неявной волатильностью принято называть такое значение волатильности исходных активов, при котором теоретически справедливая цена европейского опциона (вычисленная по формуле Блэка–Шоулза) совпадает с его рыночной ценой. Неявная волатильность не только выражает ожидаемую волатильность процесса эволюции исходного актива, но и в определенном смысле описывает предполагаемую природу развития процесса.

Довольно очевидно, что представление о волатильности как о мере изменчивости процесса тесно связано с выбранным масштабом времени, поскольку процесс может развиваться с разной интенсивностью на различных отрезках разной длины. К примеру, если известны значения процесса $P(t)$, $t \in [0, 1]$:

$$P(0) = 0, \quad P(0.1 \cdot k) = -1, \quad k = 1, 3, 5, 7, 9,$$

$$P(0.1 \cdot k) = 1, \quad k = 2, 4, 6, 8, \quad P(1) = 0,$$

и на всех последующих интервалах единичной длины процесс развивается аналогичным образом, то волатильность на единичном интервале равна нулю, так как итоговое приращение процесса за единицу времени равно нулю ($P(1) - P(0) = 0$), в то время как волатильность на промежутках времени длины 0.1 отнюдь не равна нулю (приращения процесса на таких интервалах равны либо ± 2 , либо ± 1). Понятие *вариационной волатильности* тесно связано с понятием вариации функции и несет в себе информацию о связи между скоростью изменения рассматриваемого процесса и длиной интервалов времени между последовательными наблюдениями (измерениями, отсчетами).

В данном разделе будут приведены и проинтерпретированы некоторые конкретные определения понятия волатильности с целью сравнения тра-

диционных подходов с тем, который вытекает из рассматриваемой модели. В рамках данного подхода будет предложено более наглядное толкование волатильности по сравнению с традиционными определениями. Как будет показано, величине, обычно трактуемой как коэффициент диффузии, может быть придана многомерная форма, позволяющая получить разложение волатильности на динамическую и стохастическую (трендовую и диффузионную) составляющие. Это позволяет получить более подробную информацию о природе эволюции рассматриваемого процесса.

Интерес к понятию волатильности и важность этого понятия в теории финансов являются очевидными проявлениями следующего хорошо известного принципа: новая информация может быть получена только от изменяющихся объектов (в качестве примера еще одного проявления этой закономерности можно привести принцип достаточности в математической статистике). Таким образом, изменчивость (или волатильность) процесса содержит много информации (если не всю информацию) о его развитии. На упомянутом принципе базируются эффективные методы статистического анализа данных, например, дисперсионный анализ.

Как уже отмечалось во введении, обычно при классическом подходе к анализу эволюции финансовых индексов тренды и волатильности считаются в некотором смысле противоположными понятиями и анализируются отдельно. Ниже будет продемонстрировано, что на самом деле эти понятия *нельзя* противопоставлять друг другу. Будет показано, что, если существуют нетривиальные тренды, то неизбежно существует компонента волатильности, обусловленная этими трендами. Более того, существует явно выраженная взаимосвязь трендов и волатильностей, которая может быть довольно разумно описана с помощью аппарата теории вероятностей и математической статистики.

4.2. Возможные подходы к определению и интерпретации волатильности

4.2.1. Теоретико-вероятностные характеристики изменчивости случайного процесса

Итак, собрав воедино приведенные выше толкования термина “волатильность”, можно заключить, что ближе всего по смыслу к нему, по-видимому, такие термины как “изменчивость” или “подверженность изменениям”. При

этом, говоря, что один процесс более волатилен, чем другой, мы обычно подразумеваем, что значения одного процесса сильнее разбросаны вокруг своего тренда, нежели значения другого – вокруг своего, или, что то же самое, траектория одного процесса сильнее флуктуирует вокруг некоторого тренда, нежели траектория другого процесса. В теории вероятностей наиболее употребительной мерой разбросанности значений случайной величины является *дисперсия*. Эта характеристика разброса проста, ее определение прозрачно и хорошо согласуется со здравым смыслом. Дисперсия обладает удобными аналитическими свойствами, среди которых в качестве основного можно упомянуть аддитивность по отношению к стохастически независимым случайным величинам. Переход к той же размерности, что и наблюдаемые процессы, при использовании дисперсии как меры разброса также не вызывает затруднений и вполне естественно сводится к извлечению квадратного корня из дисперсии, называемого *среднеквадратичным отклонением*. Самый простой подход к строгому определению волатильности случайного процесса средствами теории вероятностей поэтому вполне естественно сводится к отождествлению волатильности процесса с его дисперсией (или, что практически то же самое, с его среднеквадратическим отклонением). Такое определение волатильности будем называть классическим. Однако, так как дисперсия существует далеко не для всех процессов, это определение формально применимо далеко не всегда. Например, для устойчивых процессов Леви, являющихся довольно популярными моделями динамики финансовых индексов, дисперсия не определена (бесконечна).

Для описания степени изменчивости (более точно, для описания ее зависимости от времени) случайных процессов в тех ситуациях, когда возможно отсутствие дисперсии, в физике часто используется понятие “ширины диффузионного пакета”. Чтобы его определить, рассмотрим такую меру рассеяния случайной величины как *интерквартильный размах* (interquartile range). Пусть для $q \in [0, 1]$ символ $\ell_\xi(q)$ обозначает квантиль случайной величины ξ порядка q (q -квантиль случайной величины ξ), то есть число, одновременно удовлетворяющее неравенствам

$$P(\xi \leq \ell_\xi(q)) \geq q \text{ и } P(\xi \geq \ell_\xi(q)) \geq 1 - q.$$

Если функция распределения $F_\xi(x)$ случайной величины ξ непрерывна, то q -квантиль случайной величины ξ удовлетворяет уравнению

$$F_\xi(\ell_\xi(q)) = q.$$

Интерквартильный размах случайной величины ξ определяется как

$$R_\xi = \ell_\xi(\frac{3}{4}) - \ell_\xi(\frac{1}{4}).$$

В прикладных исследованиях такая мера рассеяния часто используется в качестве альтернативы среднеквадратичному отклонению, в том числе в ситуациях, когда последнее не определено из-за наличия столь тяжелых хвостов, что второй момент отсутствует. Можно сказать, что интерквартильный размах в той же мере характеризует степень разброса возможных значений случайной величины, в какой медиана характеризует ее центральное значение: внутри интервала $[\ell_\xi(\frac{1}{4}), \ell_\xi(\frac{3}{4})]$ случайная величина ξ попадает с вероятностью $\frac{1}{2}$ и с такой же вероятностью она в этот интервал не попадает. При этом среди всех возможных интервалов с таким свойством (легко видеть, что все такие интервалы имеют своими концами квантили случайной величины порядков $q - \frac{1}{4}$ и $q + \frac{1}{4}$ при $\frac{1}{4} \leq q \leq \frac{3}{4}$) интервал $[\ell_\xi(\frac{1}{4}), \ell_\xi(\frac{3}{4})]$ занимает “серединное” положение, поскольку вероятность того, что случайная величина ξ попадет левее этого интервала равна вероятности того, что случайная величина ξ попадет правее этого интервала (при этом обе эти вероятности равны $\frac{1}{4}$). Поэтому длина этого интервала вполне может быть принята в качестве меры разброса случайной величины.

Для стандартно нормально распределенной случайной величины ξ

$$R_\xi \approx 1.349.$$

В общем случае квантили определены неоднозначно. Однако можно считать, что $\ell_{a\xi+b}(q) = a\ell_\xi(q) + b$ для любой случайной величины ξ и для любых $q \in [0, 1]$, $a > 0$ и $b \in \mathbb{R}$ в том смысле, что при любом наборе чисел q , a , b всегда найдутся квантили случайных величин ξ и $a\xi + b$ порядка q , удовлетворяющие указанному равенству. Поэтому

$$R_{a\xi+b} = aR_\xi.$$

Другими словами, интерквартильный размах пропорционален параметру масштаба. К примеру, если случайная величина ξ имеет нормальное распределение с параметрами $a \in \mathbb{R}$ и $\sigma^2 > 0$, то $R_\xi = r_0\sigma$, где $r_0 \approx 1.349$.

Можно показать, что коэффициент пропорциональности r_0 в соотношении $R_\xi = r_0\sqrt{D\xi}$, где ξ – случайная величина с конечным вторым моментом, в общем случае не превосходит $2\sqrt{2} < 2.8285$. А именно, справедливо следующее утверждение.

Пусть $\mathcal{F}$ – класс всех распределений с конечным вторым моментом. Пусть $\mathcal{F}_0$ – подкласс класса $\mathcal{F}$, содержащий все унимодальные абсолютно непрерывные распределения с конечным вторым моментом и только их. Для простоты, без принципиального ограничения общности последующих рассуждений будем считать, что у распределений класса $\mathcal{F}$ математическое ожидание равно нулю, а у распределений класса $\mathcal{F}_0$ мода совпадает с математическим ожиданием и также равна нулю.

ТЕОРЕМА 4.2.1. Пусть $q \in [\frac{1}{2}, 1)$. Справедливы соотношения

$$\sup_{\mathcal{L}(\xi) \in \mathcal{F}} \frac{\sqrt{1-q}[\ell_\xi(q) - \ell_\xi(1-q)]}{\sqrt{D\xi}} = \sqrt{2} < 1.4142, \quad (4.2.1)$$

$$\sup_{\mathcal{L}(\xi) \in \mathcal{F}_0} \frac{\sqrt{1-q}[\ell_\xi(q) - \ell_\xi(1-q)]}{\sqrt{D\xi}} \leq \frac{2\sqrt{2}}{3} < 0.9429. \quad (4.2.2)$$

Д о к а з а т е л ь с т в о. Пусть $F(x)$ – функция распределения случайной величины ξ . Тогда в условиях теоремы

$$D\xi = \int_{-\infty}^{\infty} x^2 dF(x) = \int_{-\infty}^0 x^2 dF(x) + \int_0^{\infty} x^2 dF(x) \equiv \sigma_-^2 + \sigma_+^2.$$

Предположим сначала, что

$$\ell_\xi(1-q) \leq 0 \leq \ell_\xi(q). \quad (4.2.3)$$

При этом по определению квантили справедливы неравенства

$$\sigma_+^2 \geq \int_{\ell_\xi(q)}^{\infty} x^2 dF(x) \geq \ell_\xi^2(q) \int_{\ell_\xi(q)}^{\infty} dF(x) = \ell_\xi^2(q) \mathbf{P}(\xi \geq \ell_\xi(q)) \geq (1-q)\ell_\xi^2(q),$$

$$\begin{aligned} \sigma_-^2 &\geq \int_{-\infty}^{\ell_\xi(1-q)} x^2 dF(x) \geq \ell_\xi^2(1-q) \int_{-\infty}^{\ell_\xi(1-q)} dF(x) = \\ &= \ell_\xi^2(1-q) \mathbf{P}(\xi \leq \ell_\xi(1-q)) \geq (1-q)\ell_\xi^2(1-q), \end{aligned}$$

откуда

$$D\xi \geq (1-q)[\ell_\xi^2(q) + \ell_\xi^2(1-q)]. \quad (4.2.4)$$

Используя (4.2.3) и известное соотношение между средним арифметическим и средним геометрическим, получаем

$$\begin{aligned}\ell_\xi^2(q) + \ell_\xi^2(1-q) &= [\ell_\xi(q) - \ell_\xi(1-q)]^2 + 2\ell_\xi(q)\ell_\xi(1-q) \geq \\ &\geq [\ell_\xi(q) - \ell_\xi(1-q)]^2 - [\ell_\xi^2(q) + \ell_\xi^2(1-q)],\end{aligned}$$

откуда вытекает неравенство

$$\ell_\xi^2(q) + \ell_\xi^2(1-q) \geq \frac{1}{2}[\ell_\xi(q) - \ell_\xi(1-q)]^2. \quad (4.2.5)$$

Продолжая (4.2.4) с учетом (4.2.5), получаем неравенство

$$D\xi \geq \frac{1-q}{2}[\ell_\xi(q) - \ell_\xi(1-q)]^2, \quad (4.2.6)$$

откуда вытекает (4.2.1), поскольку указанная граница достигается. А именно, для случайной величины ξ , принимающей значения -1 , 0 и 1 , соответственно, с вероятностями $1-q$, $\frac{1}{2}$ и $1-q$, имеет место равенство $D\xi = \frac{1}{2}(1-q)[\ell_\xi(q) - \ell_\xi(1-q)]^2$. Отказ от условия (4.2.3), сводящийся к предположению, что квантили лежат по одну сторону от нуля, как несложно убедиться, только уменьшает мажоранту.

Чтобы доказать (4.2.2), заметим, что для распределений класса $\mathcal{F}_0$ справедливы неравенства Гаусса: для любого $x > 0$

$$x^2 P(\xi \geq x) \leq \frac{4}{9}\sigma_+^2, \quad x^2 P(\xi \leq -x) \leq \frac{4}{9}\sigma_-^2$$

(см., например, (Крамер, 1948), раздел 15.7). Поэтому для распределений класса $\mathcal{F}_0$ имеют место оценки

$$\sigma_+^2 \geq \frac{9}{4}(1-q)\ell_\xi^2(q), \quad \sigma_-^2 \geq \frac{9}{4}(1-q)\ell_\xi^2(1-q),$$

откуда так же, как и выше, вместо (4.2.6) получаем неравенство

$$D\xi \geq \frac{9}{8}(1-q)[\ell_\xi(q) - \ell_\xi(1-q)]^2,$$

откуда вытекает (4.2.2). Теорема доказана.

СЛЕДСТВИЕ 4.2.1. *Справедливы соотношения*

$$\begin{aligned}\sup_{\mathcal{L}(\xi) \in \mathcal{F}} \frac{R_\xi}{\sqrt{D\xi}} &= 2\sqrt{2} < 2.8285, \\ \sup_{\mathcal{L}(\xi) \in \mathcal{F}_0} \frac{R_\xi}{\sqrt{D\xi}} &\leq \frac{4\sqrt{2}}{3} < 1.8857.\end{aligned} \quad (4.2.7)$$

Как показывает пример нормального закона, оценка (4.2.7) довольно точна.

Пусть теперь $\xi(t)$ – центрированный самоподобный (см. раздел 3.4) случайный процесс. Предположим, что его интерквартильный размах

$$R(t) = \ell_{\xi(t)}\left(\frac{3}{4}\right) - \ell_{\xi(t)}\left(\frac{1}{4}\right)$$

удовлетворяет соотношению

$$R(t) = R(1) \cdot t^\mu, \quad t > 0, \quad (4.2.8)$$

с некоторым $\mu > 0$. Иногда величину $R(t)$ называют *шириной диффузионного пакета*. Несложно видеть, что винеровский процесс удовлетворяет соотношению (4.2.8) с $\mu = \frac{1}{2}$. Число μ при этом называется *показателем диффузии*. При этом в случае $\mu < \frac{1}{2}$ процесс $\xi(t)$ называется *субдиффузионным*, в случае $\mu > \frac{1}{2}$ – *супердиффузионным*, а равенство $\mu = \frac{1}{2}$ характеризует *нормальную диффузию* (см. (Metzler, Barkai and Klafter, 1999)). В свою очередь, супердиффузионные процессы подразделяют на *промежуточные* ($\frac{1}{2} < \mu < 1$), *баллистические* ($\mu = 1$) и *турбулентные* ($\mu > 1$). Согласно этой классификации устойчивые процессы Леви (процессы с независимыми и стационарными приращениями, имеющими устойчивые распределения, характеризующиеся показателем $\alpha \in (0, 2]$, см. разделы 1.1 и 1.2) являются промежуточными при $1 < \alpha < 2$, баллистическими – при $\alpha = 1$ (например, процесс Коши) и турбулентными – при $0 < \alpha < 1$.

Показатель μ тесно связан с показателем Херста (Hurst exponent). Здесь эта тема не будет обсуждаться, так как она представляет самостоятельный интерес из-за разнообразия подходов к определению и оцениванию показателя Херста в разных ситуациях и возникающих здесь разночтений и недоразумений.

Однако ключевым для справедливости приведенной выше классификации и эффективности соответствующего описания изменчивости процесса с помощью “ширины диффузионного пакета” является условие (4.2.8). Ясно, что на практике оно выполняется отнюдь не всегда, так как далеко не все случайные процессы обладают свойством самоподобия. Тем не менее, для описания степени изменчивости случайного процесса интерквартильным размахом можно пользоваться всегда. Однако интерквартильный размах, а также другие числовые аналитические характеристики изменчивости случайных процессов, такие, например, как среднее абсолютное отклонение $E|\xi(t) - E\xi(t)|$ и его “медианные” модификации $E|\xi(t) - \text{med}\xi(t)|$, $\text{med}|\xi(t) - \text{med}\xi(t)|$, не очень удобны с аналитической точки зрения.

Для описания степени изменчивости случайных величин и случайных процессов можно использовать более сложные характеристики – предложенные П. Леви функции концентрации. Напомним, что если ξ – случайная величина, то ее функция концентрации (функция концентрации ее распределения) $Q_\xi(x)$ определяется как

$$Q_\xi(x) = \sup_{y \in \mathbb{R}} P(y \leq \xi \leq y + x), \quad x \geq 0.$$

Свойства функций концентрации подробно описаны в книгах (Петров, 1972) и (Хенгартнер и Теодореску, 1980). В частности, из неравенства Чебышева легко вывести следующее соотношение между функцией концентрации случайной величины ξ и ее дисперсией (в случае существования последней): для любого $x \geq 0$

$$D\xi \geq \frac{1}{4} \cdot x^2 [1 - Q_\xi(x)].$$

Любая функция концентрации обладает всеми свойствами функции распределения. Недостатком применения функций концентрации при описании изменчивости случайных величин и случайных процессов является их сложность (с функциями иметь дело менее удобно, чем с числами) и связанные с этим обстоятельством затруднения при их интерпретации.

Для процесса броуновского движения с нулевым сносом (дрейфом) дисперсия и коэффициент диффузии жестко связаны. В этом случае волатильность естественно отождествляется с коэффициентом диффузии. Однако для случаев броуновского движения с ненулевым сносом или ненормальных моделей такое тождество не имеет места. В смешанном гауссовском процессе с нулевыми средними волатильность может быть охарактеризована с помощью *распределения вероятностей* (смешивающего распределения). Однако, если средние значения смешиваемых законов не равны нулю, то такой подход не годится из-за сложностей, связанных с учетом роли нетривиальных трендов.

Тем не менее, если конечномерные распределения рассматриваемого процесса имеют вид конечных сдвиг/масштабных смесей нормальных законов, оказывается вполне достаточно формально отождествить волатильность процесса с его дисперсией (или со среднеквадратическим отклонением).

В финансовой математике используются некоторые определения волатильности, каждое из которых отражает те или иные стороны изменчивости процесса и обладает большей или меньшей общностью. Они будут кратко перечислены в следующих подразделах.

4.2.2. Историческая (статистическая) волатильность

Самый простой метод вычисления исторической (статистической) волатильности заключается в непараметрическом оценивании *стандартного отклонения* рассматриваемого процесса. Возможны несколько способов статистического оценивания такой характеристики. Здесь будет рассмотрено несмещенное оценивание текущей волатильности процесса, представленного временным рядом приращений логарифма финансового индекса. Эта процедура подразумевает усреднение по времени, поэтому при ее применении необходимо задавать также, во-первых, временной диапазон (масштаб) – расстояние между наблюдениями, составляющими исходный временной ряд, – и период времени, по которому производится усреднение. При выборе такого периода естественно руководствоваться предположением о том, что в течение периода усреднения основные характеристики базового процесса остаются неизменными.

Текущая волатильность вычисляется следующим образом:

$$\sigma_n \approx \hat{s}_n,$$

где

$$\hat{s}_n = \left[\frac{1}{m-1} \sum_{i=1}^m (u_{n-i} - \bar{u})^2 \right]^{1/2}, \quad u_i = \log \left(\frac{P_i}{P_{i-1}} \right), \quad \bar{u} = \frac{1}{m} \sum_{i=1}^m u_{n-i}, \quad (4.2.9)$$

$P_i = P(t_i)$, t_i – момент времени, в который зарегистрировано i -е наблюдение (i -й элемент временного ряда), n – номер наблюдения, для которого вычисляется текущая волатильность, так что $\sigma_n = \sigma(t_n)$, число $m \leq n$ определяет длину периода, по которому осуществляется усреднение.

Если $\log P(t)$ – винеровский процесс, то $\sigma_n = \text{const}$, так что описанная процедура позволяет оценить коэффициент диффузии, характеризующий волатильность в модели Блэка–Шоулза. С этой целью, воспользовавшись самоподобием винеровского процесса и зная $T = (t_i - t_{i-1})$ – длину промежутка времени между наблюдениями, – в качестве оценки коэффициента диффузии следует взять

$$\sigma \approx \frac{\hat{s}_n}{\sqrt{T}}.$$

Формулу (4.2.9) иногда видоизменяют с целью получения (смещенной) оценки относительной изменчивости индекса:

$$\hat{s}_n = \left[\frac{1}{m} \sum_{i=1}^m v_{n-i}^2 \right]^{1/2}, \quad v_i = \frac{P_i - P_{i-1}}{P_{i-1}}, \quad \bar{u} \cong 0. \quad (4.2.10)$$

Эта оценка в большей степени интуитивна и используется для вычисления волатильности в других моделях. Заметим, что в зависимости от того, какая практическая задача решается, величину v_i с тем же успехом можно определять в терминах логарифма индекса.

Чтобы придать большее значение ближайшему прошлому, иногда применяются различные схемы взвешивания. В частности, формула (4.2.10) может быть модифицирована и приведена к виду

$$\hat{s}_n = \left[\sum_{i=1}^m \alpha_i v_{n-i}^2 \right]^{1/2}, \quad \sum_{i=1}^m \alpha_i = 1, \quad \alpha_i \geq 0, \quad \forall i \in \{1, \dots, m\}.$$

Историческая волатильность, при измерении которой в качестве исходных наблюдений берутся цены закрытия торгового дня, называется исторической закрытие-к-закрытию волатильностью (Historical Close-to-Close Volatility).

Такой подход к определению рыночной волатильности прост, не требует сложных расчетов, предоставляет возможность сравнения волатильности разных базовых активов, но имеет множество недостатков, которые подробно обсуждаются, например, в работе (Brandt and Kinlay, 2005). В этой же работе подробно описаны некоторые модификации расчета исторической волатильности, ориентированные на использование дополнительной информации. Ниже будут рассмотрены те из них, которые чаще всего используются на практике.

Историческая волатильность по Паркинсону (Historical High-Low Volatility). Физик М. Паркинсон предложил вычислять историческую волатильность по формуле

$$\hat{s}_P^2 = \frac{1}{4n \log 2} \sum_{i=1}^n \left(\log \frac{H_i}{L_i} \right)^2$$

(Parkinson, 1980). При таком подходе учитываются колебания цены в течение торгового дня за счет использования разности логарифмов максимального (H_i) и минимального (L_i) значений цены в течение i -го торгового дня вместо одной лишь цены закрытия.

Историческая волатильность по Гарману–Классу (Historical Open-High-Low-Close Volatility). Эта модификация исторической волатильности предложена в работе (Garman and Klass, 1980):

$$\hat{s}_{GK}^2 = \frac{0.511}{n} \sum_{i=1}^n \left(\log \frac{H_i}{L_i} \right)^2 - \frac{0.019}{n} \sum_{i=1}^n \left[\log \frac{C_i}{O_i} \left(\log \frac{H_i}{O_i} + \log \frac{L_i}{O_i} \right) - \right.$$

$$-2 \log \frac{H_i}{O_i} \log \frac{L_i}{O_i} \Big] - \frac{0.383}{n} \sum_{i=1}^n \left(\log \frac{C_i}{O_i} \right)^2.$$

Она учитывает “гэпы” на открытии (ситуации, когда $O_i \neq C_{i-1}$, где O_i – цена открытия в день i , а C_{i-1} – цена закрытия в день $i - 1$).

Однако при таком подходе никак не учитываются скосы (локальные рыночные тренды), и, следовательно, историческая волатильность переоценивается.

Историческая волатильность по Роджерсу–Сатчеллу (Historical Open-High-Low-Close Volatility). Л. Роджерс и С. Сатчелл (Rogers and Satchell, 1991) предложили формулу

$$\hat{s}_{RS}^2 = \frac{1}{n} \sum_{i=1}^n \left[\log \frac{H_i}{O_i} \log \frac{H_i}{C_i} + \log \frac{L_i}{O_i} \log \frac{L_i}{C_i} \right],$$

которая учитывает рыночные тренды, “гэпы” между ценой открытия текущего и ценой закрытия предыдущего дня и не переоценивает волатильность при наличии на рынке ярко выраженных трендов.

В работе (Yang and Zhang, 2000) предложена модификация формулы Роджерса–Сатчелла, согласно которой оценка волатильности представляет собой выпуклую комбинацию классической оценки исторической волатильности и оценки Роджерса–Сатчелла.

Оценки исторической волатильности Паркинсона, Гармана–Класса, Роджерса–Сатчелла, Янга–Жанга, по сути, представляют собой эвристические реализации попыток учесть временную *неоднородность* данных, когда к однородной выборке цен закрытия добавляются другие данные (цены открытия, максимальная и минимальная цены), природа которых отличается от цен закрытия. При этом сама выборка (то есть набор наблюдений, по которым рассчитывается итоговая оценка), очевидно, становится неоднородной, представляя собой *смесь* нескольких однородных выборок.

В работе (Brandt and Kinlay, 2005) представлен подробный сравнительный анализ различных методов измерения исторической волатильности. Однако все эти модификации, как видно из приведенных выше примеров, в значительной степени имеют эмпирический, искусственный характер. Поэтому на практике, несмотря на все ее недостатки, чаще всего используется обычная историческая close-to-close волатильность в силу ее простоты. Возможно, еще одним обстоятельством в пользу оценки close-to-close является то, что методы обработки неоднородных (смешанных) данных, реализованные в оценках исторической волатильности Паркин-

сона, Гармана–Класса, Роджерса–Сатчелла, Янга–Жанга и аналогичных им, все же не имеют достаточного теоретического обоснования.

Известны другие методы, основанные на тех же принципах. В качестве примера можно привести формулу, известную как ARCH(m)-модель. Идея этой модели состоит в предположении, что волатильность складывается из долговременной и текущей частей, что может быть формализовано с помощью соотношения

$$\hat{s}_n = \left[\alpha_0 \bar{s} + \sum_{i=1}^m \alpha_i v_{n-i}^2 \right]^{1/2}, \quad \sum_{i=0}^m \alpha_i = 1, \quad \alpha_i \geq 0, \quad \forall i \in \{0, \dots, m\}$$

Существуют модели, использующие аналогичные, но рекуррентные соотношения. Наиболее известными являются GARCH(p, q)-модели и, в частности, EWMA-модели. Более подробно о таких моделях см. в книге (Ширяев, 1998).

Модели типа ARCH, GARCH, EWMA используются не только для оценивания прошлого и текущего состояния рынка, но и для прогнозирования его будущего.

Кроме этого, в разделе 2.2 уже упоминался метод расчета исторической волатильности, предложенный в статье (Lo, MacKinlay and Zhang, 2002), в которой в качестве приближенного значения волатильности используется логарифм количества сделок в течение предыдущего часа.

Следует также отметить, что значение исторической (статистической) волатильности, оцененное с помощью описанных выше процедур, всегда конечно, даже если исходный временной ряд представляет собой реализацию случайного процесса с бесконечной дисперсией. Это свойство может одновременно считаться и недостатком таких характеристик волатильности, и их достоинством.

4.2.3. Неявная (подразумеваемая) волатильность

Чтобы привести определение неявной (подразумеваемой) волатильности, сначала необходимо дать описание некоторых терминов из области финансовой математики.

Опционом европейского типа называется производная ценная бумага (контракт), дающая ее обладателю право купить или продать зафиксированное в опционе количество определенных активов по указанной в опционе цене (*цене исполнения опциона*) в указанное в опционе время (*момент*

исполнения опциона, который для опционов европейского типа всегда совпадает с моментом истечения срока контракта). Если реальная цена актива в момент истечения срока контракта (так называемая *спот-цена*) не устраивает обладателя опциона (например, при контракте на продажу она выше цены исполнения опциона или при контракте на покупку она ниже цены исполнения опциона), то он может отказаться от сделки. Опцион, дающий право купить активы, называется опционом “колл”, Опцион, дающий право продать активы, называется опционом “пут”.

Таким образом, выигрыш владельца европейского опциона “колл” равен

$$d_c = \max\{P(T) - X, 0\},$$

где T – момент исполнения опциона, $P(T)$ – спот-цена, X – цена исполнения опциона.

Сами опционы, будучи ценными бумагами, также являются объектами торговли. Рассмотрим задачу определения рациональной (справедливой) цены европейского опциона “колл”. Предположим, что исходные активы обладают постоянной непрерывной дивидендной доходностью q , а их цена $P(t)$ определяется геометрическим броуновским движением:

$$\log P(t) = \log P(t_0) + \left(a - \frac{\sigma^2}{2}\right)(t - t_0) + \sigma[W(t) - W(t_0)], \quad t \geq t_0, \quad (4.2.11)$$

где $a \in \mathbb{R}$, $\sigma > 0$ – волатильность активов, понимаемая как коэффициент диффузии винеровского процесса, представленного правой частью (4.2.11), $W(t)$ – стандартный винеровский процесс. Предположим, что инвесторы могут неограниченно кредитовать и брать ссуды под безрисковую процентную ставку r при непрерывном исчислении, которая одинакова для всех сроков и не меняется с течением времени. Предположим также, что на рынке отсутствуют арбитражные возможности получения прибыли. Пусть, как и ранее, T – момент исполнения опциона, $P(T)$ – спот-цена, X – цена исполнения опциона.

Согласно хорошо известной *формуле Блэка–Шоулза*, текущая справедливая цена $c = c(\sigma)$ европейского опциона “колл” в момент времени t_0 в приведенных выше предположениях имеет вид

$$c = P(t_0)e^{-q(T-t_0)}\Phi(d_1(\sigma)) - Xe^{-r(T-t_0)}\Phi(d_2(\sigma)), \quad (4.2.12)$$

где

$$d_1(\sigma) = \frac{\log P(t_0) - \log X + (T - t_0)(r - q + \sigma^2/2)}{\sigma\sqrt{T - t_0}},$$

$$d_2(\sigma) = d_1(\sigma) - \sigma\sqrt{T - t_0},$$

$\Phi(\cdot)$ – стандартная нормальная функция распределения.

Справедливая цена с европейского опциона “колл”, выраженная формулой (4.2.12), вычислена для “идеального” рынка, описываемого приведенными выше условиями, и сама является “идеальной”, теоретической. Поскольку на реальном рынке перечисленные условия по тем или иным причинам не выполняются, рыночная цена c_m такого опциона отличается от c .

Как уже говорилось выше, *неявной* или *подразумеваемой* волатильностью σ_{implied} принято называть такое значение волатильности исходных активов (то есть значение коэффициента диффузии σ в соотношении (4.2.11)), при котором теоретически справедливая цена $c = c(\sigma)$ европейского опциона (вычисленная по формуле Блэка–Шоулза (4.2.12)) совпадает с его рыночной ценой c_m . Другими словами, σ_{implied} – это решение уравнения

$$c_m = P(t_0)e^{-q(T-t_0)}\Phi(d_1(\sigma)) - Xe^{-r(T-t_0)}\Phi(d_2(\sigma)) \quad (4.2.13)$$

относительно σ . Легко убедиться, что $c(\sigma)$ монотонно возрастает как функция σ , причем

$$\lim_{\sigma \rightarrow 0+} c(\sigma) = \max\{P(t_0)e^{-q(T-t_0)} - Xe^{-r(T-t_0)}, 0\},$$

$$\lim_{\sigma \rightarrow \infty} c(\sigma) = P(t_0)e^{-q(T-t_0)}.$$

Поэтому, если

$$\max\{P(t_0)e^{-q(T-t_0)} - Xe^{-r(T-t_0)}, 0\} \leq c_m \leq P(t_0)e^{-q(T-t_0)},$$

то существует (и притом единственное) решение уравнения (4.2.13), то есть в таком случае неявная волатильность однозначно определена.

В связи с понятием неявной волатильности необходимо упомянуть два термина: *улыбка волатильности*, известный также как *смайл-эффект*, (volatility smile, volatility skew, smile-effect) и *поверхность волатильности* (volatility surface). Улыбка волатильности получается при построении графика неявной волатильности как функции цены исполнения опциона X при фиксированном T – времени исполнения опциона. При этом функция $\sigma_{\text{implied}} = \sigma_{\text{implied}}(X, T)$ оказывается выпуклой вниз, а ее график напоминает улыбку. Смайл-эффект оказывается тесно связанным с наличием более тяжелых хвостов у реальных распределений приращений логарифмов цен

активов, нежели хвосты нормального закона. Смешанные гауссовские модели динамики финансовых индексов довольно успешно объясняют этот эффект.

Поверхность волатильности получается при построении графика функции $\sigma_{\text{implied}} = \sigma_{\text{implied}}(X, T)$ как функции двух переменных: X и T . Поверхность волатильности представляет собой оценку риска, связанного с базовым активом, и используется трейдерами с целью идентификации как самих критических моментов в будущем, так и их природы, то есть оценивания ожидаемых рисков, связанных с падением или повышением цен. Заметим, что иногда неявная волатильность, определяемая через рыночную цену опциона, оказывается косвенно подверженной влиянию баланса спроса и предложения. Например, если кто-нибудь продаст слишком много опционов, связанных с определенным активом, рыночные цены могут упасть ниже значения, определяемого соотношением (4.2.12). Однако на практике это случается редко.

4.2.4. Вариационная волатильность

Для процесса $H(t)$, $t \geq 0$, его Δ -вариация на интервале времени $[a, b]$ определяется как

$$\text{Var}_{[a,b]}(H; \Delta) = \sum |H(t_k) - H(t_{k-1})|,$$

где суммирование производится по всем целым k , для которых $a \leq t_{k-1} < t_k \leq b$ и все t_k имеют вид $k\Delta$. (Эмпирическая) Δ -вариационная волатильность процесса $H(t)$ на $[a, b]$ определяется как

$$\nu_{[a,b]}(H; \Delta) = \text{Var}_{[a,b]}(H; \Delta) \left[\frac{b-a}{\Delta} \right]^{-1}.$$

Эту характеристику разумно использовать, когда предполагается, что анализируемый процесс обладает стационарными приращениями.

За дальнейшими сведениями о характеристиках, описывающих волатильность, и их статистическом оценивании можно обратиться к книге (Ширяев, 1998).

4.2.5. Различные интерпретации волатильности

Все упомянутые выше величины, характеризующие волатильность, могут быть по-разному проинтерпретированы с разных точек зрения. С чисто ма-

тематической точки зрения, волатильность чаще всего соответствует некоему *коэффициенту диффузии*, тесно связанному с дисперсией распределения логарифмов относительных приращений биржевых цен или других изучаемых процессов. В частности, если эволюция финансового индекса описывается геометрическим броуновским движением, то волатильность отождествляется со стандартным отклонением нормальных распределений логарифмов относительных приращений за интервалы времени единичной длины. Поскольку весьма часто гауссовы модели оказываются неадекватными, вместо них используются модели, сводящиеся к не-нормальным распределениям приращений, притом часто с бесконечной дисперсией (например, модели, использующие не-нормальные устойчивые процессы). Но для таких моделей волатильность не может быть разумным образом определена как коэффициент диффузии. При этом с практической точки зрения наиболее приемлемым подходом к определению волатильности в рамках подобных моделей оказывается такой, при которой волатильность трактуется как наиболее вероятное значение абсолютной величины логарифмов относительных *приращений* финансового индекса за интервалы времени определенной длины.

В рамках подхода, связанного с неявной волатильностью, все усложняется. Если посмотреть на график поверхности волатильности только относительно оси времени, то интерпретация вполне непосредственна: неявная волатильность представляет собой меру ожидаемого риска, выраженную в терминах приращений цены, в каждый момент времени. Но стоит только посмотреть на график поверхности волатильности относительно цены исполнения опциона, и интерпретация становится уже не столь очевидной. Обычно улыбки волатильности интерпретируются таким образом, что они дают информацию об ожидаемом правдоподобии тенденции приращений финансовых индексов. К примеру, если неявная волатильность актива относительно велика как при высоких, так и при низких значениях цены исполнения опциона, то некое событие, такое, как, например, объявление квартальных итогов, может оказать как позитивное, так и негативное влияние на рассматриваемый индекс. Если же неявная волатильность актива относительно высока только при высоких значениях цены исполнения опциона, то это означает, что цены могут только повыситься, скажем, вследствие погодных условий, вызвавших неурожай. Этот пример показывает, что, в отличие от других моделей, неявная волатильность содержит компоненту, трактовка которой выходит за пределы представления об обычном коэффициенте диффузии, а именно, “динамическую” компоненту, наличие которой обусловлено существованием некоего тренда.

4.2.6. Волатильность и информация. “Структурные” и “временные” компоненты волатильности

Идея выделения специфических компонент волатильности отнюдь не нова. Во многих работах развивается подход, основанный на том, что важнейшим процессом, влияющим на изменение цен, является процесс поступления новостей. В частности, в статьях (Ross, 1989) и (Andersen, 1996) волатильность изменения биржевых цен напрямую связывается с интенсивностью потока информации, поступающей на рынок. В статье (Engle and Ng, 1993) предложена GARCH-модель, в которой некоторые параметры “отвечают” за новости, поступающие на рынок, и интерпретируются в зависимости от того, являются ли новости “хорошими” или “плохими”. При этом появление разных новостей может по-разному воздействовать на финансовые рынки, например, в зависимости от того, долго или нет (на “длинных” или “коротких” временных горизонтах) ощущаются их последствия. Типичным примером средства (и традиционно используемым инструментом) для анализа таких эффектов является лог-линейная экспоненциально-взвешенная модель Кэмпбелла (Campbell, 1991), (Campbell and Shiller, 1988), в рамках которой, к примеру, доход можно записать, с одной стороны, в виде суммы экспоненциально взвешенных линейных комбинаций ожидаемых дивидендов и доходов в терминах месяцев и дней, а с другой – в виде произведения случайной составляющей (нормально распределенной) и волатильности, имеющей вид корня квадратного из произведения двух чисел, первое из которых (зависящее как от номера дня, так и от номера месяца) характеризует величину дневных (краткосрочных) флуктуаций, второе же, не зависящее от номера дня, характеризует долгосрочную составляющую волатильности (см., например, (Engle, Ghysels and Sohn, 2006) и дальнейшие ссылки в указанной работе).

Разнообразные модели, так или иначе связанные с разложением волатильности на компоненты, отвечающие за изменчивость на различных временных горизонтах, рассматривались в работах (Ding and Granger, 1996), (Engle and Lee, 1999), (Gallant, Hsu and Tauchen, 1999), (Alizadeh, Brandt and Diebold, 2002), (Chernov, Gallant, Ghysels and Tauchen, 2003), (Engle, Sokalska and Chanda, 2005) и во многих других. Как правило, “временная” декомпозиция волатильности имеет мультипликативный характер.

Наряду с упомянутыми “временными” декомпозициями волатильности существует много моделей, в которых компоненты имеют “структурный” смысл. В частности, авторы статьи (Campbell, Lettau, Malkiel and Xu, 2001) также отмечают исключительную значимость новостей об ожидаемых де-

нежных потоках и учетных ставках для формирования цен отдельных ценных бумаг. В этой же статье выделены “идиосинкратическая” (то есть уникальная, характерная¹), “промышленная” (отраслевая²) и “рыночная”³ аддитивные компоненты волатильности различных ценных бумаг и изучена их динамика на основе данных за период с 1962 по 1997 годы. Из последовавших исследований такой структурной декомпозиции волатильности следует упомянуть работы (Malkiel and Xu, 2002), (Goyal and Santa-Clara, 2003), (Frazzini and Marsh, 2003), (Gaspar and Massa, 2004), (Bali, Cakici, Yan and Zhang, 2005), (Guo and Savickas, 2005), (Wei and Zhang, 2005), (Wei and Zhang, 2006) и другие. Структурные компоненты волатильности входят в соответствующие модели аддитивно.

В работе (Maheu and McCurdy, 2004) предложена смешанная “скачкообразно-диффузионная” GARCH-модель, в рамках которой учитываются два разных типа новостей: “нормальные” новости и “необычные” новости. При этом изменения условной дисперсии индекса, связанные с “нормальными” новостями, являются гладкими, в то время как “необычная” компонента неявно подразумеваемого процесса поступления новостей вызывает редкие, но резкие изменения как цен отдельных акций, так и интегральных финансовых индексов. В соответствии с таким представлением общая волатильность раскладывается на GARCH-компоненту и (как правило, меньшую по величине) компоненту, обусловленную наличием скачков, вызванных потоком “необычных” новостей.

Указанные декомпозиции волатильности, как правило, связаны с конкретными параметрическими моделями типа GARCH, и идентификация или интерпретация отдельных компонент волатильности при этом адекватна только в рамках конкретной модели и является довольно условной, если сама соответствующая модель оказывается неадекватной. Такие модели все же не учитывают все стохастические закономерности изменения волатильности. Эти модели, как уже говорилось, довольно искусственны и не могут дать исчерпывающий ответ на вопрос, как эволюционирует волатильность во времени.

В следующем разделе будет показано, что в рамках модели, развитой в главе 2, возможен в некотором смысле обратный подход, при котором естественно (автоматически!) выделяются *объективно существующие*

¹Эта компонента также называется идиосинкратическим риском, несистематическим или диверсифицируемым риском. В некоторых работах об этой компоненте также говорится как о “фирменной” (firm-level).

²industry-level.

³market-level.

диффузионные и динамические компоненты волатильности. Эта декомпозиция не зависит от той или иной модели взаимозависимости компонент волатильности типа ARCH или GARCH. Получаемая декомпозиция волатильности является феноменологической, так что задача интерпретации компонент волатильности, то есть наделения их тем или иным смыслом, становится вторичной. Можно сказать, что отличие подхода, развиваемого в данной книге, от описанного выше традиционного подхода заключается в том, что при традиционном подходе компоненты сначала интерпретируются за счет выбора той или иной ARCH- или GARCH-модели, а затем оцениваются статистически, а при предлагаемом подходе компоненты сначала автоматически выделяются (какие есть, такие есть!), то есть статистически оцениваются, и лишь потом – интерпретируются. При этом, по-видимому, получается более объективная картина.

4.3. Подход, основанный на модели смеси распределений вероятностей

4.3.1. Декомпозиция волатильности

Предыдущий раздел, содержащий описание традиционных подходов к определению и интерпретации волатильности, которые, вообще говоря, являются альтернативными по отношению к тому подходу, который был намечен в главе 2 и который получит дальнейшее развитие ниже, необходим для того, чтобы нагляднее продемонстрировать отличия нового подхода от использовавшихся до сих пор, а также те преимущества, которые сулит его применение.

В рамках нового подхода традиционная интерпретация исторической (статистической) волатильности будет дополнена новой информацией о трендах и диффузиях. Более того, эта новая информация окажется двоякой: с одной стороны, появится возможность учитывать направление и величину трендов, а также интенсивность диффузий. С другой стороны, станет возможным учитывать важность имеющихся трендов и диффузий. Сам факт, что о трендах и диффузиях говорится во множественном числе, указывает на нечто принципиально новое в методах анализа волатильности. А именно, для описания волатильности будет использоваться многомерное представление в терминах *компонент волатильности* и их *весов*.

До сих пор волатильность описывалась с помощью скалярной величины, интерпретируемой либо как коэффициент диффузии базового про-

цесса, либо, что ближе к реалиям, как среднее (в определенном смысле) значение абсолютной величины приращения базового процесса. После того, как в главе 2 было дано обоснование целесообразности моделирования распределений приращений рассматриваемых хаотических процессов с помощью смесей нормальных законов, можно сказать, что каждое приращение представляет собой реализацию “случайной” случайной величины. При этом смешивающее распределение задает вероятности, с какими приращение может оказаться реализацией той или иной случайной величины. Каждая из этих случайных величин имеет свое математическое ожидание, свое стандартное отклонение и соответствующий вес – меру “важности”. Таким образом, описание волатильности базового процесса сводится к описанию упомянутых параметров.

Напомним, что распределение некоторой случайной величины Z является сдвиг/масштабной смесью нормальных законов, если

$$P(Z < x) = \iint \Phi\left(\frac{x-v}{u}\right) dP(U < u, V < v) = E\Phi\left(\frac{x-V}{U}\right), \quad x \in \mathbb{R},$$

где $U \geq 0$ и V – некоторые случайные величины, или конечной смесью нормальных законов, если

$$P(Z < x) = \sum_{i=1}^k p_i \Phi\left(\frac{x-a_i}{\sigma_i}\right) \quad x \in \mathbb{R},$$

где k – натуральное число,

$$a_i \in \mathbb{R}, \quad \sigma_i > 0, \quad p_i \geq 0, \quad i = 1, \dots, k, \quad p_1 + \dots + p_k = 1.$$

Если распределение приращений индекса (или его логарифма) описывается *общей* сдвиг/масштабной смесью нормальных законов, то описание волатильности соответствует заданию распределения параметров сдвига и масштаба. Однако, как уже отмечалось в разделе 4.2.1, в таком случае интерпретация волатильности затруднена, так как та характеристика, которая традиционно считается скалярной, становится функциональной. Хотя их трактовка остается, в принципе, той же самой, *конечные* смеси более наглядно интерпретируемы: если распределение приращений индекса (или его логарифма) описывается *конечной* сдвиг/масштабной смесью нормальных законов, то волатильность можно задать с помощью вектор-функции с тремя типами компонент: первый тип соответствует весам, два других содержат, соответственно, параметры сдвига и параметры диффузии смешиваемых нормальных законов.

Из сказанного в главах 1 и 2 вытекает, что если в качестве модели неоднородного случайного блуждания, описывающего рассматриваемый процесс эволюции финансового индекса или плазменную турбулентность, используются обобщенные процессы Кокса, то вид смешивающего распределения полностью определяется поведением накопленных интенсивностей и, следовательно, статистическими закономерностями изменений мгновенных (локальных) интенсивностей элементарных (допредельных) процессов. Мгновенные интенсивности случайных блужданий с непрерывным временем тесно взаимосвязаны с мгновенными коэффициентами диффузии и мгновенными трендами. В свою очередь, мгновенные коэффициенты диффузии и мгновенные тренды характеризуют типы “механизмов” (циклических, периодических, тенденциозных или с более сложным поведением), которые явно или неявно определяют функционирование финансового рынка (следует особо отметить, что под “механизмом” здесь подразумевается не физическая или экономическая структура, но абстрактная закономерность или подпроцесс). Напротив, каждый тип “механизма” характеризуется своей собственной мгновенной интенсивностью изменений. Можно считать, что в дискретных смесях вида (3.1.3) параметры a_j и σ_j определяют j -й тип “механизмов”, в то время как вес p_j определяет долю “механизмов” j -го типа в общей структуре рынка и может считаться характеристикой силы влияния “механизмов” j -го типа на поведение рынка в целом по сравнению с другими “механизмами”.

Таким образом, статистическое оценивание параметров $\sigma_1, \dots, \sigma_k$, $a_1, \dots, a_k$ и $p_1, \dots, p_k$ дает возможность, во-первых, выделить *типичические* “механизмы” или структуры в функционировании анализируемой сложной системы и, во-вторых, оценить их относительную важность (или силу влияния).

4.3.2. Разложение классической волатильности на динамическую и диффузионную компоненты

Предположим, не вдаваясь в более подробную детализацию, что распределение некоторой случайной величины Z представляет собой сдвиг/масштабную смесь некоторой функции распределения, скажем, H . Это означает, что

$$P(Z < x) = EH \left(\frac{x - V}{U} \right) = \iint H \left(\frac{x - v}{u} \right) dP(U < u, V < v), \quad x \in \mathbb{R}, \quad (4.3.1)$$

где $U \geq 0$ и V – некоторые случайные величины. Наша цель – разложить стандартное отклонение этой случайной величины на компоненты, характеризующие случайным сдвигом и случайным масштабом, соответственно.

Основываясь на теореме Фубини, можно утверждать, что смесь (4.3.1) – это распределение случайной величины $Z \stackrel{d}{=} Y \cdot U + V$, где случайная величина Y имеет распределение $H(x)$ и стохастически независима от пары (U, V) . При этих условиях справедлив следующий результат о разложении дисперсии случайной величины Z .

ЛЕММА 4.3.1. *Предположим, что имеет место представление $Z \stackrel{d}{=} Y \cdot U + V$, в котором случайная величина Y стохастически независима от пары (U, V) , причем случайные величины Y , U и V обладают конечными моментами второго порядка. Тогда*

$$\begin{aligned} DZ &= DV + 2EY \operatorname{Cov}(U, V) + EU^2 DY + (EY)^2 DU = \\ &= DV + 2EY \operatorname{Cov}(U, V) + (EU)^2 DY + EY^2 DU. \end{aligned}$$

Доказательство. В силу независимости случайной величины Y и пары (U, V) утверждение леммы непосредственно вытекает из следующих несложных выкладок:

$$\begin{aligned} DZ &= D(UY + V) = \\ &= E[UY + V - E(UY + V)]^2 = \\ &= E[(UY - EU EY) + (V - EV)]^2 = \\ &= E(UY - EU EY)^2 + E(V - EV)^2 + 2E(UY - EU EY)(V - EV) = \\ &= DV + D(UY) + 2E(UVY - UYEV) = \\ &= DV + D(UY) + 2EY \operatorname{Cov}(U, V). \end{aligned}$$

Но для $D(UY)$ справедливо представление

$$\begin{aligned} D(UY) &= EU^2 EY^2 - (EU)^2 (EY)^2 = \\ &= EU^2 EY^2 - EU^2 (EY)^2 + EU^2 (EY)^2 - (EU)^2 (EY)^2 = \\ &= EU^2 DY + (EY)^2 DU. \end{aligned}$$

Аналогично,

$$\begin{aligned} D(UY) &= EU^2 EY^2 - (EU)^2 (EY)^2 = \\ &= EU^2 EY^2 - (EU)^2 EY^2 + (EU)^2 EY^2 - (EU)^2 (EY)^2 = \\ &= EY^2 DU + (EU)^2 DY, \end{aligned}$$

что и завершает доказательство.

Искомое разложение содержится в следующем следствии.

СЛЕДСТВИЕ 4.3.1. Пусть в дополнение к условиям Леммы 4.3.1 справедливы соотношения $EY = 0$ и $DY = 1$. Тогда

$$DZ = DV + EU^2. \quad (4.3.2)$$

Этот результат представляет собой обобщение основного соотношения дисперсионного анализа, в соответствии с которым полная дисперсия раскладывается на сумму внутригрупповой и межгрупповой дисперсий.

Конечная сдвиг/масштабная смесь нормальных законов представима в виде

$$P(Z < x) = \sum_{j=1}^k p_j \Phi\left(\frac{x - a_j}{\sigma_j}\right) = E\Phi\left(\frac{x - V}{U}\right),$$

где пара случайных величин U, V имеет дискретное распределение

$$P((U, V) = (\sigma_j, a_j)) = p_j, \quad j = 1, \dots, k,$$

так что в соотношении (4.3.2)

$$DV = \sum_{j=1}^k (a_j - \bar{a})^2 p_j, \quad EU^2 = \sum_{j=1}^k p_j \sigma_j^2, \quad (4.3.3)$$

где

$$\bar{a} = \sum_{j=1}^k a_j p_j.$$

Первое выражение в (4.3.3) зависит только от весов p_j и параметров положения (сдвига) a_j компонент и потому характеризует ту часть волатильности, которая обусловлена наличием локальных трендов, то есть “динамическую” компоненту волатильности, тогда как второе выражение в (4.3.3) зависит только от весов p_j и параметров масштаба (“коэффициентов диффузии”) σ_j компонент и потому характеризует “чисто диффузионную” компоненту волатильности.

Если вспомнить традиционное одномерное представление о волатильности как о стандартном отклонении приращения процесса, то можно заметить, что следствие 4.3.1 уточняет это представление: волатильность

процесса представляет собой корень квадратный из суммы двух компонент, первая из которых является характеристикой разбросанности локальных трендов, а вторая характеризует диффузию процесса. Если локальные тренды отсутствуют, то классическая волатильность равна корню квадратному из взвешенной суммы квадратов волатильностей компонент, причем веса компонент показывают важность соответствующей диффузионной компоненты.

Таким образом, обеспечена возможность получать больше информации по сравнению с исследованием обычной (одномерной) волатильности.

4.3.3. Преобразование волатильности при временном скейлинге

Подобно классической волатильности, многомерная волатильность зависит от длины интервала времени между последовательными отсчетами. Поэтому чрезвычайно важно иметь возможность пересчета волатильности при изменении масштаба времени, чтобы получать дополнительную информацию о поведении изучаемого процесса при больших или меньших временных масштабах, не проводя дополнительных экспериментов и/или измерений.

Предложенная выше модель, основанная на представлении распределения приращений процесса в виде конечной сдвиг/масштабной смеси нормальных законов, позволяет это сделать.

ЛЕММА 4.3.2. Пусть X и Y – две независимые случайные величины с функциями распределения

$$P(X < x) = \sum_{r=1}^m p_r \Phi\left(\frac{x - a_r}{\sigma_r}\right)$$

и

$$P(Y < x) = \sum_{j=1}^n q_j \Phi\left(\frac{x - b_j}{\delta_j}\right).$$

Тогда сумма этих случайных величин имеет распределение

$$P(X + Y < x) = \sum_{k=1}^{mn} w_k \Phi\left(\frac{x - c_k}{\gamma_k}\right),$$

где параметры w_k , c_k и γ_k удовлетворяют следующим соотношениям:

если $k = rn + j$, $r = 0, 1, \dots, m-1$ и $j = 1, \dots, n$, то

$$w_{rn+j} = p_{r+1}q_j, \quad c_{rn+j} = a_{r+1} + b_j, \quad \gamma_{rn+j} = \sqrt{\sigma_{r+1}^2 + \delta_j^2}.$$

При этом динамическая и диффузионная компоненты волатильности композиции (свертки), соответственно, имеют вид

$$\sum_{r=0}^{m-1} \sum_{j=1}^n (a_{r+1} + b_j - \bar{c})^2 p_{r+1}q_j$$

и

$$\sum_{r=0}^{m-1} \sum_{j=1}^n (\sigma_{r+1}^2 + \delta_j^2) p_{r+1}q_j,$$

где

$$\bar{c} = \sum_{r=0}^{m-1} \sum_{j=1}^n (a_{r+1} + b_j) p_{r+1}q_j.$$

Д о к а з а т е л ь с т в о. Достаточно заметить, что характеристические функции случайных величин X , Y и $X + Y$, соответственно, имеют вид

$$\psi_X(t) = \sum_{r=1}^m p_r \exp\{ia_r t - \frac{1}{2}t^2 \sigma_r^2\}, \quad \psi_Y(t) = \sum_{j=1}^m q_j \exp\{ib_j t - \frac{1}{2}t^2 \delta_j^2\}$$

и

$$\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t), \quad t \in \mathbb{R},$$

в силу независимости слагаемых.

Если случайные величины X и Y независимы и одинаково распределены, то мы получаем следующее утверждение.

СЛЕДСТВИЕ 4.3.2. В дополнение к условиям Леммы 4.3.2 предположим, что случайные величины X и Y одинаково распределены. Тогда композиция (свертка) их распределений имеет вид

$$P(X + Y < x) = \sum_{k=1}^{n^2} w_k \Phi\left(\frac{x - c_k}{\gamma_k}\right),$$

Параметры w_k , c_k и γ_k удовлетворяют следующим соотношениям: если $k = rn + j$, $r = 0, 1, \dots, n-1$ и $j = 1, \dots, n$, то

$$w_{rn+j} = p_{r+1}p_j, \quad c_{rn+j} = a_{r+1} + a_j, \quad \gamma_{rn+j} = \sqrt{\sigma_{r+1}^2 + \sigma_j^2}.$$

В частности, для совпадающих r и j имеем:

$$w_{r(n+1)+1} = p_{r+1}^2, \quad c_{r(n+1)+1} = 2a_{r+1}, \quad \gamma_{r(n+1)+1} = \sqrt{2}\sigma_{r+1}, \quad r = 0, 1, \dots, n-1.$$

При этом динамическая и диффузионная компоненты волатильности композиции (свертки), соответственно, вдвое больше динамической и диффузионной компонент волатильности слагаемых.

Пример соответствующих вычислений можно найти в разделе 7.

Утверждения, приведенные выше, позволяют пересчитывать “портреты волатильности” с меньших временных масштабов на большие. Обратную операцию, к сожалению, не всегда можно реализовать с абсолютной точностью. Действительно, легко видеть, что если число компонент смеси равно, скажем, k , то число компонент смеси, являющейся сверткой двух исходных, равно

$$n = \frac{1}{2}k(k+1). \quad (4.3.4)$$

Поэтому, если к экспериментальным данным удалось подогнать, скажем, n -компонентную смесь, то выполнить “развертку” смеси, то есть операцию, обратную вышеописанной свертке, можно как минимум, только если для этого n существует целое k , удовлетворяющее (4.3.4). Более того, по сути “развертка” заключается в отыскании *характеристической функции* вида

$$f_k(t) = \sum_{j=1}^k w_j \exp \left\{ ia_j t - \frac{\sigma_j^2 t^2}{2} \right\}, \quad (4.3.5)$$

обеспечивающей равенство

$$f_k^2(t) = f_n(t), \quad t \in \mathbb{R}, \quad (4.3.6)$$

где $f_n(t)$ – характеристическая функция подогнанной смеси. Но функция, в точности равная квадратному корню из $f_n(t)$, может не быть характеристической, даже если соотношение (4.3.4) выполнено. Поэтому на практике “развертка” сводится к отысканию псевдорешения уравнения (4.3.6), например, по следующему правилу.

По заданному n определяется

$$k = \min \left\{ k' : n \leq \frac{(k'+1)k'}{2} \right\},$$

а затем ищется функция $f_k(t)$ вида (4.3.5), исходя из условия

$$\|f_k^2(t) - f_n(t)\| \longrightarrow \min_{w_j, a_j, \sigma_j, j=1, \dots, k},$$

где $\|\cdot\|$ – какая-либо мера близости характеристических функций, например, сумма квадратов разностей их значений в каких-либо M точках $t_1, \dots, t_M$, $M \geq 3k - 1$.

4.3.4. Диффузионный спектр и предполагаемый диффузионный спектр

Если модели типа масштабных смесей нормальных законов оказываются адекватными, то смешивающее распределение можно интерпретировать как *диффузионный спектр*. Если такие модели в целом оказываются неадекватными по какому-либо критерию (например, адекватной является сдвиг-масштабная смесь или вообще иная модель, не имеющая вид смеси нормальных законов), то смешивающее распределение в наиболее правдоподобной модели типа чисто масштабной смеси нормальных законов можно интерпретировать как *предполагаемый диффузионный спектр*.

Такой подход, в определенном смысле упрощающий изучаемую ситуацию и, возможно, ее искажающий, тем не менее, может оказаться полезным, поскольку модели, имеющие вид чисто масштабных смесей, и модели типа сдвиг-масштабных смесей по-разному реагируют на их “уточнение” с помощью добавления дополнительных компонент. Для чисто масштабных смесей наблюдается эффект “насыщения”, когда введение новых компонент сверх какого-то уровня (обычно четыре-пять компонент) практически не меняет модель, в то время как модели типа сдвиг-масштабных смесей при добавлении новых компонент изменяются довольно существенно, превращаясь в своего рода аналоги ядерных оценок плотности. Эффект “насыщения” ставит естественный предел степени детализации модели, имеющей вид чисто масштабной смеси нормальных законов.

При этом некоторые компоненты предполагаемого диффузионного спектра могут быть интерпретированы как компоненты шума, которым объявляется невязка между моделью, имеющей вид чисто масштабной смеси нормальных законов, и “истинной” (адекватной) моделью, возможно, имеющей более сложный вид. В частности, если “истинная” модель имеет вид сдвиг-масштабной смеси нормальных законов, то к “зашумляющей” невязке при таком подходе относится динамическая компонента волатильности. Обсуждение этой проблемы будет продолжено в разделе 6.4.5.

Критерии адекватности или неадекватности моделей весьма относительны. Например, известно довольно много информационных критериев выбора моделей типа критерия Акаике, и в силу самого их определе-

ния они могут приводить к разным решениям относительно адекватности или неадекватности чисто масштабных-смешанных моделей. Поэтому понятие предполагаемого диффузионного спектра оказывается полезным в любом случае.

Будучи смешивающим распределением, диффузионный спектр может оказаться как непрерывным (абсолютно непрерывным), так и дискретным. Диффузионный спектр (предполагаемый диффузионный спектр) характеризует распределение волатильности (отождествляемой с, возможно, многомерным коэффициентом диффузии) по уровням в каждый момент времени. Тем самым изменение диффузионного спектра во времени характеризует перераспределение волатильности по уровням. При этом множество допустимых уровней волатильности может быть как континуальным (что соответствует непрерывному предполагаемому диффузионному спектру), так и счетным (что соответствует дискретному предполагаемому диффузионному спектру).

Преобразование дискретного и непрерывного предполагаемого диффузионного спектра при временном скейлинге удобно характеризовать в терминах случайных величин, распределение которых совпадает с диффузионным спектром. Соответствующие преобразования практически совпадают с теми, о которых шла речь в разделе 3.4, где рассматривалась замкнутость класса масштабных смесей нормальных законов относительно свертки, приводящая к “самоподобию” масштабных смесей нормальных законов.

Итак, предположим, что рассматриваемый случайный процесс имеет независимые приращения на интервалах единичной длины, причем эти приращения имеют вид масштабных смесей гауссовских законов. Тогда, как показано в разделе 3.4, все его приращения на интервалах времени с длинами, кратными длине единичного интервала, снова будут масштабными смесями гауссовских законов. Пусть, к примеру, рассматривается изменение диффузионного спектра при изменении масштаба (длины интервала времени, за который рассматриваются приращения) в n раз. Формально предположим, что $X_1, \dots, X_n$ – случайные величины со стандартным нормальным распределением, $U_1, \dots, U_n$ – неотрицательные (необязательно одинаково распределенные) случайные величины. Тогда каждое произведение $X_j \sqrt{U_j}$ интерпретируется как приращение процесса на j -м интервале единичной длины, причем его распределение имеет вид мас-

штабной смеси нормальных законов

$$H_j(x) = \int_0^{\infty} \Phi(x/\sqrt{u}) dF_j(u), \quad x \in \mathbb{R},$$

где $F_j(u) = P(U_j < u)$. Независимость приращений процесса влечет независимость в совокупности случайных величин $X_1, \dots, X_n, U_1, \dots, U_n$. Теперь рассмотрим сумму

$$X_1\sqrt{U_1} + \dots + X_n\sqrt{U_n},$$

интерпретируемую как приращение процесса на интервале длины n . В разделе 3.4 показано, что распределение этого приращения

$$H(x) = P(X_1\sqrt{U_1} + \dots + X_n\sqrt{U_n} < x)$$

имеет вид масштабной смеси нормальных законов

$$H(x) = \int_0^{\infty} \Phi(x/\sqrt{u}) dF(u), \quad x \in \mathbb{R},$$

где

$$F(u) = P(U_1 + \dots + U_n < u), \quad u \geq 0. \quad (4.3.7)$$

Соотношение (4.3.7) определяет преобразование предполагаемого диффузионного спектра при временном скейлинге в терминах случайных величин, распределения которых совпадают с предполагаемым диффузионным спектром приращений процесса на единичных интервалах. При этом преобразование предполагаемого диффузионного спектра при временном скейлинге в терминах предполагаемых диффузионных спектров приращений процесса на единичных интервалах имеет вид свертки, то есть при временном скейлинге – увеличении временного масштаба в натуральное число раз – предполагаемые диффузионные спектры сворачиваются.

Пока не предполагалось, что эти диффузионные спектры одинаковы, то есть что случайные величины $U_1, \dots, U_n$ одинаково распределены. Если же теперь дополнительно предположить, что рассматриваемый процесс имеет независимые приращения, а предполагаемые диффузионные спектры на единичном масштабе одинаковы, то есть что случайные величины $U_1, \dots, U_n$ одинаково распределены, то преобразование диффузионного спектра при временном скейлинге – увеличении временного масштаба в $n \geq 1$ раз – в терминах предполагаемых диффузионных спектров приращений процесса на единичных интервалах примет вид n -кратной свертки. А

именно, пусть $F_1(u) = P(U_1 < u)$. Тогда функция $F(u)$ в (4.3.7) – предполагаемый диффузионный спектр приращений на масштабе n – определяется как

$$F(u) = F_1^{*n}(u), \quad u \geq 0,$$

где $F_1^{*0}(u)$ – функция распределения с единственным единичным скачком в нуле, $F_1^{*1}(u) = F_1(u)$, и для $n \geq 1$

$$F_1^{*n}(u) = (F_1^{*(n-1)} * F_1)(u).$$

Если временной масштаб изменяется (в частности, уменьшается) в нецелое число раз, скажем, $d > 0$, а предполагаемый диффузионный спектр на единичном масштабе совпадает с безгранично делимым распределением, то соответствующее преобразование предполагаемого диффузионного спектра при временном скейлинге можно определить в терминах преобразований Лапласа–Стилтьеса. А именно, если функции распределения $F_1(u) = P(U_1 < u)$ соответствует безгранично делимое преобразование Лапласа–Стилтьеса

$$\psi_1(s) = \int_0^\infty e^{-su} dF_1(u), \quad s > 0,$$

то предполагаемый диффузионный спектр на масштабе d – смешивающее распределение приращений процесса на интервале времени длины d – соответствует преобразованию Лапласа–Стилтьеса

$$\psi(s) = \psi_1^d(s), \quad s > 0.$$

Например, если предполагаемый диффузионный спектр на единичном масштабе имеет вид гамма-распределения с некоторым параметром формы $\alpha > 0$ и некоторым параметром масштаба $\lambda > 0$ (что имеет место, в частности, для симметричных VG-процессов, см. раздел 2.6.3), то предполагаемый диффузионный спектр на масштабе d будет иметь вид гамма-распределения с параметром формы $d\alpha$ и тем же самым параметром масштаба $\lambda > 0$.

5.

Статистическое разделение смесей вероятностных распределений

5.1. Общая схема решения задачи

“The method employed I would gladly explain,
While I have it so clear in my head...”

Льюис Кэрролл. *“Охота на Снарка”*.

Подводя промежуточный итог сказанному выше, теперь можно сформулировать общую схему теоретического решения обсуждаемой проблемы.

(i) Асимптотический подход, основанный на предельных теоремах для обобщенных дважды стохастических пуассоновских процессов как моделей неоднородных хаотических случайных блужданий, естественно приводит к заключению о том, что аппроксимации для распределений (логарифмов) приращений процессов эволюции финансовых индексов или процессов плазменной турбулентности на интервалах времени умеренной длины следует искать в виде общих сдвиг/масштабных смесей нормальных законов, в которых смешивающий закон определяется накопленной (интегральной) интенсивностью потоков соответствующих информативных событий (элементарных скачков, “тиков”).

(ii) Проблема статистической реконструкции распределений приращений упомянутых процессов (или их логарифмов) сводится к задаче статистического оценивания смешивающего распределения, которое является па-

раметром этой задачи.

(iii) В самой общей постановке задача статистического оценивания смешивающего распределения является некорректной, так как общие сдвиг/масштабные смеси нормальных законов не являются идентифицируемыми. Таким образом, в рамках общего принципа регуляризации некорректных задач, исходная проблема заменяется задачей отыскания решения, наиболее близкого к истинному в классе конечных дискретных сдвиг/масштабных смесей нормальных законов. Эта “редуцированная” задача уже является корректной и имеет единственное решение, так как семейство конечных дискретных сдвиг/масштабных смесей нормальных законов идентифицируемо. Поскольку сдвиг/масштабные смеси нормальных законов обладают свойством устойчивости относительно смешивающего закона, эта замена оправдана и регулярна. При этом, зная оценки устойчивости, можно вычислить погрешности, образующиеся при замене исходной задачи редуцированной. При упомянутой регуляризации происходит автоматическое выделение типичных или более-менее устойчивых структур в эволюции рассматриваемых сложных систем.

(iv) Представление распределений (логарифмов) приращений процессов эволюции финансовых индексов или процессов плазменной турбулентности в виде конечных сдвиг/масштабных смесей нормальных законов естественно приводит к многомерной интерпретации волатильности рассматриваемого процесса и к возможности разложения волатильности на динамическую и диффузионные компоненты.

(v) Статистические закономерности поведения рассматриваемых процессов, формализованные в пункте (i), изменяются во времени, результатом чего является отсутствие *универсального* смешивающего закона. Таким образом, чтобы изучить динамику изменения статистических закономерностей в поведении исследуемого хаотического процесса, задача статистического разделения конечных смесей нормальных законов должна быть последовательно решена на интервалах времени, постоянно сдвигающихся в направлении “астрономического” времени. Тем самым параметры смесей (параметры сдвига (дрейфа), масштаба (диффузии), а также соответствующие веса) оцениваются как функции времени. При этом естественно возникают задачи, связанные как с выбором подходящих методов оценивания параметров сдвига и масштаба, так и с выбором оптимальных параметров вычислительных процедур, реализующих эти методы: начального приближения, ширины скользящего интервала (окна), правила остановки и др.

(vi) Наконец, для адекватной интерпретации результатов и для идентификации феноменологически выделенных (статистически оцененных) ком-

понент, то есть для адекватного сопоставления статистически оцененных компонент с реальными процессами или явлениями, необходимо из многих возможных моделей выбрать наиболее адекватную, например, проверить, является выделенная динамическая компонента волатильности статистически значимой или нет.

Обсудив в предыдущих разделах теоретические проблемы, связанные с построением аналитической модели, перейдем теперь к описанию процедур статистического анализа этих моделей. Из сказанного выше вытекает, что при *практическом* решении задачи моделирования и исследования волатильности (изменчивости) хаотических стохастических процессов ключевым этапом является статистическое разделение смесей вероятностных распределений. В первую очередь понадобятся методы оценивания параметров смесей нормальных распределений – средних значений и параметров масштаба компонент смеси, а также их весов. Будут обсуждены как сама задача разделения смесей вероятностных распределений, так и методы ее решения. Особое внимание будет уделено так называемому *ЕМ-алгоритму* – методу, который заключается в итеративном оценивании параметров и последующей максимизации правдоподобия полученной смеси. Так как хорошее поведение (сходимость к глобально наилучшему решению) этого алгоритма отнюдь не гарантировано, то потребуются его модифицировать, приспособив к нашей задаче, и с этой целью подробно исследовать его свойства. СРС-метод (*метод скользящего разделения смесей*) будет описан в нескольких модификациях. Результаты его применения будут представлены в разных вариантах. Поскольку трансформация скалярной функции в векторную не очевидна, большое внимание будет уделено наглядной визуализации получаемых результатов.

5.2. Задача разделения смесей вероятностных распределений и методы ее решения

5.2.1. Задача разделения смесей вероятностных распределений

В разделе 3.1 было подробно обсуждено понятие смеси вероятностных распределений. В модели (3.1.3) плотность $f(x, \mathbf{y}_j)$ может иметь вид

$$f(x, \mathbf{y}_j) = f_j(x, \theta_j),$$

где θ_j – значение (векторного) параметра θ . В таком случае параметр y можно считать составным, причем некоторые его компоненты определяют *тип* плотности вероятностей, в то время как остальные компоненты определяют значения параметров внутри определенного типа.

В данном разделе основное внимание будет уделено вычислительным методам, позволяющим осуществить разделение конечных смесей, то есть оценить параметры смеси плотностей вероятностей вида

$$h(x) = \sum_{i=1}^k p_i f_i(x, \theta_i). \quad (5.2.1)$$

Более того, если предположить, что компоненты смеси (5.2.1) принадлежат к одному и тому же типу распределений вероятностей и отличаются лишь значениями параметра θ_i внутри этого типа, то получается смесь

$$f(x) = \sum_{i=1}^k p_i f(x, \theta_i). \quad (5.2.2)$$

С формальной точки зрения общая проблема разделения смесей (5.2.1) выглядит следующим образом. По конечному набору (выборке) независимых наблюдений $x = (x_1, \dots, x_n)$, каждое из которых представляет собой реализацию случайной величины X , распределение которой является (или предполагается) смесью вида (5.2.1), требуется определить: (i) число k компонент смеси; (ii) плотности $f_1, \dots, f_k$; (iii) значения параметров $\theta_1, \dots, \theta_k$ и (iv) соответствующие веса $p_1, \dots, p_k$ компонент смеси.

Иногда некоторые из перечисленных пунктов оказываются избыточными, как, например, пункт (ii) в случае модели (5.2.2), включающей рассматриваемую задачу разделения конечных смесей *нормальных* законов. В таком случае задача разделения смесей сводится к задаче *оценивания* неизвестных параметров.

Задача статистического разделения смесей имеет долгую историю. С практической точки зрения она впервые рассматривалась еще в работе К. Пирсона (Pearson, 1897). Многочисленные примеры решения задач статистического разделения различных смесей можно найти в книгах (Everitt and Hand, 1981), (Titterton, Smith and Makov, 1987), (McLachlan and Basford, 1988), (Prakasa Rao, 1992), (McLachlan and Peel, 2000).

Необходимо отметить, что решение задачи статистического разделения смеси вероятностных распределений имеет смысл лишь тогда, когда смесь является идентифицируемой (см. раздел 3.2).

Ниже будут описаны различные методы решения задачи статистического разделения смесей.

Основная идея методов статистического разделения смесей заключается в том, чтобы по наблюдениям $\mathbf{x} = (x_1, \dots, x_n)$ среди всех возможных смесей вида (5.2.1) отыскать такую, которая в максимально возможной степени согласуется с неизвестным теоретическим распределением наблюдаемой случайной величины X . В разных методах понятие “согласованности” эмпирического и теоретического распределений понимается по-разному. Поэтому ниже будут также обсуждаться некоторые наиболее часто используемые критерии выбора модели, сводящиеся в рамках рассматриваемых общих моделей вида конечных смесей нормальных законов к критериям выбора оптимального числа компонент смеси и ее типа (сдвиг/масштабной или чисто масштабной смеси).

5.2.2. Метод моментов

Метод моментов был исторически первым методом анализа смесей вероятностных распределений. Этот метод использован еще в работе (Pearson, 1897) для статистического разделения смеси двух нормальных распределений. Этот метод заключается в том, что к выборочным моментам приравниваются теоретические моменты смеси, выраженные через теоретические моменты компонент и веса, а затем получаемая система уравнений решается относительно неизвестных параметров. В случае смесей нормальных законов данная задача становится чисто алгебраической. Хотя оценки метода моментов, как правило, являются состоятельными, в работе (Tan and Chang, 1971) отмечена низкая эффективность метода моментов. Некоторые примеры, связанные с решением задачи разделения смесей нормальных законов с помощью метода моментов, а также дальнейшие ссылки можно найти в книгах (Titterton, Smith and Makov, 1987), (Айвазян и др., 1989).

5.2.3. Метод максимального правдоподобия

Практически во всех работах, относящихся к применению метода максимального правдоподобия к задаче статистического разделения смесей вероятностных распределений, рассмотрены различные итерационные схемы, максимизирующие функцию правдоподобия. В частности, в работе (Hasselblad, 1966) исследованы две итерационные схемы численной максимизации функции правдоподобия в задаче статистического разделения

смеси нормальных распределений – метод наискорейшего спуска и метод Ньютона–Рафсона. Показано, что метод скорейшего спуска всегда увеличивает значение функции правдоподобия и хорошо работает при малом числе данных. Метод Ньютона–Рафсона дает хорошие результаты при большом объеме выборки, причем полученные с помощью этого метода оценки оказываются более близкими к истинным значениям параметров. Кстати, в этой работе отмечено, что дисперсии оценок весов очень сильно возрастают при уменьшении расстояния между средними значениями компонент.

Тем не менее в целом упомянутые схемы численной максимизации функции правдоподобия при оценивании параметров конечных смесей нормальных законов проигрывают в эффективности итерационной процедуре, называемой ЕМ-алгоритмом. Поэтому основное внимание в дальнейшем будет уделено именно ЕМ-алгоритму и его модификациям, которым посвящен раздел 5.3.

Информацию о других методах решения задачи статистического разделения смесей вероятностных распределений (методе минимума хи-квадрат, методе наименьших квадратов и пр.) можно найти, например, в обзоре (Исаенко и Урбах, 1976) и книгах (Everitt and Hand, 1981), (Titterington, Smith and Makov, 1987), (McLachlan and Basford, 1988), (McLachlan and Peel, 2000).

5.3. ЕМ-алгоритм и его применение к задаче оценивания неизвестных параметров смесей распределений вероятностей

ЕМ-алгоритмом принято называть довольно работоспособную схему построения процедур итерационного типа для численного решения задачи поиска экстремума целевой функции в разнообразных задачах оптимизации. В частности, в прикладной статистике эта схема довольно эффективна при поиске оценок максимального правдоподобия и родственных им в ситуациях, когда функция правдоподобия имеет сложную структуру, из-за которой другие методы оказываются неэффективными или вообще не применимыми.

По-видимому, впервые итерационная процедура типа ЕМ-алгоритма, позволяющая находить численное решение задачи максимизации функции правдоподобия при разделении смесей распределений вероятностей, была

предложена в работе (McKendrick, 1926). Затем после довольно большого перерыва эта идея вновь возникла в работах (Healy and Westmacott, 1956), (Шлезингер, 1965), (Шлезингер, 1968), (Day, 1969a), (Day, 1969b), (Wolfe, 1970), а затем развита и систематически исследована в работах (Dempster, Laird and Rubin, 1977), (Kazakos, 1977). Само название *ЕМ-алгоритм* было предложено в работе (Dempster, Laird and Rubin, 1977), посвященной применению метода максимального правдоподобия к статистическому оцениванию по неполным статистическим данным. Возможно, поэтому зарубежные источники традиционно ссылаются на эту статью (Dempster, Laird and Rubin, 1977) как на первую работу по ЕМ-алгоритму.

Основные свойства ЕМ-алгоритма описаны еще в работе (Шлезингер, 1965). Позднее в работах (Dempster, Laird and Rubin, 1977), (Everitt and Hand, 1981), (Wu, 1983), (Boyles, 1983), (Redner and Walker, 1984) эти свойства были передоказаны и развиты.

Литература по ЕМ-алгоритму и его применениям к решению задач из конкретных областей обширна. Перечисление всех работ практически невозможно. Ограничимся упоминанием лишь книг, посвященных собственно ЕМ-алгоритму (Литтл и Рубин, 1991), (McLachlan and Krishnan, 1997), книг, в которых ЕМ-алгоритму уделено значительное место (Айвазян и др., 1989), (Tanner, 1993), а также двух обстоятельных работ: (Bilmes, 1998) и (Figueiredo, 2004).

В работе (Dempster, Laird and Rubin, 1977) была предложена концепция ЕМ-алгоритма как метода работы с неполными данными (или данными с пропусками). Эта концепция удобна с методической точки зрения и хорошо проясняет смысл метода. Именно этой концепции мы и будем придерживаться при описании ЕМ-алгоритма.

Как правило, ЕМ-алгоритм применяется при решении задач двух типов. К первому можно отнести статистические задачи, связанные с анализом *действительно* неполных данных, когда некоторые статистические данные отсутствуют в силу каких-либо причин. Методы решения таких задач описаны в книге (Литтл и Рубин, 1991).

К другому типу задач можно отнести статистические задачи, в которых функция правдоподобия имеет вид, не допускающий удобных аналитических методов исследования, но допускающий серьезные упрощения, если в задачу ввести дополнительные “ненаблюдаемые” (“отсутствующие”, скрытые, латентные) величины. Примерами прикладных задач второго типа являются задачи распознавания образов, реконструкции изображений. Математическую суть этих прикладных задач составляют задачи кластерного анализа, классификации и разделения смесей вероятностных распре-

делений.

5.3.1. Предварительные сведения

Пусть $\mathbf{X}$ и $\mathbf{Y}$ – случайные величины, заданные на одном и том же измеримом пространстве $(\Omega, \mathcal{A})$. Для определенности будем считать, что $\mathbf{X}$ и $\mathbf{Y}$ принимают свои значения в $\mathbb{R}^n$ и $\mathbb{R}^m$, соответственно, где $n \geq 1$, $m \geq 1$. Другими словами, будем считать, что $(\mathbf{X}, \mathbf{Y})$ – это $(n + m)$ -мерный случайный вектор.

Предположим, что на σ -алгебре $\mathcal{A}$ задано семейство вероятностных мер $\{P_\theta : \theta \in \Theta\}$, где Θ – множество, вообще говоря, произвольной природы.

Пусть $\mathcal{B}_n$ и $\mathcal{B}_m$ – борелевские σ -алгебры на $\mathbb{R}^n$ и $\mathbb{R}^m$, соответственно, и пусть $\mu_{\mathbf{X}}$ и $\mu_{\mathbf{Y}}$ – некоторые σ -конечные меры, определенные, соответственно, на $\mathcal{B}_n$ и $\mathcal{B}_m$ (напомним, что мера ν , заданная на борелевской σ -алгебре $\mathcal{B}_k$ подмножеств множества $\mathbb{R}^k$, $k \geq 1$, называется σ -конечной, если существует последовательность множеств $A_1, A_2, \dots$ такая, что $A_i \in \mathcal{B}_k$, $\mathbb{R}^k = A_1 \cup A_2 \cup \dots$ и $\nu(A_i) < \infty$, $i \geq 1$). Пусть μ – мера, заданная на σ -алгебре $\sigma(\mathcal{B})$, порожденной множеством $\mathcal{B}$ измеримых прямоугольников,

$$\mathcal{B} = \{A \times B : A \in \mathcal{B}_n, B \in \mathcal{B}_m\},$$

и удовлетворяющая условию

$$\mu(A \times B) = \mu_{\mathbf{X}}(A) \cdot \mu_{\mathbf{Y}}(B), \quad A \in \mathcal{B}_n, B \in \mathcal{B}_m.$$

Если меры $\mu_{\mathbf{X}}$ и $\mu_{\mathbf{Y}}$ σ -конечны, то, во-первых, указанное условие определяет меру μ однозначно и, во-вторых, мера μ является σ -конечной (см., например, теорему 2 в (Халмош, 1953), с. 143). Так определенная мера μ называется *произведением* мер $\mu_{\mathbf{X}}$ и $\mu_{\mathbf{Y}}$.

Предположим, что при каждом $\theta \in \Theta$ совместное распределение $(m + n)$ -мерного случайного вектора $(\mathbf{X}, \mathbf{Y})$ абсолютно непрерывно относительно введенной выше меры μ .

Плотность распределения $(m + n)$ -мерного случайного вектора $(\mathbf{X}, \mathbf{Y})$ относительно меры μ обозначим $f_\theta(\mathbf{x}, \mathbf{y})$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$, $\theta \in \Theta$. Это означает, что при каждом $\theta \in \Theta$ для любых борелевских множеств $A \in \mathcal{B}_n$, $B \in \mathcal{B}_m$ справедливо представление

$$P_\theta(\mathbf{X} \in A, \mathbf{Y} \in B) = \int_{A \times B} f_\theta(\mathbf{x}, \mathbf{y}) \mu(d\mathbf{x} \times d\mathbf{y}). \quad (5.3.1)$$

По теореме Фубини (см., например, (Халмош, 1953), § 36) интеграл (5.3.1) можно записать в виде

$$\begin{aligned} P_\theta(\mathbf{X} \in A, \mathbf{Y} \in B) &= \int_{A \times B} f_\theta(\mathbf{x}, \mathbf{y}) \mu(d\mathbf{x} \times d\mathbf{y}) = \\ &= \int_B \left[\int_A f_\theta(\mathbf{x}, \mathbf{y}) \mu_{\mathbf{X}}(d\mathbf{x}) \right] \mu_{\mathbf{Y}}(d\mathbf{y}) = \int_A \left[\int_B f_\theta(\mathbf{x}, \mathbf{y}) \mu_{\mathbf{Y}}(d\mathbf{y}) \right] \mu_{\mathbf{X}}(d\mathbf{x}). \end{aligned} \quad (5.3.2)$$

Так как соотношения (5.3.2) справедливы для любых множеств $A \in \mathcal{B}_n$, $B \in \mathcal{B}_m$, то, полагая в них $A = \mathbb{R}^n$ или $B = \mathbb{R}^m$, соответственно, получаем, что функция

$$f_\theta^{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^m} f_\theta(\mathbf{x}, \mathbf{y}) \mu_{\mathbf{Y}}(d\mathbf{y})$$

является *маргинальной* плотностью случайной величины $\mathbf{X}$ относительно меры $\mu_{\mathbf{X}}$, а функция

$$f_\theta^{\mathbf{Y}}(\mathbf{y}) = \int_{\mathbb{R}^n} f_\theta(\mathbf{x}, \mathbf{y}) \mu_{\mathbf{X}}(d\mathbf{x})$$

является *маргинальной* плотностью случайной величины $\mathbf{Y}$ относительно меры $\mu_{\mathbf{Y}}$.

В дальнейшем в качестве мер $\mu_{\mathbf{X}}$ и $\mu_{\mathbf{Y}}$ будет рассматриваться либо мера Лебега, либо считающая мера.

Условная плотность случайной величины $\mathbf{Y}$ при условии $\mathbf{X} = \mathbf{x}$ определяется как

$$f_\theta(\mathbf{y}|\mathbf{x}) = \frac{f_\theta(\mathbf{x}, \mathbf{y})}{f_\theta^{\mathbf{X}}(\mathbf{x})}, \quad \mathbf{y} \in \mathbb{R}^m. \quad (5.3.3)$$

Выражение (5.3.3), очевидно, имеет смысл, если $f_\theta^{\mathbf{X}}(\mathbf{x}) \neq 0$. Аналогично определяется условная плотность случайной величины $\mathbf{X}$ при условии $\mathbf{Y} = \mathbf{y}$:

$$f_\theta(\mathbf{x}|\mathbf{y}) = \frac{f_\theta(\mathbf{x}, \mathbf{y})}{f_\theta^{\mathbf{Y}}(\mathbf{y})}, \quad \mathbf{x} \in \mathbb{R}^n. \quad (5.3.4)$$

Выражение (5.3.4) имеет смысл, если $f_\theta^{\mathbf{Y}}(\mathbf{y}) \neq 0$.

Соотношение (5.3.3) по традиции используется в качестве определения условной плотности, если обе случайные величины $\mathbf{X}$ и $\mathbf{Y}$ абсолютно непрерывны относительно меры Лебега. Если же случайные величины $\mathbf{X}$

и $\mathbf{Y}$ дискретны, то есть являются абсолютно непрерывными относительно считающей меры, то соотношение (5.3.3) превращается в определение условной вероятности:

$$f_{\theta}(\mathbf{y}|\mathbf{x}) = \frac{P_{\theta}(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})}{P_{\theta}(\mathbf{X} = \mathbf{x})},$$

поскольку в таком случае

$$f_{\theta}(\mathbf{x}, \mathbf{y}) = P_{\theta}(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y})$$

и

$$f_{\theta}^{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^m} f_{\theta}(\mathbf{x}, \mathbf{y}) \mu_{\mathbf{Y}}(d\mathbf{y}) = \sum_{\mathbf{y}: P_{\theta}(\mathbf{Y}=\mathbf{y})>0} P_{\theta}(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}) = P_{\theta}(\mathbf{X} = \mathbf{x}).$$

Из соотношений (5.3.3) и (5.3.4) вытекает, что

$$f_{\theta}(\mathbf{x}, \mathbf{y}) = f_{\theta}(\mathbf{y}|\mathbf{x})f_{\theta}^{\mathbf{X}}(\mathbf{x}) = f_{\theta}(\mathbf{x}|\mathbf{y})f_{\theta}^{\mathbf{Y}}(\mathbf{y}). \quad (5.3.5)$$

Будем считать, что случайная величина $\mathbf{X}$ имеет смысл наблюдаемых данных, в то время как ненаблюдаемая (скрытая) случайная величина $\mathbf{Y}$ играет вспомогательную роль.

Зная совместную плотность $f_{\theta}(\mathbf{x}, \mathbf{y})$ и значение $\mathbf{x}$ наблюдаемой величины $\mathbf{X}$, можно формально определить *полную* функцию правдоподобия

$$L(\theta; \mathbf{x}, \mathbf{y}) = f_{\theta}(\mathbf{x}, \mathbf{y}), \quad \theta \in \Theta, \quad (5.3.6)$$

При этом функцию

$$L(\theta; \mathbf{x}) = f_{\theta}^{\mathbf{X}}(\mathbf{x}), \quad (5.3.7)$$

являющуюся классической функцией правдоподобия параметра θ при заданных наблюдениях $\mathbf{x}$, можно считать *неполной* функцией правдоподобия или функцией правдоподобия параметра θ при неполных данных.

ЕМ-алгоритм предназначен для отыскания значения θ , максимизирующего функции (5.3.6) или (5.3.7) при неизвестном значении $\mathbf{Y}$. Другими словами, с помощью ЕМ-алгоритма можно найти оценки максимального правдоподобия параметра θ .

5.3.2. Общее описание ЕМ-алгоритма

В результате действия ЕМ-алгоритма, представляющего собой итерационную процедуру, вычисляется последовательность значений $\{\theta^{(m)}\}_{m \geq 1}$ параметра θ . Если задано некоторое значение $\theta^{(m)}$, то вычисление следующего значения $\theta^{(m+1)}$ можно условно подразделить на два этапа, аббревиатура наименований которых и дала название всей процедуре. Опишем эти этапы.

1. Этап вычисления математического ожидания (*Е-этап*, от английского *Expectation* – ожидание).

Определим функцию $Q(\theta; \theta^{(m)})$ как условное математическое ожидание логарифма полной функции правдоподобия при известном значении наблюдаемой компоненты $\mathbf{X}$:

$$Q(\theta; \theta^{(m)}) = \mathbf{E}_{\theta^{(m)}}[\log f_{\theta}(\mathbf{X}, \mathbf{Y}) | \mathbf{X}]. \quad (5.3.8)$$

В этом определении θ является аргументом функции $Q(\theta; \theta^{(m)})$, $\mathbf{X}$ и $\theta^{(m)}$ – параметры, так что в соотношении (5.3.8) символ $\mathbf{E}_{\theta^{(m)}}$ означает усреднение по $\mathbf{Y}$ относительно меры $\mathbf{P}_{\theta^{(m)}}$.

При известном значении $\mathbf{X} = \mathbf{x}$ функцию $Q(\theta; \theta^{(m)})$ можно вычислить по формуле

$$Q(\theta; \theta^{(m)}) = \int_{\mathbb{R}^m} [\log f_{\theta}(\mathbf{x}, \mathbf{y})] f_{\theta^{(m)}}(\mathbf{y} | \mathbf{x}) \mu_{\mathbf{Y}}(d\mathbf{y}). \quad (5.3.9)$$

2. Этап максимизации (*М-этап*, от слова *Maximization*). На этом этапе вычисляется

$$\theta^{(m+1)} = \arg \max_{\theta} Q(\theta; \theta^{(m)}).$$

Итерационный процесс останавливается в соответствии с заранее согласованным критерием остановки. Например, заранее выбирается какая-нибудь метрика $\rho(\theta_1, \theta_2)$ и фиксируется малое положительное число ϵ . Процесс останавливается на m -ом шаге, если $\rho(\theta^{(m)}, \theta^{(m-1)}) < \epsilon$.

Заметим, что иногда название “ЕМ-алгоритм” объясняют не упомянутой выше аббревиатурой английских слов *Expectation-Maximization*, но возводят к термину *Estimation-Maximization* (см., например, (Айвазян и др., 1989)). По всей вероятности, первый термин все же лучше отражает суть ЕМ-алгоритма.

5.3.3. Монотонность ЕМ-алгоритма

Свойство монотонности ЕМ-алгоритма было впервые установлено в работе (Шлезингер, 1965). Впоследствии это свойство обобщенных и модифицированных версий ЕМ-алгоритма систематически исследовалось в работах (Dempster, Laird and Rubin, 1977), (Everitt and Hand, 1981), (Boyles, 1983), (Wu, 1983), (Redner and Walker, 1984), (Jordan and Xu, 1996), (Xu and Jordan, 1996).

Недавно в работах (Neal and Hinton, 1998) и (Chrétien and Hero, 2000) было замечено, что ЕМ-алгоритм принадлежит к классу так называемых *проксимальных алгоритмов* или *РР-алгоритмов*, от английского термина Proximal Point algorithms (в статье (Neal and Hinton, 1998) отмечено соответствующее ключевое свойство ЕМ-алгоритма, но при этом ЕМ-алгоритм формально не идентифицирован как РР-алгоритм). Это замечание существенно упрощает исследование свойства монотонности ЕМ-алгоритма. При этом главную роль играет следующее представление функции $Q(\theta; \theta^{(m)})$.

Из соотношения (5.3.5) вытекает, что

$$\log f_{\theta}(\mathbf{x}, \mathbf{y}) = \log f_{\theta}(\mathbf{y}|\mathbf{x}) + \log f_{\theta}^{\mathbf{X}}(\mathbf{x}).$$

Поэтому

$$\begin{aligned} Q(\theta; \theta^{(m)}) &= \int_{\mathbf{R}^m} [\log f_{\theta}(\mathbf{x}, \mathbf{y})] f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \mu_{\mathbf{Y}}(d\mathbf{y}) = \\ &= \int_{\mathbf{R}^m} [\log f_{\theta}(\mathbf{y}|\mathbf{x}) + \log f_{\theta}^{\mathbf{X}}(\mathbf{x})] f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \mu_{\mathbf{Y}}(d\mathbf{y}) = \\ &= \log f_{\theta}^{\mathbf{X}}(\mathbf{x}) + \int_{\mathbf{R}^m} [\log f_{\theta}(\mathbf{y}|\mathbf{x})] f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \mu_{\mathbf{Y}}(d\mathbf{y}). \end{aligned} \quad (5.3.10)$$

Согласно принятой выше терминологии, обычная статистическая процедура поиска оценок максимального правдоподобия направлена на максимизацию по $\theta \in \Theta$ логарифма неполной функции правдоподобия (5.3.7), который при известном значении $\mathbf{X} = \mathbf{x}$ равен первому слагаемому в правой части (5.3.10).

5.3.4. Проксимальные алгоритмы (РР-алгоритмы) и их свойства

В общем виде проксимальный алгоритм (РР-алгоритм) можно описать следующим образом. Рассмотрим задачу максимизации функции $I(\theta) : \Theta \rightarrow \mathbb{R}$. Обобщенный проксимальный алгоритм задается рекуррентным соотношением

$$\theta^{(m+1)} = \arg \max_{\theta} \{I(\theta) - b_m \pi(\theta, \theta^{(m)})\}, \quad (5.3.11)$$

где $\{b_m\}_{m \geq 1}$ – последовательность положительных чисел, а $\pi(\theta, \theta^{(m)})$ – штрафная функция, удовлетворяющая условиям:

$$1^\circ \pi(\theta, \theta^{(m)}) \geq 0;$$

$$2^\circ \pi(\theta, \theta^{(m)}) = 0 \text{ тогда и только тогда, когда } \theta = \theta^{(m)}.$$

Самая первая версия проксимального алгоритма с $\Theta \subseteq \mathbb{R}^d$ при некотором $d \geq 1$ и $\pi(\theta, \theta^{(m)}) = \|\theta - \theta^{(m)}\|^2$ была предложена и изучена в (Martinet, 1970) и (Rockafellar, 1976). Свойства проксимальных алгоритмов подробно описаны, например, в (Васильев, 2002).

Свойство монотонности проксимального алгоритма устанавливается очень просто. Оно вытекает из определения (5.3.11) и свойств штрафной функции. Действительно, из (5.3.11) следует, что

$$\begin{aligned} I(\theta^{(m+1)}) - b_m \pi(\theta^{(m+1)}, \theta^{(m)}) &= \max_{\theta} \{I(\theta) - b_m \pi(\theta, \theta^{(m)})\} \geq \\ &\geq I(\theta^{(m)}) - b_m \pi(\theta^{(m)}, \theta^{(m)}) = I(\theta^{(m)}). \end{aligned} \quad (5.3.12)$$

Последнее равенство имеет место, так как $\pi(\theta^{(m)}, \theta^{(m)}) = 0$ в силу свойства 2° штрафной функции. Но из (5.3.12) в силу свойства 1° штрафной функции вытекает, что

$$I(\theta^{(m+1)}) - I(\theta^{(m)}) \geq b_m \pi(\theta^{(m+1)}, \theta^{(m)}) \geq 0,$$

то есть

$$I(\theta^{(m+1)}) \geq I(\theta^{(m)}),$$

что и означает наличие у любого проксимального алгоритма свойства монотонности.

5.3.5. ЕМ-алгоритм как проксимальный алгоритм

В этом разделе мы убедимся, что, если положить $I(\theta) = \log f_{\theta}^{\mathbf{X}}(\mathbf{x})$, специальным образом выбрать штрафную функцию $\pi(\theta, \theta^{(m)})$ и положить

$b_m \equiv 1$, то описанный выше ЕМ-алгоритм будет алгоритмом проксимального типа.

Рассмотрим *расстояние Кульбака–Лейблера* (Kullback and Leibler, 1951) (см. также, например, (Кульбак, 1967), (Cover and Thomas, 1991)) между условными плотностями $f_\theta(\mathbf{y}|\mathbf{x})$ и $f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x})$, определяемое как условное математическое ожидание логарифма отношения правдоподобия при известном значении наблюдаемой компоненты $\mathbf{X} = \mathbf{x}$ относительно меры $P_{\theta^{(m)}}$:

$$\mathcal{D}_{KL}[f_\theta(\cdot|\mathbf{x}); f_{\theta^{(m)}}(\cdot|\mathbf{x})] = E_{\theta^{(m)}} \left[\log \frac{f_{\theta^{(m)}}(\mathbf{Y}|\mathbf{X})}{f_\theta(\mathbf{Y}|\mathbf{X})} \middle| \mathbf{X} = \mathbf{x} \right].$$

Расстояние Кульбака–Лейблера можно вычислить по формуле

$$\mathcal{D}_{KL}[f_\theta(\cdot|\mathbf{x}); f_{\theta^{(m)}}(\cdot|\mathbf{x})] = \int_{\mathbb{R}^m} f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \log \frac{f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x})}{f_\theta(\mathbf{y}|\mathbf{x})} \mu_{\mathbf{Y}}(d\mathbf{y}).$$

При этом, применяя неравенство Йенсена, в силу вогнутости логарифмической функции легко убедиться, что

$$\begin{aligned} \mathcal{D}_{KL}[f_\theta(\cdot|\mathbf{x}); f_{\theta^{(m)}}(\cdot|\mathbf{x})] &= - \int_{\mathbb{R}^m} f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \log \frac{f_\theta(\mathbf{y}|\mathbf{x})}{f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x})} \mu_{\mathbf{Y}}(d\mathbf{y}) \geq \\ &\geq - \log \int_{\mathbb{R}^m} f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \frac{f_\theta(\mathbf{y}|\mathbf{x})}{f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x})} \mu_{\mathbf{Y}}(d\mathbf{y}) = - \log 1 = 0. \end{aligned}$$

Таким образом, расстояние Кульбака–Лейблера удовлетворяет условиям:

$$3^\circ \mathcal{D}_{KL}[f_\theta(\cdot|\mathbf{x}); f_{\theta^{(m)}}(\cdot|\mathbf{x})] \geq 0;$$

$$4^\circ \mathcal{D}_{KL}[f_\theta(\cdot|\mathbf{x}); f_{\theta^{(m)}}(\cdot|\mathbf{x})] = 0 \text{ тогда и только тогда, когда } \theta = \theta^{(m)}.$$

Положив

$$\pi(\theta, \theta^{(m)}) = \mathcal{D}_{KL}[f_\theta(\cdot|\mathbf{x}); f_{\theta^{(m)}}(\cdot|\mathbf{x})],$$

мы таким образом замечаем, что свойства 3° и 4° расстояния Кульбака–Лейблера соответствуют условиям 1° и 2°, которые определяют штрафную функцию в определении проксимального алгоритма.

Осталось убедиться, что с $I(\theta) = \log f_\theta^{\mathbf{X}}(\mathbf{x})$, $b_m \equiv 1$ и расстоянием Кульбака–Лейблера в качестве штрафной функции соотношение (5.3.11), определяющее проксимальный алгоритм, трансформируется в соотношение, определяющее М-этап ЕМ-алгоритма. Действительно, при таких $I(\theta)$ и $\pi(\theta, \theta^{(m)})$ имеем

$$I(\theta) - \pi(\theta, \theta^{(m)}) = \log f_\theta^{\mathbf{X}}(\mathbf{x}) - \mathcal{D}_{KL}[f_\theta(\cdot|\mathbf{x}); f_{\theta^{(m)}}(\cdot|\mathbf{x})] =$$

$$\begin{aligned}
&= \log f_{\theta}^{\mathbf{X}}(\mathbf{x}) - \int_{\mathbb{R}^m} f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \log \frac{f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x})}{f_{\theta}(\mathbf{y}|\mathbf{x})} \mu_{\mathbf{Y}}(d\mathbf{y}) = \log f_{\theta}^{\mathbf{X}}(\mathbf{x}) - \\
&- \int_{\mathbb{R}^m} f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \log f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \mu_{\mathbf{Y}}(d\mathbf{y}) + \int_{\mathbb{R}^m} f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \log f_{\theta}(\mathbf{y}|\mathbf{x}) \mu_{\mathbf{Y}}(d\mathbf{y}).
\end{aligned}$$

Заметим, что второе слагаемое в правой части последнего соотношения (равное дифференциальной энтропии условного распределения $f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x})$) не зависит от θ . Поэтому

$$\begin{aligned}
&\arg \max_{\theta} \{I(\theta) - \pi(\theta, \theta^{(m)})\} = \\
&= \arg \max_{\theta} \left\{ \log f_{\theta}^{\mathbf{X}}(\mathbf{x}) + \int_{\mathbb{R}^m} f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \log f_{\theta}(\mathbf{y}|\mathbf{x}) \mu_{\mathbf{Y}}(d\mathbf{y}) \right\}. \quad (5.3.13)
\end{aligned}$$

Но выражение в фигурных скобках в правой части (5.3.13) оказывается в точности равным функции $Q(\theta; \theta^{(m)})$ (см. (5.3.10)), откуда вытекает требуемое соотношение

$$\arg \max_{\theta} \{I(\theta) - \pi(\theta, \theta^{(m)})\} = \arg \max_{\theta} Q(\theta; \theta^{(m)})$$

Другими словами, рекуррентные соотношения

$$\theta^{(m+1)} = \arg \max_{\theta} \{I(\theta) - \pi(\theta, \theta^{(m)})\}$$

и

$$\theta^{(m+1)} = \arg \max_{\theta} Q(\theta; \theta^{(m)})$$

задают одну и ту же последовательность, то есть ЕМ-алгоритм является специальным проксимальным алгоритмом. А так как любой проксимальный алгоритм обладает свойством монотонности (см. раздел 5.3.4), то свойство монотонности оказывается присущим и ЕМ-алгоритму в том смысле, что, если последовательность $\{\theta^{(m)}\}_{m \geq 1}$ вычисляется в соответствии с правилами, определяющими ЕМ-алгоритм, то

$$\log f_{\theta^{(m+1)}}^{\mathbf{X}}(\mathbf{x}) \geq \log f_{\theta^{(m)}}^{\mathbf{X}}(\mathbf{x}),$$

то есть

$$L(\theta^{(m+1)}; \mathbf{x}) \geq L(\theta^{(m)}; \mathbf{x}), \quad i \geq 1.$$

5.3.6. Неподвижные точки ЕМ-алгоритма и локальные максимумы функции правдоподобия

В предыдущем разделе мы убедились, что каждая итерация ЕМ-алгоритма гарантированно увеличивает функцию правдоподобия. Однако этого недостаточно, чтобы утверждать, что последовательность оценок параметров, построенная ЕМ-алгоритмом, гарантированно сходится к локальному максимуму функции правдоподобия $L(\theta; \mathbf{x})$. Чтобы установить такую сходимость, приходится предполагать, что рассматриваемые распределения удовлетворяют дополнительным условиям регулярности и, в частности, условиям гладкости. Простейшими условиями такого типа являются условия дифференцируемости по θ плотности $f_{\theta}^{\mathbf{X}}(\mathbf{x})$ и расстояния Кульбака–Лейблера $\mathcal{D}_{KL}[f_{\theta}(\cdot | \mathbf{x}); f_{\theta^{(m)}}(\cdot | \mathbf{x})]$ (см., например, (Tanner, 1993), (Chrétien and Hero, 2000)).

Неподвижная точка ЕМ-алгоритма $\theta^{(\infty)}$ определяется условием

$$\theta^{(\infty)} = \arg \max_{\theta} \left\{ \log f_{\theta}^{\mathbf{X}}(\mathbf{x}) + \int_{\mathbf{R}^m} f_{\theta^{(\infty)}}(\mathbf{y} | \mathbf{x}) \log f_{\theta}(\mathbf{y} | \mathbf{x}) \mu_{\mathbf{Y}}(d\mathbf{y}) \right\}. \quad (5.3.14)$$

В предположении дифференцируемости указанных функций по θ это означает, что точка $\theta^{(\infty)}$ должна быть стационарной точкой функции, стоящей в фигурных скобках в соотношении (5.3.14), то есть

$$\left. \frac{\partial \log f_{\theta}^{\mathbf{X}}(\mathbf{x})}{\partial \theta} \right|_{\theta=\theta^{(\infty)}} - \left. \frac{\partial \mathcal{D}_{KL}[f_{\theta}(\cdot | \mathbf{x}); f_{\theta^{(\infty)}}(\cdot | \mathbf{x})]}{\partial \theta} \right|_{\theta=\theta^{(\infty)}} = 0.$$

Но поскольку расстояние Кульбака–Лейблера $\mathcal{D}_{KL}[f_{\theta}(\cdot | \mathbf{x}); f_{\theta^{(m)}}(\cdot | \mathbf{x})]$ имеет минимум (равный нулю) при $\theta = \theta^{(\infty)}$, его частная производная по θ в точке $\theta = \theta^{(\infty)}$ равна нулю. Отсюда вытекает, что неподвижные точки ЕМ-алгоритма являются стационарными точками функции правдоподобия, то есть при некоторых условиях регулярности последовательность оценок параметров, построенная ЕМ-алгоритмом, действительно сходится к локальному максимуму функции правдоподобия $L(\theta; \mathbf{x})$.

Одновременно свойство монотонности ЕМ-алгоритма свидетельствует о его сильной зависимости от выбора начального (стартового) приближения.

Вопросы скорости сходимости ЕМ-алгоритма обсуждались, например, в работах (Dempster, Laird and Rubin, 1977), (Wu, 1983), (Redner and Walker, 1984), (Jordan and Xu, 1996), (Xu and Jordan, 1996).

Известны многочисленные модификации ЕМ-алгоритма. В частности, такие, у которых на М-этапе функция $Q(\theta; \theta^{(m)})$ не максимизируется, но

находится некоторое $\theta^{(m+1)}$, для которого $Q(\theta^{(m+1)}; \theta^{(m)}) \geq Q(\theta; \theta^{(m)})$ (см., например, (McLachlan and Krishnan, 1997)). Соответствующие модификации ЕМ-алгоритма принято называть *обобщенными ЕМ-алгоритмами* или ГЕМ-алгоритмами (от английского Generalized EM-algorithm – обобщенный ЕМ-алгоритм). Некоторые другие модификации ЕМ-алгоритма будут рассмотрены ниже.

5.3.7. Решение задачи разделения смесей вероятностных распределений с помощью ЕМ-алгоритма

Задача отыскания наиболее правдоподобных оценок параметров смесей вероятностных распределений (задача разделения смесей), по-видимому, является одним из самых популярных приложений ЕМ-алгоритма. Скорее всего, это обусловлено тем, что другие методы решения этой задачи оказываются малоэффективными и неустойчивыми.

Для наглядности, не ограничивая общности, будем рассматривать смеси одномерных распределений.

Базовым предположением в рамках данной задачи является то, что плотность наблюдаемой случайной величины X имеет вид

$$f_{\theta}^X(x) = \sum_{i=1}^k p_i \psi_i(x; t_i), \quad (5.3.15)$$

где $k \geq 1$ – известное натуральное число, $\psi_1, \dots, \psi_k$ – известные плотности распределения, неизвестный параметр θ имеет вид $\theta = (p_1, \dots, p_k, t_1, \dots, t_k)$, причем $p_i \geq 0$, $i = 1, \dots, k$, $p_1 + \dots + p_k = 1$, t_i , $i = 1, \dots, k$, – вообще говоря, многомерные параметры. Плотности $\psi_1, \dots, \psi_k$ будем называть *компонентами* смеси (5.3.15), параметры $p_1, \dots, p_k$ будем называть *веса*ми соответствующих компонент.

Задачей разделения смеси (5.3.15) принято называть задачу статистического оценивания параметров $\theta = (p_1, \dots, p_k, t_1, \dots, t_k)$ по известным реализациям случайной величины X . Необходимым условием существования осмысленного решения задачи разделения смеси вероятностных распределений вида (5.3.15) является идентифицируемость смеси, то есть неизбежное совпадение неизвестных параметров при тождественном совпадении смесей вида (5.3.15) (как функций аргумента x) с одинаковыми известными параметрами. Более подробно об этом см. в (Teicher, 1961), (Teicher, 1963) или, например, (Айвазян и др., 1989). Многочисленные примеры, связанные с разделением различных смесей, можно найти в книгах (Everitt

and Hand, 1981), (Titterington, Smith and Makov, 1987), (McLachlan and Basford, 1988), (Prakasa Rao, 1992), (McLachlan and Peel, 2000).

Предположим, что в нашем распоряжении имеется независимая выборка значений $\mathbf{x} = (x_1, \dots, x_n)$ наблюдаемой случайной величины X , относительно которой, не ограничивая общности, будем предполагать, что ее распределение абсолютно непрерывно относительно меры Лебега (откуда необходимо вытекает, что $\psi_1, \dots, \psi_k$ – это также плотности относительно меры Лебега). В рамках модели (5.3.15) логарифм классической (неполной) функции правдоподобия параметра θ имеет вид

$$\log L(\theta; \mathbf{x}) = \log \prod_{j=1}^n f_{\theta}^X(x_j) = \sum_{j=1}^n \log \left(\sum_{i=1}^k p_i \psi_i(x_j; t_i) \right).$$

Непосредственный поиск точки максимума этой функции весьма затруднителен. Однако, если трактовать наблюдения $\mathbf{x}$ как *неполные*, то функцию правдоподобия можно записать в намного более удобном виде.

Предположим, что наряду с наблюдаемой случайной величиной X задана *ненаблюдаемая* случайная величина Y , значения которой содержат информацию о номерах компонент, в соответствии с которыми “генерируются” наблюдения $\mathbf{x} = (x_1, \dots, x_n)$. А именно, будем считать, что наблюдения организованы следующим образом. При очередном, скажем, j -ом наблюдении ($j = 1, \dots, n$) сначала реализуется значение $y_j \in \{1, 2, \dots, k\}$ ненаблюдаемой случайной величины Y . Это значение y_j имеет смысл номера той компоненты смеси, которая затем выбирается в качестве распределения наблюдаемой случайной величины X при j -ом измерении, результатом которого является значение x_j . Такая схема типична для задач кластерного или дискриминантного анализа, в которых каждое наблюдение может быть порождено одной и только одной компонентой смеси. Эта схема оказывается формально очень удобной для решения статистической задачи разделения конечных смесей вида (5.3.15), то есть задачи статистического оценивания параметров этой смеси.

Будем предполагать, что пары значений (x_j, y_j) являются стохастически независимыми реализациями пары случайных величин (X, Y) .

Совместную плотность случайных величин X и Y обозначим $f_{\theta}(x, y)$. По определению случайных величин X и Y , так как дискретная случайная величина Y абсолютно непрерывна относительно считающей меры и принимает значения $i = 1, 2, \dots, k$, то ее маргинальная плотность равна

$$f_{\theta}^Y(i) = p_i, \quad i = 1, 2, \dots, k,$$

в то время как условная плотность случайной величины X при фиксированном значении $Y = i$ равна

$$f_{\theta}(x|i) = \psi_i(x; t_i).$$

Поэтому, если бы значения $\mathbf{y} = (y_1, \dots, y_n)$ были известны, то логарифм *полной* функции правдоподобия имел бы вид

$$\begin{aligned} \log L(\theta; \mathbf{x}, \mathbf{y}) &= \log \prod_{j=1}^n f_{\theta}(x_j, y_j) = \sum_{j=1}^n \log f_{\theta}(x_j, y_j) = \\ &= \sum_{j=1}^n \log [f_{\theta}(x_j|y_j) f_{\theta}^Y(y_j)] = \sum_{j=1}^n \log [p_{y_j} \psi_{y_j}(x_j; t_{y_j})] = \\ &= \sum_{j=1}^n \log p_{y_j} + \sum_{j=1}^n \log \psi_{y_j}(x_j; t_{y_j}). \end{aligned}$$

Обратим внимание, что в этом представлении совместная плотность распределения случайных величин X и Y относительно произведения меры Лебега и считающей меры (см. раздел 2) равна

$$f_{\theta}(x, y) = p_y \psi_y(x; t_y).$$

Используя соотношение (5.3.5), нетрудно заметить, что, с другой стороны,

$$p_y \psi_y(x; t_y) = f_{\theta}(x, y) = f_{\theta}(y|x) f_{\theta}^X(x), \quad (5.3.16)$$

откуда с учетом вида модели (5.3.15) получается следующее представление условной плотности ненаблюдаемой случайной величины Y при известном значении x наблюдаемой случайной величины X :

$$f_{\theta}(y|x) = \frac{f_{\theta}(x, y)}{f_{\theta}^X(x)} = \frac{p_y \psi_y(x; t_y)}{\sum_{i=1}^k p_i \psi_i(x; t_i)}. \quad (5.3.17)$$

Учитывая определение случайной величины Y в рассматриваемой задаче, можно интерпретировать правую часть (5.3.17) как *апостериорную* вероятность того, что наблюдение x было сгенерировано в соответствии с распределением, задаваемым компонентой ψ_i смеси (5.3.15).

Следовательно, с учетом стохастической независимости реализаций $\{(x_j, y_j)\}_{j=1}^n$ пары случайных величин (X, Y) совместная условная плотность распределения набора $\mathbf{y} = (y_1, \dots, y_n)$ реализаций ненаблюдаемой

случайной величины Y при фиксированных значениях $\mathbf{x} = (x_1, \dots, x_n)$ наблюдаемой случайной величины X равна

$$f_{\theta}(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^n f_{\theta}(y_j|x_j) = \prod_{j=1}^n \frac{p_{y_j} \psi_{y_j}(x_j; t_{y_j})}{\sum_{i=1}^k p_i \psi_i(x_j; t_i)}.$$

Подставляя полученное выражение в соотношение (5.3.9), определяющее функцию $Q(\theta; \theta^{(m)})$ (m – номер итерации ЕМ-алгоритма), и вводя обозначение $\mathcal{Y} = \mathbf{N}_k^n$, где $\mathbf{N}_k = \{1, 2, \dots, k\}$, а степень n понимается в смысле декартова (прямого) произведения множеств, получаем

$$\begin{aligned} Q(\theta; \theta^{(m)}) &= \sum_{\mathbf{y} \in \mathcal{Y}} [\log f_{\theta}(\mathbf{x}, \mathbf{y})] f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) = \\ &= \sum_{\mathbf{y} \in \mathcal{Y}} \sum_{j=1}^n \log[p_{y_j} \psi_{y_j}(x_j; t_{y_j})] \prod_{r=1}^n f_{\theta^{(m)}}(y_r|x_r) = \\ &= \sum_{y_1=1}^k \sum_{y_2=1}^k \cdots \sum_{y_n=1}^k \sum_{j=1}^n \log[p_{y_j} \psi_{y_j}(x_j; t_{y_j})] \prod_{r=1}^n f_{\theta^{(m)}}(y_r|x_r) = \\ &= \sum_{y_1=1}^k \sum_{y_2=1}^k \cdots \sum_{y_n=1}^k \sum_{j=1}^n \sum_{l=1}^k \delta_{l,y_j} \log[p_l \psi_l(x_j; t_l)] \prod_{r=1}^n f_{\theta^{(m)}}(y_r|x_r) = \\ &= \sum_{l=1}^k \sum_{j=1}^n \log[p_l \psi_l(x_j; t_l)] \sum_{y_1=1}^k \sum_{y_2=1}^k \cdots \sum_{y_n=1}^k \delta_{l,y_j} \prod_{r=1}^n f_{\theta^{(m)}}(y_r|x_r), \end{aligned} \quad (5.3.18)$$

где $\delta_{a,b}$ – символ Кронекера: $\delta_{a,b} = 1$, если $a = b$, и $\delta_{a,b} = 0$, если $a \neq b$.

Чтобы упростить правую часть (5.3.18), заметим, что для $l \in \mathbf{N}_k$

$$\begin{aligned} &\sum_{y_1=1}^k \sum_{y_2=1}^k \cdots \sum_{y_n=1}^k \delta_{l,y_j} \prod_{r=1}^n f_{\theta^{(m)}}(y_r|x_r) = \\ &= \left(\sum_{y_1=1}^k \cdots \sum_{y_{j-1}=1}^k \sum_{y_{j+1}=1}^k \cdots \sum_{y_n=1}^k \prod_{\substack{r=1 \\ r \neq j}}^n f_{\theta^{(m)}}(y_r|x_r) \right) f_{\theta^{(m)}}(l|x_j) = \\ &= \prod_{\substack{r=1 \\ r \neq j}}^n \left(\sum_{y_r=1}^k f_{\theta^{(m)}}(y_r|x_r) \right) f_{\theta^{(m)}}(l|x_j) = f_{\theta^{(m)}}(l|x_j), \end{aligned} \quad (5.3.19)$$

так как в соответствии с соотношением (5.3.16)

$$\sum_{i=1}^k f_{\theta^{(m)}}(i|x_r) = 1.$$

Используя (5.3.19), выражение (5.3.18) для $Q(\theta; \theta^{(m)})$ можно переписать в виде

$$\begin{aligned} Q(\theta; \theta^{(m)}) &= \sum_{l=1}^k \sum_{j=1}^n f_{\theta^{(m)}}(l|x_j) \log[p_l \psi_l(x_j; t_l)] = \\ &= \sum_{l=1}^k \sum_{j=1}^n f_{\theta^{(m)}}(l|x_j) \log p_l + \sum_{l=1}^k \sum_{j=1}^n f_{\theta^{(m)}}(l|x_j) \log \psi_l(x_j; t_l). \end{aligned} \quad (5.3.20)$$

Теперь видно, что для отыскания максимума функции $Q(\theta; \theta^{(m)})$ по $\theta = (p_1, \dots, p_k, t_1, \dots, t_k)$ можно максимизировать слагаемые в правой части (5.3.20) независимо друг от друга, так как они зависят от *разных* параметров: первое зависит только от весов $p_1, \dots, p_k$, а второе – только от параметров $t_1, \dots, t_k$ компонент смеси.

Найдем значения $p_1, \dots, p_k$, максимизирующие первое слагаемое в правой части (5.3.20). С этой целью, учитывая очевидное ограничение

$$\sum_{l=1}^k p_l = 1,$$

воспользуемся методом неопределенных множителей Лагранжа, согласно которому искомые значения $p_1, \dots, p_k$ удовлетворяют уравнениям

$$\frac{\partial}{\partial p_l} \left[\sum_{l=1}^k \sum_{j=1}^n f_{\theta^{(m)}}(l|x_j) \log p_l + \lambda \left(\sum_{l=1}^k p_l - 1 \right) \right] = 0, \quad l = 1, \dots, k,$$

очевидно, эквивалентным уравнениям

$$\frac{1}{p_l} \sum_{j=1}^n f_{\theta^{(m)}}(l|x_j) + \lambda = 0, \quad l = 1, \dots, k.$$

Просуммировав эти уравнения по l , приходим к заключению, что $\lambda = -n$, откуда вытекает, что искомые значения весов равны

$$p_l^* = \frac{1}{n} \sum_{j=1}^n f_{\theta^{(m)}}(l|x_j), \quad l = 1, \dots, k. \quad (5.3.21)$$

Подразумевая, что значение $\theta^{(m)} = (p_1^{(m)}, \dots, p_k^{(m)}, t_1^{(m)}, \dots, t_k^{(m)})$ параметра θ на m -ой итерации ЕМ-алгоритма известно, и учитывая (5.3.17), теперь можно выписать значения весов $p_1^{(m+1)}, \dots, p_k^{(m+1)}$ на $(m+1)$ -ой итерации ЕМ-алгоритма. А именно, с учетом (5.3.21) имеем

$$p_l^{(m+1)} = p_l^* = \frac{1}{n} \sum_{j=1}^n \frac{p_l^{(m)} \psi_l(x_j; t_l^{(m)})}{\sum_{i=1}^k p_i^{(m)} \psi_i(x_j; t_i^{(m)})}, \quad l = 1, \dots, k. \quad (5.3.22)$$

Оптимальные значения параметров $t_1, \dots, t_k$, реализующие максимум второго слагаемого в правой части (5.3.20), зависят от конкретного аналитического выражения для $\psi_i(x; t_i)$, $i = 1, \dots, k$. В следующем разделе будет подробно рассмотрена ситуация, в которой $\psi_i(x; t_i)$ – нормальные плотности.

5.3.8. Разделение конечных смесей нормальных распределений с помощью ЕМ-алгоритма

В данном разделе ЕМ-алгоритм будет рассматриваться применительно к анализу специального случая модели (5.3.15), в котором

$$\psi_i(x; t_i) = \frac{1}{\sigma_i} \phi\left(\frac{x - a_i}{\sigma_i}\right), \quad x \in \mathbb{R},$$

где

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, \quad x \in \mathbb{R},$$

– плотность стандартного нормального распределения вероятностей, $t_i = (a_i, \sigma_i^2)$, $a_i \in \mathbb{R}$, $\sigma_i > 0$, $i = 1, \dots, k$. Этот случай, пожалуй, чаще всего встречается в прикладных работах, связанных с применением ЕМ-алгоритма.

Для данного случая апостериорная вероятность того, что наблюдение x_j было сгенерировано в соответствии с распределением, задаваемым i -ой компонентой смеси (5.3.15), (правая часть (5.3.17)), имеет вид

$$f_{\theta}(i|x_j) = \frac{\frac{p_i}{\sigma_i} \phi\left(\frac{x_j - a_i}{\sigma_i}\right)}{\sum_{r=1}^k \frac{p_r}{\sigma_r} \phi\left(\frac{x_j - a_r}{\sigma_r}\right)}.$$

Пусть значение

$$\theta^{(m)} = (p_1^{(m)}, \dots, p_k^{(m)}, a_1^{(m)}, \dots, a_k^{(m)}, \sigma_1^{(m)}, \dots, \sigma_k^{(m)})$$

параметра θ на m -ой итерации ЕМ-алгоритма известно. Обозначим

$$\begin{aligned} g_{ij}^{(m)} &= f_{\theta^{(m)}}(i|x_j) = \frac{\frac{p_i^{(m)}}{\sigma_i^{(m)}} \phi\left(\frac{x_j - a_i^{(m)}}{\sigma_i^{(m)}}\right)}{\sum_{r=1}^k \frac{p_r^{(m)}}{\sigma_r^{(m)}} \phi\left(\frac{x_j - a_r^{(m)}}{\sigma_r^{(m)}}\right)} = \\ &= \frac{\frac{p_i^{(m)}}{\sigma_i^{(m)}} \exp\left\{-\frac{1}{2}\left(\frac{x_j - a_i^{(m)}}{\sigma_i^{(m)}}\right)^2\right\}}{\sum_{r=1}^k \frac{p_r^{(m)}}{\sigma_r^{(m)}} \exp\left\{-\frac{1}{2}\left(\frac{x_j - a_r^{(m)}}{\sigma_r^{(m)}}\right)^2\right\}}. \end{aligned}$$

Тогда в соответствии с (5.3.21) и (5.3.22) уточненное значение параметра p_i на $(m+1)$ -ой итерации ЕМ-алгоритма примет вид

$$p_i^{(m+1)} = \frac{1}{n} \sum_{j=1}^n g_{ij}^{(m)}. \quad (5.3.23)$$

В то же время в рамках рассматриваемой модели конечной смеси нормальных законов второе слагаемое в правой части (5.3.20) с учетом введенных обозначений оказывается равным

$$\begin{aligned} &\sum_{i=1}^k \sum_{j=1}^n f_{\theta^{(m)}}(i|x_j) \log \psi_i(x_j; t_i) = \\ &= - \sum_{i=1}^k \sum_{j=1}^n g_{ij}^{(m)} \left[\frac{1}{2} \log 2\pi + \log \sigma_i + \frac{1}{2} \left(\frac{x_j - a_i}{\sigma_i} \right)^2 \right] = \\ &= -\frac{1}{2} \log 2\pi \sum_{i=1}^k \sum_{j=1}^n g_{ij}^{(m)} - \sum_{i=1}^k \log \sigma_i \sum_{j=1}^n g_{ij}^{(m)} - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^n g_{ij}^{(m)} \left(\frac{x_j - a_i}{\sigma_i} \right)^2. \end{aligned} \quad (5.3.24)$$

Заметим, что первое слагаемое в правой части (5.3.24) не зависит от параметров a_i и σ_i . Обозначим

$$\Psi = \sum_{i=1}^k \log \sigma_i \sum_{j=1}^n g_{ij}^{(m)} + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^n g_{ij}^{(m)} \left(\frac{x_j - a_i}{\sigma_i} \right)^2.$$

Чтобы найти значения a_i и σ_i , $i = 1, \dots, k$, которые доставляют максимум второму слагаемому в правой части (5.3.20), продифференцируем Ψ по a_i

и σ_i , $i = 1, \dots, k$, приравняем эти частные производные нулю и решим полученные уравнения относительно a_i и σ_i , $i = 1, \dots, k$. В результате получим

$$a_i^* = \frac{1}{\sum_{j=1}^n g_{ij}^{(m)}} \sum_{j=1}^n g_{ij}^{(m)} x_j,$$

$$\sigma_i^* = \left[\frac{1}{\sum_{j=1}^n g_{ij}^{(m)}} \sum_{j=1}^n g_{ij}^{(m)} (x_j - a_i^*)^2 \right]^{1/2}, \quad i = 1, \dots, k.$$

Таким образом, уточненные значения параметров a_i и σ_i на $(m + 1)$ -ой итерации ЕМ-алгоритма имеют вид

$$a_i^{(m+1)} = \frac{1}{\sum_{j=1}^n g_{ij}^{(m)}} \sum_{j=1}^n g_{ij}^{(m)} x_j,$$

$$\sigma_i^{(m+1)} = \left[\frac{1}{\sum_{j=1}^n g_{ij}^{(m)}} \sum_{j=1}^n g_{ij}^{(m)} (x_j - a_i^{(m+1)})^2 \right]^{1/2}, \quad i = 1, \dots, k.$$

Эти соотношения в совокупности с (5.3.23) определяют ЕМ-алгоритм для разделения конечных смесей нормальных законов. Отметим, что фактически указанные рекуррентные соотношения реализуют оба этапа (Е-этап и М-этап) ЕМ-алгоритма *одновременно*.

В работе (Xu and Jordan, 1996) приведены некоторые теоретические результаты, показывающие (вопреки эвристическому заключению, данному в статье (Redner and Walker, 1984)), что ЕМ-алгоритм оказывается более эффективным средством решения задачи разделения конечной смеси нормальных законов, нежели другие процедуры численной оптимизации, например, такие как методы градиентного восхождения или методы сопряженных градиентов.

Необходимо отметить, что, хотя в прикладных работах ЕМ-алгоритм чаще всего применяется к исследованию модели (5.3.15) с нормальными компонентами, именно эта модель не удовлетворяет условиям, гарантирующим правильную работу ЕМ-алгоритма. А именно, во всех упоминавшихся выше работах, в которых исследовалась сходимость ЕМ-алгоритма, его сходимость доказана при обязательном условии ограниченности логарифма функции правдоподобия (не говоря уж о том, что для сходимости проксимальных алгоритмов нужно, чтобы оптимизируемая ими функция обладала свойствами выпуклости (Васильев, 2002)). Для смесей нормальных распределений указанные условия, вообще говоря, не выполняются.

Чтобы убедиться в этом, достаточно рассмотреть случай двухкомпонентной смеси, для которой логарифм функции правдоподобия имеет вид

$$\begin{aligned} & \log L(a_1, a_2, \sigma_1, \sigma_2, p; \mathbf{x}) = \\ & = \sum_{j=1}^n \log \left[\frac{p}{\sqrt{2\pi}\sigma_1} \exp \left\{ -\frac{1}{2} \left(\frac{x_j - a_1}{\sigma_1} \right)^2 \right\} + \frac{1-p}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{1}{2} \left(\frac{x_j - a_2}{\sigma_2} \right)^2 \right\} \right]. \end{aligned}$$

Если $a_i \neq x_j$ для $i = 1, 2$ и $j = 1, \dots, n$, то, обозначив $z_{ij} = |x_j - a_i|/\sigma_i$, получим

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left\{ -\frac{1}{2} \left(\frac{x_j - a_i}{\sigma_i} \right)^2 \right\} = \frac{z_{ij}}{\sqrt{2\pi}\sigma_i z_{ij}} e^{-z_{ij}^2/2} \leq \\ & \leq \frac{1}{\sqrt{2\pi}|x_j - a_i|} \max_{z>0} z e^{-z^2/2} = \frac{1}{\sqrt{2\pi e}|x_j - a_i|}. \end{aligned}$$

Поэтому равномерно по p , σ_1 и σ_2

$$\begin{aligned} \log L(a_1, a_2, \sigma_1, \sigma_2, p; \mathbf{x}) & \leq \sum_{j=1}^n \log \left(\frac{1}{\sqrt{2\pi e}|x_j - a_i|} \right) = \\ & = -\frac{n}{2} [\log(2\pi) + 1] - \sum_{j=1}^n \log |x_j - a_i|. \end{aligned}$$

Однако, полагая, к примеру, $a_1 = x_1$, сразу замечаем, что $\log L(a_1, a_2, \sigma_1, \sigma_2, p; \mathbf{x})$ неограниченно возрастает как $-\log \sigma_1$ при $\sigma_1 \rightarrow 0$.

Эта особенность в совокупности с наличием большого числа стационарных точек (локальных максимумов) логарифма функции правдоподобия для модели (5.3.15) с большим ($k \geq 2$) числом нормальных компонент ведет к таким неприятным свойствам ЕМ-алгоритма в рамках задачи разделения смесей нормальных законов как повышенная чувствительность (он находит не “правильные” оценки параметров, но наиболее правдоподобные, причем правдоподобие найденных ЕМ-алгоритмом оценок оказывается заметно выше правдоподобия “правильных” оценок) и большая неустойчивость по отношению к начальному приближению и исходным данным. Эти свойства и некоторые способы противодействия им будут рассмотрены ниже.

Основная трудность в применении ЕМ-алгоритма заключается в том, что не существует критерия, позволяющего достоверно распознать природу достигнутого максимума – является ли он глобальным или локальным.

В силу этого обстоятельства необходимы другие критерии, позволяющие принять результат вычислений по ЕМ-алгоритму в качестве окончательного или отвергнуть его.

Другая очень важная проблема заключается в правильном выборе начальных приближений. Важность этой проблемы обусловлена двумя обстоятельствами: во-первых, решения, полученные с помощью ЕМ-алгоритма, могут сильно зависеть от начальных приближений и, во-вторых, при разных начальных приближениях алгоритм может сходиться с разной скоростью.

Обе эти проблемы оказываются крайне важными для применения ЕМ-алгоритма в “скользящем” режиме в рамках СРС-метода.

5.3.9. Выбор начальных приближений для ЕМ-алгоритма

При реализации ЕМ-алгоритма выбор начального (стартового) приближения представляет собой отдельную серьезную проблему. Эта проблема будет затронута при обсуждения реализации ЕМ-алгоритма в “скользящем” режиме в разделе 6.3.2. Здесь же следует упомянуть, что в работе (Королев, 2007b) предложено в качестве стартового приближения для ЕМ-алгоритма использовать приближенные решения задачи разделения смесей, полученные с применением идеологии наименьших квадратов или наименьших модулей. При таком подходе удастся избежать применения численных процедур итерационного типа, а алгоритм отыскания приближенного решения сводится к реализации симплекс-метода.

5.3.10. Правила остановки ЕМ-алгоритма

Необходимо иметь разумные критерии для остановки итерационного ЕМ-алгоритма. Существует несколько подходов, которые и будут перечислены здесь.

Выберем малое число $\delta > 0$.

(i) Условимся останавливать ЕМ-алгоритм, если расстояние между эмпирической функцией распределения и теоретической смесью функций распределения, в которую подставлены параметры, вычисленные с помощью ЕМ-алгоритма, становится меньше δ :

$$\|F^{(m)} - F_{\text{emp}}\| < \delta, \quad (5.3.25)$$

где

$$F^{(m)}(x) = \sum_{j=1}^k p_j^{(m)} \Phi\left(\frac{x - a_j^{(m)}}{\sigma_j^{(m)}}\right)$$

– теоретическая смесь нормальных функций распределения, в которую подставлены параметры, вычисленные с помощью ЕМ-алгоритма на m -итерации,

$$F_{\text{emp}}(x) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}_{(-\infty, x)}(x_j)$$

– эмпирическая функция распределения ($\mathbb{I}_{(-\infty, x)}(x_j) = 1$, если $X_j < x$, и $\mathbb{I}_{(-\infty, x)}(x_j) = 0$ в противном случае).

(ii) ЕМ-алгоритм можно останавливать, когда разность между значениями функции правдоподобия, вычисленными на соседних итерациях, мала, то есть дальнейшая работа ЕМ-алгоритма практически не увеличит текущее правдоподобие:

$$|L(\theta^{(m)}; \mathbf{x}) - L(\theta^{(m-1)}; \mathbf{x})| < \delta. \quad (5.3.26)$$

(iii) ЕМ-алгоритм можно останавливать, когда расстояние между оценками параметров, вычисленными на соседних итерациях, мало, то есть дальнейшая работа ЕМ-алгоритма практически не изменит уже полученные оценки:

$$\|\theta^{(m)} - \theta^{(m-1)}\| < \delta. \quad (5.3.27)$$

В соотношениях (5.3.25) и (5.3.27) нормы $\|\cdot\|$ можно определять по-разному.

У всех перечисленных критериев остановки есть и достоинства, и недостатки.

Критерий (5.3.25), основанный на близости эмпирической функции распределения и теоретической смеси функций распределения, в которую подставлены параметры, вычисленные с помощью ЕМ-алгоритма, дает хорошую возможность судить о том, насколько хороша полученная аппроксимация. Однако ЕМ-алгоритм может никогда не достичь таких значений параметров теоретической функции распределения, которые обеспечивают хорошее согласие с имеющимися наблюдениями.

Критерий (5.3.26), ориентирующийся на функцию правдоподобия, естественно согласуется с методом, используемым в ЕМ-алгоритме, но как бы сводит все имеющиеся возможности, так сказать, степени свободы алгоритма, к единственному значению. Согласно этому критерию алгоритм может быть остановлен, если последовательно вычисленные значения L

близки, но параметры далеки от точки максимума (функция правдоподобия может принимать близкие значения на двух достаточно отдаленных друг от друга множествах значений параметров).

Критерий (5.3.27), использующий расстояние между последовательно полученными оценками вектора параметров, конечно же остановит ЕМ-алгоритм, если последний достигнет точки максимума функции правдоподобия. К сожалению, этот критерий также остановит алгоритм в случае, когда алгоритм достигнет области, далекой от точки максимума, но в которой скорость сходимости очень низка.

5.4. Модификации ЕМ-алгоритма

5.4.1. Медианные модификации ЕМ-алгоритма

Как было экспериментально установлено, ЕМ-алгоритм обладает сильной неустойчивостью по начальным данным. Например, в случае четырехкомпонентной смеси нормальных законов при объеме выборки 200–300 наблюдений замена лишь одного наблюдения другим может кардинально изменить итоговые оценки, полученные с помощью ЕМ-алгоритма. Возможно, эта неустойчивость обусловлена тем, что стандартные (наиболее правдоподобные для случая нормального распределения) оценки математического ожидания и дисперсии (среднее арифметическое и выборочная дисперсия) при “засорении” (*контаминации*) выборки “посторонними” или “паразитными” наблюдениями становятся заметно менее эффективными по сравнению со, скажем, выборочной медианой. Этот эффект обнаружен Дж. Тьюки (Tukey, 1960) и описан, например, в (Айвазян и др., 1983) и (Королев, 2006). Формально модель контаминации Тьюки сводится к тому, что вместо “чистого” модельного распределения, интерпретируемого как *однородная* модель, в качестве модельного распределения рассматривается неоднородная модель, имеющая вид смеси исходного “чистого” распределения и некоторого другого закона, описывающего “засоряющие” наблюдения. В задаче разделения смесей по самой сути модели, когда оцениваются параметры одной компоненты смеси, наблюдения с распределениями, соответствующими другим компонентам, являются “загрязняющими”. Это обстоятельство может сыграть особенно важную роль при реализации SEM-алгоритма, описываемого ниже.

Для противодействия указанной неустойчивости ЕМ-алгоритма мож-

но использовать его медианные модификации. Смысл этих модификаций в том, что наиболее “неустойчивые” этапы выполнения ЕМ-алгоритма заменяются более устойчивыми. В частности, на М-этапе неустойчивые моментные оценки наибольшего правдоподобия (которые для нормальных компонент минимизируют квадратичный риск) заменяются более устойчивыми (робастными) оценками медианного типа, оптимальными в смысле среднего абсолютного отклонения.

Опишем две возможные медианные модификации М-этапа ЕМ-алгоритма. В рамках этих модификаций параметры a_i оцениваются одинаково. Различными являются лишь оценки параметров σ_i .

Пусть числа $g_{ij}^{(m)}$ известны. По числам $g_{ij}^{(m)}$ определим “вероятности” $p_{ij}^{(m)}$ по правилу

$$p_{ij}^{(m)} = g_{ij}^{(m)} \left(\sum_{j=1}^n g_{ij}^{(m)} \right)^{-1}, \quad i = 1, \dots, k; \quad j = 1, \dots, n$$

(n – объем выборки, k – число компонент смеси). Пусть $\mathbf{x} = (x_1, \dots, x_n)$ – выборка. Тогда число $p_{ij}^{(m)}$ можно интерпретировать как вероятность того, что наблюдение x_j имеет распределение, определяемое i -й компонентой смеси.

Введем “фиктивные” случайные величины $\xi_i^{(m)}$, $i = 1, \dots, k$, которые, соответственно, принимают значение x_j с вероятностями $p_{ij}^{(m)}$, $i = 1, \dots, k$, $j = 1, \dots, n$ (несложно видеть, что $\sum_{j=1}^n p_{ij}^{(m)} = 1$). При этом оценка параметра сдвига i -й компоненты смеси на $(m+1)$ -й итерации, приведенная в предыдущем разделе, оказывается в точности равной математическому ожиданию случайной величины $\xi_i^{(m)}$:

$$a_i^{(m+1)} = \frac{1}{\sum_{j=1}^n g_{ij}^{(m)}} \sum_{j=1}^n g_{ij}^{(m)} x_j = \sum_{j=1}^n p_{ij}^{(m)} x_j = E_{\theta^{(m)}} \xi_i^{(m)}.$$

Для того чтобы построить модификацию ЕМ-алгоритма, более устойчивую по отношению к наличию “засоряющих” наблюдений (а при оценивании параметров какой-либо компоненты смеси наблюдения, распределения которых соответствуют другим компонентам, неизбежно будут “засоряющими” по отношению к оцениваемой компоненте), в качестве оценки параметра a_i на $(m+1)$ -й итерации предлагается взять медиану $\text{med } \xi_i^{(m)}$ случайной величины $\xi_i^{(m)}$, которую можно вычислить так. Переупорядочим значения $x_1, \dots, x_n$ случайной величины $\xi_i^{(m)}$ по неубыванию. Получим

вариационный ряд $x_{(1)}, \dots, x_{(n)}$. Ясно, что одно и то же переупорядочение имеет место для значений всех случайных величин $\xi_i^{(m)}$. Одновременно переставятся и вероятности $p_{ij}^{(m)}$, соответствующие значениям каждой случайной величины $\xi_i^{(m)}$. Пусть $\hat{p}_{ij}^{(m)}$ – это та из вероятностей $p_{ij}^{(m)}$, которая соответствует значению $x_{(j)}$ случайной величины $\xi_i^{(m)}$. Положим

$$J_i = \min\{j : \hat{p}_{i1}^{(m)} + \hat{p}_{i2}^{(m)} + \dots + \hat{p}_{ij}^{(m)} \geq \frac{1}{2}\}.$$

Тогда

$$a_i^{(m+1)} = \text{med } \xi_i^{(m)} = x_{(J_i)}. \quad (5.4.1)$$

Для оценивания параметра σ_i на $(m+1)$ -ой итерации сначала по указанной выше схеме вычислим медиану случайной величины $|\xi_i^{(m)} - a_i^{(m+1)}|$,

$$\hat{m}_i^{(m+1)} = \text{med} |\xi_i^{(m)} - a_i^{(m+1)}|.$$

Затем введем “фиктивную” случайную величину $\zeta_i^{(m+1)}$ с функцией распределения

$$P_{\theta^{(m+1)}}(\zeta_i^{(m+1)} < x) = \Phi\left(\frac{x - a_i^{(m+1)}}{\sigma_i^{(m+1)}}\right),$$

то есть распределение случайной величины $\zeta_i^{(m+1)}$ является i -й компонентой смеси. “Эмпирическим” аналогом случайной величины $\zeta_i^{(m+1)}$ является случайная величина $\xi_i^{(m)}$, введенная ранее. В идеале (при достаточно большом m и при большом n) должно быть справедливо приближенное равенство $P_{\theta^{(m+1)}}(\zeta_i^{(m+1)} < x) \approx P_{\theta^{(m+1)}}(\xi_i^{(m)} < x)$, $-\infty < x < +\infty$.

Таким образом, отыскав эмпирическую медиану $\hat{m}_i^{(m+1)}$ (то есть медиану случайной величины $|\xi_i^{(m)} - a_i^{(m+1)}|$), в соответствии с идеологией метода моментов можно заметить, что она близка к медиане $\mu_i^{(m+1)}$ случайной величины $|\zeta_i^{(m+1)} - a_i^{(m+1)}|$.

Медиана $\mu_i^{(m+1)}$ случайной величины $|\zeta_i^{(m+1)} - a_i^{(m+1)}|$ определяется из условия

$$P_{\theta^{(m+1)}}(|\zeta_i^{(m+1)} - a_i^{(m+1)}| \leq \mu_i^{(m+1)}) = \frac{1}{2}.$$

Но

$$\begin{aligned}
& P_{\theta^{(m+1)}}(|\zeta_i^{(m+1)} - a_i^{(m+1)}| \leq \mu_i^{(m+1)}) = \\
& = P_{\theta^{(m+1)}}(-\mu_i^{(m+1)} \leq \zeta_i^{(m+1)} - a_i^{(m+1)} \leq \mu_i^{(m+1)}) = \\
& = P_{\theta^{(m+1)}}(a_i^{(m+1)} - \mu_i^{(m+1)} \leq \zeta_i^{(m+1)} \leq a_i^{(m+1)} + \mu_i^{(m+1)}) = \\
& = \Phi\left(\frac{(a_i^{(m+1)} + \mu_i^{(m+1)}) - a_i^{(m+1)}}{\sigma_i^{(m+1)}}\right) - \Phi\left(\frac{(a_i^{(m+1)} - \mu_i^{(m+1)}) - a_i^{(m+1)}}{\sigma_i^{(m+1)}}\right) = \\
& = \Phi\left(\frac{\mu_i^{(m+1)}}{\sigma_i^{(m+1)}}\right) - \Phi\left(-\frac{\mu_i^{(m+1)}}{\sigma_i^{(m+1)}}\right) = 2\Phi\left(\frac{\mu_i^{(m+1)}}{\sigma_i^{(m+1)}}\right) - 1.
\end{aligned}$$

Следовательно, справедливо соотношение

$$2\Phi\left(\frac{\mu_i^{(m+1)}}{\sigma_i^{(m+1)}}\right) - 1 = \frac{1}{2},$$

то есть

$$\Phi\left(\frac{\mu_i^{(m+1)}}{\sigma_i^{(m+1)}}\right) = \frac{3}{4},$$

что эквивалентно соотношению

$$\frac{\mu_i^{(m+1)}}{\sigma_i^{(m+1)}} = u_{3/4},$$

где $u_{3/4}$ – квантиль порядка $3/4$ стандартного нормального закона. В таблицах находим $u_{3/4} \approx 0.6745$. Следуя идеологии метода моментов, приравняем эмпирическую медиану $\hat{m}_i^{(m+1)}$ теоретической медиане $\mu_i^{(m+1)}$ и окончательно получим уравнение для оценки параметра σ_i на $(m+1)$ -ой итерации:

$$\sigma_i^{(m+1)} = \frac{\hat{m}_i^{(m+1)}}{u_{3/4}} = 1.4826 \hat{m}_i^{(m+1)}. \quad (5.4.2)$$

Оценки $p_i^{(m+1)}$ весов p_i в модели (5.3.15) ищутся, как и ранее, по формулам (5.3.23). Числа же $g_{ij}^{(m+1)}$ на каждой итерации переназначаются так же, как и прежде, а именно,

$$g_{ij}^{(m+1)} = \frac{\frac{p_i^{(m+1)}}{\sigma_i^{(m+1)}} \exp\left\{-\frac{1}{2}\left(\frac{x_j - a_i^{(m+1)}}{\sigma_i^{(m+1)}}\right)^2\right\}}{\sum_{r=1}^k \frac{p_r^{(m+1)}}{\sigma_r^{(m+1)}} \exp\left\{-\frac{1}{2}\left(\frac{x_j - a_r^{(m+1)}}{\sigma_r^{(m+1)}}\right)^2\right\}}. \quad (5.4.3)$$

Итак, соотношения (5.3.23), (5.4.1), (5.4.2) и (5.4.3) определяют первую медианную модификацию ЕМ-алгоритма.

Вторая медианная модификация ЕМ-алгоритма отличается от первой лишь способом оценивания параметров σ_i . А именно, вычислим $E_{\theta^{(m+1)}} |\zeta_i^{(m+1)} - a_i^{(m+1)}|$. Имеем

$$\begin{aligned} E_{\theta^{(m+1)}} |\zeta_i^{(m+1)} - a_i^{(m+1)}| &= \int_{-\infty}^{\infty} |x - a_i^{(m+1)}| d_x \Phi \left(\frac{x - a_i^{(m+1)}}{\sigma_i^{(m+1)}} \right) = \\ &= 2 \int_0^{\infty} x d_x \Phi \left(\frac{x}{\sigma_i^{(m+1)}} \right) = \sigma_i^{(m+1)} \sqrt{\frac{2}{\pi}}. \end{aligned}$$

Эмпирическим аналогом “теоретического” математического ожидания $E_{\theta^{(m+1)}} |\zeta_i^{(m+1)} - a_i^{(m+1)}|$ является величина

$$s_i^{(m+1)} = E_{\theta^{(m)}} |\xi_i^{(m)} - a_i^{(m+1)}| = \sum_{j=1}^n p_{ij}^{(m)} |x_j - a_i^{(m+1)}|.$$

Реализуя метод моментов и приравнивая величину $E_{\theta^{(m+1)}} |\zeta_i^{(m+1)} - a_i^{(m+1)}|$ ее эмпирическому аналогу, получаем еще одну оценку для параметра σ_i на $(m+1)$ -ой итерации:

$$\sigma_i^{(m+1)} = \sqrt{\frac{\pi}{2}} \cdot s_i^{(m+1)} = 1.2533 \cdot s_i^{(m+1)}. \quad (5.4.4)$$

Таким образом, вторая медианная модификация ЕМ-алгоритма определяется соотношениями (5.3.23), (5.4.1), (5.4.3) и (5.4.4).

Заметим, что вторая модификация более соответствует духу так называемой L_1 -теории устойчивого оценивания в силу известного свойства

$$\arg \min_a E_{\theta^{(m+1)}} |\zeta_i^{(m+1)} - a| = \text{med } \zeta_i^{(m+1)} \quad (\approx \text{med } \xi_i^{(m)} = a_i^{(m+1)}).$$

Другие популярные модификации ЕМ-алгоритма (SEM-алгоритм, СЕМ-алгоритм, МСЕМ и SAЕМ-алгоритмы), а также статистические критерии выбора “правильной” модели смеси подробно описаны в следующих разделах.

5.4.2. SEM-алгоритм

Как уже отмечалось, классический ЕМ-алгоритм относится к категории так называемых “жадных” алгоритмов (greedy algorithms) в том смысле, что он “бросается” на первый попавшийся локальный максимум. Другими словами, будучи методом *локальной оптимизации*, он приводит не к глобальному максимуму функции правдоподобия, а к тому локальному максимуму, который, грубо говоря, является ближайшим к начальному приближению.

Самый простой способ противодействия этому свойству заключается в том, чтобы, не ограничиваясь единственным начальным приближением и, соответственно, единственной траекторией ЕМ-алгоритма, реализовать несколько траекторий, задавая (например, случайно) нескольких различных начальных приближений, а затем выбрать тот из результатов, для которого правдоподобие является наибольшим среди всех реализованных траекторий ЕМ-алгоритма. Однако при таком подходе остается неясным ответ на вопрос о том, каким механизмом разумнее всего пользоваться при переходе от одного начального приближения к другому. В частности, когда начальное приближение задается случайно, без дополнительной информации нельзя исчерпывающим образом определить распределение вероятностей, в соответствии с которым следует генерировать очередное начальное приближение.

Другой, оказавшийся весьма эффективным, способ заключается как бы в случайном, но целенаправленном “встряхивании” наблюдений (выборки) на каждой итерации. Этот способ лежит в основе SEM-алгоритма, название которого является аббревиатурой термина Stochastic EM-algorithm (стохастический (или случайный) ЕМ-алгоритм). SEM-алгоритм, предложенный в работах (Broniatowski, Celeux and Diebolt, 1983), (Celeux and Diebolt, 1984), (Celeux and Diebolt, 1985), весьма прост.

Чтобы описать SEM-алгоритм, представим ненаблюдаемую информацию в иной форме (однако, по сути, эквивалентной старой форме). А именно, будем считать, что каждому наблюдению x_j соответствует *вектор* $\vec{y}_j = (y_{1j}, y_{2j}, \dots, y_{kj})$, $j = 1, \dots, n$, где k – число компонент смеси, n – объем выборки. При этом

$$y_{ij} = \begin{cases} 1, & \text{если наблюдение } x_j \text{ порождено } i\text{-й компонентой смеси,} \\ 0, & \text{в противном случае.} \end{cases}$$

Связь между “старой” y_j и “новой” $\vec{y}_j$ формами представления ненаблюдаемой информации такова: y_j равен такому номеру i , для которого $y_{ij} = 1$.

При каждом j единице равна только одна из компонент вектора $\vec{y}_j$, остальные компоненты этого вектора равны нулю.

В терминах величин $\mathbf{y} = \{\vec{y}_j = (y_{1j}, y_{2j}, \dots, y_{kj}), j = 1, \dots, n\}$ логарифм полной функции правдоподобия для модели (5.1) принимает вид

$$\begin{aligned} \log L(\theta; \mathbf{x}, \mathbf{y}) &= \sum_{j=1}^n \sum_{i=1}^k y_{ij} \log[p_i \psi_i(x_j; t_i)] = \\ &= \sum_{i=1}^k \log p_i \sum_{j=1}^n y_{ij} + \sum_{i=1}^k \sum_{j=1}^n y_{ij} \log \psi_i(x_j; t_i). \end{aligned} \quad (5.4.5)$$

Векторы $\vec{y}_j = (y_{1j}, y_{2j}, \dots, y_{kj})$, $j = 1, \dots, n$, разбивают исходную наблюдаемую выборку $\mathbf{x}$ на k классов (кластеров) $\mathcal{K}_1, \dots, \mathcal{K}_k$:

$$\mathbf{x} = \mathcal{K}_1 \cup \dots \cup \mathcal{K}_k.$$

Для каждого $i = 1, \dots, k$ с формальной точки зрения $\mathcal{K}_i$ – это множество тех наблюдений x_j , каждому из которых соответствует $y_{ij} = 1$. При этом каждое наблюдение x_j входит ровно в один кластер, то есть $\mathcal{K}_i \cap \mathcal{K}_j = \emptyset$ при $i \neq j$. Пусть ν_i – это число наблюдений, попавших в кластер $\mathcal{K}_i$, $i = 1, \dots, k$,

$$\nu_i = \sum_{j=1}^n y_{ij}.$$

Очевидно, что $\nu_1 + \dots + \nu_k = n$. Тогда, продолжая (5.4.5), для логарифма полной функции правдоподобия в модели (5.1) мы получаем представление

$$\log L(\theta; \mathbf{x}, \mathbf{y}) = \sum_{i=1}^k \nu_i \log p_i + \sum_{i=1}^k \sum_{j: x_j \in \mathcal{K}_i} \log \psi_i(x_j; t_i). \quad (5.4.6)$$

Если бы величины y_{ij} были известны, то искать значение θ , максимизирующее функцию правдоподобия (5.4.6), можно было бы, максимизируя по θ каждое из слагаемых в правой части (5.4.6), поскольку эти слагаемые зависят только от “своих” групп параметров. А именно, с помощью метода неопределенных множителей Лагранжа несложно убедиться, что максимум первого слагаемого по набору $p_1, \dots, p_k$ при очевидном ограничении $p_1 + \dots + p_k = 1$ достигается при

$$p_i^* = \frac{\nu_i}{n} \quad (5.4.7)$$

(см. (5.7)). Далее заметим, что

$$\sum_{j: x_j \in \mathcal{K}_i} \log \psi_i(x_j; t_i) = \log \prod_{j: x_j \in \mathcal{K}_i} \psi_i(x_j; t_i) \equiv \log L_i(t_i; \mathcal{K}_i),$$

где $L_i(t_i; \mathcal{K}_i)$ – это функция правдоподобия параметра t_i , построенная по подвыборке (кластеру) $\mathcal{K}_i$ в предположении, что каждый элемент подвыборки имеет плотность распределения $\psi_i(x; t_i)$. Отсюда видно, что, полагая

$$t_i^* = \arg \max_t L_i(t; \mathcal{K}_i), \quad i = 1, \dots, k, \quad (5.4.8)$$

мы доставляем максимум второму слагаемому в правой части (5.4.6). Легко видеть, что соотношение (5.4.8) определяет обычные оценки наибольшего правдоподобия для параметров i -ой компоненты смеси (5.1), построенные по подвыборке наблюдений, распределение которых равно этой компоненте, то есть по кластеру $\mathcal{K}_i$.

Таким образом, если бы величины y_{ij} были известны, то оценки наибольшего правдоподобия параметров модели (5.1) определялись бы соотношениями (5.4.7) и (5.4.8). Однако на практике величины y_{ij} не известны. Идея SEM-алгоритма заключается в том, что эти величины определяются с помощью специального имитационного моделирования. Итерационный SEM-алгоритм определяется так.

Предположим, что известны значения $g_{ij}^{(m)}$ апостериорных вероятностей принадлежности наблюдения x_j к кластеру $\mathcal{K}_i$, $i = 1, \dots, k$; $j = 1, \dots, n$; m – номер итерации (отметим, что

$$\sum_{i=1}^k g_{ij}^{(m)} = 1$$

для каждого j и при каждом m).

На первом этапе SEM-алгоритма (S-этапе, от слов Stochastic или Simulation) для каждого $j = 1, \dots, n$ генерируются векторы $\vec{y}_j^{(m+1)} = (y_{1j}^{(m+1)}, y_{2j}^{(m+1)}, \dots, y_{kj}^{(m+1)})$ как реализации случайных векторов с полиномиальным распределением с параметрами 1 и $g_{1j}^{(m)}, \dots, g_{kj}^{(m)}$ ($g_{ij}^{(m)}$ – это вероятность того, что величина $y_{ij}^{(m+1)}$ равна единице). По векторам $\vec{y}_j^{(m+1)}$ определяется разбиение выборки $\mathbf{x} = (x_1, \dots, x_n)$ на кластеры $\mathcal{K}_1^{(m+1)}, \dots, \mathcal{K}_k^{(m+1)}$ и соответствующие числа $\nu_1^{(m+1)}, \dots, \nu_k^{(m+1)}$ (численности кластеров) на $(m+1)$ -й итерации. (Можно сказать, что на S-этапе реализуется случайное “встряхивание” исходной выборки, о котором говорилось выше.)

На втором этапе (М-этапе), этапе максимизации, в соответствии с формулами (5.4.7) и (5.4.8) вычисляются оценки максимального правдоподобия компонент параметра θ :

$$p_i^{(m+1)} = \frac{\nu_i^{(m+1)}}{n}, \quad (5.4.9)$$

$$t_i^{(m+1)} = \arg \max_t L_i(t; \mathcal{K}_i^{(m+1)}), \quad i = 1, \dots, k. \quad (5.4.10)$$

Наконец, на третьем этапе (Е-этапе), переназначаются вероятности g_{ij} . Название этого этапа восходит к слову Expectation. Это обусловлено тем, что, если $\vec{Y}_j^{(m+1)} = (Y_{1j}^{(m+1)}, \dots, Y_{kj}^{(m+1)})$ – это случайный вектор, реализацией которого является вектор $\vec{y}_j^{(m+1)}$, а $\vec{X} = (X_1, \dots, X_n)$ – это случайный вектор, реализацией которого является выборка $\mathbf{x} = (x_1, \dots, x_n)$, то по определению

$$g_{ij}^{(m+1)} = \mathbb{E}_{\theta^{(m+1)}}(Y_{ij}^{(m+1)} | X_j)$$

(очевидно, $Y_{ij}^{(m+1)}$ – это индикатор (случайного) события $\{X_j \in \mathcal{K}_i^{(m+1)}\}$, а математическое ожидание индикатора случайного события равно вероятности этого события). При известном значении $X_j = x_j$ в соответствии с (5.3) мы имеем

$$g_{ij}^{(m+1)} = \frac{p_i^{(m+1)} \psi_i(x_j; t_i^{(m+1)})}{\sum_{r=1}^k p_r^{(m+1)} \psi_r(x_j; t_r^{(m+1)})}. \quad (5.4.11)$$

Для случая смеси нормальных распределений, в которой

$$\psi_i(x; t_i) = \frac{1}{\sigma_i} \phi\left(\frac{x - a_i}{\sigma_i}\right), \quad x \in \mathbb{R},$$

(см. главу 6) SEM-алгоритм выглядит так. Соотношение (5.4.9) остается без изменений, соотношение (5.4.10) трансформируется в пару соотношений

$$a_i^{(m+1)} = \frac{1}{\nu_i^{(m+1)}} \sum_{j=1}^n y_{ij}^{(m+1)} x_j = \frac{1}{\nu_i^{(m+1)}} \sum_{j: x_j \in \mathcal{K}_i^{(m+1)}} x_j, \quad (5.4.12)$$

$$\begin{aligned} \sigma_i^{(m+1)} &= \left[\frac{1}{\nu_i^{(m+1)}} \sum_{j=1}^n y_{ij}^{(m+1)} (x_j - a_i^{(m+1)})^2 \right]^{1/2} = \\ &= \left[\frac{1}{\nu_i^{(m+1)}} \sum_{j: x_j \in \mathcal{K}_i^{(m+1)}} (x_j - a_i^{(m+1)})^2 \right]^{1/2}. \end{aligned}$$

Соотношение же (5.4.11) принимает вид (7.3).

Свойства SEM-алгоритма были подвергнуты эмпирическому исследованию в (Celeux and Diebolt, 1984) и теоретическому исследованию в (Diebolt and Celeux, 1993), (Ip, 1994) и (Diebolt and Ip, 1996). В частности, в этих работах для многих достаточно общих конкретных случаев отмечено, что построенная SEM-алгоритмом последовательность $\{\theta^{(m)}\}_{m \geq 1}$, вообще говоря, не сходится с вероятностью единица, но образует цепь Маркова, которая при некоторых дополнительных условиях регулярности довольно быстро сходится к стационарному распределению. Стационарность достигается после довольно продолжительного периода “приработки” алгоритма. При этом получаемые с помощью SEM-алгоритма оценки параметров смеси являются асимптотически несмещенными в том смысле что оценка максимального правдоподобия параметров смеси является асимптотически эквивалентной математическому ожиданию $\theta^{(m)}$ относительно стационарного распределения. Поэтому в качестве “окончательной” оценки $\tilde{\theta}^{(m)}$ параметра θ после m итераций SEM-алгоритма в упомянутых работах предлагается использовать “выборочное среднее”

$$\tilde{\theta}^{(m)} = \tilde{\theta}^{(m)}(m_0) = \frac{1}{m - m_0} \sum_{r=m_0+1}^m \theta^{(r)},$$

где m_0 – настолько большое число, что при $r > m_0$ цепь Маркова $\theta^{(r)}$ близка к стационарному режиму.

Многочисленные реализации SEM-алгоритма показали, что он работает относительно быстро по сравнению с другими методами, результаты его работы практически не зависят от начального приближения, позволяет избегать выхода на неустойчивые локальные максимумы анализируемой функции правдоподобия за счет постоянного случайного “встряхивания” выборки и, более того – как правило, приводит к глобальному максимуму этой функции. Кроме того, SEM-алгоритм легко модифицировать с целью отыскания числа k компонент смеси, если оно заранее не известно (об этом см. ниже).

Так как на каждой итерации SEM-алгоритма в каждый из кластеров $\mathcal{K}_1^{(m+1)}, \dots, \mathcal{K}_k^{(m+1)}$ могут попасть “лишние” наблюдения, фактически распределенные в соответствии с другими компонентами смеси, то в силу причин, о которых говорилось в предыдущем разделе, устойчивые медианные модификации могут быть весьма перспективными для SEM-алгоритма. Медианные модификации SEM-алгоритма определяются следующим образом.

Упорядочим элементы выборки $\mathbf{x} = (x_1, \dots, x_n)$, попавшие в кластер $\mathcal{K}_i^{(m+1)}$, по неубыванию. Полученный в результате набор обозначим $\mathcal{K}_i^{(m+1)} = \{x_{i,1}^{(m+1)}, \dots, x_{i,\nu_i}^{(m+1)}\}$. Тогда в случае смеси нормальных компонент вместо (5.4.12) можно использовать более устойчивую оценку

$$a_i^{(m+1)} = \begin{cases} \frac{1}{2}(x_{i,\nu_i^{(m+1)}/2}^{(m+1)} + x_{i,\nu_i^{(m+1)}/2+1}^{(m+1)}), & \text{если } \nu_i^{(m+1)} - \text{четное,} \\ x_{i, [\nu_i^{(m+1)}/2]+1}^{(m+1)}, & \text{если } \nu_i^{(m+1)} - \text{нечетное,} \end{cases}$$

где символ $[z]$ обозначает целую часть числа z . Другими словами, в качестве оценки параметра a_i на $(m+1)$ -й итерации SEM-алгоритма можно использовать выборочную медиану кластера $\mathcal{K}_i^{(m+1)}$.

Исходя из тех же рассуждений, что и в предыдущем разделе, в качестве оценки параметра σ_i на $(m+1)$ -й итерации SEM-алгоритма можно взять

$$\sigma_i^{(m+1)} = \sqrt{\frac{\pi}{2}} \cdot S_i^{(m+1)} = 1.2533 \cdot S_i^{(m+1)},$$

где

$$S_i^{(m+1)} = \frac{1}{\nu_i^{(m+1)}} \sum_{j=1}^{\nu_i^{(m+1)}} |x_{i,j}^{(m+1)} - a_i^{(m+1)}|$$

– выборочное среднее абсолютное отклонение, вычисленное для кластера $\mathcal{K}_i^{(m+1)}$.

5.4.3. СЕМ-алгоритм

В работах (Celeux and Govaert, 1991), (Celeux and Govaert, 1992) был предложен так называемый *классификационный* ЕМ-алгоритм или СЕМ-алгоритм (аббревиатура английского названия Classification EM-algorithm). Этот алгоритм совпадает с SEM-алгоритмом за исключением того, что вместо S-этапа для определения величин $\vec{y}_j = (y_{1j}, y_{2j}, \dots, y_{kj})$ используется детерминированное правило, эквивалентное классификации по принципу максимума апостериорной вероятности, согласно которому для каждого j наблюдение x_j приписывается к тому из кластеров $\mathcal{K}_1, \dots, \mathcal{K}_k$, вероятность принадлежности x_j к которому является наибольшей. А именно, на $(m+1)$ -й итерации СЕМ-алгоритма равной единице полагается та из компонент вектора $\vec{y}_j^{(m+1)} = (y_{1j}^{(m+1)}, y_{2j}^{(m+1)}, \dots, y_{kj}^{(m+1)})$, номер которой равен номеру наибольшего из чисел $g_{1j}^{(m+1)}, g_{2j}^{(m+1)}, \dots, g_{kj}^{(m+1)}$.

Формулы для оценок параметров на очередной итерации СЕМ-алгоритма (в том числе и его медианной модификации) идентичны соответствующим формулам для SEM-алгоритма.

5.4.4. МСЕМ и SAЕМ-алгоритмы

Классический МСЕМ-алгоритм

При решении некоторых прикладных задач с помощью классического ЕМ-алгоритма вычисление функции $Q(\theta; \theta^{(m)})$ по формуле (2.2) на Е-этапе может быть весьма трудоемким из-за того, что соответствующий интеграл невозможно привести к замкнутому аналитическому виду в терминах элементарных функций. Чтобы обойти это препятствие, в работе (Wei and Tanner, 1990) предложен МСЕМ-алгоритм (Monte-Carlo EM-algorithm). Суть этого алгоритма заключается в том, что функция $Q(\theta; \theta^{(m)})$ заменяется ее аналогом $\bar{Q}(\theta; \theta^{(m)})$, вычисленным с помощью имитационного моделирования как

$$\bar{Q}(\theta; \theta^{(m)}) = \frac{1}{N_m} \sum_{r=1}^{N_m} \log f_{\theta}(\mathbf{x}, \mathbf{y}_r),$$

где $\mathbf{y}_1, \dots, \mathbf{y}_{N_m}$ – независимые псевдослучайные реализации ненаблюдаемой величины $\mathbf{y}$, вычисленные (симулированные) в соответствии с условным распределением $f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x})$. При этом в силу закона больших чисел при достаточно больших N_m имеет место приближенное равенство

$$\bar{Q}(\theta; \theta^{(m)}) \approx Q(\theta; \theta^{(m)}). \quad (5.4.13)$$

Далее на М-этапе $(m+1)$ -й итерации МСЕМ-алгоритма ищется такое значение $\theta^{(m+1)}$ параметра θ , для которого

$$\bar{Q}(\theta^{(m+1)}; \theta^{(m)}) \geq \bar{Q}(\theta; \theta^{(m)})$$

для любого $\theta \in \Theta$.

Хотя аппроксимация (5.4.13) в целом позволяет преодолеть вычислительные сложности на Е-этапе, она привносит в метод дополнительную погрешность

$$\varepsilon^{(m)}(N_m) = \left| \int [\log f_{\theta}(\mathbf{x}, \mathbf{y})] f_{\theta^{(m)}}(\mathbf{y}|\mathbf{x}) \mu_{\mathbf{Y}}(d\mathbf{y}) - \frac{1}{N_m} \sum_{r=1}^{N_m} \log f_{\theta}(\mathbf{x}, \mathbf{y}_r) \right|,$$

величина которой зависит от объема N_m имитируемой выборки значений ненаблюдаемой величины $\mathbf{y}$. В принципе, величина $\varepsilon^{(m)}(N_m)$ может быть сделана произвольно малой за счет простого увеличения N_m . Однако на практике вычислительные ресурсы (время, мощность процессора, объем памяти) всегда ограничены. Более того, как отмечено в работе (Wei and Tanner, 1990), для гарантированной сходимости МСЕМ-алгоритма допустимы умеренные значения N_m на начальных итерациях, но на заключительном этапе необходима все бóльшая и бóльшая точность, то есть все меньшие и меньшие значения $\varepsilon^{(m)}(N_m)$ и, следовательно, все бóльшие и бóльшие значения N_m . Вопросам оптимального распределения объема симуляции и вычислительных ресурсов по итерациям МСЕМ-алгоритма посвящены работы (Booth and Hobert, 1999), (Levine and Casella, 2001), (Levine and Fan, 2004), (Caffo et al., 2005).

МС-модификация SEM-алгоритма

Легко видеть, что с формальной точки зрения SEM-алгоритм, описанный в разделе 7.2, является специальным МСЕМ-алгоритмом для решения задачи разделения смесей вероятностных распределений с $N_m = 1$, $m \geq 1$. Однако возможны и нетривиальные (с $N_m > 1$) МС-версии SEM-алгоритма для разделения смесей.

Действительно, в отличие от классического SEM-алгоритма, в рамках которого на $(m + 1)$ -й итерации генерируется лишь одна реализация $\mathbf{y}^{(m+1)} = \{\vec{y}_1^{(m+1)}, \dots, \vec{y}_n^{(m+1)}\}$ векторов, определяющих разбиение выборки $\mathbf{x}$ на кластеры (см. раздел 7.2), можно на каждой итерации генерировать не одну, но $N_{m+1} > 1$ реализаций указанного вектора $\mathbf{y}_1^{(m+1)}, \dots, \mathbf{y}_{N_{m+1}}^{(m+1)}$. Обозначим кластер $\mathcal{K}_i$ при r -й реализации $\mathbf{y}_r^{(m+1)}$ указанного вектора на $(m + 1)$ -й итерации через $\mathcal{K}_{i,r}^{(m+1)}$, а количество элементов выборки $\mathbf{x}$, попавших в этот кластер, через $\nu_{i,r}^{(m+1)}$, $r = 1, \dots, N_{m+1}$. Тогда в качестве оценок весов p_i следует взять

$$p_i^{(m+1)} = \frac{1}{nN_{m+1}} \sum_{r=1}^{N_{m+1}} \nu_{i,r}^{(m+1)}.$$

При этом, если рассматривается модель (5.1) с нормальными компонентами, то в качестве оценок параметров a_i и σ_i на $(m + 1)$ -й итерации следует взять

$$a_i^{(m+1)} = \frac{1}{N_{m+1}\nu_i^{(m+1)}} \sum_{r=1}^{N_{m+1}} \sum_{j: x_j \in \mathcal{K}_{i,r}^{(m+1)}} x_j,$$

$$\sigma_i^{(m+1)} = \left[\frac{1}{N_{m+1} \nu_i^{(m+1)}} \sum_{r=1}^{N_{m+1}} \sum_{j: x_j \in \mathcal{K}_{i,r}^{(m+1)}} \left(x_j - a_i^{(m+1)} \right)^2 \right]^{1/2}.$$

Величины же $g_{ij}^{(m+1)}$ переназначаются, как и в классическом SEM-алгоритме, по формуле (7.3).

Асимптотические свойства SEM- и MCEM-алгоритмов описаны в статье (Nielsen, 2000).

SAEM-алгоритм

Существуют алгоритмы, родственные MCEM-алгоритмам, которые, в противоположность классическому MCEM-алгоритму, сходятся при постоянных (не увеличивающихся) и довольно умеренных объемах N_m симуляции на каждой итерации. Таким свойством, в частности, обладает так называемый SAEM-алгоритм.

Рассмотрим последовательность $\{\gamma_m\}_{m \geq 1}$ положительных чисел такую, что

$$\sum_{m \geq 1} \gamma_m = \infty, \quad \sum_{m \geq 1} \gamma_m^2 < \infty.$$

Определим последовательность функций $W^{(m)}(\theta)$ с помощью рекуррентного соотношения

$$\begin{aligned} W^{(m)}(\theta) &= (1 - \gamma_m) W^{(m-1)}(\theta) + \gamma_m \bar{Q}(\theta; \theta^{(m)}) = \\ &= W^{(m-1)}(\theta) + \gamma_m \left(\frac{1}{N_m} \sum_{r=1}^{N_m} \log f_{\theta}(\mathbf{x}, \mathbf{y}_r) - W^{(m-1)}(\theta) \right). \end{aligned} \quad (5.4.14)$$

При этом последовательность $\{\theta^{(m)}\}_{m \geq 1}$ формируется в соответствии с правилом

$$\theta^{(m)} = \arg \max_{\theta} W^{(m)}(\theta), \quad m \geq 1. \quad (5.4.15)$$

Рекуррентное соотношение (5.4.14) совпадает с соотношением, определяющим метод оценивания, называемый *стохастической аппроксимацией*, который был предложен в работе (Robbins and Monro, 1951), см. также (Вазан, 1972), (Невельсон и Хасьминский, 1972). Поэтому разновидность EM-алгоритма, определяемая соотношениями (5.4.14) и (5.4.15), называется SAEM-алгоритмом (Stochastic Approximation EM-algorithm), см. (Celeux and Diebolt, 1992), (Delyon et al., 1999).

Одна из замечательных особенностей SAEM-алгоритма заключается в том, что он сходится при постоянных значениях N_m (Delyon et al., 1999). Более того, представляется вполне разумным, что при вычислении очередного значения $\theta^{(m)}$ SAEM-алгоритм использует информацию о *всех* симулированных данных на *всех* предшествующих итерациях.

В отличие от MCEM-алгоритма, в котором решение о значении N_m принимается на *каждой* итерации, в рамках SAEM-алгоритма решение о выборе параметров метода $\{\gamma_m\}_{m \geq 1}$ принимается *единовременно*, обычно до начала работы алгоритма. В работе (Polyak and Juditsky, 1992) показано, что скорость сходимости SAEM-алгоритма оптимальна, если шаг метода γ_m выбирается в соответствии с соотношением $\gamma_m \propto 1/m^\alpha$ при $\frac{1}{2} < \alpha \leq 1$. Тем не менее необходимо заметить, что при больших шагах (то есть $\alpha \approx \frac{1}{2}$) SAEM-алгоритм быстро попадает в окрестность решения, однако одновременно возрастает имитационная погрешность $\varepsilon^{(m)}(N_m)$ (см. предыдущий раздел). С другой стороны, при малых шагах (то есть $\alpha \approx 1$) имитационная погрешность быстро убывает, однако сильно уменьшается скорость сходимости алгоритма (Jank, 2004).

SEM-, SAEM- и MCEM-алгоритмы были подвергнуты систематическому сравнительному анализу в работе (Celeux, Chauveau and Diebolt, 1995). В этой работе отмечено, что многочисленные численные эксперименты показали, что, как правило, SEM-алгоритм оказывался более эффективным. Однако на некоторых смесях ни один из этих алгоритмов не дал разумного решения. Тем не менее, авторы упомянутой работы считают, что и в таких случаях SEM-алгоритм может быть использован в качестве средства разведочного анализа, позволяющего обозначить возможные точки максимума функции правдоподобия с целью их дальнейшего исследования другими средствами.

5.5. ЕМ-алгоритм с большим числом компонент как средство построения непараметрических оценок плотности

Если число компонент смеси (5.3.15) очень велико и сопоставимо с объемом выборки, то, даже если истинная плотность распределения отлична от (5.3.15), результат действия ЕМ-алгоритма может быть использован как

непараметрическая статистическая оценка плотности $f(x)$:

$$f(x) \approx \hat{f}(x) = \sum_{i=1}^k p_i \psi_i(x; t_i),$$

например, с ядрами вида

$$\psi_i(x; t_i) = \frac{1}{\sigma_i} \phi\left(\frac{x - a_i}{\sigma_i}\right), \quad x \in \mathbb{R},$$

где

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, \quad x \in \mathbb{R},$$

– плотность стандартного нормального распределения вероятностей, $t_i = (a_i, \sigma_i^2)$, $a_i \in \mathbb{R}$, $\sigma_i > 0$, $i = 1, \dots, k$.

Очевидным недостатком такого подхода является слишком продолжительное время вычислений вследствие большого числа оцениваемых параметров. Однако такой подход имеет неоспоримое преимущество: при применении классических ядерных оценок плотности на практике слабым местом является определение значения параметра гладкости оценок (“ширины окна”), в то время как при предлагаемом подходе параметр $\sigma = (\sigma_1, \dots, \sigma_k)$, регламентирующий гладкость оценки (“ширину окна”), во-первых, не один (многомерен) и, во-вторых, не выбирается произвольно, но оценивается *автоматически*. Другими словами, получается гибкий адаптивный алгоритм построения оценки плотности.

При этом представляет большой интерес исследование соотношения между объемом выборки и возможным числом компонент аппроксимирующей смеси.

5.6. Разделение смесей многомерных нормальных распределений с помощью ЕМ-алгоритма

Пусть $d \geq 1$ – натуральное число, $\mathbf{a} = (a_1, \dots, a_d)^\top$ – d -мерный вектор-столбец, Σ – положительно определенная симметрическая $(d \times d)$ -матрица. Многомерное (d -мерное) нормальное распределение вероятностей $\Phi_{\mathbf{a}, \Sigma}$ определяется своей плотностью вероятностей

$$\varphi_{\mathbf{a}, \Sigma}(\mathbf{x}) = \pi^{-d/2} (\text{Det} \Sigma)^{-1} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{a})^\top \Sigma^{-1} (\mathbf{x} - \mathbf{a}) \right\}, \quad \mathbf{x} \in \mathbb{R}^d. \quad (5.6.1)$$

Если d -мерный случайный вектор $\mathbf{X}$ имеет нормальное распределение $\Phi_{\mathbf{a}, \Sigma}$, то

$$\mathbf{E}\mathbf{X} = \mathbf{a}, \quad \text{Cov}\mathbf{X} = \Sigma.$$

Рассмотрим теперь многомерную смешанную нормальную модель, задаваемую распределением

$$F(A) = \sum_{j=1}^k p_j \Phi_{\mathbf{a}_j, \Sigma_j}(A), \quad (5.6.1)$$

где $A \subseteq \mathbb{R}^d$ – произвольное борелевское множество, $k \geq 1$ – натуральное число, $\mathbf{a}_1, \dots, \mathbf{a}_k$ – d -мерные случайные векторы, $\Sigma_1, \dots, \Sigma_k$ – положительно определенные симметрические $(d \times d)$ -матрицы, $p_1, \dots, p_k$ – неотрицательные числа, $p_1 + \dots + p_k = 1$.

Пусть $\mathbf{x}_1, \dots, \mathbf{x}_n$ – независимая выборка реализаций случайного вектора $\mathbf{X}$, распределение которого имеет вид (5.6.1). ЕМ-алгоритм для статистического оценивания параметров $p_1, \dots, p_k$, $\mathbf{a}_1, \dots, \mathbf{a}_k$ и $\Sigma_1, \dots, \Sigma_k$ по выборке $\mathbf{x}_1, \dots, \mathbf{x}_n$ определяется во многом аналогично одномерному случаю.

Пусть $m \geq 0$ – номер итерации. При известных значениях $p_1^{(m)}, \dots, p_k^{(m)}$, $\mathbf{a}_1^{(m)}, \dots, \mathbf{a}_k^{(m)}$ и $\Sigma_1^{(m)}, \dots, \Sigma_k^{(m)}$ параметров модели (5.6.1) на m -й итерации вычисляются величины

$$g_{ij}^{(m)} = \frac{(\text{Det}\Sigma_i^{(m)})^{-1} \exp \left\{ -\frac{1}{2}(\mathbf{x}_j - \mathbf{a}_i^{(m)})^\top (\Sigma_i^{(m)})^{-1} (\mathbf{x}_j - \mathbf{a}_i^{(m)}) \right\}}{\sum_{i=1}^k (\text{Det}\Sigma_i^{(m)})^{-1} \exp \left\{ -\frac{1}{2}(\mathbf{x}_j - \mathbf{a}_i^{(m)})^\top (\Sigma_i^{(m)})^{-1} (\mathbf{x}_j - \mathbf{a}_i^{(m)}) \right\}}.$$

С помощью этих величин затем вычисляются оценки параметров $p_1, \dots, p_k$, $\mathbf{a}_1, \dots, \mathbf{a}_k$ и $\Sigma_1, \dots, \Sigma_k$ на $(m+1)$ -й итерации:

$$\begin{aligned} \mathbf{a}_i^{(m+1)} &= \frac{1}{\sum_{j=1}^n g_{ij}^{(m+1)}} \sum_{j=1}^n g_{ij}^{(m)} \mathbf{x}_j, \\ \Sigma_i^{(m+1)} &= \frac{1}{\sum_{j=1}^n g_{ij}^{(m+1)}} \sum_{j=1}^n g_{ij}^{(m)} (\mathbf{x}_j - \mathbf{a}_i^{(m+1)})(\mathbf{x}_j - \mathbf{a}_i^{(m+1)})^\top, \\ p_i^{(m+1)} &= \frac{1}{n} \sum_{j=1}^n g_{ij}^{(m+1)}, \quad i = 1, \dots, k. \end{aligned}$$

Общее число параметров этой задачи равно $k[1 + \frac{1}{2}d(d+3)] - 1$ и квадратично возрастает с ростом размерности. Поэтому ЕМ-алгоритм может давать правдоподобные результаты лишь при очень большом объеме выборки, но при этом, естественно, сильно ухудшается его быстродействие.

5.7. Приближенное разделение конечных смесей с помощью метода фиксированных компонент для выбора начального приближения ЕМ-алгоритма

5.7.1. Основная идея метода фиксированных компонент

Идея, которая лежит в основе метода разделения смесей, описываемого в данном разделе, очень близка к идее гармонического анализа, когда периодическая функция раскладывается в ряд Фурье, то есть представляется взвешенной комбинацией (рядом), возможно, бесконечно большого числа синусов и косинусов с различными (но кратными) периодами (то есть с различными параметрами масштаба – частотами). Возможность приближения исходной функции с помощью такого разложения обоснована тем обстоятельством, что семейство синусов и косинусов с указанными свойствами образует базис, то есть полную систему (линейно независимых) функций в пространстве (регулярных) периодических функций.

Аналогия будет видна более отчетливо, если мы рассмотрим семейство чисто масштабных смесей нормальных законов с нулевым средним. Поскольку при некоторых условиях регулярности каждое распределение вероятностей, сосредоточенное на неотрицательной полуоси, может быть приближено решетчатым распределением с произвольно высокой точностью (скажем, в метрике Леви, метризирующей слабую сходимость), мы можем заключить, что семейство нормальных распределений $N_{\text{rat}} = \{\mathcal{N}(0, s) : s - \text{рациональное}\}$ образует счетную полную систему функций в пространстве масштабных смесей нормальных законов с нулевым средним. То есть для любой масштабной смеси нормальных законов и произвольно малого $\epsilon > 0$ существует конечная линейная комбинация распределений из семейства N_{rat} такая, что расстояние (скажем, расстояние Леви) между смесью и линейной комбинацией не превышает $\epsilon > 0$. Поскольку число слагаемых в такой линейной комбинации конечно и все параметры масштаба рациональны, найдется такое (минимальное) значение параметра масштаба, что параметры масштаба всех членов рассматриваемой линейной комбинации будут кратными этому минимальному значению параметра масштаба.

Рассмотрим смесь функций распределения вида

$$F(x) = \sum_{i=1}^k p_i \Phi\left(\frac{x - a_i}{\sigma_i}\right), \quad x \in \mathbb{R}, \quad (5.7.1)$$

где $k \geq 1$ – целое. В классической задаче разделения смесей параметрами, подлежащими статистическому оцениванию, являются тройки (p_i, a_i, σ_i) , $i = 1, \dots, k$, где $a_i \in \mathbb{R}$, $\sigma_i > 0$, $p_i \geq 0$, $p_1 + \dots + p_k = 1$.

Предположим, что заранее известны числа $\underline{a}$, $\bar{a}$ и $\bar{\sigma}$ такие, что $\underline{a} \leq a_i \leq \bar{a}$ и $\sigma_i \leq \bar{\sigma}$ при всех $i = 1, \dots, k$. Другими словами, известны диапазоны изменения неизвестных параметров a_i и σ_i .

Идея, лежащая в основе рассматриваемого подхода, заключается в замене интервалов $[\underline{a}, \bar{a}]$ и $(0, \bar{\sigma}]$ возможных значений неизвестных параметров масштаба σ_i и сдвига a_i дискретными множествами известных точек. Эти точки могут быть выбраны, например, исходя из следующих соображений.

Пусть ε_a и ε_σ – положительные числа, определяющие априорные требования к точности оценивания параметров a_i и σ_i :

$$\max_i |a_i - \hat{a}_i| \leq \varepsilon_a, \quad \max_i |\sigma_i - \hat{\sigma}_i| \leq \varepsilon_\sigma, \quad (5.7.2)$$

где $\hat{a}_i$ и $\hat{\sigma}_i$ – искомые оценки параметров. Числа ε_a и ε_σ также можно интерпретировать как пороги различимости возможных значений параметров: значения a' , a'' и σ' , σ'' , соответственно, считаются неразличимыми, если

$$|a' - a''| \leq \varepsilon_a, \quad |\sigma' - \sigma''| \leq \varepsilon_\sigma. \quad (5.7.3)$$

Положим $k_a = [(\bar{a} - \underline{a})/\varepsilon_a] + 1$, $k_\sigma = [\bar{\sigma}/\varepsilon_\sigma] + 1$, где символ $[z]$ обозначает целую часть числа z . Для $r = 1, 2, \dots, k_a + 1$ положим $\tilde{a}_r = \underline{a} + (r - 1)\varepsilon_a$. Аналогично, для $l = 1, 2, \dots, k_\sigma$ положим $\tilde{\sigma}_l = l\varepsilon_\sigma$. Тогда точки с координатами $(\tilde{a}_r, \tilde{\sigma}_l)$ образуют узлы конечной равномерной сети, накрывающей прямоугольник $\{(a, \sigma) : \underline{a} \leq a \leq \bar{a}, 0 \leq \sigma \leq \bar{\sigma}\}$, который представляет собой множество возможных значений параметров сдвига и масштаба компонент смеси (5.7.1) (чтобы избежать возможной некорректности, мы исключили возможность равенства параметра масштаба нулю). Число узлов полученной сети равно $K = (k_a + 1)k_\sigma$. Для удобства записи и упрощения обозначений перенумеруем каким-либо образом узлы указанной сети, вводя *единый* индекс i для координат $(\tilde{a}_i, \tilde{\sigma}_i)$ узла с номером i после перенумерации, $i = 1, \dots, K$.

Сетка, накидываемая на множество значений параметров компонент, совсем не обязана быть равномерной, то есть расстояние между соседними узлами сети не обязательно должно быть одинаковым для всех узлов. В разделе 6.4 будет описан способ выбора узлов неравномерной сетки с целью экономии временных и вычислительных ресурсов за счет расположения узлов с учетом доступной информации о возможных значениях параметров компонент смеси.

Базовая посылка рассматриваемого подхода заключается в аппроксимации смеси (5.7.1) смесью с заведомо бóльшим числом *известных* компонент:

$$F(x) = \sum_{i=1}^k p_i \Phi\left(\frac{x - a_i}{\sigma_i}\right) \approx \sum_{i=1}^K \tilde{p}_i \Phi\left(\frac{x - \tilde{a}_i}{\tilde{\sigma}_i}\right) \equiv \tilde{F}(x), \quad x \in \mathbb{R}. \quad (5.7.4)$$

Такое приближение практически допустимо, поскольку в силу соотношений (5.7.2) и (5.7.3) для любой пары (a_r, σ_r) параметров компоненты смеси (5.7.1) обязательно найдется практически не отличимая от нее пара $(\tilde{a}_i, \tilde{\sigma}_i)$ параметров компоненты смеси $\tilde{F}(x)$. Веса же остальных компонент смеси $\tilde{F}(x)$, для параметров которых не найдется “близкой” пары параметров (a_r, σ_r) компоненты смеси (5.7.1), можно считать равными нулю. Действительно, если бы в соотношении (5.7.4) вместо *приближенного* было бы *точное* равенство, то в силу идентифицируемости семейства конечных смесей нормальных законов, в полном соответствии с определением идентифицируемости конечных смесей с точностью до переиндексации были бы справедливы равенства:

$$k = K, \quad p_i = \tilde{p}_i, \quad a_i = \tilde{a}_i, \quad \sigma_i = \tilde{\sigma}_i, \quad i = 1, \dots, k.$$

Заметим, что неизвестными параметрами смеси $\tilde{F}(x)$ являются *только* веса $\tilde{p}_1, \dots, \tilde{p}_K$.

Пусть $\mathbf{x} = (x_1, \dots, x_n)$ – (независимая) выборка наблюдений, каждое из которых представляет собой реализацию случайной величины с функцией распределения $F(x)$, задаваемой соотношением (5.7.1). Пусть $(x_{(1)}, \dots, x_{(n)})$ и $F_n(x)$ – соответственно, вариационный ряд и эмпирическая функция распределения, построенные по выборке $\mathbf{x}$,

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}(x_j < x), \quad x \in \mathbb{R}.$$

При этом очевидно, что

$$F_n(x_{(j)}) = \frac{j}{n}, \quad j = 1, \dots, n. \quad (5.7.5)$$

В силу теоремы Гливенко при больших n равномерно по $x \in \mathbb{R}$ выполняется соотношение

$$F_n(x) \approx F(x). \quad (5.7.6)$$

Обозначим

$$\Phi_{ij} = \Phi\left(\frac{x_{(j)} - \tilde{a}_i}{\tilde{\sigma}_i}\right), \quad j = 1, \dots, n; i = 1, \dots, K.$$

Отметим, что величины Φ_{ij} известны.

Из (5.7.4) и (5.7.5) вытекает, что при больших n

$$\tilde{F}(x) \approx F_n(x), \quad x \in \mathbb{R},$$

откуда с учетом (5.7.6) мы получаем приближенное соотношение

$$\tilde{F}(x_{(j)}) = \sum_{i=1}^K \tilde{p}_i \Phi_{ij} \approx \frac{j}{n}, \quad j = 1, \dots, n. \quad (5.7.7)$$

Соотношение (5.7.7) можно использовать для отыскания оценок параметров $\tilde{p}_1, \dots, \tilde{p}_K$ с помощью метода наименьших квадратов и метода наименьших модулей, рассматриваемых в последующих разделах. Реализация этих методов в рассматриваемом случае довольно проста, поскольку $\tilde{F}(x)$ зависит от параметров $\tilde{p}_1, \dots, \tilde{p}_K$ линейно, так что мы не выходим за рамки линейной модели регрессионного анализа.

5.7.2. Разделение конечных смесей вероятностных распределений с фиксированными компонентами при помощи метода наименьших квадратов

Рассматриваемую задачу удобно записать в векторно-матричном виде. Обозначим

$$\mathbf{p} = \begin{pmatrix} \tilde{p}_1 \\ \tilde{p}_2 \\ \dots \\ \tilde{p}_K \end{pmatrix}, \quad \mathbf{\Phi} = \begin{pmatrix} \Phi_{11} & \Phi_{21} & \dots & \Phi_{K1} \\ \Phi_{12} & \Phi_{22} & \dots & \Phi_{K2} \\ \dots & \dots & \dots & \dots \\ \Phi_{1n} & \Phi_{2n} & \dots & \Phi_{Kn} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} \frac{1}{n} \\ \frac{2}{n} \\ \dots \\ \frac{n-1}{n} \\ 1 \end{pmatrix}.$$

Тогда соотношение (5.7.7) можно переписать в матричной форме:

$$\mathbf{\Phi p} \approx \mathbf{u}$$

или

$$\Phi \mathbf{p} + \mathbf{d} = \mathbf{u},$$

где

$$\mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ \dots \\ d_n \end{pmatrix}$$

– вектор-столбец невязок с компонентами

$$d_j = \frac{j}{n} - \sum_{i=1}^K \tilde{p}_i \Phi_{ij}.$$

Предположим, что $n \geq K$. В этом случае ранг случайной (в силу случайности выборки $\mathbf{x} = (x_1, \dots, x_n)$) матрицы Φ с вероятностью единица равен K . Тогда из общей теории метода наименьших квадратов для линейных моделей вытекает, что решение

$$\hat{\mathbf{p}}^* = \begin{pmatrix} \hat{p}_1^* \\ \hat{p}_2^* \\ \dots \\ \hat{p}_K^* \end{pmatrix}$$

безусловной задачи наименьших квадратов

$$\arg \min_{\mathbf{p}} \mathbf{d}^\top \mathbf{d} = \arg \min_{\mathbf{p}} \sum_{j=1}^n \left[\sum_{i=1}^K \tilde{p}_i \Phi_{ij} - \frac{j}{n} \right]^2 \quad (5.7.8)$$

с вероятностью единица существует, единственно и имеет вид

$$\hat{\mathbf{p}}^* = (\Phi^\top \Phi)^{-1} \Phi^\top \mathbf{u}. \quad (5.7.9)$$

При этом сумма компонент вектора $\hat{\mathbf{p}}^*$ отнюдь не обязана быть равной единице. Обозначим

$$\gamma = 1 - \sum_{i=1}^K \hat{p}_i^*. \quad (5.7.10)$$

Пусть $\mathbf{a}$ – K -мерный вектор-столбец, все компоненты которого равны единице:

$$\mathbf{a}^\top = (1, 1, \dots, 1).$$

Тогда соотношение (5.7.10) можно записать в виде

$$\gamma = 1 - \mathbf{a}^\top \hat{\mathbf{p}}^*.$$

Но в рассматриваемом случае компоненты искомого вектора $\mathbf{p}$, будучи вероятностями в дискретном распределении, связаны очевидным условием

$$\tilde{p}_1 + \dots + \tilde{p}_K = 1. \quad (5.7.11)$$

В терминах вектора $\mathbf{a}$ условие (5.7.11) запишется как

$$\mathbf{a}^\top \mathbf{p} = 1. \quad (5.7.12)$$

Хорошо известно, что, если ранг матрицы Φ равен K , то решение $\hat{\mathbf{p}}$ задачи (5.7.8) при условии (5.7.12) имеет вид

$$\hat{\mathbf{p}} = \hat{\mathbf{p}}^* + (\Phi^\top \Phi)^{-1} \mathbf{a} [\mathbf{a}^\top (\Phi^\top \Phi)^{-1} \mathbf{a}]^{-1} (1 - \mathbf{a}^\top \hat{\mathbf{p}}^*) \quad (5.7.13)$$

(см., например, с. 85–86 в (Себер, 1980)), где вектор $\hat{\mathbf{p}}^*$ определен в (5.7.9). Несложно видеть, что выражение $\mathbf{a}^\top (\Phi^\top \Phi)^{-1} \mathbf{a}$ равно сумме всех элементов матрицы $(\Phi^\top \Phi)^{-1}$. Обозначим эту сумму S . Тогда с учетом (5.7.10) соотношение (5.7.13) примет вид

$$\hat{\mathbf{p}} = \hat{\mathbf{p}}^* + \frac{\gamma}{S} \cdot (\Phi^\top \Phi)^{-1} \mathbf{a}. \quad (5.7.14)$$

Несложно видеть, что i -ая компонента s_i вектор-столбца $(\Phi^\top \Phi)^{-1} \mathbf{a}$ равна сумме всех элементов i -ой строки матрицы $(\Phi^\top \Phi)^{-1}$. Таким образом, из (5.7.14) мы окончательно получаем, что

$$\hat{p}_i = \hat{p}_i^* + \frac{\gamma s_i}{S}, \quad i = 1, \dots, K. \quad (5.7.15)$$

Условие $n \geq K$, упрощающее вычисления, определяет выбор числа $k_a + 1$ возможных значений параметров сдвига и числа k_σ возможных значений параметров масштаба компонент исходной смеси (5.7.1). А именно, чтобы обеспечить существование единственного решения $\hat{\mathbf{p}}$ условной задачи наименьших квадратов, параметры сетки k_a и k_σ должны быть связаны с объемом выборки n соотношением

$$(k_a + 1)k_\sigma \leq n.$$

Случай $n < K$ нуждается в более тщательном анализе, поскольку в таком случае оценки наименьших квадратов определены неоднозначно.

Если оценки $\hat{p}_i$ удовлетворяют неравенствам

$$0 \leq \hat{p}_i \leq 1 \quad (5.7.16)$$

для всех $i = 1, \dots, k$, то задача решена. Однако может случиться так, что некоторые из последних неравенств не реализуются на полученном векторе. В таком случае можно идти несколькими путями.

Во-первых, в работе (Waterman, 1974) показано, что задачу наименьших квадратов с ограничениями типа неравенств можно свести к задаче последовательного решения нескольких задач наименьших квадратов без ограничений. Вытекающий из этого результата алгоритм решения рассматриваемой задачи таков.

Если для некоторых $i = 1, \dots, k$ соотношения (5.7.16) не выполнены, то минимум функции

$$SS(\mathbf{p}) = \sum_{j=1}^n \left[\sum_{i=1}^K \tilde{p}_i \Phi_{ij} - \frac{j}{n} \right]^2$$

на множестве

$$\mathcal{P} = [0, 1]^K \cap \{\mathbf{p} : \tilde{p}_1 + \dots + \tilde{p}_K = 1\}$$

достигается в одной из граничных точек, в которых хотя бы для одного i выполнено равенство $p_i = 0$ (напомним, что $\hat{p}_1 + \dots + \hat{p}_K = 1$ по построению). Пусть I – некоторое подмножество множества $\{1, 2, \dots, K\}$. Как известно, всего таких подмножеств 2^K . Для каждого такого подмножества I мы находим $\hat{\mathbf{p}}_I$ – точку минимума функции

$$SS_I(\mathbf{p}_I) = \sum_{j=1}^n \left[\sum_{i \in I} \tilde{p}_i \Phi_{ij} - \frac{j}{n} \right]^2$$

при условии $\mathbf{p}_I^\top \mathbf{p}_I = 1$ по формулам, аналогичным (5.7.14) и (5.7.15) (здесь $\mathbf{p}_I$ – это вектор $\mathbf{p}$, компоненты которого с номерами, не попавшими в множество I , равны нулю). Если при этом $\hat{\mathbf{p}}_I \in \mathcal{P}$, то вычисляется значение $SS_I(\hat{\mathbf{p}}_I)$. Перебрав все возможные подмножества I , можно найти решение исходной задачи – точку

$$\hat{\mathbf{p}} = \arg \min_I SS_I(\hat{\mathbf{p}}_I).$$

Однако при таком подходе требуется решить 2^K задач наименьших квадратов с линейным ограничением, что при больших K занимает чрезвычайно много времени.

Во-вторых, в работе (Judge and Takayama, 1966) показано, что задачу наименьших квадратов с ограничениями типа неравенств можно свести к задаче квадратичного программирования, и предложена итерационная процедура – модифицированный симплекс-метод – для ее решения. Эта процедура реализована в виде встроенного средства Optimization Toolbox системы MATLAB и использована, в частности, в работе (Королев, Ломской, Пресняков и Рэй, 2005) для решения исходной задачи наименьших квадратов с ограничениями типа неравенств и равенства. Однако при этом объем вычислений трудно оценить, а статистические свойства оценок, получаемых при таком подходе, чрезвычайно трудно исследовать.

Наконец, исходную задачу наименьших квадратов с ограничениями типа неравенств и равенства можно заменить приближенной, накиннув конечную сетку и на множество возможных значений вектора $(p_1, \dots, p_k)$, сведя таким образом решение задачи к простому перебору.

5.7.3. Разделение конечных смесей вероятностных распределений с фиксированными компонентами при помощи метода наименьших модулей

Решение в смысле минимума sup-нормы

Соотношение (5.7.7) допускает различные конкретизации, связанные с различными нормами невязки. Решение задачи по методу наименьших квадратов, рассмотренное в предыдущем разделе, связано с минимизацией обычной евклидовой нормы невязки.

В данном разделе мы рассмотрим решение, связанное с минимизацией sup-нормы вектора невязок

$$\mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ \dots \\ d_n \end{pmatrix},$$

где, как и ранее,

$$d_j = \frac{j}{n} - \sum_{i=1}^K \tilde{p}_i \Phi_{ij}.$$

А именно, мы будем искать решение задачи

$$\mathbf{p}^* = \arg \min_{\mathbf{p}} \max_{1 \leq j \leq n} |d_j|. \quad (5.7.17)$$

Как известно, задача (5.7.17) может быть сведена (см., например, с. 91 в (Васильев и Иваницкий, 2003) и с. 137 в (Ашманов и Тимохов, 1991)) к задаче линейного программирования вида

$$f(\mathbf{p}, \theta) = \theta \longrightarrow \min_{(\mathbf{p}, \theta) \in \mathcal{Q}'}, \quad (5.7.18)$$

$$\mathcal{Q}' = \left\{ (\mathbf{p}, \theta) \in \mathbb{R}_+^{K+1} : \tilde{p}_1 + \dots + \tilde{p}_K \geq 1; \tilde{p}_1 + \dots + \tilde{p}_K \leq 1; \right. \\ \left. \theta \geq \sum_{i=1}^K \tilde{p}_i \Phi_{ij} - \frac{j}{n}; \theta \geq - \sum_{i=1}^K \tilde{p}_i \Phi_{ij} + \frac{j}{n}, j = 1, \dots, n \right\}. \quad (5.7.19)$$

Последняя задача, как известно, решается с помощью стандартного симплекс-метода, заключающегося в направленном переборе угловых точек множества $\mathcal{Q}'$ (см., например, главу 1 в (Васильев и Иваницкий, 2003)).

Решение в смысле минимума L_1 -нормы

В данном разделе мы рассмотрим решение, связанное с минимизацией L_1 -нормы вектора невязок $\mathbf{d}$. А именно, мы будем искать решение задачи наименьших модулей

$$\mathbf{p}^* = \arg \min_{\mathbf{p}} \sum_{j=1}^n |d_j|. \quad (5.7.20)$$

Обозначим $\vec{\theta} = (\theta_1, \dots, \theta_n)^\top$. Как известно, задача (5.7.20) может быть сведена (см., например, с. 91 в (Васильев и Иваницкий, 2003) и с. 137 в (Ашманов и Тимохов, 1991)) к задаче линейного программирования вида

$$f(\mathbf{p}, \theta) = \sum_{j=1}^n \theta_j \longrightarrow \min_{(\mathbf{p}, \vec{\theta}) \in \mathcal{Q}''}, \quad (5.7.21)$$

$$\mathcal{Q}'' = \left\{ (\mathbf{p}, \vec{\theta}) \in \mathbb{R}_+^{K+n} : \tilde{p}_1 + \dots + \tilde{p}_K \geq 1; \tilde{p}_1 + \dots + \tilde{p}_K \leq 1; \right. \\ \left. \theta_j \geq \sum_{i=1}^K \tilde{p}_i \Phi_{ij} - \frac{j}{n}; \theta_j \geq - \sum_{i=1}^K \tilde{p}_i \Phi_{ij} + \frac{j}{n}, j = 1, \dots, n \right\}. \quad (5.7.22)$$

Последняя задача также решается с помощью стандартного симплекс-метода, заключающегося в направленном переборе угловых точек множества $\mathcal{Q}''$ (см., например, главу 1 в (Васильев и Иваницкий, 2003)).

Вычислительные реализации задач линейного программирования (5.7.18), (5.7.19) и (5.7.21), (5.7.22) в популярных системах МATHCAD и MATLAB обладают намного более высоким быстродействием, нежели вычислительные реализации ЕМ-алгоритма.

С содержательной точки зрения, рассмотренные в данном разделе задачи построения оценок вектора весов $\mathbf{p}$ вполне равноправны с задачей, рассмотренной в предыдущем разделе. Более того, как известно, оценки наименьших модулей более устойчивы по отношению к наличию резко выделяющихся наблюдений, нежели оценки наименьших квадратов. Поэтому оценки минимума $\sup$ -нормы и минимума L_1 -нормы в определенном смысле предпочтительнее оценок наименьших квадратов.

Конечно, рассмотренные методы дают лишь приближенное решение задачи разделения смесей. Однако при его реализации удастся избежать итерационных процедур типа ЕМ-алгоритма для поиска решений экстремальных задач. Решение, полученное такими методами, вполне можно использовать как начальное приближение для ЕМ-алгоритма или его модификаций с целью последующего получения более точного решения.

При этом, имея решение (5.7.15) задачи разделения смесей, начальное приближение ЕМ-алгоритма можно выбирать следующим образом. Упорядочим полученные оценки весов $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_K$ по убыванию и получим набор $\hat{p}_{i_1}, \hat{p}_{i_2}, \dots, \hat{p}_{i_K}$ такой, что $\hat{p}_{i_1} \geq \hat{p}_{i_2} \geq \dots \geq \hat{p}_{i_K}$. Выберем порог $\delta > 0$ из тех соображений, что на практике, если $p_{i_j} < \delta$, то мы будем считать, что компонента смеси, соответствующая весу p_{i_j} , отсутствует. Положим

$$k = \max\{j : p_{i_j} \geq \delta\}.$$

Тогда в качестве начального приближения для ЕМ-алгоритма возьмем параметры a_{i_j} и $\sigma_{i_j}^2$, соответствующие весам $\hat{p}_{i_1}, \hat{p}_{i_2}, \dots, \hat{p}_{i_k}$. Особо следует отметить, что при таком подходе число компонент смеси определяется автоматически по заданному порогу пренебрежимости δ .

Обратим внимание, что в рамках рассмотренного в данном разделе подхода возможно классическое детерминистическое истолкование точности в отношении приближения параметров a_i и σ_i , при котором точность характеризуется *одним* числом, задаваемым шагом дискретной сетки, накидываемой на множества значений указанных параметров. В отношении же весовых параметров $p_1, \dots, p_k$, чтобы охарактеризовать точность приближения, приходится пользоваться статистическим истолкованием, при котором одного числа недостаточно, а нужно задавать еще и надежность вывода (коэффициент доверия или уровень значимости).

5.7.4. Разделение конечных смесей вероятностных распределений с фиксированными компонентами при помощи “усеченного” ЕМ-алгоритма

Для решения задачи отыскания оценок весов $\tilde{p}_i$, $i = 1, \dots, K$, смеси $\tilde{F}(x)$ (см. (5.7.4)) в методе фиксированных компонент можно использовать “усеченный” ЕМ-алгоритм.

В разделе 5 показано, что при применении ЕМ-алгоритма к решению задачи разделения смесей вероятностных распределений (5.1) оценки весов компонент можно искать независимо от оценок параметров самих компонент (см. представление (5.6)). Более того, при известных значениях параметров компонент (или их оценок) оценки весов имеют довольно простой вид (5.8). Это позволяет указать несложную итерационную процедуру оценивания весов $\tilde{p}_i$ смеси $\tilde{F}(x)$. Эта процедура по сути представляет собой “усеченный” вариант ЕМ-алгоритма. В этом варианте за ненужностью отсутствует шаг, связанный с переназначением оценок параметров компонент смеси, поскольку в методе фиксированных компонент значения параметров компонент смеси $\tilde{F}(x)$ известны.

Пусть, как и ранее, $\mathbf{x} = (x_1, \dots, x_n)$ – выборка, $(\tilde{a}_i, \tilde{\sigma}_i)$ – узлы сетки, накидываемой на множество значений параметров компонент, $i = 1, \dots, K$. Для удобства обозначим

$$x_{ji} = \frac{x_j - \tilde{a}_i}{\tilde{\sigma}_i}, \quad \phi_{ij} = \frac{1}{\tilde{\sigma}_i} \phi(x_{ji}), \quad j = 1, \dots, n; \quad i = 1, \dots, K,$$

где, как и ранее, $\phi(x)$ – стандартная нормальная плотность. Тогда обсуждаемый “усеченный” вариант ЕМ-алгоритма для оценивания весов $\tilde{p}_i$ смеси $\tilde{F}(x)$ определяется с помощью следующего рекуррентного соотношения, вытекающего из (5.8):

$$\tilde{p}_i^{(m+1)} = \frac{\tilde{p}_i^{(m)}}{n} \sum_{j=1}^n \frac{\phi_{ij}}{\sum_{r=1}^K \tilde{p}_r^{(m)} \phi_{rj}} \quad j = 1, \dots, n; \quad i = 1, \dots, K. \quad (5.7.23)$$

Здесь $m \geq 1$ – номер итерации.

Поскольку функция правдоподобия параметров $\mathbf{p} = (\tilde{p}_1, \dots, \tilde{p}_K)^\top$ регулярна (в частности, обладает свойствами гладкости), то последовательность $\{\tilde{p}_i^{(m)}\}_{m \geq 1}$, определенная соотношением (5.7.23), сходится к соответствующим оценкам максимального правдоподобия весов в смеси $\tilde{F}(x)$. Однако, как уже отмечалось выше, функция правдоподобия в исходной задаче разделения смесей $F(x)$ нормальных распределений (см. (5.7.4)) обладает весьма нерегулярным рельефом. К тому же узлы $(\tilde{a}_i, \tilde{\sigma}_i)$ сетки могут

не совпасть с “истинными” значениями (a_j, σ_j) параметров компонент исходной смеси. Поэтому сходимость рекурсии (5.7.23), максимизирующей правдоподобие оценок весов в смеси $\tilde{F}(x)$, к оценкам максимального правдоподобия параметров p_i в исходной смеси $F(x)$, к сожалению, отнюдь не гарантирована.

5.7.5. Разделение конечных смесей вероятностных распределений с фиксированными компонентами при помощи байесовской классификации

Если объем выборки n намного превосходит число K компонент смеси $\tilde{F}(x)$, то для оценивания весов $\tilde{p}_i$, $i = 1, \dots, K$, можно использовать совсем простую вычислительную процедуру. Из общей теории байесовских правил классификации (см., например, (Андерсон, 1963)) известно, что оптимальное решение задачи классификации наблюдений $\mathbf{x} = (x_1, \dots, x_n)$ на K классов (совокупностей), определяемых распределениями $\Phi((x - a_i)/\sigma_i)$, $i = 1, \dots, K$, которое минимизирует максимальную вероятность ошибочной классификации, имеет следующий вид.

Обозначим

$$k(j) = \arg \max_{1 \leq i \leq K} \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ -\frac{(x_j - a_i)^2}{2\sigma_i^2} \right\}.$$

В соответствии с описываемым правилом классификации наблюдение x_j приписывается к классу с номером $k(j)$, $j = 1, \dots, n$.

Пусть

$$\nu_i = \sum_{j: k(j)=i} 1$$

– количество наблюдений, отнесенных к классу с номером i , $i = 1, \dots, K$. Тогда в качестве оценки веса $\tilde{p}_i$ следует взять величину

$$\tilde{p}_i^* = \frac{\nu_i}{n}, \quad i = 1, \dots, K.$$

5.8. Выбор модели (определение типа и числа компонент смеси)

Для конкретных прикладных задач, математически формализованных с помощью смесей вероятностных (в частности, гауссовых) распределений,

огромную важность имеет интерпретация полученных результатов, например, возможность отождествления полученных компонент смеси с теми или иными реальными объектами и/или процессами. Очевидно, что такая интерпретация может быть разумной только лишь, если рассматриваемая формальная модель адекватна исследуемому процессу или явлению.

Включая в модель все больше и больше параметров, можно добиться все большего и большего ее согласия с анализируемыми данными, то есть увеличить ее адекватность. Поэтому, казалось бы, при использовании математических моделей типа конечных смесей нормальных законов следует включать в смеси как можно больше компонент, и единственным обстоятельством, определяющим выбор модели, является лишь мощность используемой вычислительной системы (производительность, память и другие ресурсы). Однако на самом деле ситуация намного сложнее.

К примеру, когда исследуются компоненты волатильности хаотических стохастических процессов, в качестве моделей распределений их приращений можно рассматривать два типа смесей. Первый тип – чисто масштабные смеси, в которых смешивание производится только по параметру масштаба, а параметры сдвига равны нулю (например, если исследователь заведомо заинтересован лишь в чисто стохастической интерпретации волатильности). И второй – общие сдвиг-масштабные смеси (которые позволяют выделить также и динамическую составляющую волатильности). Оказывается, что на практике два этих типа смесей принципиально по-разному ведут себя по отношению к увеличению числа их компонент. Для чисто масштабных смесей наблюдается *эффект насыщения*, когда заметное повышение согласия модели с данными наблюдается лишь при увеличении числа компонент смеси до определенного порога (как правило, довольно небольшого и равного 4 или 5), а при дальнейшем увеличении числа компонент (повышении “проработанности” модели) изменения согласия незначительны и практически несущественны (см. раздел 6.4.5). Для моделей типа общих сдвиг-масштабных смесей наблюдается принципиально иной *эффект перетекания волатильности* из диффузионной (стохастической) составляющей в динамическую: при малом (2–3) числе компонент диффузионная составляющая превалирует над динамической, а при увеличении числа компонент соотношение этих составляющих принципиально изменяется в пользу динамической (см. раздел 6.5). С формальной точки зрения эффект перетекания волатильности сводится к тому, что модели типа смесей превращаются во все более и более точные непараметрические оценки плотности, аналогичные ядерным оценкам. Интерпретация таких оценок, в силу их почти непараметрической природы, весьма проблематична.

Таким образом, хотелось бы заранее решить, модель какого типа использовать в той или иной конкретной ситуации. В частности, хотелось бы заранее ответить на вопрос, значима ли динамическая составляющая волатильности, или ее присутствие обусловлено лишь чисто стохастическими флуктуациями результатов измерений вокруг их среднего значения. Если отличие оценок параметров сдвига компонент от нуля на самом деле вызвано не систематическими, но чисто стохастическими причинами, то, как следует из сказанного выше, использование формально более точной модели типа сдвиг-масштабной смеси может сильно исказить интерпретацию результатов. Аналогично, к неправильной, искаженной интерпретации результатов может привести и использование формально более точной модели с завышенным числом компонент или неточной модели с заниженным числом компонент.

Некоторые критерии выбора подходящей модели (типа смеси и количества компонент) рассмотрены в данном разделе.

5.8.1. Некоторые сведения из теории проверки сложных статистических гипотез

Здесь мы приведем некоторые основные определения и нужные нам результаты из теории проверки сложных статистических гипотез.

В математической статистике базовым понятием, формализующим имеющиеся в распоряжении исследователя данные, является понятие *статистической структуры*, под которой подразумевается тройка объектов $(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$, где $\mathbb{R}^n$ – это n -мерное евклидово пространство, которому принадлежат возможные значения $\mathbf{x} = (x_1, \dots, x_n)$ наблюдаемого случайного вектора $\mathbf{X} = (X_1, \dots, X_n)$ (выборки), $\mathcal{B}_n$ – это борелевская σ -алгебра подмножеств n -мерного евклидова пространства, $\mathcal{P}$ – семейство вероятностных мер, содержащее ту, которая задает неизвестное “истинное” распределение наблюдаемого случайного вектора $\mathbf{X}$. Обычно семейство $\mathcal{P}$ задается в виде

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}, \quad (5.8.1)$$

где Θ – некоторое множество, имеющее, вообще говоря, более простую структуру, нежели $\mathcal{P}$.

Если говорить нестрого, то под *статистической гипотезой* H обычно подразумевают некое утверждение, описывающее неизвестное распределение наблюдаемого случайного вектора $\mathbf{X}$, то есть сужающее множество $\mathcal{P}$. С формальной же точки зрения, статистическая гипотеза является под-

множеством множества $\mathcal{P}$. Пусть $\mathcal{P}_0$ – некоторое подмножество множества $\mathcal{P}$ (возможно, совпадающее с $\mathcal{P}$). Объединяя неформальное объяснение с математически корректным формализмом, мы будем обозначать статистическую гипотезу о том, что распределение наблюдаемого случайного вектора $\mathbf{X}$ определяется какой-либо мерой из подмножества $\mathcal{P}_0$, в виде $H: \mathbf{P} \in \mathcal{P}_0$. Если семейство $\mathcal{P}$ задано в параметрическом виде (5.8.1), то под статистической гипотезой мы будем подразумевать подмножество Θ_0 множества Θ . Для такой параметрической гипотезы мы будем использовать обозначение $H: \theta \in \Theta_0$. В дальнейшем мы будем иметь дело только с параметрическими гипотезами. Если гипотеза $H: \theta \in \Theta_0$ описывает распределение наблюдаемого случайного вектора $\mathbf{X}$ однозначно, то есть множество Θ_0 содержит единственный элемент, то гипотеза $H: \theta \in \Theta_0$ называется *простой*. Гипотеза, не являющаяся простой, называется *сложной*.

Предположим, что существует σ -конечная мера μ , относительно которой все меры из семейства (5.8.1) являются абсолютно непрерывными. Соответствующие плотности случайного вектора $\mathbf{X}$ обозначим $f_{\theta}^{\mathbf{X}}(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$. Предположим, что компоненты $X_1, \dots, X_n$ случайного вектора $\mathbf{X}$ стохастически независимы и имеют одно и то же распределение. Тогда функция правдоподобия параметра θ примет вид

$$L(\theta; \mathbf{x}) \equiv f_{\theta}^{\mathbf{X}}(\mathbf{x}) = \prod_{j=1}^n f_{\theta}(x_j), \quad \theta \in \Theta,$$

где $f_{\theta}(x)$, $x \in \mathbb{R}$, – плотность распределения случайной величины X_1 (относительно меры μ).

В дальнейшем мы будем предполагать, что $\Theta \subseteq \mathbb{R}^r$ при некотором $r \geq 1$, то есть мы считаем, что $\theta = (\theta_1, \dots, \theta_r)$, причем $\theta_i \in \mathbb{R}$, $i = 1, \dots, r$. Нас будут интересовать сложные (“цилиндрические”) статистические гипотезы следующего вида. Пусть s – целое число, $1 \leq s < r$. Зафиксируем числа $t_1, \dots, t_{r-s}$. Назовем статистическую гипотезу

$$H_s: \theta \in \Theta_s \equiv \{\theta = (\theta_1, \dots, \theta_r) \in \Theta : \theta_i = t_{i-s}, i = \overline{s+1, r}\}$$

цилиндрической сложной гипотезой размерности s , порожденной числами $t_1, \dots, t_{r-s}$. Несложно видеть, что $\Theta_s \subset \Theta$.

Обозначим

$$\hat{\theta} = \arg \max_{\theta \in \Theta} L(\theta; \mathbf{x}), \quad (5.8.2)$$

$$\hat{\theta}^s = \arg \max_{\theta \in \Theta_s} L(\theta; \mathbf{x}). \quad (5.8.3)$$

Для проверки сложной гипотезы $H_s: \theta \in \Theta_s$ используется критерий отношения правдоподобия, основанный на статистике

$$\lambda_s(\mathbf{X}) = \frac{L(\hat{\theta}^s; \mathbf{X})}{L(\hat{\theta}; \mathbf{X})}. \quad (5.8.4)$$

Очевидно, что всегда $0 \leq \lambda_s(\mathbf{x}) \leq 1$. Чем ближе значение статистики $\lambda_s(\mathbf{x})$ к нулю, тем менее правдоподобна гипотеза $H_s: \theta \in \Theta_s$ по сравнению с гипотезой $H: \theta \in \Theta$.

Чтобы придать последнему утверждению строгую форму и получить возможность сделать более или менее обоснованный вывод о том, есть ли *систематические* причины отвергнуть гипотезу $H_s: \theta \in \Theta_s$ или согласиться с ней, можно использовать следующий фундаментальный результат, доказанный С. Уилксом еще в 1938 г. (Wilks, 1938) и развитый в работе (Chernoff, 1954), см. также (Уилкс, 1967), раздел 13.8. Заметим, что стремление к нулю величины $\lambda_s(\mathbf{x})$ эквивалентно неограниченному возрастанию величины $-\log \lambda_s(\mathbf{x})$.

ТЕОРЕМА 5.8.1. Пусть компоненты вектора $\mathbf{X} = (X_1, \dots, X_n)$ – независимы и одинаково распределены, причем функция правдоподобия дважды непрерывно дифференцируема по $\theta \in \Theta$, а статистика $\lambda_s(\mathbf{X})$ определена в соответствии с (5.8.4). Тогда равномерно по $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P_\theta(-2 \log \lambda_s(\mathbf{X}) < x) = \frac{1}{2^{(r-s)/2} \Gamma\left(\frac{r-s}{2}\right)} \int_0^x z^{(r-s)/2} e^{-z/2} dz$$

для любого $\theta \in \Theta_s$, где $\Gamma(\cdot)$ – гамма-функция Эйлера. Другими словами, распределение статистики $-2 \log \lambda_s(\mathbf{X})$ при гипотезе $H_s = \Theta_s$ при неограниченно увеличивающемся объеме выборки n стремится к распределению хи-квадрат с числом степеней свободы, равным $(r-s)$.

Основанный на теореме 5.8.1 критерий проверки сложной цилиндрической гипотезы $H_s: \theta \in \Theta_s$ против сложной альтернативы $H: \theta \in \Theta$ выглядит так. Заранее фиксируется малое положительное число α – уровень значимости. Оно характеризует требования к качеству вывода (уровень значимости – это вероятность ошибочно отвергнуть верную гипотезу). На основе имеющейся выборки $\mathbf{x}$ по приведенным выше формулам (5.8.2), (5.8.3), (5.8.4) вычисляется значение статистики $\Lambda_s(\mathbf{x}) = -2 \log \lambda_s(\mathbf{x})$. Если полученное значение $\Lambda_s(\mathbf{x})$ окажется большим $\chi_{r-s}(1-\alpha)$ – квантили порядка $(1-\alpha)$ распределения хи-квадрат с числом степеней свободы, равным $(r-s)$, – то гипотеза $H_s: \theta \in \Theta_s$ отвергается. Если же оказывается,

что $\Lambda_s(\mathbf{x}) \leq \chi_{r-s}(1 - \alpha)$, то оснований отвергать гипотезу $H_s: \theta \in \Theta_s$ нет. При этом вероятность ошибочно отвергнуть гипотезу $H_s: \theta \in \Theta_s$ в том случае, когда она на самом деле верна, примерно равна α .

5.8.2. Проверка значимости динамической составляющей волатильности с помощью критерия отношения правдоподобия

Несложно убедиться, что если распределение некоторой случайной величины X имеет вид конечной сдвиг/масштабной смеси нормальных законов, то есть

$$P(X < x) = \sum_{i=1}^k p_i \Phi\left(\frac{x - a_i}{\sigma_i}\right), \quad (5.8.5)$$

где $p_i \geq 0$, $a_i \in \mathbb{R}$, $\sigma_i > 0$, $i = 1, \dots, k$, $p_1 + \dots + p_k = 1$, то дисперсия случайной величины X может быть представлена в виде суммы двух слагаемых

$$DX = \sum_{i=1}^k (a_i - \bar{a})^2 p_i + \sum_{i=1}^k p_i \sigma_i^2, \quad (5.8.6)$$

где

$$\bar{a} = \sum_{i=1}^k a_i p_i.$$

При этом первое выражение в (5.8.6) характеризует ту часть дисперсии, которая обусловлена наличием ненулевых сдвигов, то есть “динамическую” составляющую дисперсии (или волатильности), тогда как второе выражение в (5.8.6) характеризует “чисто диффузионную” составляющую дисперсии.

Как уже говорилось, при анализе волатильности реальных хаотических процессов модели типа чисто масштабных смесей с $a_1 = \dots = a_k = 0$ и модели общего типа, в которых хотя бы некоторые a_i отличны от нуля, обладают разными свойствами. Более того, интерпретация моделей этих двух разных типов может быть принципиально разной. Поэтому чрезвычайную важность приобретает вопрос о возможности проверки гипотезы о том, что наблюдаемая волатильность имеет чисто стохастическую, диффузионную природу, то есть гипотезы $H_0: a_1 = \dots = a_k = 0$. Если эта гипотеза отвергается, то возникает принципиально иное объяснение наблюдаемой изменчивости исследуемого процесса. Приведенная в предыдущем

разделе теорема 5.8.1 позволяет построить вполне разумный критерий для проверки упомянутой гипотезы.

Наряду с моделью (5.8.5) рассмотрим модель

$$P(X < x) = \sum_{i=1}^k p_i \Phi\left(\frac{x}{\sigma_i}\right). \quad (5.8.7)$$

Пусть в нашем распоряжении имеется выборка $\mathbf{x} = (x_1, \dots, x_n)$ независимых реализаций случайной величины X , в отношении распределения которой имеются два предположения (5.8.5) и (5.8.7). В соответствии с моделями (5.8.5) и (5.8.7) рассмотрим две функции правдоподобия

$$L_0(p_1, \dots, p_k, \sigma_1, \dots, \sigma_k; \mathbf{x}) = \prod_{j=1}^n \left[\sum_{i=1}^k \frac{p_i}{\sigma_i} \phi\left(\frac{x_j}{\sigma_i}\right) \right]$$

и

$$L(p_1, \dots, p_k, a_1, \dots, a_k, \sigma_1, \dots, \sigma_k; \mathbf{x}) = \prod_{j=1}^n \left[\sum_{i=1}^k \frac{p_i}{\sigma_i} \phi\left(\frac{x_j - a_i}{\sigma_i}\right) \right],$$

где, как обычно, $\phi(x) = (2\pi)^{-1/2} e^{-x^2/2}$ – стандартная нормальная плотность. Пусть

$$\begin{aligned} (p_1^\circ, \dots, p_k^\circ, \sigma_1^\circ, \dots, \sigma_k^\circ) &= \arg \max_{\substack{p_1, \dots, p_k \\ \sigma_1, \dots, \sigma_k}} L_0(p_1, \dots, p_k, \sigma_1, \dots, \sigma_k; \mathbf{x}), \\ (\hat{p}_1, \dots, \hat{p}_k, \hat{a}_1, \dots, \hat{a}_k, \hat{\sigma}_1, \dots, \hat{\sigma}_k) &= \\ &= \arg \max_{\substack{p_1, \dots, p_k \\ a_1, \dots, a_k \\ \sigma_1, \dots, \sigma_k}} L(p_1, \dots, p_k, a_1, \dots, a_k, \sigma_1, \dots, \sigma_k; \mathbf{x}). \end{aligned}$$

Положим

$$\Lambda(\mathbf{x}) = -2 \log \left(\frac{L_0(p_1^\circ, \dots, p_k^\circ, \sigma_1^\circ, \dots, \sigma_k^\circ; \mathbf{x})}{L(\hat{p}_1, \dots, \hat{p}_k, \hat{a}_1, \dots, \hat{a}_k, \hat{\sigma}_1, \dots, \hat{\sigma}_k; \mathbf{x})} \right).$$

В соответствии с теоремой 5.8.1, если справедлива гипотеза H_0 : $a_1 = \dots = a_k = 0$, то при $n \rightarrow \infty$ распределение случайной величины $\Lambda(\mathbf{X})$ стремится к распределению хи-квадрат с k степенями свободы. Поэтому критерий проверки гипотезы H_0 : $a_1 = \dots = a_k = 0$ устроен так. Пусть α – малое положительное число (уровень значимости), $\chi_k(1 - \alpha)$ – $(1 - \alpha)$ -квантиль распределения хи-квадрат с k степенями свободы. Если

$$\Lambda(\mathbf{x}) > \chi_k(1 - \alpha),$$

то гипотеза $H_0: a_1 = \dots = a_k = 0$ отвергается, то есть отличие динамической компоненты волатильности от нуля статистически значимо. Если же

$$\Lambda(\mathbf{x}) \leq \chi_k(1 - \alpha),$$

то систематических (“системных”) оснований отвергнуть гипотезу H_0 в пользу предположения о том, что $a_i \neq 0$ для некоторых $i \in \{1, 2, \dots, k\}$, нет. В таком случае при некоторых дополнительных условиях, смысл которых в том, что оценкам максимального правдоподобия “запрещается” близко приближаться к границе параметрического множества (см. главу 6), при больших n вероятность ошибочно отвергнуть гипотезу о том, что динамическая компонента волатильности равна нулю, должна быть примерно равной α .

5.8.3. “Последовательный” критерий отношения правдоподобия для определения числа компонент смеси

Во многих реальных прикладных задачах, математическая формализация которых сводится к задаче статистического разделения смесей, число компонент смеси заранее не известно. Существует несколько более или менее обоснованных методов экспериментального отыскания числа компонент. К сожалению, во всех методах экспериментального определения числа компонент смеси в той или иной степени присутствует элемент произвола. Некоторые из этих методов будут рассмотрены в данном и следующих разделах.

Предположим, что в модели (5.3.15) $\psi_1(x; t) \equiv \dots \equiv \psi_k(x; t)$, то есть все компоненты смеси принадлежат к одному и тому же аналитическому типу и различаются лишь значением параметра t , так что модель (5.3.15) трансформируется к виду

$$f_{\theta}^X(x) = \sum_{i=1}^k p_i \psi(x; t_i). \quad (5.8.8)$$

Пусть, как и ранее, $\mathbf{x} = (x_1, \dots, x_n)$ – исходная выборка, являющаяся наблюдаемым значением случайного вектора $\mathbf{X} = (X_1, \dots, X_n)$, компоненты которого являются независимыми копиями случайной величины, плотность распределения которой имеет вид (5.8.8). Соответствующая этой мо-

дели функция правдоподобия имеет вид

$$L_k(\theta; \mathbf{x}) = \prod_{j=1}^n \left[\sum_{i=1}^k p_i \psi(x_j; t_i) \right], \quad (5.8.9)$$

где $\theta = (p_1, \dots, p_k, t_1, \dots, t_k)$.

Чтобы подчеркнуть зависимость размерности параметра $\theta = (p_1, \dots, p_k, t_1, \dots, t_k)$ модели (5.8.8) от числа компонент k , мы будем этот параметр и его оценку максимального правдоподобия обозначать $\theta(k)$ и $\hat{\theta}(k)$, соответственно. Предположим, что значения этих оценок для $k = 1, 2, \dots$ уже найдены. Другими словами, известны значения

$$\hat{\theta}(k) = \arg \max_{\theta(k)} L_k(\theta(k); \mathbf{x}), \quad k = 1, 2, \dots,$$

где функция правдоподобия $L_k(\theta(k); \mathbf{x})$ определена в (5.8.9).

Для статистического определения числа компонент смеси (5.8.8) можно воспользоваться теоремой 5.8.1, согласно которой при некоторых дополнительных условиях регулярности на функцию правдоподобия $L_k(\theta(k); \mathbf{x})$ распределение статистики критерия отношения правдоподобия

$$\Lambda_{k+1}(\mathbf{X}) = -2 \log \left(\frac{L_k(\hat{\theta}(k); \mathbf{X})}{L_{k+1}(\hat{\theta}(k+1); \mathbf{X})} \right) \quad (5.8.10)$$

при $n \rightarrow \infty$ (n – объем выборки), вычисленное при справедливости гипотезы H_k о том, что “истинное” число компонент смеси равно k , стремится к распределению хи-квадрат с числом степеней свободы, равным $l + 1$, где l – размерность параметра t в модели (5.8.9). Число $l + 1$ оказывается равным разности размерностей параметров $\theta(k + 1)$ и $\theta(k)$: действительно, если число компонент смеси увеличивается на единицу, то к параметру $\theta(k)$ добавляются l параметров новой компоненты и один весовой параметр p_{k+1} .

Тогда формальная процедура статистического определения числа компонент смеси может быть сведена к последовательной проверке гипотезы H_k против альтернативы H_{k+1} с помощью критерия (5.8.10) (см., например, (Айвазян и др., 1989)). Зададим заранее уровень значимости α (понимаемый как вероятность ошибочно отвергнуть верную гипотезу H_k). Для $k = 1, 2, \dots$ последовательно вычисляем значение статистики (5.8.10) и сравниваем его с критическим значением, равным $(1 - \alpha)$ -квантили $\chi_{l+1}(1 - \alpha)$ распределения хи-квадрат с числом степеней свободы, равным

$l+1$. Если статистика (5.8.10) оказывается большей, чем указанная $(1-\alpha)$ -квантиль, то гипотеза H_k отвергается. В качестве итоговой оценки числа k компонент смеси (5.8.8) принимается такое значение k , при котором гипотеза H_k впервые оказывается не отвергнутой. Обозначим такую оценку $\hat{k}_n$.

В работе (Орлов, 1983) показано, что построенная указанным образом оценка $\hat{k}_n$ оказывается несколько завышенной. А именно, асимптотическое (при $n \rightarrow \infty$) распределение случайной величины $\hat{k}_n - k$ (относительно любой из вероятностных мер $P_{\theta(k)}$, соответствующей “истинному” значению числа компонент k) является геометрическим с параметром α :

$$\lim_{n \rightarrow \infty} P_{\theta(k)}(\hat{k}_n - k < 0) = 0,$$

$$\lim_{n \rightarrow \infty} P_{\theta(k)}(\hat{k}_n - k = r) = (1 - \alpha)\alpha^r, \quad r = 0, 1, 2, \dots$$

Несложно видеть, что при этом

$$\lim_{n \rightarrow \infty} E_{\theta(k)} \hat{k}_n = k + \frac{\alpha}{1 - \alpha},$$

то есть асимптотическое смещение оценки $\hat{k}_n$ равно $\alpha/(1 - \alpha)$.

К сожалению, модели типа конечной смеси нормальных распределений не удовлетворяют условиям регулярности (см. главу 6). Поэтому реальный уровень значимости при проверке гипотезы H_k против альтернативы H_{k+1} может существенно отличаться от α вследствие отличия асимптотического распределения статистики критерия отношения правдоподобия от распределения хи-квадрат с указанным выше числом степеней свободы. Примеры вычисления предельных законов, имеющих вид распределений взвешенных сумм независимых хи-квадрат-распределенных случайных величин, для статистик модифицированных тестов типа отношения правдоподобия и дальнейшие ссылки можно найти, например, в недавних работах (Lo, Mendell and Rubin, 2001), (Lo, 2005) и (Chen and Kalbfleisch, 2005).

5.8.4. Определение числа компонент смеси с помощью SEM-алгоритма

Для определения числа компонент смеси может быть приспособлен SEM-алгоритм. Соответствующая его модификация определяется следующим образом.

До начала работы алгоритма фиксируются два числа: K – максимально возможное число компонент и ν_0 – минимально допустимое число наблюдений в одном кластере. По поводу обоснования выбора этих чисел можно заметить следующее. Значение порога K в значительной мере определяется соображениями практической интерпретируемости полученного результата. Порог же ν_0 определяется соображениями минимально допустимой значимости компонент. Действительно, в соответствии с формулой (5.4.7), вес компоненты оказывается прямо пропорциональным численности соответствующего кластера. Поэтому если из каких-либо соображений нецелесообразно считать значимыми компоненты смеси, веса которых меньше некоторого заранее задаваемого (малого) значения, скажем, α , то ν_0 , соответственно, определяется как $\nu_0 = \alpha \cdot n$, где n – объем выборки.

В рассматриваемой версии SEM-алгоритма величина k – число компонент смеси – является параметром, который может изменяться на каждой итерации. Поэтому набор начальных приближений включает и значение $k^{(0)} = K$.

Согласно работе (Celeux and Diebolt, 1984) в рассматриваемой версии SEM-алгоритма S-этап $(m + 1)$ -й итерации дополняется следующими действиями. Если после симуляции величин $\mathbf{y}$ какие-либо из соответственно полученных кластеров оказываются малочисленными, то есть при каких-либо i выполняются неравенства

$$\nu_i^{(m+1)} \leq \nu_0, \quad (5.8.11)$$

то соответствующий кластер аннулируется. Процедура аннулирования (изъятия) малочисленных кластеров на $(m + 1)$ -й итерации заключается в том, что:

- 1°. Переопределяются апостериорные вероятности $g_{ij}^{(m)}$: вводятся обновленные с учетом изъятия малочисленных кластеров величины

$$\tilde{g}_{ij}^{(m)} = \begin{cases} 0, & \text{если выполнено условие (5.8.11),} \\ \frac{g_{ij}^{(m)}}{\sum_{i \in \mathcal{M}(m+1)} g_{ij}^{(m)}}, & \text{в противном случае,} \end{cases}$$

где $\mathcal{M}(m + 1)$ – множество номеров всех “достаточно представительных” кластеров (то есть тех кластеров, для которых условие (5.8.11) не выполнено) на $(m + 1)$ -й итерации.

- 2°. Наблюдения, попавшие в аннулированные кластеры, перераспределяются по “достаточно представительным” кластерам, для чего для каждого из таких наблюдений генерируются векторы $\vec{y}_j^{(m+1)} = (y_{1j}^{(m+1)}, y_{2j}^{(m+1)}, \dots, y_{k^{(m)}j}^{(m+1)})$ как реализации случайных векторов с полиномиальным распределением с параметрами 1 и $\tilde{g}_{1j}^{(m)}, \dots, \tilde{g}_{k^{(m)}j}^{(m)}$. “Непристроенное” наблюдение x_j приписывается к тому из “достаточно представительных” кластеров, номер которого совпадает с номером компоненты вектора $\vec{y}_j^{(m+1)}$, равной единице.
- 3°. Переопределяется число компонент смеси: $k^{(m+1)} = k^{(m)} - l_0^{(m+1)}$, где $l_0^{(m+1)}$ – число малочисленных кластеров (то есть удовлетворяющих условию (5.8.11)) на $(m+1)$ -й итерации.

Таким образом, в соответствии с описанной процедурой аннулирования малочисленных кластеров, в ходе работы данной версии SEM-алгоритма число компонент смеси может только уменьшиться.

5.8.5. Информационные критерии выбора модели (числа компонент и типа смеси)

Информационный критерий Акаике (AIC)

Этот критерий предложен Хиротугу Акаике в 1971 г., впервые описан в работе (Akaike, 1973) и исследован в работах (Akaike, 1974), (Акаике, 1983), (Bozdogan, 1987) и многих других. Хотя в работе (Akaike, 1974) автор пояснил, что предложенное им название этого критерия AIC является аббревиатурой английского термина “An Information Criterion”, в работах других авторов это название расшифровывается как “Akaike Information Criterion”.

Пусть $\mathbf{x} = (x_1, \dots, x_n)$ – исходная наблюдаемая выборка, являющаяся реализацией случайного вектора $\mathbf{X} = (X_1, \dots, X_n)$, компоненты которого независимы и одинаково распределены. Пусть $f^{\mathbf{X}}(\mathbf{x})$ – неизвестная “истинная” плотность распределения случайного вектора $\mathbf{X}$. Для удобства обозначений в дальнейшем мы будем опускать верхний индекс у плотности и вместо $f^{\mathbf{X}}(\mathbf{x})$ будем писать просто $f(\mathbf{x})$.

Пусть $f_M(\mathbf{x}; \theta)$ – модельная (гипотетическая) плотность распределения случайного вектора $\mathbf{X}$ (рассматриваемые плотности понимаются относительно некоторой σ -конечной меры μ). Предположим, что размерность параметра θ равна d , то есть $\theta = (\theta_1, \dots, \theta_d)$. Пусть $\hat{\theta} = \hat{\theta}(\mathbf{x})$ – оценка максимального правдоподобия параметра θ , построенная по выборке $\mathbf{x}$.

В качестве показателя качества модели X . Акаике предложил рассматривать расстояние Кульбака–Лейблера между “истинной” $f(\cdot)$ и наиболее правдоподобной “модельной” $f_M(\cdot; \hat{\theta})$ плотностями (см. раздел 4.2). Модель тем лучше, чем меньше указанное расстояние.

При фиксированном значении $\hat{\theta} = \hat{\theta}(x)$ это расстояние записывается в следующем виде. Пусть $\mathbf{Z}$ – “фиктивный” случайный вектор, имеющий такую же плотность распределения $f(x)$, как и вектор $\mathbf{X}$. Тогда

$$\begin{aligned} \mathcal{D}_{KL}[f(\cdot); f_M(\cdot; \hat{\theta}(x))] &= \mathbb{E} \left[\log \frac{f(\mathbf{Z})}{f_M(\mathbf{Z}; \hat{\theta}(\mathbf{X}))} \middle| \mathbf{X} = x \right] = \\ &= \mathbb{E}[\log f(\mathbf{Z}) | \mathbf{X} = x] - \mathbb{E}[\log f_M(\mathbf{Z}; \hat{\theta}(\mathbf{X})) | \mathbf{X} = x], \end{aligned} \quad (5.8.12)$$

где математическое ожидание берется по “истинному” распределению $f(x)$. Расстояние Кульбака–Лейблера по форме тесно связано с таким понятием теории информации, как энтропия. Теоретико-информационное название расстояния Кульбака–Лейблера (5.8.12) – *негэнтропия* (отрицательная энтропия) распределения $f(\cdot)$ по отношению к распределению $f_M(\cdot; \hat{\theta}(x))$. Негэнтропию также называют *информацией по Кульбаку*, см., например, (Акаике, 1983), (Кульбак, 1967). Эти обстоятельства обусловили терминологию, согласно которой обсуждаемые критерии качества модели называют *информационными*.

Обратим внимание, что первое слагаемое в правой части (5.8.12) не зависит от гипотетической модели, и потому всегда постоянно, так что качество модели, понимаемое в смысле расстояния Кульбака–Лейблера между модельным и истинным распределениями, характеризуется вторым слагаемым в правой части (5.8.12). Однако практическое вычисление второго слагаемого

$$-\mathbb{E}[\log f_M(\mathbf{Z}; \hat{\theta}(\mathbf{X})) | \mathbf{X} = x] = - \int \log f_M(\mathbf{z}; \hat{\theta}(x)) f(\mathbf{z}) \mu(d\mathbf{z}) \quad (5.8.13)$$

невозможно, поскольку “истинное” распределение $f(\mathbf{z})$ не известно.

Вместо неизвестной величины (5.8.13) можно использовать ее статистическую оценку $-\log f_M(x; \hat{\theta}(x))$. Однако при этом серьезную помеху представляет то обстоятельство, что статистика $-\log f_M(\mathbf{X}; \hat{\theta}(\mathbf{X}))$ не является несмещенной оценкой величины (5.8.13). Действительно, с одной стороны, в соответствии с теоремой 5.8.1, при некоторых условиях регулярности, обеспечивающих асимптотическую нормальность оценки максимального правдоподобия параметра θ и существование такого θ , при котором $f(x) \equiv f_M(x; \theta)$, при $n \rightarrow \infty$ справедливо соотношение

$$2 \log f_M(\mathbf{X}; \hat{\theta}(\mathbf{X})) - 2 \log f(\mathbf{X}) \sim \chi_d^2, \quad (5.8.14)$$

где запись $Y \sim \chi_d^2$ означает, что случайная величина Y имеет распределение хи-квадрат с d степенями свободы.

С другой стороны, при условиях теоремы 5.8.1 распределение оценки максимального правдоподобия, вычисленное при “истинном” значении параметра θ , является асимптотически нормальным с математическим ожиданием, равным “истинному” значению параметра, и ковариационной матрицей, выражающейся в терминах фишеровской информации. Х. Акаике заметил, что этому утверждению можно придать такой вид: при $n \rightarrow \infty$

$$E[2 \log f(\mathbf{Z}) - 2 \log f_M(\mathbf{Z}; \hat{\theta}(\mathbf{X})) | \mathbf{X}] \sim \chi_d^2, \quad (5.8.15)$$

где математическое ожидание берется в соответствии с “истинным” распределением $f(\mathbf{x})$, а $\mathbf{Z}$ – введенный выше случайный вектор, плотность распределения которого совпадает с $f(\mathbf{x})$ (см. (Akaike, 1973), (Акаике, 1983), (Cavanaugh, 1997)).

Возьмем теперь математические ожидания обеих частей в соотношениях (5.8.14) и (5.8.15). Соответственно, получим асимптотические (при $n \rightarrow \infty$) равенства

$$2E \log f_M(\mathbf{X}; \hat{\theta}(\mathbf{X})) - 2E \log f(\mathbf{X}) = d \quad (5.8.16)$$

и, с учетом равенства $EE[\log f(\mathbf{Z}) | \mathbf{X}] = E \log f(\mathbf{X})$,

$$2E \log f(\mathbf{X}) - 2EE[\log f_M(\mathbf{Z}; \hat{\theta}(\mathbf{X})) | \mathbf{X}] = d, \quad (5.8.17)$$

Сложив (5.8.16) и (5.8.17), мы окончательно получим асимптотическое (при $n \rightarrow \infty$) равенство

$$E\{-2E[\log f_M(\mathbf{Z}; \hat{\theta}(\mathbf{X})) | \mathbf{X}]\} = E\{-2 \log f_M(\mathbf{X}; \hat{\theta}(\mathbf{X})) + 2d\}.$$

Таким образом, статистика

$$AIC = -2 \log f_M(\mathbf{X}; \hat{\theta}(\mathbf{X})) + 2d \quad (5.8.18)$$

является асимптотически несмещенной оценкой удвоенной величины (5.8.13) в том смысле, что математическое ожидание разности этих двух величин, вычисленное в соответствии с “истинным” распределением $f(\mathbf{x})$, стремится к нулю при $n \rightarrow \infty$ при указанных выше условиях регулярности.

Приведенные выше рассуждения дают основания считать величину AIC (5.8.18) критерием качества модели $f_M(\mathbf{x}; \theta)$, точнее, мерой “несогласия” модели и реальных данных: чем меньше AIC , тем лучше модель.

При этом ясно, что включение в модель дополнительных параметров может только увеличить правдоподобие модели и, стало быть, уменьшить первое слагаемое в AIC . Однако при этом увеличивается второе слагаемое, играющее роль “штрафа” за использование дополнительных параметров.

Пусть теперь $\mathbf{x} = (x_1, \dots, x_n)$ – исходная наблюдаемая выборка, элементы которой являются независимыми реализациями случайной величины, плотность распределения которой имеет вид

$$f(x) = \sum_{i=1}^k \frac{p_i}{\sigma_i} \phi\left(\frac{x - a_i}{\sigma_i}\right), \quad x \in \mathbb{R}, \quad (5.8.19)$$

где $\phi(x)$ – стандартная нормальная плотность. Соответствующая этой модели функция правдоподобия имеет вид

$$L_k(\theta; \mathbf{x}) = \prod_{j=1}^n \left[\sum_{i=1}^k \frac{p_i}{\sigma_i} \phi\left(\frac{x_j - a_i}{\sigma_i}\right) \right],$$

где $\theta = (p_1, \dots, p_k, a_1, \dots, a_k, \sigma_1, \dots, \sigma_k)$. Размерность параметра θ равна $d = 3k - 1$. Поэтому с использованием критерия Акаике ответ на вопрос о том, какое значение числа k компонент смеси является “оптимальным”, имеет вид

$$k_{\text{opt}} = \arg \min_k \{-\log L_k(\hat{\theta}(k); \mathbf{x}) + 3k - 1\},$$

где $\hat{\theta}(k)$ – оценка максимального правдоподобия параметра θ , построенная в соответствии с моделью (5.8.19).

Точно так же, в соответствии с критерием Акаике вопрос о том, какая из моделей смесей – чисто масштабная или общая сдвиг-масштабная – лучше, при фиксированном k решается путем сравнения двух значений:

$$\frac{1}{2}AIC_{sl} + 1 = -\log L_k(\hat{\theta}(k); \mathbf{x}) + 3k,$$

соответствующего модели, имеющей вид общей сдвиг-масштабной смеси (5.8.19), и

$$\frac{1}{2}AIC_s + 1 = -\log L_k^\circ(\hat{\theta}^\circ(k); \mathbf{x}) + 2k,$$

соответствующего модели

$$f(x) = \sum_{i=1}^k \frac{p_i}{\sigma_i} \phi\left(\frac{x}{\sigma_i}\right), \quad x \in \mathbb{R},$$

где

$$L_k^\circ(\theta^\circ; \mathbf{x}) = \prod_{j=1}^n \left[\sum_{i=1}^k \frac{p_i}{\sigma_i} \phi\left(\frac{x_j}{\sigma_i}\right) \right], \quad \theta^\circ = (p_1, \dots, p_k, \sigma_1, \dots, \sigma_k),$$

$$\hat{\theta}^\circ(k) = \arg \max_{\theta^\circ} L_k^\circ(\theta^\circ; \mathbf{x}).$$

При этом следует выбрать ту модель, которая соответствует меньшему из чисел AIC_{sl} и AIC_s .

Известны многие модификации критерия Акаике. В частности, в работе (Bozdogan, 1987) предложен *состоятельный* информационный критерий Акаике ($CAIC$) вида

$$CAIC = -2 \log f_M(\mathbf{x}; \hat{\theta}) + d(\log n + 1), \quad (5.8.20)$$

где n – объем выборки. Состоятельность такого критерия, по аналогии с классической трактовкой этого свойства в математической статистике, заключается в том, что вычисленная в соответствии с “истинной” моделью вероятность правильного выбора модели с помощью этого критерия при упоминавшихся выше условиях регулярности, обеспечивающих асимптотическую нормальность оценки максимального правдоподобия параметра θ и существование такого θ , при котором $f(\mathbf{x}) \equiv f_M(\mathbf{x}; \theta)$, стремится к единице при $n \rightarrow \infty$. В работе (Bozdogan, 1994) также рассматривались модифицированные информационные критерии Акаике вида

$$AIC_3 = -2 \log f_M(\mathbf{x}; \hat{\theta}) + 3d$$

и

$$AIC_4 = -2 \log f_M(\mathbf{x}; \hat{\theta}) + 4d.$$

Известны и другие имеющие простой вид модификации критерия Акаике: критерий Акаике для малых выборок (Corrected Akaike Information Criterion)

$$AIC_c = AIC + \frac{2d(d+1)}{n-d+1}$$

(Sugiura, 1978), (Hurvich and Tsai, 1989), (Hurvich and Tsai, 1995), (Cavanaugh, 1997); критерий Хэннана–Куинна

$$HQ = -2 \log f_M(\mathbf{x}; \hat{\theta}) + 2d \log(\log n)$$

(Hannan and Quinn, 1979); критерии минимальной длины описания модели (Minimum Description Length)

$$MDL_2 = -2 \log f_M(\mathbf{x}; \hat{\theta}) + 2d \log n,$$

$$MDL_5 = -2 \log f_M(\mathbf{x}; \hat{\theta}) + 5d \log n$$

(Liang et al., 1992) (по поводу принципа минимальной длины описания модели см. работу (Rissanen, 1976)). В некоторых модификациях критерия Акаике используются более сложные штрафные функции, зависящие от фишеровской информации (Takeuchi, 1976), (Bozdogan, 1987), (Bozdogan, 1994) и других характеристик (Ishiguro et al., 1997).

Из всех формальных модификаций критерия Акаике наиболее используемой является байесовский информационный критерий, рассматриваемый в следующем разделе.

Байесовский информационный критерий (*BIC*)

Пусть, как и ранее, $\mathbf{x} = (x_1, \dots, x_n)$ – исходная наблюдаемая выборка, элементы которой являются независимыми реализациями случайной величины, плотность распределения которой имеет вид (5.1). Соответствующая этой модели функция правдоподобия имеет вид

$$L_k(\theta; \mathbf{x}) = \prod_{j=1}^n \left[\sum_{i=1}^k p_i \psi_i(x_j; t_i) \right], \quad (5.8.21)$$

где $\theta = (p_1, \dots, p_k, t_1, \dots, t_k)$. В соответствии с принятой нами терминологией такая функция правдоподобия является “неполной”.

В работе (Schwartz, 1978), посвященной оцениванию размерности модели (то есть числа задействованных в ней параметров) предложен следующий критерий, названный *байесовским информационным критерием*. В соответствии с этим критерием следует вычислить величину

$$BIC(k, n) = 2 \log L_k(\hat{\theta}; \mathbf{x}) - M_k \log n,$$

где $\hat{\theta}$ – оценка максимального правдоподобия параметра θ , функция $L_k(\hat{\theta}; \mathbf{x})$ определена соотношением (5.8.21), M_k – число независимых параметров модели (5.1) при фиксированном k , причем само число компонент смеси не является независимым параметром, n – объем выборки. Как видно, байесовский информационный критерий асимптотически (при $n \rightarrow \infty$) аналогичен состоятельному информационному критерию Акаике (5.8.20). Однако при обосновании величины *BIC* использована байесовская идеология (также см. (Akaike, 1978), (Akaike, 1983)). Если априори все возможные значения k одинаково правдоподобны, то величина $BIC(k, n)$ с точностью до аддитивной константы пропорциональна апостериорной вероятности того, что наблюдения $\mathbf{x}$ соответствуют модели (5.1) с данным

k . Поэтому, чем больше значение $BIC(k, n)$, тем больше правдоподобие модели (5.1) с соответствующим k .

Как мы уже отмечали, согласие модели (5.1) с данными $\mathbf{x}$ может только возрасти, если в модель будут добавлены дополнительные параметры. Следовательно, сама функция правдоподобия $L_k(\hat{\theta}; \mathbf{x})$ не может быть использована в качестве критерия адекватности модели. Поэтому, как и в информационном критерии Акаике, в байесовском информационном критерии, определяемом величиной $BIC(k, n)$, к функции правдоподобия добавляется второе слагаемое, интерпретируемое как “штраф” или наказание за использование дополнительных параметров.

Как отмечено в (Schwartz, 1978), критерий $BIC(k, n)$ можно использовать не только для выбора k , но и для сравнения разных моделей типа (5.1), определяемых разными параметризациями.

Размерность M_k параметра θ в модели (5.8.19) равна $3k - 1$. Поэтому с использованием байесовского информационного критерия решение задачи об “оптимальном” значении числа k компонент смеси имеет вид

$$k_{\text{opt}} = \arg \max_k \{2 \log L_k(\hat{\theta}(k); \mathbf{x}) - (3k - 1) \log n\},$$

где $\hat{\theta}(k)$ – оценка максимального правдоподобия параметра θ , построенная в соответствии с моделью (5.8.19), по выборке $\mathbf{x} = (x_1, \dots, x_n)$.

Точно так же, в соответствии с байесовским информационным критерием вопрос о том, какая из моделей смесей – чисто масштабная или общая сдвиг-масштабная – лучше, при фиксированном k решается путем сравнения двух значений:

$$BIC_{sl} = 2 \log L_k(\hat{\theta}(k); \mathbf{x}) - (3k - 1) \log n,$$

соответствующего модели, имеющей вид общей сдвиг-масштабной смеси (5.8.19), и

$$BIC_s = 2 \log L_k^\circ(\hat{\theta}^\circ(k); \mathbf{x}) - (2k - 1) \log n,$$

соответствующего модели (5.8.19), в которой $a_1 = \dots = a_k = 0$. При этом следует выбрать ту модель, которая соответствует большему из чисел BIC_{sl} и BIC_s .

Общепринятая шкала значений критерия $BIC(k, n)$ такова (см., например, (Kass and Raftery, 1995), (Fraley and Raftery, 1998a), (Fraley and Raftery, 1998b)).

Если $BIC(k, n) < 2$, то согласие модели и данных плохое.

Если $2 \leq BIC(k, n) < 6$, то согласие модели и данных удовлетворительное.

Если $6 \leq BIC(k, n) < 10$, то согласие модели и данных хорошее.

Если $BIC(k, n) \geq 10$, то согласие модели и данных отличное.

5.9. Примеры применения ЕМ-алгоритма в “статическом” режиме

В данном разделе будут приведены примеры применения ЕМ-алгоритма к анализу реальных данных. Это преследует двойную цель: во-первых, продемонстрировать довольно высокую эффективность ЕМ-алгоритма и, во-вторых, продемонстрировать очень высокую адекватность моделей типа конечных смесей нормальных законов, описанных в разделах 1–3.

Первые примеры применения ЕМ-алгоритма в “статическом” режиме были приведены на рис. 1.

Для дальнейшего анализа были использованы данные об эволюции индекса немецкой фондовой биржи XETRA DAX. Рассматривались дневные, часовые и получасовые данные. Вот результат анализа получасовых данных, приведенных на рис. 2.

Соответствующий ряд приращений логарифмов индекса DAX изображен на рис. 3.

Так как абсолютные значения приращений оказались очень маленькими (порядка 10^{-4} , ..., 10^{-6}), то ряд лог-разностей был умножен на 1000. После этого был применен ЕМ-алгоритм. Получены следующие результаты:

$$\begin{aligned} p_1 &= 0.277167, & \mu_1 &= 0.136550, & \sigma_1 &= 0.582462 \\ p_2 &= 0.331305, & \mu_2 &= 0.404432, & \sigma_2 &= 2.611836 \\ p_3 &= 0.033119, & \mu_3 &= -0.995228, & \sigma_3 &= 11.117036 \\ p_4 &= 0.358409, & \mu_4 &= -0.434266, & \sigma_4 &= 1.611024 \end{aligned}$$

Результаты подгонки конечной смеси нормальных законов с указанными выше параметрами к таким данным приведены на рис. 4.

На этом рисунке изображены графики смеси нормальных законов, подогнанной с помощью ЕМ-алгоритма к получасовым приращениям логарифмов индекса XETRA DAX, а также компонент смеси (с учетом весовых коэффициентов), плотности нормального распределения с параметрами,

равными выборочному среднему и выборочной дисперсии, и гистограмма приращений логарифмов исследуемого ряда.

Согласие подогнанной модельной смеси с эмпирическим распределением было проверено по критерию Колмогорова–Смирнова. При этом уровень значимости характеризует полученное P -значение 0.991601. Заметим, что, строго говоря, критерием Колмогорова–Смирнова нельзя пользоваться для проверки сложных гипотез согласия, когда параметры гипотетического распределения оцениваются по той же выборке, так как при этом изменяется предельное распределение статистики критерия, вследствие чего “истинное” P -значение – вероятность того, что статистика критерия примет такое же или еще большее значение при гипотетическом распределении (другими словами, вероятность правильности решения о справедливости проверяемой гипотезы, то есть достоверность вывода) – не совпадает с “объявленным”. Тем не менее в этом случае “объявленное” P -значение теста Колмогорова–Смирнова все же можно рассматривать как некую характеристику согласия: чем это значение ближе к единице, тем лучше согласие гипотетического распределения с эмпирическим.

Рассмотрим теперь дневные наблюдения по индексу XETRA DAX. Соответствующий временной ряд изображен на рис. 5.

Для анализа был использован ряд приращений, полученный методом взятия логарифмических разностей. Так как абсолютные значения получились очень маленькими (порядка 10^{-4} , ..., 10^{-6}), то ряд лог-разностей был умножен на 1000. После этого был применен ЕМ-алгоритм. Получены следующие результаты:

$$\begin{array}{lll} p_1 = 0.267695, & \mu_1 = -2.731578, & \sigma_1 = 10.052026 \\ p_2 = 0.307873, & \mu_2 = 5.273676, & \sigma_2 = 12.597191 \\ p_3 = 0.194618, & \mu_3 = -15.085179, & \sigma_3 = 22.003638 \\ p_4 = 0.229814, & \mu_4 = 8.260053, & \sigma_4 = 28.098410 \end{array}$$

При этом уровень значимости характеризуется P -значением 0.965153. Результаты подгонки конечной смеси нормальных законов с указанными выше параметрами к таким данным приведены на рисунке 6.

На этом рисунке изображены графики смеси нормальных законов, подогнанной с помощью ЕМ-алгоритма к дневным приращениям логарифмов индекса XETRA DAX, а также компонент смеси (с учетом весовых коэффициентов), плотности нормального распределения с параметрами, равными выборочному среднему и выборочной дисперсии, и гистограмма приращений логарифмов исследуемого ряда.

Видно, что распределение дневных приращений логарифмов индекса XETRA DAX больше похоже на нормальное, нежели распределение получасовых приращений. Так проявляется эффект “уноормализования”, теоретически описанный в разделе 2.2.

С помощью ЕМ-алгоритма также анализировались данные об эволюции фондовых индексов американских фондовых бирж Nasdaq, Standard & Poor 500. Рассматривались дневные, часовые и получасовые данные. Результаты анализа указанных финансовых инструментов оказались аналогичными приведенным выше (естественно, с другими параметрами смесей).

Другие примеры применения ЕМ-алгоритма к анализу хаотических процессов содержатся в разделе 7.2, где этот метод применяется к данным, полученным в экспериментах с плазменной турбулентностью. Забегая вперед, скажем, что и в том случае наблюдается удивительно высокое согласие модели – конечной смеси нормальных законов – с эмпирическими распределениями приращений анализируемых процессов.

6.

Скользящее (динамическое) разделение смесей вероятностных распределений. СРС-метод: описание и практическое применение

“Не есть ли жизнь только взбаламученный нуль?”

К. Э. Циолковский. *“Нирвана”*.

6.1. Общая идея метода скользящего разделения смесей

Общая идея метода скользящего (динамического) разделения смесей (СРС-метода) такова. Как уже отмечалось, есть все основания считать, что не существует универсальной смеси нормальных законов, которая может адекватно описать состояние рынка. По всей видимости, все параметры, включая k – количество компонент смеси, то есть число параметров смеси само по себе, – изменяются во времени. Следовательно, предметом исследования должно стать описание характера изменения параметров смеси во времени. При этом следует опираться на принцип “локальной эргодичности”, согласно которому усреднение вероятностных характеристик наблюдаемого процесса “по ансамблю” (то есть по пространству

элементарных исходов) можно заменить усреднением этих характеристик по времени. К такому принципу приходится вынужденно прибегать при применении практически любого метода анализа временных рядов, ведь в наличии имеется *только одна* реализация случайного процесса – временной ряд. С этой целью параметры смеси будут оцениваться по выборке, представляющей собой некоторый отрезок временного ряда, содержащий наблюдения, накопленные в течение некоторого интервала времени фиксированной длины. Затем интервал будет сдвигаться (скользить) в направлении астрономического времени, и процедура оценивания параметров смеси будет повторяться всякий раз при очередном сдвиге интервала. Таким образом, используемый метод будет называться *методом скользящего (динамического) разделения смесей*.

При этом для интерпретации получаемых результатов очень важно, что, коль скоро был принят принцип “локальной эргодичности”, то статистические характеристики изучаемого процесса (вынужденно) считаются *неизменными* внутри каждого интервала времени – отрезка временного ряда, используемого для текущего оценивания параметров смеси. Практическая адекватность этого принципа определяет длину скользящего интервала, также называемого *окном* (при этом говорят не о *длине* интервала, но о *ширине* окна). Очевидно, что, чем ширина окна меньше, тем больше оснований считать принцип “локальной эргодичности” выполненным. К сожалению, также довольно очевидно, что при уменьшении ширины окна одновременно уменьшается точность получаемых оценок параметров смеси, так как уменьшается объем выборки. Отыскание требуемого баланса – это прерогатива исследователя в зависимости от целей анализа.

Основным объектом проводимого анализа является временной ряд (реализация процесса), состоящий из приращений логарифма индекса, наблюдаемого в равноотстоящие моменты времени

$$\tilde{x}_n = \log(x_n) - \log(x_{n-1}), \quad n \in \{2, \dots, M\}. \quad (6.1.1)$$

Для полноты исследования и проверки базовой модели рассмотрим также временной ряд, состоящий из остатков исходного временного ряда после удаления из него кратковременной тенденции (тренда), оцененной с помощью обычного скользящего среднего. Такое преобразование исходных данных типично для рекомендаций по анализу экономических или финансовых данных во многих источниках. Итак, будем также рассматривать временной ряд

$$\tilde{x}_n = x_n - \bar{x}_n, \quad \text{где} \quad \bar{x}_n = \frac{1}{K} \sum_{i=1}^K x_{n-i+1}, \quad n \in \{K, \dots, M\}. \quad (6.1.2)$$

Здесь M – длина исходного временного ряда (объем выборки), K – параметр сглаживания (“порядок” скользящего среднего).

Оба процесса $\{\tilde{x}_n\}$ должны иметь практически нулевое среднее значение на интервале времени достаточно большой длины. Далее будем иметь дело лишь с временными рядами указанной структуры.

6.2. Оценивание предполагаемого диффузионного спектра с помощью СРС-метода

Поскольку заранее нет информации о том, как эволюционируют параметры рассматриваемых процессов, до поры до времени придется руководствоваться некоторыми упрощающими предположениями. Несмотря на то что уже имелась возможность (например, на рис. 1) убедиться в том, что смешиваемые распределения (компоненты смеси) не всегда имеют нулевые или почти нулевые средние, сначала предположим, что смешивающие распределения имеют нулевые математические ожидания. В обоснование этого предположения можно заметить, что, как правило, в реальных процессах, описывающих более или менее стабильное развитие рассматриваемой системы, компоненты с ненулевыми средними играют второстепенные роли (их веса относительно невелики). Другими словами, сейчас внимание будет сосредоточено на изучении предполагаемых диффузионных спектров рассматриваемых процессов и будем оценивать лишь диффузионные компоненты волатильности в предположении отсутствия (равенства нулю) динамических компонент. Это существенно упростит визуализацию полученных результатов. Несмотря на то что такое предположение неизбежно ведет к некоторой потере общности, полученных в рамках такого ограничения результатов будет все же достаточно для отладки СРС-метода и иллюстрации его возможностей. С другой стороны, в разделе 4.3.6 уже было показано, что иногда такое упрощение может оказаться весьма полезным с точки зрения интерпретации результатов. Позднее вернемся к общей ситуации.

Сделанное предположение позволяет в некоторой степени упростить вычисления в рамках ЕМ-алгоритма и не проводить оценивание средних значений компонент. Упрощенная таким образом итерационная процедура имеет вид

$$p_i^{(r+1)} = \frac{1}{n} \sum_{j=1}^n g_{ij}^{(r)},$$

$$\sigma_i^{(r+1)} = \left[\frac{1}{\sum_{j=1}^n g_{ij}^{(r)}} \sum_{j=1}^n g_{ij}^{(r)} X_j^2 \right]^{1/2}, \quad i \in \{1, \dots, k\},$$

где

$$g_{ij}^{(r)} = \frac{\frac{p_i^{(r)}}{\sigma_i^{(r)}} \exp \left\{ -\frac{X_j^2}{2(\sigma_i^{(r)})^2} \right\}}{\sum_{\ell=1}^k \frac{p_\ell^{(r)}}{\sigma_\ell^{(r)}} \exp \left\{ -\frac{X_j^2}{2(\sigma_\ell^{(r)})^2} \right\}}, \quad i \in \{1, \dots, k\}, \quad j \in \{1, \dots, n\}.$$

В общей ситуации, конечно же, такое упрощение довольно опасно, поскольку, поступая так, приходится считать, что вся волатильность имеет лишь диффузионную природу. Тем не менее качество получаемых результатов будет все же достаточно высоким, так как упомянутое ограничение оказывает не очень существенное влияние на реализацию СРС-метода, поскольку довольно маловероятно, что ЕМ-алгоритм в итоге припишет существенные веса тем значениям a_i , которые существенно отличны от нуля. Здесь под “существенностью” подразумевается то, что соответствующие веса p_i превышают некоторый (довольно малый) порог, например, $p_{\text{threshold}} = 0.05$.

В качестве результатов получим процессы (последовательности) диффузионных компонент волатильности и весов вида

$$\{p_n^{(1)}, \dots, p_n^{(k)}\}_{n \in \{L+1, M\}}, \quad \{\sigma_n^{(1)}, \dots, \sigma_n^{(k)}\}_{n \in \{L+1, M\}}$$

для процесса (6.1.1) и

$$\{p_n^{(1)}, \dots, p_n^{(k)}\}_{n \in \{K+L-1, M\}}, \quad \{\sigma_n^{(1)}, \dots, \sigma_n^{(k)}\}_{n \in \{K+L-1, M\}}$$

для процесса (6.1.2). Здесь L – длина скользящего интервала (ширина окна), k – максимальное число компонент смеси.

Поскольку определение ЕМ-алгоритм допускает определенную свободу в выборе начальных приближений и правил остановки, необходимо исследовать влияние введенного выше ограничения на выбор этих параметров и проанализировать следующие задачи.

- *Устойчивость ЕМ-алгоритма.* Теорема 5.2.1 гарантирует сходимость алгоритма лишь к локальному максимуму. Поэтому следует найти критерий приемлемости полученного решения.

- *Начальные точки.* Необходимо найти способ определения начальных приближений, обеспечивающих удовлетворительное функционирование ЕМ-алгоритма, и исследовать, насколько сильно этот выбор влияет на окончательный результат.
- *Число параметров.* Необходимо выяснить, как задавать число параметров, обеспечивающее приемлемую адекватность модели.
- *Способ визуализации результатов.* Изменение параметров смеси во времени необходимо представить наглядно и в легко интерпретируемом виде.
- *Скорость сходимости.* Необходимо изучить зависимость скорости сходимости алгоритма от выбора начального приближения, числа параметров (компонент смеси) и ширины окна.

6.3. СРС-метод с оцениванием значений компонент волатильности

Рассмотрим несколько модификаций СРС-метода. Можно выделить два основных класса модификаций СРС-метода. В рамках первого предусматривается возможность оценивания значений параметров масштаба (диффузионных компонент) и их весов при каждом фиксированном окне. Второй класс содержит модификации СРС-метода, в которых оцениванию подлежат только веса диффузионных компонент, а множество (достаточно богатое) возможных значений самих параметров диффузии заранее фиксировано. При этом внутри обоих классов возможны модификации метода за счет изменения правил выбора начальных приближений и остановки. В частности, будет описана реализация СРС-метода с ядерными оценками волатильности.

Начнем с описания первого из упомянутых типов СРС-метода на примере его применения к анализу временного ряда, состоящего из $M = 700$ наблюдений за эволюцией индекса DAX, делавшихся через каждые полчаса. График исходного ряда приведен на рис. 2. При получении ряда остатков, отфильтрованного с помощью скользящего среднего (6.1.2), будет использоваться скользящее среднее порядка $K = 9$, что соответствует половине рабочего дня на немецком биржевом рынке. Длина скользящего интервала (ширина окна) в СРС-методе полагается равной $L = 180$, что

соответствует двум торговым неделям. Такой выбор параметра L обусловлен тем, что, во-первых, интервал длины L содержит достаточно данных для того, чтобы ЕМ-алгоритм дал приемлемые и довольно точные результаты, и, во-вторых, окончательный результат не будет искажен теми наблюдениями, которые лежат довольно далеко в прошлом. Влияние выбора различной ширины окна на окончательный результат будет затронуто ниже.

6.3.1. Критерий остановки

В качестве критерия остановки мы будем использовать комбинацию правил (5.3.25) и (5.3.27), которая обеспечивает либо выбор алгоритмом приемлемого распределения, либо окончание вычислений в случае, когда алгоритм “завяз” в какой-либо точке по причине достижения им локального или глобального максимума функции правдоподобия или когда скорость сходимости падает ниже допустимого предела.

6.3.2. Зависимость СРС-метода от выбора начальных приближений

Что же касается выбора начальных приближений, то необходимо отметить, что это весьма серьезная проблема. Постольку-поскольку ничего не известно о том, как будет или должен выглядеть окончательный результат, начальные приближения придется выбирать с помощью датчика псевдослучайных чисел. В обоснование такого подхода можно заметить, что, так как используется несколько параметров, то даже при относительно малом числе реализаций случайно задаваемых начальных приближений можно проверить большую часть различных векторов без риска систематического выбора приближений, которые могут исказить окончательный результат. Под термином “различные векторы” будем подразумевать различные распределения компонент волатильности: как различные распределения вектора весов при фиксированных значениях диффузионных компонент волатильности, так и более общий случай, когда сами значения диффузионных компонент также меняются.

Выбрать начальные приближения весов можно было бы и детерминированно (заметим, правда, что при этом, если начальное значение каждого веса полагается равным $1/k$, где k – число компонент, то следует иметь в виду, что значения $g_{ij} = 1/k$ являются неподвижными точками описанной выше итерационной процедуры). Однако, не желая связывать себя

никакими детерминированными стереотипами и стремясь всесторонне исследовать поведение ЕМ-алгоритма, чтобы по возможности получить полную картину, сосредоточимся на использовании варьируемых начальных приближений. А при выборе начальных приближений для компонент волатильности будем руководствоваться тем, что они не должны быть слишком малыми, чтобы не вызвать деление на ноль при выполнении вычислений по ходу работы ЕМ-алгоритма. Они также не должны существенно превосходить оцененное стандартное отклонение $\bar{\sigma}$ всего процесса.

Итак, в качестве начальных приближений будут использоваться следующие псевдослучайные векторы:

$$\vec{p} = (p_1, \dots, p_k) = \frac{1}{\sum_{i=1}^k u_i} \cdot (u_1, \dots, u_k), \text{ где } u_i \sim \mathcal{U}(0.1, 1) \text{ и независимы;}$$

$$\vec{\sigma} = (\sigma_1, \dots, \sigma_k) = \bar{\sigma} \cdot (v_1, \dots, v_k), \text{ где } v_i \sim \mathcal{U}(0.25, 1.75) \text{ и независимы}$$

(здесь символом $\xi \sim \mathcal{U}(a, b)$ обозначен тот факт, что случайная величина ξ имеет равномерное распределение на отрезке $[a, b]$).

Случайный выбор начальных значений можно осуществлять как сразу для всего алгоритма (то есть одни и те же начальные значения, сгенерированные в самом начале, используются затем при каждом сдвиге скользящего интервала), так и для каждого нового положения окна (то есть при каждом сдвиге скользящего интервала начальные приближения генерируются заново). При этом можно по-разному организовывать выбор вектора весов и вектора компонент волатильностей. При тестировании СРС-метода были использованы все возможные комбинации таких возможностей. Они перечислены в следующей таблице:

	Веса	Компоненты волатильности
Метод 1	r реализаций для всего алгоритма	s реализаций для всего алгоритма
Метод 2	r реализаций для каждого окна	s реализаций для всего алгоритма
Метод 3	r реализаций для всего алгоритма	s реализаций для каждого окна
Метод 4	r реализаций для каждого окна	s реализаций для каждого окна

Следует пояснить, почему важно иметь возможность комбинировать указанные выше способы задания начальных приближений. При таком комбинировании сравнение разных способов дает информацию о “систематическом” поведении ЕМ-алгоритма, то есть информацию о том: (а) приводит ли ЕМ-алгоритм к систематически хорошим или плохим результатам при “глобальном” определении начального приближения; (б) возможно ли получить результаты, на которые “приятно” смотреть, то есть результаты, которые допускают наглядную визуализацию; (с) устойчив ли ЕМ-алгоритм по отношению к начальным приближениям, то есть насколько

сильно отличаются результаты при одном и том же положении скользящего интервала и при “соседних” положениях интервала.

При сравнении результатов целесообразно использовать нормированную логарифмическую функцию правдоподобия $\log L(\Theta)$. Чтобы получить “сравнимые” значения функции правдоподобия, будем вычислять оценки весов и компонент волатильности нормированной версии исходного процесса на каждом интервале:

$$(x_{n-L+1}^{\text{int}}, \dots, x_n^{\text{int}}) = \frac{1}{\hat{\sigma}_{n-L+1,n}}(x_{n-L+1}, \dots, x_n),$$

где $\hat{\sigma}_{n-L+1,n}$ обозначает выборочное стандартное отклонение процесса на интервале $[n-L+1, n]$. Такое преобразование иногда называют *деволатилизацией* исходного процесса. Подчеркнем, что “деволатилизованный” процесс будет использоваться лишь при вычислении функции правдоподобия на каждом интервале, оценки же весов и диффузионных компонент волатильности вычисляются по исходному процессу. Таким образом, поскольку используются интервалы фиксированной длины, то имеется возможность сравнивать правдоподобие как всех результатов для одного и того же процесса, так и результатов, получаемых для разных процессов (6.1.1) и (6.1.2).

Чтобы обсудить зависимость получаемых результатов от начальных приближений, сначала сравним между собой указанные четыре способа задания начальных приближений. На рис. 7 и 8 приведены точечные графики всех оценок, полученных ЕМ-алгоритмом с $r = 10$ и $s = 10$, соответственно, для процессов (6.1.1) и (6.1.2). Все значения нанесены на один и тот же график. При этом возможно наложение совпадающих точек. Для наглядности те значения диффузионных компонент волатильности, которым соответствуют веса, меньшие $p_{\text{threshold}} = 0.05$, не приводятся. Заметим, что для метода 1 использовано лишь 10 начальных приближений (Каждой псевдослучайной реализации значений диффузионных компонент волатильности соответствует одна псевдослучайная реализация значений весов). Для методов 2, 3 и 4 использовано $10 \cdot 504 = 5040$ псевдослучайных реализаций, чего, по-видимому, вполне достаточно, чтобы дать возможность оценить скорость сходимости алгоритма.

Из рис. 7 и 8 видно, что во всех случаях алгоритм сходил к практически одинаковым предельным значениям. Однако на некоторых графиках появляются изолированные точки, что говорит о довольно сильной зависимости алгоритма от выбора начальных приближений.

Тот факт, что, как правило, большинство вычисленных значений кон-

центрируется вокруг одних и тех же точек, оставляет мало сомнений в том, что ЕМ-алгоритм находит глобальный максимум или, по крайней мере, тот локальный максимум, который лежит близко к глобальному. Отметим, что данные (6.1.2), полученные из исходного процесса фильтрацией скользящего среднего, ЕМ-алгоритм обрабатывает не так успешно, как данные (6.1.1), полученные из исходного процесса взятием логарифмических разностей. Это особенно хорошо видно на рис. 8 для значений параметра n (имеющего смысл номера наблюдения, то есть времени), попадающих внутрь промежутка между $n = 250$ и $n = 400$, где видны “облака” точек, свидетельствующие о том, что при разных начальных приближениях ЕМ-алгоритм сходил к существенно разным предельным значениям. Это вызывает подозрения, что используемая модель, имеющая вид конечной смеси нормальных законов, демонстрирует плохое соответствие с реальными данными (6.1.2), получаемым довольно традиционным способом – взятием остатков после сглаживания исходного ряда с помощью скользящего среднего. Ниже можно будет убедиться, что эти подозрения вполне оправданны.

Чтобы систематизировать различия результатов, получаемых четырьмя версиями СРС-метода, сравним методы попарно:

Методы 1 и 2	Только небольшие различия. На некоторых интервалах значения, получаемые вторым методом, более разбросанны, в то время как значения, получаемые первым методом, более похожи на линии (когда применяется первый метод, алгоритм часто сходится к одинаковым значениям на соседних интервалах).
Методы 1 и 3	Небольшие различия. Тот же эффект, что описан выше, но с более заметными различиями.
Методы 2 и 3	Хотя два графика очень похожи, все же заметны различия: Значения, полученные вторым методом, образуют более гладкую картину, в то время как третий метод (особенно когда значения концентрируются в двух и более точках) дает более разбросанные результаты.
Методы 2 и 4	Те же различия, что между методами 2 и 3.
Методы 3 и 4	Практически нет различий.

Сопоставляя графики, приведенные на рис. 7, можно заключить, что ЕМ-алгоритм проявляет большую неустойчивость по отношению к выбору начального приближения для диффузионных компонент волатильности $\vec{\sigma}$, нежели по отношению к выбору начального приближения для весов $\vec{p}$: заметно различие между методами 2 и 3, различие между результатами методов 1 и 3 сильнее, чем различие между результатами методов 1 и 2, и, наконец, между методами 3 и 4 едва ли можно заметить какое-нибудь

различие, тогда как результаты методов 2 и 4 различаются весьма существенно (это верно лишь по отношению к оценкам диффузионных компонент волатильности $\vec{\sigma}$).

Относительно выбора числа параметров (то есть числа компонент подгоняемой смеси нормальных законов) следует заметить, что максимальное число точек концентрации вычисленных значений диффузионных компонент волатильности $\vec{\sigma}$ оказалось равным трем для любого момента времени n . Поэтому число k компонент подгоняемой смеси нормальных законов можно было бы положить равным трем без особого риска получить результат, существенно отличающийся от представленного здесь. Тем не менее нет гарантии, что выбор $k = 3$ вполне целесообразен, так как все же возможно, что в какой-либо момент времени существует больше доминирующих компонент волатильности.

Пока что рассматривалось влияние тех или иных обстоятельств на поведение СРС-метода, судя лишь по получаемым оценкам диффузионных компонент волатильности $\vec{\sigma}$ и не уделяя практически никакого внимания оценкам соответствующих весов. Единственный критерий, использовавшийся в отношении весов, заключался в том, что рассматривались лишь “существенные” компоненты волатильности, то есть те компоненты, веса которых превосходили некоторый порог.

Поведение СРС-метода по отношению к весам нельзя сравнивать с его поведением по отношению к компонентам волатильности. К примеру, если две изначально различные компоненты волатильности по окончании работы ЕМ-алгоритма оказываются лежащими в одной точке, одна из них может быть просто удалена, а веса просуммированы. На рис. 9 можно видеть, что, если принять во внимание то обстоятельство, что веса, соответствующие одной и той же компоненте волатильности, складываются (то есть несколько компонент волатильности совпадают), то между результатами методов 1 и 2 нет практически никакой разницы. Однако можно заметить, что изначальный выбор весов влияет на характер изменения оценок компонент волатильности. В случае метода 1, где начальные компоненты волатильности и соответствующие веса совпадают для каждого положения скользящего окна, явно прослеживается автономное поведение компонент волатильности в том смысле, что все компоненты волатильности с одним и тем же изначальным значением так и развиваются, образуя одну группу, в течение всей работы СРС-метода. В случае метода 2 все происходит иначе: хотя начальные приближения для компонент волатильности для всех положений скользящего окна были равны, компоненты с равными начальными значениями собираются в различные группы.

Укажем на еще один способ выбора начальных приближений. Он заключается в том, что при каждом новом сдвиге окна в качестве начального приближения для оцениваемых параметров выбирается вектор финальных оценок, полученный при предыдущем положении окна. Целесообразность такого выбора будет обсуждаться в разделе 6.3.5, где речь пойдет о том, сколь сильно результат работы СРС-метода зависит от ширины скользящего окна.

Обсудим качество полученных результатов. На рис. 10 приведены графики логарифмов функции правдоподобия $L(\Theta)$. На этом рисунке разница в качестве подгонки модели, имеющей вид конечной смеси нормальных законов, к данным, организованным разными способами, становится особенно наглядной. На всей временной оси упомянутая модель демонстрирует большее правдоподобие по отношению к процессу логарифмических разностей индекса (6.1.1), нежели по отношению к процессу остатков после отфильтровывания скользящего среднего (6.1.2). Более того, за несколькими исключениями, алгоритм сходил к одинаково или почти одинаково правдоподобным предельным значениям в случае процесса (6.1.1), тогда как при анализе процесса (6.1.2) довольно часто имела место сходимости к лучшим или худшим пределам в смысле их правдоподобия. Графики функций правдоподобия подтверждают практически все сказанное выше и дают некоторую новую информацию: в случае процесса (6.1.1) большинство начальных приближений приводило к одинаково правдоподобным результатам, за исключением немногих случаев, когда достигался локальный максимум с меньшим правдоподобием (методы 3 и 4).

В случае процесса отфильтрованных остатков (6.1.2) иногда возникают отклонения (в большую или меньшую сторону) от тех значений функции правдоподобия, к которым приводит основная часть начальных приближений. По-видимому, это объясняется тем, что для такого процесса статистическое разделение модельных смесей нормальных законов на компоненты по исходным данным вида (6.1.2) осуществляется не так хорошо, как для процесса (6.1.1), в силу того, что модель типа конечной смеси нормальных законов не адекватна для так организованных данных. Причина этой неадекватности обсуждаться не будет, так как вполне достаточно того, что СРС-метод дает существенно правдоподобные результаты на данных типа (6.1.1), как, собственно, и должно быть согласно теории, развитой в разделах 1–3.

Графики гистограмм и плотностей, приведенные на рис. 11, показывают, насколько подогнанные смеси нормальных плотностей близки к эмпирическим распределениям. Приведенные примеры соответствуют наиболь-

шим и наименьшим значениям двух функций правдоподобия. Четыре примера иллюстрируют качество модели, имеющей вид конечной смеси нормальных законов, для данных типа логарифмических приращений индекса (6.1.1) и данных типа остатков после фильтрации (6.1.2). Для $n = 126 \div 305$ – интервала максимума (по n) функции правдоподобия – гистограмма приращений логарифмов индекса и график подогнанной смеси нормальных плотностей очень похожи. Точно так же для $n = 306 \div 585$ – интервала минимума (по n) функции правдоподобия – различия между гистограммой приращений логарифмов индекса и графиком подогнанной смеси нормальных плотностей очень невелики. Для данных типа остатков после отфильтровывания скользящего среднего картина несколько иная: для интервала максимума (по n) функции правдоподобия ($n = 78 \div 257$) графики схожи, однако для интервала минимума (по n) функции правдоподобия ($n = 398 \div 577$) различия между гистограммой и подогнанной смесью очевидны.

Кстати, графики для $n = 126 \div 305$ особенно наглядно показывают, что островершинность (эксцесс) реальных распределений оказывается существенно выше эксцесса нормального распределения.

Все сказанное подтверждает вывод о том, что для наших целей нет совсем никакой необходимости в анализе процессов типа (6.1.2). Таким образом, далее будут рассматриваться только процессы типа (6.1.1), поскольку, как уже говорилось, конечные смеси нормальных законов намного лучше согласуются с исходными данными, представленными в виде (6.1.1), нежели с данными типа (6.1.2).

6.3.3. Визуализация результатов

Теперь, после довольно подробного исследования поведения ЕМ-алгоритма в рамках СРС-метода, можно сосредоточиться на обсуждении вопросов удобства представления и наглядности визуализации получаемых результатов. Самый простой способ одновременного наглядного изображения процессов диффузионных компонент волатильности и соответствующих весов заключается в использовании стандартного трехмерного “стерженькового” графика (stem plot), процедура построения которого имеется во многих продвинутых математических пакетах компьютерных программ. Пример такого графика приведен на рис. 12. Здесь точка, в которую помещено основание стерженька, соответствует значению диффузионной компоненты волатильности, а высота стерженька – соответствующему весу.

Некоторые части графика читаются хорошо, особенно те, где процессы, соответствующие различным компонентам волатильности, хорошо различимы. Однако на некоторых участках графика трудно понять, какой именно компоненте соответствует то или иное значение волатильности и соответствующий вес (главным образом, при $n = 50 \div 100$).

Другая проблема заключается в том, что процессы компонент-весов имеют скачки. В то время как процессы компонент с малыми значениями коэффициентов диффузии, как правило, непрерывны, процессы компонент с высокими значениями коэффициентов диффузии довольно часто исчезают, опять появляются и претерпевают скачки. Конечно, если интересоваться только одним значением n , то это не проблема. Однако если есть желание исследовать и интерпретировать поведение диффузионных компонент волатильности на всем интервале времени, то использование стандартного трехмерного “стерженькового” графика может вызвать затруднения или по крайней мере неудобства.

Другой способ представления результатов заключается в использовании плоских контурных (топографических) графиков (contour plots). Соответствующие примеры приведены на рис. 13.

Чтобы построить контурный график, итоговые данные следует экстраполировать, так что образ предварительно размывается (используется стандартная для графических пакетов операция “blur”). При этом линии одного цвета ограничивают участки плоскости, содержащие значения компонент волатильности с одинаковыми весами, что соответствует обычной топографической карте, на которой изотопы разных цветов характеризуют участки местности с разными высотами. В нашем случае “высота” – это величина веса. Получаемые изображения размыты и потому более не соответствуют абсолютно результатам вычислений. Тем не менее размывание образа все же оправдано, так как полученные значения компонент волатильности и соответствующих весов – это не точные их значения, а оценки, которые не обязаны в точности совпадать с реальными значениями параметров. Обоснованию использования “размытых” изображений будет уделено внимание в следующем разделе, посвященном СРС-методу с ядерными оценками компонент волатильности.

“Температурные” графики (рис. 14 и 15) выглядят похоже и получены с помощью стандартной для графических пакетов операции гауссовского размывания (Gaussian blur filter) графика, полученного нанесением на плоскость точек, соответствующих значениям диффузионных компонент, интенсивность (яркость) которых соответствует их весам. Трехмерное изображение на рис. 15 может дать очень хорошее представление о

балансе значимости тех или иных компонент волатильности и хорошо читается, поскольку одновременно существуют не так много доминирующих компонент волатильности. “Температурные” графики типа рис. 14 могут оказаться наиболее удобными как при изображении решений задач, связанных с описанием эволюции финансовых индексов, так и задач из области физики плазмы. Прослеживается прямая аналогия терминов “температурный” график и температура рынка, температура плазмы. Не будем здесь далее развивать эти лингвистические аналогии, предоставив это специалистам из соответствующих областей. Заметим, что принцип выбора цветов на “температурных” графиках (рис. 14 и 15) аналогичен принципу раскраски “физических” географических карт: уровень моря соответствует уровню нулевых весов (или “почти нулевых” пренебрежимо малых весов) и имеет синий цвет, затем, с увеличением высоты (величины веса компоненты волатильности) цвет изменяется от синего к голубому, зеленому, желтому, оранжевому и красному, которым окрашены области значений диффузионных компонент волатильности с наибольшими весами.

6.3.4. СРС-метод с ядерными оценками компонент волатильности

То, что ЕМ-алгоритм достиг *локального* максимума функции правдоподобия, можно интерпретировать как отыскание ЕМ-алгоритмом более-менее правдоподобных оценок параметров смеси. Так как локальных максимумов функции правдоподобия может быть несколько, то может быть и несколько более-менее правдоподобных решений задачи. При этом естественно возникает проблема выбора наиболее правдоподобного решения.

Самый простой и почти детерминированный алгоритм решения указанной проблемы заключается в последовательной реализации ЕМ-алгоритма с несколькими псевдослучайно заданными начальными приближениями и выборе того из полученных векторов оценок, для которого функция правдоподобия оказалась наибольшей среди всех осуществленных реализаций ЕМ-алгоритма. Сходными по замыслу и немного более сложно реализуемыми средствами решения этой задачи в рамках того же подхода являются SEM- и MSEM-алгоритмы (см., например, (Королев, 2007а)). Однако у такого подхода, несмотря на его логическую обоснованность, есть некоторые недостатки, главным из которых является то, что абсолютный максимум функции правдоподобия может-таки быть все же не достигнут осуществленными реализациями ЕМ-алгоритма. Далее, функция правдоподобия зависит от случайной выборки (наблюдений, соответ-

ствующих текущему положению окна) как от параметров. Стало быть, решение задачи отыскания параметров смеси зависит от случая. Другими словами, из-за необходимости дать *единственный* ответ, то есть привести *один* вектор параметров смеси, при “нетипичной” выборке может быть получена “нетипичная” смесь и, стало быть, “нетипичный”, то есть неадекватный портрет волатильности.

Чтобы избежать возможной потери информации, рассмотрим еще один подход к проблеме выбора наиболее правдоподобного ответа из нескольких более-менее правдоподобных. Этот подход сродни подходу, ведущему к понятию нечеткого множества. Некоторые аспекты такого подхода уже были обозначены в предыдущем разделе в связи с использованием “размытых”, то есть нечетких изображений для большей наглядности представления полученных результатов. Как следует из сказанного выше, удобство визуализации результатов – важный, но не единственный аргумент в пользу использования “размытых” изображений при представлении полученных результатов. Операция “размывания” результатов также может быть использована с целью сохранения всех имеющихся результатов, пусть и не самых правдоподобных. При этом можно использовать хорошо известный в непараметрической статистике метод построения ядерных оценок плотностей, восходящий к работам (Rosenblatt, 1956), (Parsen, 1962) и (Bartlett, 1963).

А именно, при каждом фиксированном положении окна следует несколько раз, скажем, N , указанным выше способом задать вектор начальных приближений для вектора диффузионных компонент волатильности $\vec{\sigma}_0$ и вектора соответствующих весов $\vec{p}_0$. Точнее, следует задать N независимых пар начальных приближений $(\vec{\sigma}_0, \vec{p}_0)$.

Для каждого из N указанных начальных приближений следует реализовать ЕМ-алгоритм и получить, соответственно, N наборов оцененных параметров смеси $\sigma_i^j, p_i^j, i = 1, \dots, k, j = 1, \dots, N$ (нижний индекс i соответствует номеру компоненты смеси, верхний индекс j соответствует номеру реализации ЕМ-алгоритма).

Пусть $\psi(u), u \in \mathbb{R}$, – некоторая известная и фиксированная плотность распределения, которую назовем *ядром* оценки диффузионной компоненты волатильности. Пусть $h > 0$ – некоторое число, которое назовем *параметром гладкости* ядерной оценки диффузионной компоненты волатильности. *Ядерной оценкой диффузионной компоненты волатильности*

при фиксированном положении скользящего окна назовем функцию

$$v(\sigma) = \frac{1}{Nh} \sum_{j=1}^N \sum_{i=1}^k p_i^j \psi\left(\frac{\sigma - \sigma_i^j}{h}\right), \quad \sigma > 0. \quad (6.3.1)$$

Несложно видеть, что, так как по предположению ядро $\psi(u)$ – некоторая плотность вероятностей, то и ядерная оценка диффузионной компоненты волатильности $v(\sigma)$ является плотностью вероятностей как функция аргумента σ при фиксированных параметрах h и σ_i^j , p_i^j , $i = 1, \dots, k$, $j = 1, \dots, N$. С помощью ядерной оценки диффузионной компоненты волатильности, тем самым, можно получить абсолютно непрерывную оценку смешивающего распределения в смеси нормальных законов, подогнанной к эмпирическому распределению приращений логарифмов рассматриваемого индекса. При этом, если ядро ψ гладкое (например, если $\psi(u)$ – стандартная нормальная плотность), то и ядерная оценка диффузионной компоненты волатильности $v(\sigma)$ будет гладкой. Кстати, заметим, что использование нормального ядра в ядерных оценках диффузионной компоненты волатильности аналогично использованию операции гауссовского “размывания” (Gaussian blur) в упомянутых выше способах визуализации результатов, но, в отличие от операции гауссовского “размывания”, параметры которой определяются исключительно наглядностью получаемого представления, ядерная оценка диффузионной компоненты волатильности имеет еще и некоторые теоретические обоснования в виде общей теории ядерных оценок плотностей (см., например, (Devroye and Györfi, 1985)).

По поводу выбора ядра $\psi(u)$ можно заметить следующее. При использовании ядерных оценок плотности главными проблемами являются выбор ядра и выбор параметра гладкости. Как правило, используются ядра, удовлетворяющие условиям

$$\int_{-\infty}^{\infty} \psi(u) du = 1, \quad \int_{-\infty}^{\infty} u \psi(u) du = 0, \quad \int_{-\infty}^{\infty} u^2 \psi(u) du = 1.$$

Первое из этих условий вытекает из требования, чтобы функция $\psi(u)$ была плотностью распределения, второе условие означает, что случайная величина с плотностью распределения $\psi(u)$ имеет нулевое математическое ожидание, а третье условие означает, что дисперсия этой случайной величины равна единице. Чаще всего в качестве $\psi(u)$ используются равномерная плотность $\psi(u) = \mathbb{I}_{[-\sqrt{3}, \sqrt{3}]}(u)$ (в этом случае получается непрерывная оценка для смешивающей функции распределения компонент волатильно-

сти, но ступенчатая оценка для *плотности* распределения компонент волатильности) или стандартная нормальная плотность $\psi(u) = \phi(u)$. Некоторые исследователи отмечают, что хорошие, наглядные результаты дает применение квадратичного ядра

$$\psi(u) = \begin{cases} 0 & \text{при } u < -2.5, \\ -\frac{576u^2}{390625} + \frac{144}{15625} & \text{при } -2.5 \leq u \leq 2.5, \\ 0 & \text{при } u > 2.5. \end{cases}$$

При малых значениях параметра гладкости h ядерная оценка имеет много довольно часто расположенных острых зубцов. При увеличении параметра h ядерная оценка становится все более и более гладкой. При этом в качестве окончательного значения h выбирается то, при котором вид ядерной оценки распределения компонент волатильности (смешивающего распределения в смеси нормальных законов) в наибольшей степени устраивает исследователя. Другими словами, выбор параметра сглаживания на практике – это в большей степени искусство или шаманство, нежели математика.

Введенная выше ядерная оценка диффузионной компоненты волатильности $v(\sigma)$ была определена при конкретном *фиксированном* положении скользящего окна. Обозначим правый конец скользящего окна через t . Тогда, очевидно, $v(\sigma) = v(\sigma, t)$. Если ядерную оценку диффузионной компоненты волатильности $v(\sigma, t)$ последовательно вычислять, сдвигая окно, для всех возможных значений t и построить трехмерный график полученной функции двух переменных над плоскостью (t, σ) , то в итоге получится картина, подобная приведенной на рис. 15. При этом при фиксированном t функция $v(\sigma, t)$, введенная в (6.3.1), представляет собой как бы сечение рельефа, изображенного на этом рисунке, плоскостью, перпендикулярной оси времени и параллельной оси значений компонент волатильности σ . Для наглядного изображения результатов, полученных с помощью ядерных оценок компонент волатильности, можно использовать и двумерные графики, аналогичные двумерному “температурному” изображению, приведенному на рис. 14.

По поводу вроде бы имеющего место несоответствия гладкости смешивающего распределения, получаемого с помощью указанной выше процедуры ядерного оценивания компонент волатильности, и дискретности смешивающего распределения в модельной смеси можно заметить следующее. Конечность числа компонент смеси используется (i) для удобства

интерпретации получаемых результатов и (ii) для того, чтобы можно было использовать довольно эффективные вычислительные процедуры (ЕМ-алгоритм). Тем самым, конечные смеси возникают как удобные *аппроксимации* для “истинных” смесей нормальных законов, которые к тому же являются лишь *асимптотическими аппроксимациями* для истинных распределений приращений логарифмов индекса. Таким образом, использование описанной выше процедуры ядерного оценивания компонент волатильности в некотором смысле смягчает эффект возможных ошибок, накапливающихся при такой двукратной замене “истинного” распределения более удобным (пусть и с надлежащим теоретическим обоснованием).

6.3.5. О влиянии ширины окна на результат применения СРС-метода

Теперь рассмотрим вопросы, связанные с длиной скользящего интервала (шириной окна), и ее взаимосвязь с интервалом времени между наблюдениями (шагом временного ряда).

Как уже отмечалось, интерес представляют такие значения ширины скользящего окна, которые кратны торговой неделе, поскольку при таком выборе ширины окна нет риска получить результаты, искаженные периодическими колебаниями интенсивностей торгов в разные торговые дни, а также тем обстоятельством, что число рабочих дней в течение скользящего интервала может меняться в зависимости от его положения. Целесообразность такого выбора обусловлена тем, что по некоторым дням средняя активность торгов действительно систематически выше.

Когда выбирается длина скользящего интервала, приходится находить компромисс между гладкостью получаемого результата и его актуальностью. Если выбрать малую длину интервала, то для получения оценок будут использованы данные, лежащие не так далеко в прошлом, но при этом оценки могут быть подвержены сильному влиянию лишь нескольких изменяющихся значений, то есть при сдвиге только на один шаг можно получить существенно отличную картину. Напротив, если использовать интервал большой длины, то резких изменений оценок при сдвигах на малое число шагов как правило не будет, но при этом оценки будут зависеть от большого числа наблюдений, далеко отстоящих от текущего момента в прошлое и потому, возможно, утративших информативность.

При выборе ширины окна и частоты, с какой измеряется рассматриваемый индекс, необходимо учитывать два обстоятельства: (i) число наблюде-

ний, попадающих внутрь скользящего интервала, должно быть достаточным для приемлемого оценивания параметров смеси (обычно более ста) и (ii) интервал между каждыми двумя соседними наблюдениями достаточно велик, чтобы можно было считать, что в течение такого интервала заключается большое число контрактов.

Рис. 16 показывает различие в результатах, полученных при четырех разных значениях ширины окна, равных одной, двум, трем и четырем торговым неделям. При оценивании используются получасовые данные (шаг временного ряда – интервал между наблюдениями – равен 30 минутам). Сравнение четырех графиков говорит о том, что при выборе ширины окна следует обратить внимание на две характеристики: гладкость доминирующего процесса низкой волатильности (самые темные линии) и степень детализации остальных компонент волатильности.

При сравнении первого и второго графиков (интервалы длиной в одну и две недели) видно, что процесс, описывающий компоненту низкой волатильности, на втором графике более гладкий, нежели на первом, тогда как поведение процессов, описывающих компоненты высокой волатильности, примерно одинаково на обоих графиках. Более того, заметим, что выборка объемом лишь 90 наблюдений все же не позволяет получить достаточно надежные (точные) оценки параметров смеси, поэтому второй случай (двухнедельный скользящий интервал) представляется более предпочтительным. Третий и четвертый графики отличаются незначительно, компоненты волатильности на них практически одинаково гладкие. При этом ясно, что из этой пары третий случай (трехнедельный скользящий интервал) более предпочтителен, так как в нем задействовано меньше наблюдений из “далекого” прошлого.

Таким образом, главный вопрос заключается в том, какой из двух интервалов – двух- или трехнедельный – следует предпочесть. Ответ зависит от целей исследования. Двухнедельный интервал представляется более пригодным для исследования свойств ЕМ-алгоритма, рассмотренных в предыдущих разделах (зависимость от начальных приближений, скорость сходимости и т. п.), так как, по-видимому, такой интервал представляет собой “более серьезное” испытание для ЕМ-алгоритма: имеются некоторые довольно резкие переходы, в которых, по-видимому, ЕМ-алгоритм оканчивает работу в разных локальных максимумах. Эта же длина интервала представляется наиболее пригодной для прогнозирования, которое будет предметом одного из следующих разделов. Дело в том, что, опять-таки, при таком выборе ширины окна используется не так много прошлых значений (что существенно для прогнозирования) и результат получается

не слишком гладким, то есть результат прогнозирования оказывается не слишком тривиальным. Использование двухнедельного окна также более целесообразно с точки зрения практического применения: если исследователя интересует текущее состояние рынка, то, очевидно, лучше использовать как можно меньше прошлых значений. Однако, если исследователь заинтересован в описании эволюции состояния рынка за более или менее продолжительный период времени, то целесообразнее использовать трехнедельное окно, поскольку при такой длине скользящего интервала процессы, описывающие динамику различных компонент волатильности, более заметны в силу их большей гладкости.

Если СРС-метод используется для прогнозирования, другими словами, если использование результатов применения СРС-метода ориентировано на будущее, тогда, конечно же, при реализации метода рабочее окно должно от начального (крайнего левого, если руководствоваться графическим представлением, в котором времени соответствует ось абсцисс) положения сдвигаться вправо, то есть в направлении естественного возрастания времени. Однако, если СРС-метод используется с целью анализа прошлого, то можно рекомендовать применить СРС-метод также и в обратном направлении, то есть как бы обращая время вспять. Сопоставление результатов, полученных при применении “прямой” и “обращенной” версий СРС-метода может убрать эффект инерции, присущий СРС-методу с большой шириной рабочего окна.

6.3.6. О влиянии масштаба (длины интервала времени между наблюдениями) на результат применения СРС-метода

Для человеческого ума непонятна абсолютная непрерывность движения. Человеку становятся понятны законы какого бы то ни было движения только тогда, когда он рассматривает произвольно взятые единицы этого движения. Но вместе с тем из этого-то произвольного деления непрерывного движения на прерывные единицы проистекает большая часть человеческих заблуждений.

Л. Н. Толстой. *Война и мир*. Т. 3.

Рассмотрим теперь некоторые вопросы, связанные с влиянием длины интервала времени между наблюдениями на получаемый результат. Эти вопросы совсем не просты. Периодичность измерения значений индекса играет чрезвычайно важную роль для интерпретации результатов. Эта про-

блема уже затрагивалась в разделе 4.1. Ее суть заключается в том, что коэффициент диффузии характеризует стандартное отклонение приращения (логарифма) индекса *за вполне определенный промежуток времени* – интервал времени между наблюдениями, то есть стандартное отклонение разности значений (логарифма) индекса в равноотстоящие моменты времени. При этом характер эволюции процесса *между* последовательными наблюдениями (измерениями, отсчетами) остается неизвестным.

Более того, при неадекватном выборе временного масштаба – длины интервала времени между наблюдениями – может быть получена картина, совершенно не соответствующая действительности. В этом легко убедиться с помощью следующего тривиального примера. Рассмотрим функцию $X(t) = \sin t$. Сначала предположим, что она “наблюдается” в точках, кратных периоду, скажем, точках вида $t_k = t_0 + 2\pi k$, $k = 0, 1, 2, \dots$. Очевидно, что тогда “временной ряд” $x_k = X(t_k) = \text{const}(k) \equiv \sin t_0$ абсолютно не содержит никакой информации об изменении “наблюдаемой” функции. Но если та же функция $X(t) = \sin t$ “наблюдается” в точках вида $t_k = t_0 + (2\pi + \epsilon)k$, $k = 0, 1, 2, \dots$, где $\epsilon > 0$ – малое число, то “временной ряд” $x_k = X(t_k) = \sin(t_0 + (2\pi + \epsilon)k)$ в зависимости от величины ϵ может вместо гладкой плавной кривой выглядеть как пила с неровными и довольно острыми зубцами, причем в точках оси времени, соответствующих зубцам, запросто можно предположить (а с помощью соответствующих статистических процедур и “научно обосновать”) отсутствие производной у исходной функции, что, естественно, не имеет никакого отношения к реальности. К сожалению, выбор интервала времени между наблюдениями (измерениями) часто не зависит от исследователя-статистика.

Как уже говорилось, от выбора интервала времени между наблюдениями сильно зависит интерпретация результатов. Предположим, что этот интервал достаточно мал для того, чтобы быть уверенным, что процесс не слишком сильно “виляет” между соседними точками отсчетов. Забегая вперед, на время отойдем от принятого ранее упрощения, заключающегося в том, что средние значения компонент смеси равны нулю, и будем считать, что ЕМ-алгоритм на каждом шаге оценивает все параметры смеси. Так как изучаемый временной ряд представляет собой *приращения* (логарифма) индекса (см. (6.1.1)), то оцененные средние значения компонент смеси (параметры положения, математические ожидания нормальных законов, составляющих смесь) являются *средними* (по “ансамблю”, то есть по элементарным исходам) перемещениями (логарифма) индекса. Но так как интервалы времени между всеми отсчетами одинаковы, то, приняв длину этих интервалов за единицу времени, можно заключить, что среднее

значение каждой компоненты, будучи средним перемещением за единицу времени, есть не что иное как *усредненная по “ансамблю”* средняя (по времени – за принятую единицу времени) скорость изменения соответствующей составляющей исходного процесса. Таким образом, при оценивании параметров смеси на каждом шаге автоматически получается (статистически оценивается) *распределение скоростей* изменения компонент той системы, которую характеризует временной ряд значений рассматриваемого индекса. Так, наблюдая финансовый индекс, с помощью СРС-метода реализуется возможность оценивать изменение во времени скоростей изменения компонент финансового рынка. Ясно, что понятие *средней скорости* сильно зависит от длины интервала времени, усреднение по которому проводится. Таким образом, залогом успеха каждого применения СРС-метода служит адекватность длины интервала времени между наблюдениями характерному масштабу времени, соответствующему интенсивности изучаемых изменений рассматриваемой системы.

Перейдем от проблем, связанных с интерпретацией результатов, к “технической” стороне дела. Сравнивая два графика на рис. 17, можно заметить разницу в результатах анализа компонент волатильности по данным, полученным при разной длине интервала времени между наблюдениями. Хотя оба графика фактически соответствуют одному и тому же процессу, можно заметить, что тот график, где используются получасовые данные, имеет большее разрешение, нежели график, соответствующий часовым данным. На этих двух графиках особенно различается поведение компонент с высокой волатильностью.

Также можно заметить, что на больших интервалах между отсчетами компоненты с высокой волатильностью более значимы (имеют большие веса).

Отметим, однако, что выбор большой ширины рабочего окна для ЕМ-алгоритма может оказаться опасным вследствие возможности получения неадекватных результатов из-за возрастающей инерции СРС-метода.

Соотношение между распределениями вероятностей, соответствующими данным разных типов (то есть с разной длиной интервала времени между последовательными отсчетами), обсуждалось в разделе 4.3.3 (см. лемму 4.3.1). Теоретически предсказанные с помощью леммы 4.3.1 и реальные (статистически оцененные) значения можно сравнить с помощью рис. 18. На нем изображены два графика. Левый демонстрирует теоретические значения компонент волатильности и соответствующих весов, вычисленные по формулам свертки получасовых данных, представленным в следствии 4.3.2. Правый график демонстрирует статистически оценен-

ный портрет волатильности того же процесса, полученный применением СРС-метода с длиной скользящего интервала $L = 360$ к часовым данным о динамике того же процесса. В целом наблюдается аналогичная картина. Возможно, за счет специального выбора параметров СРС-метода (в частности, специального выбора начальных приближений) сходство может быть усилено. Но и имеющиеся аналогии говорят в пользу подхода, которому посвящена данная книга.

6.3.7. Промежуточные итоги

Подводя промежуточный итог обсуждений свойств СРС-метода и результатов, получаемых с его помощью при анализе реальных данных, необходимо отметить следующее.

Применение СРС-метода к реальным данным показывает, что распределения приращений логарифмов исследуемых финансовых индексов отличны от нормальных.

Эффект, который может быть интерпретирован как стохастичность *одномерной* волатильности – коэффициента диффузии в модели геометрического броуновского движения, – в рамках рассматриваемого подхода должен проявляться как наличие *единственной* кривой в графике диффузионных компонент волатильности, получаемой при применении СРС-метода. Другими словами, стохастичность одномерной волатильности проявляется в существовании *единственной* диффузионной компоненты волатильности. Однако применение СРС-метода к реальным данным показывает, что указанный эффект наблюдается лишь при очень малой ширине окна и при специальном выборе начальных приближений для ЕМ-алгоритма при очередном сдвиге скользящего окна, когда в качестве таковых берутся итоговые оценки, полученные при предыдущем положении окна. Во всех остальных случаях (даже при очень малой ширине окна и случайном выборе начальных приближений для ЕМ-алгоритма) наблюдается *многомерная* диффузионная картина.

В портретах диффузионных компонент волатильности, получаемых с помощью СРС-метода, четко наблюдаются *несколько* довольно гладко меняющихся компонент волатильности. Другими словами, дискретное смешивающее распределение с конечным носителем демонстрирует высокую адекватность. Можно было ожидать, что смешивающее распределение будет иметь не столь явно конечный носитель и окажется более гладким. Однако получаемые результаты свидетельствуют об обратном. Это означает, что в каждый момент состояние рынка определяется конечным (доволь-

но малым) числом некоторых “механизмов”, и СРС-метод, по-видимому, позволяет их выделять.

6.4. СРС-метод с фиксированными значениями компонент волатильности

6.4.1. Основная идея СРС-метода с фиксированными значениями компонент волатильности

Идея, которая лежит в основе модификации СРС-метода, описываемой в данном разделе, уже обсуждалась в разделе 5.8.

Выше было введено одно упрощение (правда, несколько ограничивающее общность) – были зафиксированы нулевые параметры положения компонент подгоняемой смеси, и функции СРС-метода были редуцированы к оцениванию только значений диффузионных компонент волатильности и их весов. Приведенные выше рассуждения позволяют, руководствуясь уже не соображениями удобства изложения материала (ради чего, собственно, и вводилось упомянутое ограничение), а имея в виду реальное упрощение метода, без *существенной* потери общности, пойти еще дальше и зафиксировать (довольно большое) конечное множество известных значений диффузионных компонент волатильности и свести статистический анализ волатильности с помощью СРС-метода к оцениванию *только* весов компонент. Ясно, что такое упрощение сулит большую экономию вычислительных ресурсов и может заметно ускорить вычисления в рамках СРС-метода. Именно о такой версии СРС-метода и пойдет речь в данном разделе.

В отличие от раздела 5.8, здесь будут рассмотрены способы выбора приемлемой неравномерной конечной сетки, которую можно набросить на (возможно, континуальное) множество возможных значений диффузионной волатильности. По сравнению с реализацией ЕМ-алгоритма в обычном, статическом, режиме, при его использовании в скользящем (динамическом) режиме в рамках СРС-метода проблема быстродействия и экономии вычислительных ресурсов становится еще более острой. “Приемлемая” конечная сетка, с одной стороны, должна обеспечивать хорошую точность и не приводить к искажению “истинного” портрета волатильности описанной выше версией СРС-метода с фиксированными значениями компонент волатильности, совпадающими с узлами этой сетки. С другой стороны, она

должна обеспечивать высокую производительность СРС-метода, в частности, содержать не так много узлов. Возможные способы отыскания такого компромисса описываются ниже.

6.4.2. Дискретизация пространства компонент волатильности

Как уже отмечалось, применение СРС-метода в значительной мере представляет собой акт творчества, направленного на более наглядное представление (визуализацию) получаемых результатов. С формальной точки зрения, описанные выше алгоритмы *должны* работать при произвольно большом числе компонент. Более того, в соответствии со сказанным в разделе 3.3, чем больше число компонент, тем выше точность метода, то есть тем ближе получаемые результаты к оцениваемой волатильности, или, в терминах искусства, тем больше портрет волатильности, созданный с помощью СРС-метода, похож на оригинал. Однако СРС-метод должен *работать*, то есть он должен быть достаточно быстродействующим, особенно когда используется для прогнозирования. Можно показать, что производительность (быстродействие) алгоритма экспоненциально убывает при увеличении числа k^{discr} “активных” компонент волатильности. Поэтому, в противоположность требованиям повышения точности, требования повышения производительности предполагают, что k^{discr} должно быть как можно меньшим. Подобные рассуждения мотивируют рассмотрение механизма трансформации континуального пространства возможных значений компонент волатильности в дискретное множество с целью оптимального выбора таких k значений, каждое из которых будет участвовать в окончательном представлении результатов, причем это окончательное представление будет достаточно подробным. Другими словами, ни один из весов, соответствующих выбранным k компонентам, не должен оставаться равным нулю на протяжении всего анализируемого промежутка времени, и ни один значимый процесс, описывающий компоненту волатильности, не будет “наильно” сдвинут на слишком высокий или слишком низкий уровень из-за слишком разреженной дискретизации.

С этой целью сначала применим ЕМ-алгоритм ко всему временному ряду. Вместо того, чтобы взять $k^{\text{cont}} = 4$ начальных значений, в такой “глобальной” реализации ЕМ-алгоритма будем использовать несколько больше компонент, чтобы с некоторой гарантией идентифицировать все компоненты, участвующие в процессе. Для рассматриваемого метода дискретизации можно взять довольно большое число компонент, поскольку, как

будет видно в дальнейшем, начиная с некоторого “порога насыщения” добавление новых компонент не оказывает практически никакого влияния на результат. На рис. ?? показан результат для процесса типа (6.1.1) с $k^{\text{cont}} = 10$ компонентами (точки соответствуют весам, смысл кривой будет объяснен позднее). Интересно, что значения компонент волатильности концентрируются вокруг четырех точек, хотя скользящая реализация ЕМ-алгоритма (СРС-метод) выявляет более широкое распределение компонент волатильности. В результате получается вектор оценок компонент волатильности

$$\vec{\sigma}^{\text{cont}} = (\sigma_1, \dots, \sigma_{k^{\text{cont}}})$$

(для определенности предполагается, что элементы этого вектора *упорядочены по возрастанию*) и вектор оценок соответствующих весов

$$\vec{p}^{\text{cont}} = (p_1, \dots, p_{k^{\text{cont}}}).$$

Самая простая дискретизация может быть получена путем распределения k^{discr} уровней (значений) компонент волатильности равномерно по некоторому интервалу, включающему минимальное (σ_1) и максимальное ($\sigma_{\text{cont}} = \sigma_{k^{\text{cont}}}$) из полученных значений волатильности. Для некоторых $\delta_1, \delta_2 \geq 0$ положим

$$\vec{\sigma}^{\text{unif}} = (\sigma_1^{\text{unif}}, \dots, \sigma_{k^{\text{cont}}}^{\text{unif}}),$$

где

$$\sigma_i^{\text{unif}} = (1 - \delta_1)\sigma_1 + \frac{i - 1}{k^{\text{discr}} - 1}\Delta, \quad i = 1, \dots, k^{\text{cont}},$$

и

$$\Delta = [(1 + \delta_2)\sigma_{\text{cont}} - (1 - \delta_1)\sigma_1].$$

У этого метода есть недостаток: вполне возможно, что какие-либо соседние оцененные значения σ_j и σ_{j+1} окажутся отстоящими довольно далеко друг от друга, причем на протяжении всего рассматриваемого промежутка времени ни одна компонента волатильности не попадет в интервал $[\sigma_j, \sigma_{j+1}]$. Вместе с тем при довольно большом значении k^{discr} и равномерном распределении уровней в указанный интервал могут попасть несколько фиксированных уровней, из-за чего появятся точки, которым, скорее всего, не будут приписаны доминирующие веса, что просто увеличит время сходимости алгоритма, не давая никакой дополнительной информации. Поэтому, по-видимому, более целесообразно использовать *неравномерную* сетку значений фиксированных уровней волатильности и концентрировать их в соответствии с их важностью вокруг значений $\sigma_1, \dots, \sigma_{\text{cont}}$.

Следуя сказанному, распределим фиксированные уровни волатильности вокруг глобально доминирующих значений $\sigma_1, \dots, \sigma_{\text{cont}}$ в соответствии с “нормальным” законом. Однако носитель классического нормального закона бесконечен и включает отрицательную полуось, в то время как фиксированные уровни волатильности должны принадлежать ограниченному интервалу, левая граница которого положительна. Чтобы совместить это ограничение с желанием иметь “нормальное” распределение фиксированных уровней вокруг глобально доминирующих компонент, можно поступить двумя способами.

Во-первых, можно реализовать СРС-метод с ядерными оценками компонент волатильности, в котором рабочий интервал совпадает со всем интервалом времени, на котором задан исходный временной ряд, и нормальным ядром. В результате получится гладкая ядерная оценка $G(\sigma)$ смешивающего распределения – глобального распределения компонент волатильности (функцию $G(\sigma)$ можно определить как интеграл от ядерной оценки (6.3.1)). Затем в качестве фиксированных уровней волатильности надо выбрать квантили функции распределения $G(\sigma)$ равноотстоящих порядков (при этом соблюдая условие положительности наименьшего уровня). Недостаток такого подхода заключается в необходимости осуществления нескольких (возможно, многих) предварительных глобальных реализаций ЕМ-алгоритма для ядерного оценивания компонент волатильности. Это может свести на нет все преимущества, которые имеет СРС-метод с фиксированными уровнями компонент волатильности с точки зрения его быстродействия.

Чтобы не быть связанными необходимостью осуществления нескольких предварительных глобальных реализаций ЕМ-алгоритма, рассмотрим второй способ получения “псевдонормального” распределения фиксированных уровней вокруг глобально доминирующих компонент. Зафиксируем некоторые числа $\tau > 0$ и $\delta > 0$ и введем функцию распределения

$$F(x) = \sum_{i=1}^{k_{\text{cont}}} \left[p_i \cdot \frac{\Phi\left(\frac{x-\sigma_i}{\tau}\right) - \Phi\left(-\frac{\sigma_i}{\tau}\right)}{\Phi\left(\frac{(1+\delta)\sigma_{\text{cont}}-\sigma_i}{\tau}\right) - \Phi\left(-\frac{\sigma_i}{\tau}\right)} \right], \quad x \in [0, (1+\delta)\sigma_{\text{cont}}],$$

где $\Phi(x)$ – стандартная нормальная функция распределения. Функция распределения $F(x)$ приписывает “нормальные” веса тем точкам множества значений компонент волатильности, в которых наиболее правдоподобно (в соответствии с глобально вычисленными оценками компонент волатильности и их весов) размещение доминирующей части распределения компонент волатильности. Заметим, что все точки роста функции распределения $F(x)$ сосредоточены в интервале $[0, (1+\delta)\sigma_{\text{cont}}]$. Плотность вероятностей,

соответствующая функции распределения $F(x)$, и глобально оцененные компоненты волатильности с их весами изображены на рис. ???. В рассматриваемом случае использовались значения параметров $\tau = \frac{1}{30}(1 + \delta)\sigma_{\text{cont}}$ и $\delta = 0.2$.

Зафиксируем произвольное малое число $\epsilon > 0$. Теперь в качестве фиксированных уровней волатильности выберем квантили функции распределения $F(x)$ равноотстоящих порядков:

$$\vec{\sigma}^{\text{opt}} = (\sigma_1^{\text{opt}}, \dots, \sigma_{k^{\text{diskr}}}^{\text{opt}}),$$

где

$$\begin{aligned} \sigma_1^{\text{opt}} &= F^{-1}(\epsilon), \quad \sigma_{k^{\text{diskr}}}^{\text{opt}} = F^{-1}(1 - \epsilon), \\ \sigma_i^{\text{opt}} &= F^{-1}\left(\epsilon + \frac{i - 1}{k^{\text{diskr}} - 1}(1 - 2\epsilon)\right), \quad i = 2, \dots, k^{\text{diskr}} - 1, \end{aligned}$$

В следующем разделе будут приведены дополнительные доводы в пользу целесообразности указанного способа дискретизации множества компонент волатильности.

6.4.3. Сравнительная эффективность СРС-метода с фиксированными значениями компонент волатильности

От модификации СРС-метода, описанной выше, вполне можно ожидать увеличения производительности (в первую очередь, быстродействия), поскольку здесь ЕМ-алгоритм не должен оценивать значения уровней волатильности (σ_j), так как они известны и фиксированы.

Впрочем, некоторые проблемы все еще остаются и, более того, появляются новые. Фиксация значения уровней волатильности в определенном смысле удаляет от предположений, обеспечивающих сходимость ЕМ-алгоритма к оценкам максимального правдоподобия. В противоположность предыдущему случаю, когда фиксировались только (нулевые) значения a_i , теперь фиксируются те значения, которые имеют большое влияние на сходимость алгоритма, увеличивая риск сходимости к локальному максимуму функции правдоподобия и даже расходимости алгоритма, так как теперь оптимизируется только одно слагаемое в разложении функции правдоподобия (см. (5.3.20)).

Решение о принятии результата действия ЕМ-алгоритма должно основываться на сравнении со значением функции правдоподобия, полученным

с помощью ЕМ-алгоритма в разделе 6.3. Важно отметить, что, с одной стороны, ценой повышения быстродействия при переходе от переменных уровней (компонент) волатильности к фиксированным может оказаться определенное уменьшение правдоподобия, поскольку свобода выбора уровней волатильностей теперь сильно ограничивается дискретизацией. С другой стороны, вполне возможно, что правдоподобие возрастет, поскольку увеличивается число (пусть и фиксированных) параметров. Например, в предыдущем разделе было четыре возможных значения, теперь – двадцать.

Существуют и другие алгоритмы, способные находить экстремумы нелинейных функций более чем одной переменной. Для сравнения с модифицированным (усеченным) ЕМ-алгоритмом здесь будет рассмотрен один из них – `fmincon` – стандартная процедура пакета Matlab. Эта процедура ориентирована на отыскание условного минимума скалярной функции нескольких переменных с помощью метода последовательного квадратичного программирования. Ее описание приведено в (Fletcher and Powell, 1963), (Goldfarb, 1970) (также см. Matlab Help и раздел 8.2 в (Королев, 2007b)). Этой процедуре, однако, присущи те же недостатки, что и ЕМ-алгоритму. В частности, нет гарантии, что найденный локальный экстремум окажется глобальным.

Наряду со сравнением результатов двух различных алгоритмов необходимо также уделить внимание выбору начальных приближений для весов. С этой целью будут использованы те факты, которые уже известны.

Для выбора начальных приближений будет использоваться метод, аналогичный использованному в предыдущем разделе. Однако вместо случайного выбора значения волатильности в соответствии с ее нормальным распределением, теперь для каждого уровня волатильности соответствующий вес будет распределяться по соседним уровням с помощью нормальной плотности:

$$p_i^{\text{discr}} = \sum_{j=1}^{k^{\text{cont}}} \left[\frac{p_j^{\text{cont}} \varphi\left(\frac{\sigma_i^{\text{discr}} - \sigma_j^{\text{cont}}}{\tau}\right)}{\sum_{\ell=1}^{k^{\text{discr}}} \varphi\left(\frac{\sigma_{\ell}^{\text{discr}} - \sigma_j^{\text{cont}}}{\tau}\right)} \right]. \quad (6.4.1)$$

В качестве альтернативы рассмотрим равномерное начальное приближение:

$$p_i = \frac{1}{k^{\text{discr}}} \quad \text{для } i = \{1, \dots, k^{\text{discr}}\}.$$

Сравнительный анализ результатов несколько успокаивает, поскольку он явно свидетельствует в пользу модифицированного ЕМ-алгоритма.

Финальные функции правдоподобия, полученные тремя процедурами типа ЕМ-алгоритма (ЕМ-алгоритм с переменными уровнями волатильности, ЕМ-алгоритм с фиксированными уровнями волатильности и случайным выбором начальных приближений для весов на основе (6.4.1), ЕМ-алгоритм с фиксированными уровнями волатильности и равномерными начальными приближениями для весов), очень похожи, и ни в одном случае не зафиксирован результат работы усеченного ЕМ-алгоритма, который был бы значимо хуже результата, полученного обычным ЕМ-алгоритмом.

Как и предполагалось, на некоторых участках временного ряда метод с двадцатью фиксированными значениями волатильности дал лучшие результаты. При этом в терминах функции правдоподобия нет никакой заметной разницы в результатах работы этого метода с двумя описанными выше способами выбора начальных приближений: соответствующие графики полученных весов почти идентичны.

В случае **fmincon**-алгоритма дела обстоят совсем не так. На большинстве позиций окна результаты его работы практически аналогичны результатам работы модифицированного ЕМ-алгоритма, но имеются такие участки, на которых **fmincon**-алгоритм сходил к результатам, весьма далеким от оптимальных. Более того, при сравнении графиков весов, полученных ЕМ-алгоритмом и **fmincon**-алгоритмом, становится заметен еще один недостаток последнего метода: в то время как процессы оценок отдельных весов, полученные ЕМ-алгоритмом, почти всегда непрерывны, те же процессы, полученные **fmincon**-алгоритмом, имеют много разрывов, производя впечатление “случайно” распределенных по соседним уровням волатильности.

Более того, при сравнении правдоподобия оценок, полученных модифицированным ЕМ-алгоритмом и **fmincon**-алгоритмом, можно заметить, что правдоподобие оценок, полученных **fmincon**-алгоритмом, всюду, по крайней мере, не лучше правдоподобия оценок, полученных модифицированным ЕМ-алгоритмом. Это наблюдение убеждает в том, что модифицированный ЕМ-алгоритм в самом деле дает хорошие результаты.

Относительно выбора начальных приближений можно заметить следующее. В случае модифицированного ЕМ-алгоритма адаптированный метод выбора начальных приближений для весов с помощью соотношения (6.4.1) немного ускоряет сходимость алгоритма, но практически не влияет на окончательный результат. В случае **fmincon**-алгоритма на некоторых участках адаптированный метод дает немного лучшие результаты, чем метод “равномерного распределения” начальных приближений (см., например, рис. ??, $n = 480 - 490$), однако точки, в которых результаты

плохие, просто распределены по-другому.

Процессы, представляющие веса различных компонент волатильности, изображены на рис. ???. График показывает, что наш способ выбора (фиксированных) уровней волатильности вполне хорош: за исключением двух низких уровней волатильности все уровни соответствуют тем компонентам волатильности, которые доминируют на протяжении достаточно продолжительного периода времени.

Для фиксированных параметров скорость сходимости главным образом зависит от критерия останова алгоритма, то есть от его толерантности (чувствительности). При этом скорости сходимости алгоритмов с двумя методами выбора начальных приближений различаются незначительно. Задавая разные пороги чувствительности алгоритма, можно убедиться, что он относительно быстро сходится к приближенному решению, но заметно замедляется при поиске более точного решения. Что касается параметров, то скорость сходимости главным образом зависит от их числа и объема выборки (длины скользящего интервала – ширины окна). Указать более-менее явный вид зависимости скорости сходимости алгоритма от его параметров практически невозможно, поскольку она зависит не только от числа компонент волатильности и длины скользящего интервала, но также и от самих данных.

6.4.4. Сравнение версий СРС-метода с оцениваемыми и фиксированными значениями компонент волатильности

Версии СРС-метода с оцениваемыми и фиксированными значениями компонент волатильности не только дают результаты разной природы, но также позволяют интерпретировать эти результаты с разных точек зрения.

Когда используется версия с оцениваемыми компонентами волатильности, в каждый момент времени можно сказать, какая смесь нормальных законов в наилучшей степени соответствует наблюдаемым данным. В частности, оценив значения параметров σ_i и p_i , можно утверждать, что с вероятностью p_i очередное наблюдение будет реализацией нормально распределенной случайной величины с нулевым средним и дисперсией σ_i^2 . Более того, можно проследить эволюцию компонент волатильности во времени (в том числе проследить изменение самих параметров σ_i) и их весов – показателей их важности. Многочисленные эксперименты с реальными данными подтверждают это, поскольку, как видно на приводимых в этой книге ри-

сунках, оцененные компоненты волатильности получаются довольно гладкими и в основном изменяются плавно, а их разрывы довольно регулярны (другими словами, компоненты волатильности, напоминающие функцию Дирихле, как правило, не возникают).

В случае версии с фиксированными компонентами интерпретация вычисленных весовых процессов иная. Здесь они описывают эволюцию вероятности того, что очередное наблюдение будет реализацией нормально распределенной случайной величины с нулевым средним и стандартным отклонением, попадающим в соответствующий заранее фиксированный интервал. В некотором смысле множество возможных уровней волатильности разбивается на категории и с помощью СРС-метода отслеживается присутствие и относительная важность каждой из них.

Однако, если фиксированных уровней волатильности много, то картина, получаемая версией СРС-метода с фиксированными компонентами, будет мало отличаться от картины, получаемой версией СРС-метода с оцениваемыми компонентами, за счет эффекта, наблюдаемого в телевидении, где при увеличении числа строк в развертке изображения на экране увеличивается четкость изображения.

6.4.5. Эффект “насыщения” (эффект “достаточности” числа компонент) масштабных смесей, оцененных СРС-методом. Еще раз о том, что такое волатильность

Расщепляя смеси с помощью ЕМ-алгоритма, вообще говоря, можно ошибочно получить вместо одной компоненты смеси две или более компонент с одинаковыми или почти одинаковыми параметрами сдвига и масштаба. СРС-метод с фиксированной сеткой значений параметров смеси позволяет, в определенном смысле упрощая вычислительную задачу, избежать путаницы с выделением таким мнимых компонент и отслеживать изменение лишь весов компонент с фиксированными параметрами сдвига и масштаба. Как уже говорилось, идея такого подхода сродни идее спектрального анализа, при котором фиксируется сетка значений частот и отслеживаются веса соответствующих гармоник.

При этом, казалось бы, увеличивая число узлов сетки, то есть делая ее более частой, можно подобрать аппроксимирующую смесь все более и более точно.

Однако нельзя не отметить следующий довольно неожиданный, но

стабильно наблюдающийся при анализе реальных данных удивительный эффект: при подгонке чисто масштабных смесей, даже если фиксировалась довольно частая сетка (порядка 30–100 значений), *значимыми оказывались не более пяти компонент смеси, веса же остальных компонент оказывались нулевыми!* Другими словами, наблюдался эффект “насыщения” масштабных смесей или эффект достаточности числа компонент таких смесей: начиная с некоторого (довольно умеренного) числа компонент, добавление новых возможных компонент не приводило к повышению точности аппроксимации: их веса оставались нулевыми!

К сожалению, общие сдвиг-масштабные смеси ведут себя принципиально противоположным образом, превращаясь при измельчении шага сетки в непараметрические оценки плотностей типа ядерных оценок.

Тем не менее эффект “насыщения” масштабных смесей позволяет еще раз вернуться к обсуждению вопроса о том, что же такое волатильность, и предложить остановиться на таком ее определении, которое отождествляет волатильность процесса с его дисперсией, вычисленной в предположении, что распределение приращений процесса *в точности* равно подогнанной с помощью ЕМ-алгоритма конечной смеси нормальных законов. Во-первых, так как смесь конечна, то дисперсия (а стало быть, и волатильность, так определенная) всегда существует. Во-вторых, как было показано выше, такое определение естественно приводит к информативной и многомерной интерпретации волатильности и выделению ее динамической и диффузионных компонент. В-третьих, эффект достаточности компонент масштабных смесей, упомянутый выше, делает это определение практически не зависящим от числа компонент подогнанной смеси при их адекватности и позволяет довольно естественную интерпретацию в терминах предполагаемого диффузионного спектра при их неадекватности (см. раздел 4.4).

Такое определение, конечно же, является статистическим, то есть в качестве волатильности рассматривается статистика (функция от конкретных данных). Тем самым подразумевается зависимость волатильности от длины интервалов времени между наблюдениями: изменяя эти интервалы, мы получаем другие данные, и стало быть, другую волатильность. Наконец, нельзя не сказать о том, что при таком определении прослеживается явная аналогия с подходом, связанным с “подразумеваемой” (implied) волатильностью (см. раздел 4.2.3). Определяя волатильность как дисперсию подгоняемой смеси, мы тоже в некотором смысле заменяем ее “истинное” значение “предполагаемой” величиной – дисперсией “предполагаемой” конечной смеси.

6.5. СРС-метод с оцениванием диффузионных и динамической компонент волатильности

Выше суть метода скользящего разделения смесей была продемонстрирована на примере оценивания только диффузионных компонент волатильности (предполагаемого диффузионного спектра). При этом динамическая компонента волатильности (см. раздел 4.3.2) предполагалась равной нулю. Другими словами, возможное существование локальных трендов игнорировалось. Полная же версия СРС-метода наряду с оцениваем дисперсий компонент подразумевает также оценивание параметров сдвига компонент смесей нормальных законов, то есть оценивание и локальных трендов. При этом, поскольку в качестве исходных данных для СРС-метода используется ряд *приращений* (логарифмов) временного ряда, динамическая компонента волатильности допускает как минимум две возможные интерпретации, первая из которых важна для анализа эволюции финансовых индексов, а вторая важна в физических приложениях.

А именно, если в момент времени t распределение приращений (логарифмов) временного ряда имеет вид

$$F_t(x) = \sum_{i=1}^{k(t)} p_i(t) \Phi\left(\frac{x - a_i(t)}{\sigma_i(t)}\right),$$

то, во-первых, вектор-функция $\mathbf{a}(t) = (a_1(t), \dots, a_{k(t)}(t))$ может быть интерпретирована как “разложение” приращений “основного” тренда изучаемого процесса на компоненты, соответствующие тенденциям действия основных структур финансового рынка (основных групп участников рынка). С этой точки зрения большой практический интерес представляет задача покомпонентного “интегрирования” функции $\mathbf{a}(t)$ с целью получения разложения тренда не приращений, но самого исходного процесса. Эта задача не так проста, поскольку, с одной стороны, размерность вектор-функции $\mathbf{a}(t)$ может меняться во времени и, с другой стороны, необходимо правильно идентифицировать компоненты этой функции при переходе от предыдущего момента времени к следующему.

Во-вторых, компоненты вектор-функции $\mathbf{a}(t) = (a_1(t), \dots, a_{k(t)}(t))$ представляют собой *ожидаемые*, то есть *средние* значения “компонент” приращений исходного процесса. Но исходный ряд представляет собой значения изучаемого процесса, наблюдаемые в *равноотстоящие* моменты

времени, расстояние между которыми равно одному отсчету, то есть *единице времени*, характеризующей рассматриваемый временной масштаб. Следовательно, компоненты вектор-функции $\mathbf{a}(t) = (a_1(t), \dots, a_{k(t)}(t))$ представляют собой *средние* значения “компонент” приращений исходного процесса за единицу времени, то есть *средние скорости* изменения рассматриваемого процесса. Таким образом, в каждый момент времени функции $\mathbf{a}(t)$ и $\mathbf{p}(t) = (p_1(t), \dots, p_{k(t)}(t))$ определяют *распределение средних скоростей* по структурам в рассматриваемой сложной системе.

При использовании СРС-метода с оцениванием как диффузионных, так и динамической компонент волатильности можно использовать две модификации СРС-метода: с фиксированными сетками значений параметров сдвига и масштаба компонент и с оцениванием этих значений. Практическое применение обеих модификаций показывает, что предполагаемый диффузионный спектр как правило адекватно передает динамику диффузионных компонент волатильности даже тогда, когда динамические компоненты не равны нулю.

При этом наблюдается эффект перетекания волатильности из стохастической в динамическую компоненту при увеличении числа компонент смеси. В противоположность эффекту насыщения смеси, когда оценивается только предполагаемый диффузионный спектр (см. раздел 6.4.5), в случае, где оцениваются все параметры компонент смеси, при увеличении числа компонент смеси соотношение между динамической и диффузионной компонентами волатильности изменяется в пользу динамической. Можно сказать, что при этом из средства классификации – разделения смесей – СРС-метод превращается в средство непараметрического оценивания плотности.

Этот эффект можно использовать для выбора числа компонент, если заранее задать среднее соотношение динамической и диффузионной компонент волатильности, которое фактически определяет ту желаемую долю динамической волатильности, которая содержится в предполагаемом диффузионном спектре, то есть определяет ту долю динамической волатильности, которой можно было бы пренебречь при рассмотрении только предполагаемого диффузионного спектра. Другими словами, это соотношение задает степень “проработанности” динамической модели.

6.6. Исследование динамики корреляционной структуры хаотических стохастических процессов с помощью скользящего разделения смесей многомерных наблюдений

СРС-метод позволяет выявить динамику корреляционной структуры выделяемых им компонент. Для этого исходный временной ряд $x_1, \dots, x_N$ надо разбить на группы по $d \geq 1$ элементов в каждой. Разбиение можно проводить как на непересекающиеся группы

$$(x_1, \dots, x_d), (x_{d+1}, \dots, x_{2d}), (x_{2d+1}, \dots, x_{3d}), \dots,$$

так и на пересекающиеся группы

$$(x_1, \dots, x_d), (x_2, \dots, x_{2+(d-1)}), (x_3, \dots, x_{3+(d-1)}), \dots$$

Затем каждая группа из d элементов интерпретируется как d -мерный вектор

$$\mathbf{x}_j = (x_{(j-1)d+1}, \dots, x_{jd})^\top, \quad j \geq 1,$$

для первого способа разбиения или

$$\mathbf{x}_j = (x_j, \dots, x_{j+(d-1)})^\top, \quad j \geq 1,$$

для второго способа разбиения исходного временного ряда. К полученной выборке $\mathbf{x}_1, \dots, \mathbf{x}_n$, где $n = \lfloor N/d \rfloor$ в первом случае и $n = N - d$ во втором случае, применяется ЕМ-алгоритм разделения смесей d -мерных нормальных распределений, описанный в разделе 5.3.13.

Оба способа разбиения исходного ряда имеют свои достоинства и недостатки. В частности, при первом способе разбиения сильно (в d раз) уменьшается объем выборки, доступной СРС-методу, однако сохраняется независимость наблюдений (в той мере, в которой она была присуща исходному ряду), что является важным условием сходимости итераций ЕМ-алгоритма к разумным оценкам параметров задачи. При втором же способе обоснованность результатов снижается, однако может получиться более интересная картина корреляционных связей внутри выделяемых компонент.

При этом выбор $d = 2$ позволяет оценивать ковариации соседних наблюдений в выделяемых компонентах, выбор $d = 3$ позволяет оценивать как ковариации соседних наблюдений, так и отстоящих друг от друга на два отсчета, и так далее. Однако при таком подходе неизбежно возникает

конфликт между желанием сузить скользящее окно с целью оперативного отслеживания динамики корреляционной структуры рассматриваемых процессов и требованием максимально возможно его расширить с целью получения сколько-нибудь точных и правдоподобных результатов.

6.7. Примеры применения СРС-метода к анализу влияния информационных интервенций на состояние финансовых рынков

6.7.1. Взрывы в лондонском метро в июле 2005 года

Когда работа над этой книгой была в самом разгаре, произошли события, которые дали возможность протестировать возможности применения СРС-метода к анализу существенно нестационарной динамики финансовых рынков, обусловленной информационными интервенциями. В июле 2005 года террористы произвели серию взрывов в лондонском метро и в автобусах. Эти взрывы унесли жизни более пятидесяти человек и парализовали транспортную систему Лондона.

Сообщения о лондонских взрывах вызвали заметную панику на европейских финансовых рынках. На рис. ?? а) представлен график изменения индекса DAX, поминутно измерявшегося в течение нескольких дней, предшествовавших взрывам, в день взрывов и в течение нескольких дней после взрывов, а также график поминутных приращений этого индекса (в верхней части рисунка) и синхронизированный по времени портрет диффузионных компонент волатильности, полученный с помощью СРС-метода.

В четверг 7 июля ясно виден взрыв волатильности. Этот взрыв характеризуется возникновением новых компонент очень высокой волатильности и исчезновением тех компонент низкой волатильности, которые существовали до 9:00 утра. Это может быть интерпретировано как *полное* изменение характера рынка.

В противоположность такому изменению, в понедельник и вторник также наблюдались довольно заметные всплески волатильности, однако их структура совершенно иная и характеризуется тем, что *в дополнение* к существовавшим компонентам волатильности возникают *новые*, которые затем существуют какое-то (довольно непродолжительное) время. Это означает, что в момент таких “небольших” всплесков волатильности рынок продолжает оставаться устойчивым, причем устойчивость рынка обеспе-

чивается тем, что основные, доминирующие компоненты низкой или умеренной волатильности “продолжают существовать” и почти не изменяются во время всплесков волатильности.

Следует отметить, что всплеск волатильности, по силе визуально равная тем, что регистрировались в понедельник и вторник, наблюдалась и в пятницу. Но, как видно на портрете предполагаемой диффузионной волатильности на рис. ?? с), ее структура отличается от структуры всплесков волатильности в понедельник и вторник. После лондонских взрывов в четверг рынок пришел в возбужденное состояние, поэтому новая информация вызывает мгновенную трансформацию компонент низкой волатильности в компоненты высокой волатильности. Другими словами, в пятницу проявления стабильности рынка незаметны.

Видно, что рынок вернулся в его исходное состояние к следующему понедельнику, когда те же самые базовые компоненты низкой волатильности снова начали играть доминирующую роль.

Рис. ?? d) отображает поведение динамической компоненты волатильности того же временного ряда, содержащего поминутно измеренные значения индекса DAX. Здесь картина не менее интересна и информативна. Несмотря на то что “суммарный” тренд процесса приращений практически горизонтален (ожидаемые значения приращений почти равны нулю), в каждый момент времени этот “нулевой” тренд приращений складывается из, вообще говоря, нескольких довольно существенно разнонаправленных ненулевых тенденций. Именно этим обстоятельством обусловлен выбор афоризма, принадлежащего К. Э. Циолковскому, в качестве эпиграфа к данной книге. Рис. ?? d) наглядно иллюстрирует возможности СРС-метода “раскладывать ноль на ненулевые компоненты”.

В рассматриваемой ситуации эволюция динамических компонент волатильности во времени во многом аналогична соответствующему поведению диффузионных компонент. Если до появления информации о терактах заметны непрерывные “ниточки” регулярных положительных или отрицательных тенденций, то есть рынок живет обычной “спокойной” жизнью, то после терактов в четверг и пятницу непрерывный характер развития тенденций сменяется “пунктирным” хаотическим режимом, что может свидетельствовать о возбужденном и менее устойчивом состоянии рынка, более чувствительном к новой информации. Непрерывные “ниточки” регулярных тенденций снова возникают лишь в понедельник, когда рынок успокаивается.

Сравнивая рис. ?? с) и ?? d), можно заметить, что зависимость диффузионных компонент волатильности от времени имеет существенно более

гладкий, непрерывный характер, нежели зависимость от времени трендов компонент смеси.

Интересно отметить, что рис. ?? d) также позволяет объяснить локальные всплески волатильности во вторник и среду возникновением противоборства двух или нескольких заметных разнонаправленных тенденций, в то время как локальная всплеска волатильности в пятницу оказывается обусловленной стохастическими причинами, так как рынок возбужден. Это подтверждается графиками, приведенными на рис. ?? b).

Равенство нулю “суммарного” тренда приращений, казалось бы, могло бы свидетельствовать о том, что динамическая компонента волатильности должна быть пренебрежимо малой по сравнению с диффузионной. Однако на самом деле это далеко не так. На рис. ?? b) приведены графики диффузионной компоненты волатильности индекса DAX

$$V_{diff}(t) = \sum_{j=1}^{k(t)} p_j(t) \sigma_j^2(t)$$

(более светлая кривая) и ее динамической компоненты

$$V_{dyn}(t) = \sum_{j=1}^{k(t)} p_j(t) [a_j(t) - \bar{a}(t)]^2$$

(более темная кривая), где t – астрономическое время,

$$\bar{a}(t) = \frac{1}{k(t)} \sum_{j=1}^{k(t)} p_j(t) a_j(t),$$

$k(t)$ – число компонент в смеси нормальных законов, подогнанной к распределению поминутных приращений логарифмов индекса DAX на момент t (правая граница окна СРС-метода равна t), $a_j(t)$, σ_j^2 и $p_j(t)$ – соответствующие оценки параметров сдвига (математических ожиданий), параметров масштаба (дисперсий) и весов компонент смеси (см. раздел 4.3.2). Из этих графиков ясно видно, что динамическая и диффузионная компоненты волатильности вносят примерно одинаковый вклад в результирующую волатильность, причем в разные моменты времени доминирующую роль играют разные компоненты, то есть в разные моменты времени волатильность индекса может быть обусловлена разными причинами: наличием нескольких разно- или однонаправленных тенденций или же причинами чисто

стохастического характера. При этом в конкретной рассматриваемой ситуации взрыва волатильности как следствия информационной интервенции диффузионную компоненту можно интерпретировать как волатильность, обусловленную субъективными факторами панического характера, определяющими поведение отдельных участников рынка, в то время как динамическая компонента обусловлена наличием (или появлением) тенденций в поведении групп участников, стратегии которых направлены на использование возникшей паники.

6.7.2. Падение российского рынка в период с февраля по март 2007 года

В феврале–марте 2007 года в течение двух недель можно было наблюдать сильное движение вниз на рынках развивающихся стран. В частности, падение на российском рынке составило около 13% (для сравнения, в июле 2005 года атака террористов обрушила рынки всего на 3%). Падение было вызвано рядом последовательных событий: падение китайского фондового рынка, укрепление японской йены относительно доллара и, как следствие, закрытие большого количества так называемых carry trades. Далее, проблемы на американском рынке проблемных ипотечных кредитов, заявления ряда крупных компаний о финансовых трудностях, которые рынок воспринял как признание банкротства, а также общие опасения замедления экономического роста США и возможность глобального кризиса на мировом фондовом рынке. События, вызвавшие такую реакцию рынка, были менее радикальны и более ожидаемы рынком, чем атака террористов, и были обусловлены экономическими причинами. Несмотря на то что негативные данные по американской экономике, ослабление японской йены и проблемы на американском рынке недвижимости прогнозировались аналитиками, рынок воспринял их панически, что и выразилось в ряде последовательных скачков индекса вниз. На рис. ?? представлен график изменения индекса РТС, измерявшегося в течение периода падения (слева), и соответствующие логарифмические разности этого индекса (справа). На рис. ?? показаны синхронизированные по времени портреты диффузионных и динамических компонент волатильности, полученные с помощью СРС-метода.

Перед обвалом рынка и диффузионная, и динамическая компонента явно вносили примерно равный вклад в суммарную волатильность (рис. ??). В это время диффузионную компоненту (см. левый рис. ??) порождают три четко прослеживаемые компоненты (темная – самая весома

и две светлые – менее весомые). Одновременно с выходом первых негативных новостей наблюдается взрыв волатильности, характеризующийся возникновением компонент высокой волатильности, имеющих на графике прерывистый характер (левый рис. ??, позиции 100–200 шкалы времени). Кстати, этот так называемый “взрыв” не затронул компоненту низкой волатильности, существовавшую до него, но прервал доминировавшую компоненту и перераспределил веса: прежде наиболее весомая составляющая чуть утратила свой вес. Интересно, что немного ранее, чем началось основное резкое падение, диффузионный портрет опять стал ясным: существуют три компоненты, третья из которых – новая, существующая четко до конца падения индекса. Поэтому логично предположить, что эта компонента как раз отражает те факторы на рынке, которые привели к падению, в то время как две другие обусловлены чисто российским спокойным, стабильным информационным фоном. Новая компонента появилась непосредственно перед провалом рынка, что свидетельствует о том, что обвал был заранее ожидаем. Компоненты низкой волатильности, существовавшие до обвала, остаются, и, более того, их вклад в суммарную диффузионную компоненту остается примерно тем же на протяжении всего наблюдаемого периода. Здесь четко прослеживается реальная информационная картина российской экономики того периода: с одной стороны, умеренно позитивный российский экономический климат и резко негативный новостной фон глобальных рынков – с другой стороны. Если компонента, отражающая ситуацию на внутреннем российском рынке, остается постоянной равновесной, низковолатильной и непрерывной на протяжении всего анализируемого периода, то компоненты внешнего негативного новостного фона резко возникают, сохраняют высокую волатильность, а затем по мере выхода позитивных новостей и общей усталости инвесторов от падения сходят на нет. Базовая компонента волатильности прерывается второй раз в период разворота рынка, когда участники не могут понять, кончилось ли падение.

Правый рис. ?? отображает поведение динамических компонент волатильности индекса РТС за тот же период времени. Эта картина не менее интересна. До обвала динамическая компонента включает в себя нейтральные, но наиболее весомые тенденции низкой волатильности (позиции 0–100) наравне с более волатильной и менее значимой непрерывной составляющей, обусловленной отрицательным внешнеэкономическим климатом. В момент же выхода одной за другой негативных новостей картина меняется, отрицательная составляющая приближается к нейтральным, чуть увеличивая вес. Все тенденции непрерывны за исключением момента

неопределенности перед обвалом. В момент разворота индекса появляется кратковременная высоковолатильная положительная составляющая, объясняемая общей усталостью инвесторов от падения и ожиданием роста индекса на фоне положительного внутриэкономического климата. Интересно, что портрет динамической (трендовой) составляющей волатильности намного более спокойный, чем диффузионной.

Из последнего графика (рис. ??) ясно видно, что динамическая компонента волатильности (темный график) на протяжении почти всего периода вносит больший вклад в результирующую волатильность. Преобладание достаточно спокойной динамической компоненты в рассматриваемом случае объясняется тем, что негативный новостной фон был строго внешним по отношению к российской экономике и подавляющее большинство инвесторов понимали, что происходит, и пытались извлечь из этого выгоду, что выразилось в массовых продажах и затем в массовых покупках. Паника же непрофессиональных инвесторов, отраженная в диффузионных компонентах, носила локальный характер.

Из портретов составляющих волатильности ясно видно, что, несмотря на масштабы падения, ситуация не была близка к лондонским взрывам. Не наблюдается ни равных по силе всплесков волатильности, ни сколь угодно значимого прерывания основных составляющих. Это объясняется тем, что вызвавшие соответствующую реакцию рынка причины, отраженные в компонентах волатильности, не были столь радикальными.

6.7.3. События в США 11 сентября 2001 года

Еще одно трагическое событие, потрясшее весь мир и в то же время давшее нам почву для анализа, произошло 11 сентября 2001 года. В 8:45 по местному времени реактивный пассажирский самолет Боинг-767 врезался в небоскреб номер один Всемирного торгового центра в Нью-Йорке (рис. ?? – пункт 700 на шкале времени индекса CAC 40 (слева), пункт 650 на шкале времени индекса DAX (справа)).

Уже по первым кадрам CNN было видно, что зданию нанесен огромный ущерб. Через 18 минут после первой авиакатастрофы во вторую башню Всемирного торгового центра врезался другой пассажирский самолет. В 9:45 круживший над зданием Пентагона двухмоторный самолет упал на территорию военного ведомства США (Пентагон) и взорвался. В 10:00 в результате повреждений, причиненных терактом, в южной башне Всемирного торгового центра в Нью-Йорке произошел взрыв, и башня обрушилась. Затем поступила информация о еще нескольких взрывах около ад-

министративных зданий и о том, что по предварительным данным угнано от 5 до 11 самолетов. К 11 часам дня Пентагон объявил всеобщую мобилизацию, и в то же время прекратились все крупные финансовые операции в Америке. Трагедия унесла тысячи жизней и нанесла удар по экономике США. К вечеру, помимо новостей о масштабах разрушений, вышло сообщение о том, что американские биржи The American Stock Exchange, The NASDAQ и The New York Stock Exchange объявили, что и на следующий день они не будут функционировать. Таким образом, для анализа ситуации с помощью CPC-метода взять наиболее интересные американские индексы (например S&P 500 или NASDAQ) оказалось невозможным. Поэтому были выбраны индексы французской и немецкой бирж – CAC 40 (левый рис. ??) и DAX (правый рис. ??). Значения индексов брались с интервалом в 1 минуту с 10 по 13 сентября 2001 года.

По графикам индексов наблюдается удивительное согласие тенденций. Это свидетельствует о серьезности события не только для американского, но и для мирового рынка. Рис. ??, отображающий значения относительных приращений рассматриваемых индексов, помогает осознать масштаб бедствия. За один день, а вернее, в течение трех часов, французский индекс упал на 11%, немецкий – на 13%. Для большего числа европейских рынков эта неделя стала худшей за три предшествовавших года.

Портреты диффузионных компонент волатильности (рис. ??) отображают состояние рынка до, после и во время выхода новостей о терактах. До выхода новостей о взрывах французский рынок формировали три примерно равновесные тенденции, немецкий – четыре. Почти одновременно с вестью о терактах на обоих предполагаемых диффузионных спектрах появляется кратковременная прерывистая высоковолатильная но маловесная компонента. Она очень быстро исчезает, но сразу после этого общий рисунок существовавших компонент радикально меняется. Все компоненты приближаются друг к другу, а о непрерывности нет и речи. Интересным моментом данной ситуации является и тот факт, что реакция на новости о терактах в Америке выражается не только в появлении новых компонент высокой волатильности, но и в смещении акцентов. Хорошо видно доминирование средней компоненты, хотя прежде веса были распределены примерно равномерно. Такие картины диффузионной составляющей волатильности свидетельствуют о том, что в спокойное время рынок формировали несколько разных примерно равнозначных по силе тенденций. В период критического события основной тенденцией, определяющей поведение индекса, становится само это событие, которое подавляет все ранее существовавшие компоненты. Уже через 1,5 дня, с середины 12 сентября

(левый рис. ?? – отметка 1250, правый рис. 19 – отметка 1300) картина рынка начинает восстанавливаться: прежде сгрудившиеся три тенденции начинают расходиться и обретать свою прежнюю значимость. Заметим, что как для САС 40, так и для DAX имеются компоненты волатильности, которые не были затронуты паникой. Другими словами, не все ранее существовавшие факторы, влиявшие на формирование индексов, потеряли значение на фоне событий в США.

Рассмотрим теперь динамическую составляющую. Здесь картина, существовавшая до появления новостей, в момент выхода последних вообще не меняется. Более того, сгусток новых компонент появляется, когда рынок на самом дне (левый рис. ?? – пункты 800–1000, правый рис. ?? – пункты 900–1250). Из этого понятно, что основные группы инвесторов активизировались в конце падения. В сам же момент падения паники среди основных игроков не возникло. Когда рынок на дне, есть хорошая возможность для получения прибыли – удачный момент покупки акций по низкой цене. Таким образом, на портретах динамической составляющей прослеживаются кратковременные новые компоненты. Но ни общая картина составляющих волатильности, ни соотношение сил почти не меняются. Это можно объяснить и тем, что, несмотря на масштаб бедствия, последнее происходило в Америке, а значит, хоть и сильно повлияло на мировой рынок, все же не перекрыло все остальные рыночные факторы каждого из анализируемых рынков. Очевидно, картина американских индексов была бы совсем иной, и, по всей видимости, и динамическая, и диффузионная составляющие были бы намного более беспорядочны.

На графиках соотношения динамической и диффузионной компонент (рис. ??), полученных с помощью СРС-метода, видно, что на всем периоде превалировала диффузионная составляющая волатильности. Динамическая превышала последнюю лишь в течение короткого периода после основного падения индекса. Это подтверждает ранее сделанные выводы об активизации главных инвесторов с целью получения прибыли.

Интересно отметить, что портрет предполагаемого диффузионного спектра волатильности индекса РТС в начале 2007 года (левый рис. ??) является почти зеркальным отражением портрета предполагаемого диффузионного спектра волатильности индекса САС 40 10–13 сентября 2001 года (левый рис. ??). В первом случае “паника” была вызвана ожиданием падения, и, как только падение, наконец, началось, рынок успокоился в том смысле, что вместо хаотического разброса компонент волатильности появились три четко выраженные компоненты – три четкие структуры. Во втором случае все наоборот: “паника” стала следствием начавшегося

падения.

7.

Применение СРС-метода к изучению турбулентности

7.1. Гидродинамическая и плазменная турбулентность

Уже сравнительно давно замечена аналогия между стохастическими свойствами процессов, протекающих на финансовых рынках, и процессов, характеризующих турбулентность в физике. В связи с этим необходимо упомянуть работы (Ghashghaie et al., 1996), (Mantegna and Stanley, 1996), (Mantegna and Stanley, 1997), (Barndorff-Nielsen, 1998b), описывающие взаимопроникновение идеологий теории турбулентности и теории финансовых рынков, а также вероятностных моделей и методов, разработанных в рамках соответствующих теорий. Данная глава посвящена применению вероятностных моделей и статистических методов, в частности, метода скользящего разделения смесей, разработанных выше для анализа эволюции финансовых индексов, к изучению структурной плазменной турбулентности.

Здесь будут кратко описаны турбулентные системы и рассмотрены критерии, выделяющие плазменную турбулентность среди других видов турбулентности.

Системы, проявляющие турбулентные свойства, могут быть как в твердом, так и в жидком и газообразном состояниях. Все такие системы характеризуются взаимодействием многих частиц и могут быть описаны (с уменьшением степени детализации) на микроскопическом (кинетическом) или макроскопическом (гидродинамическом) уровнях. Частицы, составля-

ющие эти системы, могут быть электрически нейтральными или же могут нести электрический заряд. В последнем случае система представляет собой *плазму*. В плазме важную роль играют коллективные эффекты взаимодействия электрических зарядов отдельных частиц.

Плазменные системы могут быть самых разных масштабов – лабораторных, геофизических и астрономических. На лабораторном масштабе они возникают, например, в металлах, полупроводниках, ионизированных газах, ионизированных жидких структурах в таких установках как осцилляторы, плазменные усилители, преобразователи энергии, установки ядерного синтеза и др. На геофизическом масштабе плазма встречается в областях, окружающих земную атмосферу, – в ионосфере и экзосфере. На астрономическом масштабе с плазмой связаны многие эффекты, наблюдаемые в межпланетном и межзвездном пространстве.

Турбулентные эффекты наблюдаются на разных масштабах в нейтральных и плазменных системах. Турбулентность возникает в нейтральной жидкости, текущей в трубе или обтекающей некий объект, если скорость потока превышает некое критическое значение, определяемое размерами трубы или объекта и вязкостью жидкости. Наоборот, на некоторых скоростях, зависящих от размера объекта и вязкости жидкости, объект, перемещающийся в жидкости, вызывает турбулентные эффекты в попутном потоке. Инжекция пучка электронов в плазму ионизированного газа порождает условия турбулентности, зависящие от скорости и размера пучка и от плотности электронов. Воздействие мощным микроволновым излучением на плазму вызывает турбулентность, когда амплитуда микроволнового излучения превышает некоторый уровень. В присутствии статического магнитного поля плазма в туннельной камере, возбужденная электрическим током, резко проявляет турбулентные свойства, если магнитное поле увеличивается до критического значения.

Во всех приведенных выше примерах в силу нелинейности турбулентность возникает, когда некоторый физический параметр превышает некий критический уровень. Турбулентность может возникнуть внезапно или проявиться постепенно. Она может быть относительно однородной в пространстве или проявляться в виде локальных или периодических вспышек. Это разнообразие эффектов обусловлено проявлением многих нелинейных факторов, которые могут стать причинами возникновения турбулентности. Истоки турбулентности лежат в скрытых неустойчивостях, которые становятся существенными при определенном уровне возбуждения системы. Поскольку существует много типов неустойчивостей на микро- и макро-уровне, естественно ожидать, что существует много разных типов

турбулентности. Еще одной характеристической чертой, также проявляющейся в природе лежащих в основе турбулентности неустойчивостей, является тип перехода к турбулентному состоянию. Если переход осуществляется постепенно, то процесс возникновения турбулентности называется *мягким*. Когда же этот переход внезапен, то этот процесс называется *жестким*. Также следует отметить, что возникновение турбулентности может привести к определенным аномалиям в проявлении макроскопических транспортных свойств, таких как электропроводимость и теплопроводность.

Причины и формы плазменной турбулентности чрезвычайно разнообразны. Поэтому, хотя все примеры применения CPC-метода, приводимые ниже, связаны с плазменной турбулентностью, при описании вероятностно-статистических моделей турбулентности ограничимся турбулентностью гидродинамической, на примере которой можно довольно наглядно проиллюстрировать основные статистические закономерности, наблюдаемые в турбулентных системах. При этом становится возможным избежать подробного обсуждения задачи о взаимодействии многих частиц и соответствующих кинетических уравнений и выделить главные эффекты, присущие турбулентным системам, – перемежаемость и (локальную) масштабную инвариантность, – а также описать главные свойства соответствующих вероятностных моделей.

7.2. Феноменология турбулентности. Перемежаемость

А “веселые призраки” – это некая опасная турбулентция, имеющая место в некоторых районах Зоны

А. и Б. Стругацкие. *“Пикник на обочине”*.

В книге (Фриш, 1998) говорится: “По-видимому, в уравнении Навье–Стокса содержится вся [гидродинамическая] турбулентность... С другой стороны, мир турбулентных явлений почти так же разнообразен, как живая природа.” Полное понимание природы турбулентности пока не достигнуто, и в настоящее время его достижение представляет собой нерешенную проблему.

Общепринятого строгого определения турбулентного потока пока не существует. Турбулентные потоки принято характеризовать низкой импульсной диффузией, высокой импульсной конвекцией и быстрым измене-

нием давления и скорости во времени и в пространстве. Поток, не являющийся турбулентным, называется ламинарным. Безразмерная величина – число Рейнольдса

$$R = \frac{UL}{\nu}$$

– характеризует условия, при которых состояние потока тяготеет к турбулентному или ламинарному. Здесь U – характерная скорость, L – характерный линейный масштаб, ν – вязкость среды (жидкости). Рост числа Рейнольдса увеличивает турбулентный характер потока. Предельное состояние, соответствующее бесконечному числу Рейнольдса, называется полностью развитой турбулентностью.

Хотя основное уравнение гидродинамики – уравнение Навье–Стокса – известно с 1823 г. (Navier, 1823), его общее решение для больших чисел Рейнольдса, то есть для случая турбулентности, не известно. Для этого случая даже использование мощных компьютеров не позволяет найти точное решение. Природа турбулентности до конца не понята даже в идеальном изотропном и однородном случае. Нет всеобъемлющей теории, позволяющей с единых позиций описать многие эффекты, наблюдаемые в турбулентности. Как следствие, возникли многие феноменологические модели, ориентированные на описание конкретных аспектов турбулентной динамики. Большинство этих моделей можно классифицировать согласно тем физическим объектам, эффектам и явлениям, для описания которых они предназначены. Важнейшими из таких объектов и явлений являются поле скоростей и процесс диссипации энергии.

Статистический анализ разнообразных временных рядов выявил несколько универсальных эффектов в однородных и изотропных турбулентных потоках. В первую очередь, это эффект *перемежаемости*. Другие эффекты – универсальности и наличие скейлинговых (масштабных) соотношений для статистических закономерностей (см., например, (Meneveau and Sreenivasan, 1991), (Sreenivasan and Antonia, 1997) и ссылки в этих работах) – не связаны напрямую с тематикой данной книги и потому ниже будут обсуждаться лишь вскользь. Скейлинговые соотношения должны выполняться для полностью развитых турбулентных потоков и практически не наблюдаются для потоков с малыми или умеренными числами Рейнольдса. Перемежаемость наблюдается для намного более широкого диапазона чисел Рейнольдса, см., например, (Castaing et al., 1990), (Vincent and Meneguzzi, 1991), (Stolovitzki et al., 1992), (Hosokawa et al., 1994), (Zhu et al., 1995), (Barndorff-Nielsen, P. Blæsild and Schmiegell, 2004), (Barndorff-Nielsen and Schmiegell, 2005).

Примерно в 50-е годы прошлого столетия было замечено, что, в то время как распределение компонент скорости в однородных турбулентных полях в газах и жидкостях хорошо согласуется с нормальной моделью, разности и производные этих компонент как по времени, так и по пространственным координатам имеют распределения, заметно отличающиеся от нормальных. Обсуждение этого феномена и соответствующие ссылки на более ранние работы можно найти в (Batchelor, 1956). Как и финансовые данные, распределения характеристик, измеренных в турбулентных полях, демонстрируют точно такие же “отклонения от нормальности”, вытекающей из модели однородного случайного блуждания, что и распределения финансовых данных: (i) эмпирические плотности вероятности имеют более острые вершины и более тяжелые (почти экспоненциально убывающие) хвосты; (ii) отсутствует масштабная (по времени) инвариантность в том смысле, что распределения приращений процессов за интервалы времени большей длины намного ближе к нормальным, нежели распределения приращений процессов за интервалы меньшей длины. Островершинность распределений вероятностей в турбулентных полях объясняется наличием у регистрируемых в них характеристик свойства перемежаемости (см., например, (Фриш, 1998)). Справедливости ради следует отметить, что аналогия между стохастическими характеристиками турбулентных полей и финансовых рынков не является полной. В частности, в работе (Arneodo et al., 1996) отмечено принципиальное отличие корреляционных свойств процессов, регистрируемых в физических турбулентных системах и на финансовых рынках.

С вышеупомянутым эффектом не-нормальности распределений приращений скоростей, отмеченным в (Batchelor, 1956) и (Ghashghaie et al., 1996), хорошо согласуются измерения локальных характеристик турбулентности как при умеренных, так и при высоких значениях числа Рейнольдса. Более того, некоторые эксперименты показывают, что и распределения самих компонент скоростей могут отличаться от нормального.

Как отмечено в работе (Barndorff-Nielsen, 1979), к сожалению, стандартные методы построения аналитических аппроксимаций для таких распределений, основанные на разложениях Грама–Шарлье или Винера–Эрмита, приводят к чрезмерной сложности получаемых конструкций и дают лишь некоторое представление о сути явления.

Перемежаемость турбулентного поля скоростей представляет огромный интерес в исследованиях в области турбулентности (см., например, (Фриш, 1998), глава 8). Под перемежаемостью обычно понимают нерегулярное чередование участков, на которых закономерности развития про-

цесса сохраняются примерно неизменными. Суть перемежаемости турбулентного потока заключается в появлении и исчезновении (прохождении) когерентных структур различных размеров (вихрей) в поле скоростей. С теоретико-вероятностной и статистической точек зрения с перемежаемостью связано “не-гауссовское” поведение распределений вероятностей приращений скорости при уменьшении масштаба. Типичная картина такова. На больших временных масштабах это распределение хорошо аппроксимируется гауссовским, на умеренных (промежуточных) масштабах хвосты этого распределения превращаются в экспоненциальные, а на малых масштабах – в так называемые *растянутые* (stretched) экспоненциальные (Castaing et al., 1990), (Vincent and Meneguzzi, 1991).

В 1941 г. А. Н. Колмогоров сформулировал гипотезу универсальности: в пределе при неограниченно увеличивающихся числах Рейнольдса все статистические свойства на малых масштабах однозначно и универсально определяются масштабом и средней скоростью диссипации энергии (Колмогоров, 1941). Из этой гипотезы вытекает “закон 2/3” Колмогорова, согласно которому в полностью развитой турбулентности средний квадрат приращений скоростей изменяется пропорционально масштабу (длине интервала, за который берутся приращения) в степени 2/3. (Очень хороший обзор работ А. Н. Колмогорова по турбулентности с теоретико-вероятностных позиций дан А. Н. Ширяевым, см. (Shiryaev, 2006).)

Однако для моментов приращений скоростей более высоких порядков наблюдается более сложная (мультифрактальная) зависимость. Подробный обзор соответствующих результатов приведен в книге (Фриш, 1998). В частности, из результатов работ (Kolmogorov, 1962) и (Obukhov, 1962) вытекает, что, если универсальные статистические закономерности описываются логнормальным распределением, то для моментов n -го порядка приращений скоростей показатель степени при скейлинге равен

$$k_n = \frac{n}{3} - \mu \cdot \frac{n(n-3)}{18}, \quad n > 2, \quad 1/4 < \mu < 1/2.$$

Величину $-\frac{\mu}{18}n(n-3)$ в этом соотношении называют *поправкой перемежаемости*, поскольку из такого представления показателя степени при скейлинге вытекает, что при изменении масштаба распределение приращений скоростей неизбежно меняет тип от нормального на больших масштабах до островершинного и тяжелохвостого на малых масштабах. Отсюда, в частности, вытекает несостоятельность и неадекватность самоподобных моделей и, в частности, устойчивых процессов Леви в качестве моделей характеристик, наблюдаемых в турбулентных потоках.

Аппарат неоднородных случайных блужданий, порожденных обобщенными дважды стохастическими пуассоновскими процессами (обобщенными процессами Кокса) (см. раздел 1.3), оказывается очень удобным для описания эффекта перемежаемости на сверхмалых временных масштабах. Более того, обобщенные процессы Кокса позволяют напрямую связать наличие перемежаемости с отличием траекторий управляющих процессов от линейных функций за счет нестационарности интенсивностей потоков элементарных перемещений. Перемежаемость проявляется в непостоянстве волатильности, что, согласно сказанному в разделе 2.5, приводит к островершинности соответствующих распределений. Ниже будут приведены некоторые конкретные вероятностные модели, хорошо описывающие такие эффекты.

7.3. Параметрические смешанные гауссовские модели турбулентности

Возможность применения смесей нормальных законов к описанию распределения скоростей в турбулентной среде обсуждалась еще в работе (Dutton, 1970), где в качестве физического обоснования такого подхода использовался неоднородный, перемежающийся характер турбулентности. В данном разделе мы приведем примеры конкретных параметрических смешанных гауссовских моделей распределений скоростей в турбулентных потоках, использовавшихся в различных исследованиях.

ПРИМЕР 7.3.1. Нормальное\логнормальное распределение. Многие авторы рассматривали задачу моделирования распределения приращений компонент скоростей в турбулентных потоках с помощью смесей нормальных законов, см., например, (Barndorff-Nielsen, 1979), (Castaing et al., 1990). Основываясь на работах (Kolmogorov, 1962), (Obukhov, 1962), при этом некоторые исследователи предлагают в качестве распределения волатильности, то есть смешивающего распределения в соответствующих смешанных гауссовских моделях распределений характеристик турбулентных потоков использовать *логнормальное* распределение, см., например, (Castaing et al., 1990). В разделе 2.2 мы уже отмечали, что применительно к анализу динамики финансовых индексов аналогичные нормально\логнормальные смешанные модели рассматривались в работах (Clark, 1970), (Clark, 1973).

ПРИМЕР 7.3.2. Обобщенное гиперболическое распределение. В работе О. Барндорфф-Нильсена (Barndorff-Nielsen, 1979) сделана довольно убедительная попытка аппроксимировать распределения приращений компонент скоростей в турбулентных атмосферных потоках с помощью обобщенного гиперболического распределения, задаваемого плотностью

$$\begin{aligned}
 h(x; \alpha, \beta, \gamma_1, \gamma_2, \mu, \delta) = \\
 = C(\alpha, \beta, \gamma_1, \gamma_2, \delta) \exp \left\{ -\frac{1}{2} \left[\alpha \left(\sqrt{\delta^2 + (x - \mu)^2} - (x - \mu) \right)^{\gamma_1} + \right. \right. \\
 \left. \left. + \beta \left(\sqrt{\delta^2 + (x - \mu)^2} + (x - \mu) \right)^{\gamma_2} \right] \right\}, \quad (7.3.1) \\
 x \in \mathbb{R}, \quad \mu \in \mathbb{R}, \quad \alpha, \beta, \delta, \gamma_1, \gamma_2 > 0,
 \end{aligned}$$

$C(\alpha, \beta, \gamma_1, \gamma_2, \delta)$ – нормирующая постоянная. Частными случаями распределения (7.3.1) являются нормальное, NIG- (см. примеры 2.2.3 и 2.6.4) и гиперболическое (см. примеры 2.2.4, 2.6.2 и 2.6.3) распределения. “Предельными” случаями семейства (7.3.1) являются распределения Коши, Лапласа и так называемые экспоненциально-степенные распределения.

Подгонка распределения (7.3.1) к экспериментальным данным, связанным с изучением атмосферной турбулентности, демонстрирует удивительно высокое согласие (см. (Barndorff-Nielsen, 1979)). При этом накопленные данные позволили хорошо описать эмпирические распределения в пределах центральной области значений аргумента, ограниченной примерно двадцатью среднеквадратичными отклонениями. Тем не менее в качестве “теоретических” аргументов в пользу применения модели (7.3.1) для описания распределений характеристик турбулентности в упомянутой статье приводятся лишь соображения удобства, заключающиеся в гладкости, унимодальности, возможной несимметричности этой модели. “Физические” обоснования модели (7.3.1) отсутствуют.

Хорошее согласие модели (7.3.1) может быть объяснено большим количеством свободно подбираемых параметров. Кроме того, распределения семейства (7.3.1) допускают представление в виде специальных сдвиг/масштабных смесей нормальных законов. В соответствии со сказанным выше это обстоятельство является, пожалуй, самым серьезным объяснением адекватности модели (7.7.1) для описания распределений вероятностей, возникающих в хаотических процессах. С другой стороны, физическая интерпретация ее параметров неочевидна и весьма затруднена, что в определенном смысле снижает ее теоретическое значение.

Результаты статьи (Barndorff-Nielsen, 1979) получили продолжение в работах (Barndorff-Nielsen, Jensen and Sørensen, 1990) и (Barndorff-Nielsen,

P. Blæsild and Schmiegel, 2004). В первой из этих работ с целью изучения эффекта долговременной памяти процессов, регистрируемых в турбулентной среде, построена модель авторегрессионного типа со стохастическими членами, имеющими распределения с лог-линейными хвостами (в частности, лапласовское и обобщенное логистическое).

ПРИМЕР 7.3.3. Нормальное\\обратное гауссовское распределение. В работах (Barndorff-Nielsen, P. Blæsild and Schmiegel, 2004) и (Barndorff-Nielsen and Schmiegel, 2006) показано, что практически на всех амплитудах и масштабах распределения вероятностей приращений скорости в турбулентных потоках очень хорошо описываются единым классом распределений – нормальным\\обратным гауссовским (NIG-) семейством, допускающим довольно удобную интерпретацию. Поскольку в разделе 2.6.3 уже довольно подробно говорилось о NIG-распределениях и процессах Леви, здесь на свойствах NIG-моделей турбулентности останавливаться не будем.

Последующий анализ поведения параметров NIG-распределений, статистически подгоняемых к самым разным временным рядам, полученным в результате наблюдения за турбулентными потоками с самыми разными числами Рейнольдса, позволил авторам статьи (Barndorff-Nielsen, P. Blæsild and Schmiegel, 2004) сформулировать “закон универсальности”: развитие турбулентного поля скоростей во времени подчиняется “внутренним” часам, которые зависят от условий эксперимента, но в терминах которых одномерные маргинальные распределения приращений скоростей становятся не зависящими от условий эксперимента. По сути, этот закон является проявлением и подтверждением адекватности вероятностных моделей динамики основных параметров турбулентных потоков, имеющих вид специальных суперпозиций процесса броуновского движения и субординатора – внутреннего хронометра (случайного времени). Субординатор определяется факторами, являющимися внешними по отношению к той системе, поведение которой описывается броуновским движением. При этом внешние факторы зависят от условий эксперимента. При фиксированном субординаторе условные распределения таких моделей являются гауссовскими и от условий эксперимента тем самым не зависят.

ПРИМЕР 7.3.4. Дробно-устойчивые распределения. Как уже говорилось в разделе 1.2, для описания турбулентных потоков в статьях (Shlesinger, West and Klafter, 1987), (Klafter, Blumen and Shlesinger, 1987), (Batanov et al., 1998) использовалась схема однородных случайных блужданий с непрерывным временем. При этом, как следствие, в качестве модели процессов, регистрируемых в турбулентных средах, возникают про-

цессы дробной аномальной диффузии вида суперпозиции $Z(t) = W(M(t))$ стандартного винеровского процесса и независимого от него процесса *обратного случайного времени*

$$M(t) = \min\{\tau : T(\tau) \leq t\}, \quad t > 0,$$

где $T(\tau)$ – процесс Леви с неубывающими траекториями такой, что $T(0) = 0$ и $T(1)$ имеет одностороннее устойчивое распределение $G(x; \beta)$ с параметром $\beta \in (0, 1)$. Конечномерные распределения таких процессов описаны в теореме 1.2.1 и имеют вид масштабных смесей нормальных законов, в которых смешивающие распределения – это “обратные” устойчивые законы.

Процессы дробной аномальной субдиффузии и присущие им дробно-устойчивые распределения вероятностей в некоторых случаях оказываются довольно подходящими для описания локальных флуктуаций частиц в магнитоактивной плазме (Pedrosa et al., 1996), (Connor et al., 1999), (Batanov et al., 1998).

Процессы дробной аномальной субдиффузии соответствуют наглядной физической модели, согласно которой блуждающая частица время от времени попадает в ловушку, где задерживается на какое-то время, причем перескок из ловушки в ловушку происходит мгновенно.

В экспериментах с плазменной турбулентностью для некоторых временных рядов локальных потоков в установке Л-2М¹ наблюдалось хорошее согласие эмпирических плотностей приращений с модельными дробно-устойчивыми законами. В качестве примера приведем аппроксимацию двух локальных потоков в последовательных разрядах, когда поверхность, на которой происходят измерения, сдвигается на 2 мм по радиусу (рис. ??). Для разряда 46 805 блуждание частицы описывается зависимостью $x \propto t^{0.38}$. Для разряда 46 804 блуждание частицы описывается зависимостью $x \propto t^{0.042}$. По критерию хи-квадрат гипотеза о согласии эмпирических плотностей приращений с соответствующими дробно-устойчивыми законами принимается. Однако разница в законах блуждания частицы в этих разрядах очень велика. В первом разряде частица движется лишь несколько медленнее броуновской, а в следующем разряде частица практически не смещается. Смещение частиц в разрядах 46 804 и 46 805 за первые 10 временных шагов показано на рис. ??, на этом же рисунке представлено смещение броуновской частицы.

Такие противоречивые результаты по смещению частицы при моделировании локального потока процессом дробной аномальной субдиффу-

¹Характеристики этой и других экспериментальных установок приведены в (Батанов, Бенинг, Королев и др., 2002).

зии (см. раздел 1.2) с дробно-устойчивыми распределениями приращений, указывают на необходимость добавления каких-то дополнительных условий в такое моделирование. Заметим, что для отдельных локальных потоков промоделировать плотность распределения вероятности приращений дробно-устойчивым распределением вообще не удалось (оцененные значения параметров выпали за пределы области допустимых значений). Вероятно, это связано не только с невозможностью мгновенных перескоков частиц между ловушками, но и тем, что ансамбль частиц плазмы нельзя считать однородным.

ПРИМЕР 7.3.5. В разделе 2.6 уже упоминалось, что в работе (Holdom, 1998) идея “турбулентной” перемежаемости используется для моделирования финансовых данных. При этом в качестве распределения волатильности используется распределение с плотностью

$$g(x; \theta) = \frac{2\theta^\theta}{\Gamma(\theta)} x^{2\theta-1} e^{-\theta x^2}, \quad x \geq 0.$$

В частности, при $\theta = \frac{1}{2}$ плотность $g(x; \theta) = g(x; \frac{1}{2})$ имеет вид

$$g(x; \frac{1}{2}) = \frac{\sqrt{2}}{\sqrt{\pi}} e^{-x^2/2}, \quad x \geq 0,$$

и совпадает с плотностью вероятностей максимума стандартного винеровского процесса на отрезке $[0, 1]$. Соответствующее смешанное гауссовское распределение задается плотностью вида

$$\begin{aligned} p(x; \theta) &= \int_0^\infty \frac{g(s; \theta)}{s} \phi\left(\frac{x}{s}\right) ds = \\ &= \frac{2^{(3-2\theta)/4} \theta^{(1+2\theta)/4}}{\sqrt{\pi} \Gamma(\theta)} |x|^{\theta-1/2} K_{\theta-1/2}(\sqrt{2\theta}|x|), \quad x \in \mathbb{R}, \end{aligned} \quad (7.3.2)$$

где $K_\nu(x)$ – модифицированная функция Бесселя второго рода. При $\theta = 1$ плотность $p(x; \theta)$ задает двойное показательное распределение (распределение Лапласа)

$$p(x; 1) = \frac{\sqrt{2}}{2} \exp\{-\sqrt{2}|x|\}, \quad x \in \mathbb{R}.$$

Семейство распределений типа (7.3.2) обладает свойством замкнутости относительно операции свертки. В работе (Holdom, 1998) отмечено, что с помощью распределения (7.3.2) можно хорошо промоделировать следующий

эффект, наблюдаемый как в турбулентных потоках, так и при анализе финансовых данных: хотя наблюдаемая автокорреляция самих изменений логарифмов биржевых цен пренебрежимо мала, автокорреляция некоторых функций от изменений логарифмов биржевых цен довольно заметна.

ПРИМЕР 7.3.6. Распределение Лапласа. Распределение Лапласа оказалось довольно адекватной и информативной моделью для описания статистических закономерностей, которым удовлетворяют приращения дрейфового потока в плазме линейной установки ТАУ-1 (Королев и Скворцова, 2003). Как уже неоднократно отмечалось, распределение Лапласа (с точностью до параметра масштаба) может быть представлено в виде масштабной смеси гауссовских законов, где в качестве смешивающего распределения выступает экспоненциальное (показательное) распределение. Такое распределение возникает в качестве предельного для неоднородных случайных блужданий с непрерывным временем, описываемых обобщенными дважды стохастическими пуассоновскими процессами (обобщенными процессами Кокса), в которых накопленная интенсивность имеет, например, геометрическое или экспоненциальное распределение. Отсюда вытекает предположение о том, что каждое приращение является результатом классической (броуновской) диффузии из точки в точку со своим коэффициентом диффузии. Это означает, что случайный процесс, описывающий поведение приращений локального потока во времени, является диффузионным, причем на последовательных интервалах оцифровки коэффициент диффузии изменяется случайным образом. Таким образом, в данном диффузионном процессе, описывающем поведение приращений потока, коэффициенты диффузии – случайные величины с экспоненциальным распределением. В этой связи можно заметить, что распределение Гаусса обладает наибольшей энтропией среди всех законов с конечным вторым моментом и соответствует устойчивым состояниям в замкнутых системах. Экспоненциальное распределение обладает наибольшей энтропией среди всех законов с конечным первым моментом, сосредоточенных на неотрицательной полуоси, и также соответствует устойчивым состояниям. Любая плазменная установка является открытой системой, динамическое равновесие по макропараметрам плазмы в которой существенно превышает характерные времена флуктуаций. Отсюда следует, что если приращения потока имеют распределение Лапласа, то выборка соответствует максимально неопределенным (“самым случайным”) условиям проведения эксперимента. Тогда измеренное распределение для приращений локального потока в некотором приближении в ТАУ-1 соответствует наиболее устойчивому стационарному процессу для открытой плазменной

системы. Гистограмма приращений сводится к распределению Лапласа (распределению с максимальной энтропией) только для коротких выборок локального потока при временах оцифровки сигнала менее 1–5 мкс. Время оцифровки (1–5 мкс) совпадает со временем автокорреляции приращений, назовем его характеристическим для локального дрейфового потока. Для дрейфовых движений, раскачиваемых дрейфово-диссипативной неустойчивостью, в условиях эксперимента на ТАУ-1 были измерены периоды колебаний, линейные инкременты, времена выхода на насыщение интенсивности, автокорреляционные времена турбулентных пульсаций и т. д. Оказалось, что характеристическое время изменения дрейфового потока частиц соответствует времени срыва дрейфового волнового пакета. При удлинении выборки или при увеличении времени оцифровки форма распределения вероятностей приращений начинает отличаться от лапласовской существенно, критерии согласия с моделью уменьшаются, а далее она просто отвергается. Заметим, что любые измерения стохастических плазменных сигналов в реальном физическом эксперименте сопровождаются множеством наводок разного уровня: тепловые шумы детекторов, шумы усилителей и АЦП, шумы источников питания компьютеров и т. п. Как правило, временные выборки приращений шумовых сигналов (наводок) являются нормально распределенными. При удлинении временной выборки относительный уровень таких процессов падает по сравнению с анализируемым сигналом. Короткие временные выборки приращений потоков вместе с шумами – это практически бесконечный набор приращений совокупности случайных процессов с непрерывным масштабом времени (плазменных сигналов и шумов), именно такие выборки можно промоделировать распределением Лапласа (процессом с максимальной энтропией для открытых физических систем).

Задачу об определении “характерного времени” или “характерного масштабного времени” процесса с помощью подгонки распределения Лапласа к временному ряду можно сформулировать в терминах дисперсионных гамма-процессов, обсуждавшихся в разделе 2.6.3 (см. пример 2.6.5). Договоримся под процессом Лапласа понимать процесс Леви $Z(t)$ такой, что $Z(1)$ имеет распределение Лапласа (двойное экспоненциальное). Это частный случай дисперсионного гамма-процесса (VG-процесса) с параметром сдвига (и вызванной им асимметрии распределения) $\theta = 0$ и масштабным параметром ν . Вопрос о выборе “характерного времени” сводится тем самым к определению такого временного масштаба t , при котором гамма-распределение с параметром формы t/ν превращается в экспоненциальное, что по определению VG-процесса (см. пример 2.6.5) может быть только то-

гда, когда $t = \nu$, то есть когда параметр масштаба ν гамма-распределения равен временному масштабу t .

Представление процесса диффузии частиц в структурной турбулентности на малых интервалах непрерывным процессом Лапласа оказалось плодотворным. Во-первых, не отвергнутый в коротких выборках процесс Лапласа позволил впервые определить характеристическое время диффузии частиц, самое короткое время структурной турбулентности. Например, для дрейфового потока – это время срыва дрейфово-диссипативной неустойчивости. Во-вторых, известно, что около 3% всех величин, подчиняющихся распределению Лапласа, выходят за рамки трех среднеквадратичных отклонений. Таким образом, можно оценить, что не менее одного из каждых тридцати приращений в локальном потоке частиц в структурной турбулентности являются редким (аномально быстрым) событием, что на порядок выше, чем частота появления редких событий в нормальном процессе (это нижняя оценка частоты появления редких событий). Корреляционный и спектральный анализ величин локального потока на основе равноотстоящей выборки временных приращений амплитуд потока дает возможность определить характеристическое («динамическое») время процесса локального потока в плазме Л-2М и ТАУ-1; в обоих случаях это время оказалось на порядок меньше корреляционного времени пульсаций в структурной турбулентности.

Как будет видно ниже, поведение турбулентной плазмы все же отличается большей стохастичностью, нежели поведение финансовых индексов. Здесь также удастся выделить локальные тенденции, однако, в отличие от финансовых индексов, поведение самих этих локальных тенденций более стохастично (что, в частности, проявляется в том, что диффузионная компонента волатильности превалирует над динамической). Поэтому в целом идея поиска разумной физически обоснованной параметрической аналитической модели для описания возникающих распределений вероятностей здесь имеет больший смысл, нежели в исследовании динамики финансовых индексов. Тем не менее чрезвычайная трудоемкость статистического оценивания параметров моделей типа логнормально/нормальной, обобщенной гиперболической (7.3.1) или модели (7.3.2) вкупе с неочевидной «физической» интерпретируемостью параметров этих моделей делает весьма неэффективной идею использования упомянутых параметрических моделей в их «скользящем» варианте с целью отслеживания зависимости ее параметров от времени.

7.4. “Статическая” подгонка смесей нормальных законов к эмпирическим распределениям, полученным в экспериментах с турбулентной плазмой

В данном разделе будут рассмотрены примеры подгонки конечных смесей нормальных законов к экспериментальным данным. Сначала приведем результаты исследования временного ряда флуктуаций плотности плазмы в стеллараторе LHD².

Длина стационарной выборки флуктуаций плотности плазмы, которые измерялись в области СВЧ нагрева, не превышала $5 \cdot 10^3$ точек (10 мс). Исследование флуктуаций в LHD проводилось в самой горячей плазме в области ввода мощного электромагнитного излучения (до нескольких мегаватт); в таких условиях временные ряды турбулентных пульсаций являются наиболее неравновесными (неоднородными) (Батанов, Колик и др., 2003).

В ряде разностей первого порядка выделены два отрезка длиной 1600 и 2000 точек: средний спокойный интервал (наблюдения с 9400 по 11000) и правая часть, с большим разбросом (наблюдения с 14000 по 16000), см. рис. ??.

При расщеплении смеси на четыре компоненты по первому отрезку временного ряда к двум основным структурам добавляются незначительные (веса не более 0.05) компоненты, выявляющие два состояния, отличные от основных. Первое из них (среднеквадратичное отклонение порядка 0.0016) характеризуется самым коротким временем (в плазме такое время называется инкрементом развития неустойчивости). Другими словами, вклад неустойчивости с малым временем в суммарную турбулентность не превышает 3%. Последняя (большое по абсолютному значению математическое ожидание четвертой компоненты) может быть связана с наводками на магнитную систему установки (т.е. является “паразитной”). Параметры этой четырехкомпонентной смеси таковы:

²В настоящее время сверхпроводящий стелларатор LHD является самой современной термоядерной установкой стеллараторного класса с диаметром плазмы около 150 см, электронной плотностью $n \sim 10^{13} \div 10^{14} \text{ см}^{-3}$ и температурой электронов $T_e \sim 10^3 \div 10^4 \text{ эВ}$ (Motojima et al., 2000).

$$\begin{aligned}
p_1 &= 0.0260, & a_1 &= -0.0004, & \sigma_1 &= 0.0016, \\
p_2 &= 0.3236, & a_2 &= 0.0004, & \sigma_2 &= 0.0060 & \text{(основная компонента)}, \\
p_3 &= 0.6062, & a_3 &= 0, & \sigma_3 &= 0.0112 & \text{(основная компонента)}, \\
p_4 &= 0.0442, & a_4 &= -0.0027, & \sigma_4 &= 0.0117.
\end{aligned}$$

Согласие с экспериментальными данными очень хорошее (P -значение теста Колмогорова–Смирнова равно 0.9938).

Заметим, что отвергать нормальность в этом случае нет оснований (P -значение равно 0.1412, параметры нормального распределения при этом следующие: $a = 0$, $\sigma^2 = 0.0097$). Но даже визуально из рис. ?? и ?? понятно, что полученные с помощью ЕМ-алгоритма смеси лучше описывают островершинность гистограммы экспериментальных данных.

Флуктуации с большим разбросом амплитуды (ряд приращений с 14000 по 16000 наблюдение) не являются нормальными (P -значение теста Колмогорова–Смирнова равно 0.0017, $a = -0.0001$, $\sigma^2 = 0.0294$). Для этого ряда распределение приращений амплитуд промоделировано смесью из трех нормальных компонент, каждая из которых существенна:

$$\begin{aligned}
p_1 &= 0.5731, & a_1 &= -0.0006, & \sigma_1 &= 0.0355, \\
p_2 &= 0.3072, & a_2 &= 0.0006, & \sigma_2 &= 0.0134, \\
p_3 &= 0.1197, & a_3 &= 0.0003, & \sigma_3 &= 0.0291.
\end{aligned}$$

P -значение теста Колмогорова–Смирнова равно 0.9825 (комментарии по поводу правомерности применения критерия Колмогорова–Смирнова в подобной ситуации приведены в разделе 5.3.3).

Как видно из графиков, приведенных на рис. ?? и ??, подогнанные смеси демонстрируют высокое согласие с реальными данными и могут дать много информации о том, в каком состоянии находится плазма. В то же время приведенные примеры показывают, что нет “универсальной” четырехкомпонентной смеси и, более того, не видно даже “шаблона” или общего правила, по которому смешиваются нормальные законы, обеспечивающие наилучшее согласие, хотя оба примера относятся к одному и тому же временному ряду. Отметим, что для данного примера выбран самый сложный для анализа с помощью ЕМ-алгоритма временной ряд турбулентных пульсаций. Результаты анализа иных выборок приведены в (Скворцова, Королев и др., 2005).

В физике плазмы существует еще одна возможность исследования взаимодействия ансамблей частиц в плазменной турбулентности – это прямые измерения. Такие измерения можно проводить в низкотемпературной

плазме. В частности, измерения локальных потоков проводились в краевой плазме стелларатора Л-2М и в линейной установке ТАУ-1. Значения локального потока во времени представляет собой временной ряд, элементы которого не являются некоррелированными, в отличие от элементов временного ряда его приращений. Поэтому и в случае анализа локальных потоков корректно использовать ряд разностей первого порядка.

Длина временного ряда локального потока, измеренного в краевой плазме стелларатора Л-2М, составила около 5000 точек (5 мс).

Согласно статистическим тестам, приращения краевого потока (размер анализируемой выборки – 3000 точек с 1000 по 4000 наблюдение) не являются ни выборкой из нормального распределения, ни выборкой из двухкомпонентной смеси нормальных распределений. Однако уже с 97% достоверностью можно утверждать, что в краевом потоке Л-2М присутствуют как минимум три различные структуры. Параметры нормальных компонент смеси и их веса, полученные с помощью ЕМ-алгоритма, таковы:

$$\begin{aligned} p_1 &= 0.3955, & a_1 &= 0.1416, & \sigma_1 &= 1.8930, \\ p_2 &= 0.5171, & a_2 &= -0.3335, & \sigma_2 &= 6.0149, \\ p_3 &= 0.0874, & a_3 &= 1.3239, & \sigma_3 &= 19.7521. \end{aligned}$$

При этом P -значение теста Колмогорова-Смирнова равно 0.9747. Отметим, что третья компонента, несмотря на маленький вес (0.087), все равно вносит существенный вклад в смесь, например, в дисперсию смеси (ее среднеквадратичное отклонение намного больше, чем у других компонент) и математическое ожидание (оно также намного больше по абсолютному значению, чем у остальных компонент).

Наконец, рассмотрим временную выборку локального потока частиц в низкотемпературной плазме самой маленькой линейной плазменной установки ТАУ-1, расположенной в ИОФ РАН им. А. М. Прохорова в Москве, которая наиболее сильно отличается от гауссовой.

Даже на коротком отрезке времени (0.5 мс, 500 точек) данные не являются нормальными (P -значение равно лишь 0.02). Немного лучше (с P -значением, равным 0.4) островершинная плотность приращений процесса на коротком интервале аппроксимируется распределением Лапласа. Однако с гораздо большей достоверностью соответствующее эмпирическое распределение согласуется с плотностью смеси четырех нормальных законов. С достоверностью 99,75% (по критерию Колмогорова-Смирнова) выборке приращений потока частиц с 20 000 по 20 500 наблюдение соответствует модель смеси со следующими параметрами:

$$\begin{aligned}
p_1 &= 0.2036, & a_1 &= 0.6844, & \sigma_1 &= 1.2284, \\
p_2 &= 0.0483, & a_2 &= 5.9305, & \sigma_2 &= 0.6541, \\
p_3 &= 0.2192, & a_3 &= -0.7849, & \sigma_3 &= 3.0276, \\
p_4 &= 0.5289, & a_4 &= -0.4877, & \sigma_4 &= 6.6860.
\end{aligned}$$

В заключение приведем параметры смеси нормальных распределений, которая с достоверностью 99% соответствует ряду приращений длиной в 20 000 точек:

$$\begin{aligned}
p_1 &= 0.0693, & a_1 &= 0.0588, & \sigma_1 &= 0.8016, \\
p_2 &= 0.2718, & a_2 &= -0.2217, & \sigma_2 &= 2.1183, \\
p_3 &= 0.4742, & a_3 &= 0.0146, & \sigma_3 &= 4.3499, \\
p_4 &= 0.1847, & a_4 &= 0.2656, & \sigma_4 &= 8.6550.
\end{aligned}$$

Неоднородные (во времени) случайные блуждания с непрерывным временем, описываемые обобщенными процессами Кокса, оказались удобными и адекватными феноменологическими математическими моделями процессов плазменной турбулентности на микроуровне. Для анализа физических процессов, протекающих в плазменной турбулентности, возможность разделения смеси на конечное число характерных процессов оказалась очень плодотворной. Например, оказалось, что в высокотемпературной и низкотемпературной плазме в центре и на краю шнура в стеллараторе Л-2М турбулентность описывается сдвиг/масштабными смесями нормальных законов с одними и теми же количествами компонент смеси. Этот факт подтверждает единую природу турбулентных пульсаций, связанную с существованием ансамблей взаимодействующих стохастических структур, по всему объему замагниченной плазмы в тороидальной ловушке. В плазме стелларатора TJ-II удалось определить, что турбулентность горячей плазмы определяется развитием единственной неустойчивости дрейфово-диссипативной природы (Skvortsova, Batanov, Kolik et al., 2005). Это особенно важно для анализа плазменной турбулентности, потому что теоретически в плазме возможно развитие множества низкочастотных (лежащих в одних и тех же интервалах частот) неустойчивостей разного типа (см., например, (Михайловский, 1971), (Михайловский, 1978)). Поскольку каждый тип динамических структур плазменной турбулентности характеризуется своим коэффициентом диффузии, то успешное разделение смеси на конечное число компонент означает, что стохастический характер наблюдаемой плазменной турбулентности обусловлен конечным числом динамических структур (конечным числом основных взаимодействующих процессов). Это было подтверждено моделированием локального потока частиц в низкочастотной турбулентности также конечными

смесями нормальных законов. На установке ТАУ-1 удалось сопоставить смесь четырех нормальных распределений, которой моделируется временная выборка приращений локального потока, четырем плазменным процессам: стохастическому выносу частиц в нелинейной дрейфовой волне и солитоне, блужданию частиц в поле дрейфовых волновых пакетов и ионно-звуковых солитонов, а также их блужданию при нелинейном взаимодействии солитонов в ионно-звуковой турбулентности (Скворцова, Королев и др., 2005). Представление плазменной турбулентности как неоднородного случайного блуждания с непрерывным временем позволяет обосновать новую физическую модель перемежаемости турбулентных пульсаций без вариации локального градиента плотности и температуры плазмы (см., например, (Каррерас и др., 1996)). Сумма влияний нормальных процесса, которые задают ансамбли инкрементов и декрементов в турбулентности (управляющего дискретного процесса), смогут приводить к тяжелым хвостам в распределениях вероятности их суммарного во времени общего инкремента, что совершенно естественно. Тогда перемежаемость турбулентности может существовать в стационарной фазе разряда плазмы с неизменными, так называемыми каноническими профилями плотности и температуры (см., например, (Днестровский и др., 2002)).

7.5. Результаты применения метода скользящего разделения смесей

Этот мир я сравнил бы с простым фонарем,
Солнце – с свечкой, пылающей жарким огнем.
Мы блуждаем, как тени, в загадочном мире,
Ничего достоверно не зная о нем.

Омар Хайям.

В данном разделе будут приведены некоторые (фактически, *первые*) результаты применения метода скользящего разделения смесей к статистическому анализу временных рядов локального потока в краевой низко-температурной плазме двух тороидальных установок: стелларатора Л-2М и токамака ФТ-2. Отметим, что в токамаке ФТ-2, в отличие от Л-2М, плазма может существовать в двух режимах: Н-режим (High) отличается от L-режима (Low) существенно улучшенными характеристиками удержания частиц в ловушке, которое сопровождается увеличением плотности и температуры плазмы с одновременным уменьшением турбулентного потока частиц на стенку камеры (Lashkul et al., 2005). Главное внимание будет

уделено измерению потоков в ФТ-2. Эти примеры очень актуальны: совсем недавно принято решение о постройке полномасштабного токамака ITER во Франции.

Главная идея, лежащая в основании используемого подхода, такова. В каждый момент времени состояние плазмы есть результат взаимодействия некоторого числа типов структур (вихрей, потоков и т. п.) Каждый тип структур характеризуется, во-первых, своими локальными скоростями и, во-вторых, своей интенсивностью флуктуаций (диффузией). Более того, возможно появление агрегированных структур, влияние которых изменяется во времени. Плотность распределения приращений турбулентного потока в L-режиме моделировалась смесью со всего двумя компонентами. При переходе в H-режим число компонент смеси возрастало до трех (с достоверностью 98–99% для временных рядов длиной 2000 точек) (Skvortsova, Korolev, Batanov et al., 2006).

Уже отмечалось, что уравнения, описывающие плазменную турбулентность, так или иначе включают значение коэффициента диффузии. Среди специалистов по физике плазмы стало расхожим мнение, что при практических расчетах значение коэффициента диффузии выписывается в значительной степени “от лукавого”. Поэтому результаты численного моделирования иногда сильно отличны от реальности. Отмеченная неустойчивость обусловлена тем, что существующие модели слишком упрощенно и потому грубо интерпретируют изменчивость рассматриваемых стохастических хаотических процессов как некий постоянный скаляр – число, которое наделяется смыслом “коэффициента диффузии”. Как уже было видно на примере анализа финансовых данных, в реальных хаотических процессах изменчивость, во-первых, имеет существенно многомерную структуру и, во-вторых, изменчива сама. Как было показано, в не-гауссовых моделях изменчивость превращается в довольно сложное понятие, тем не менее поддающееся анализу и интерпретации с помощью моделей типа конечных смесей гауссовских законов. Не стремясь сразу предложить альтернативные уравнения, описывающие поведение турбулентной плазмы, мы лишь представим результаты применения СРС-метода. Эти результаты весьма убедительно свидетельствуют против упрощенного толкования изменчивости и, прежде всего, диффузии в турбулентной плазме.

Предъявим результаты практического разложения волатильности рассматриваемых процессов на диффузионную и динамическую компоненты.

Особо следует остановиться на том, что динамическая компонента волатильности (точнее, ее элементы в представлении распределений приращений рассматриваемого процесса) позволяет явно определить (вычис-

лить) распределение (по ансамблю) средних (по времени) скоростей, присущих структурам, составляющим анализируемую систему. Это распределение задается набором пар $\{(a_i, p_i), i = 1, \dots, k\}$, где a_i – математическое ожидание i -й компоненты смеси, имеющее смысл среднего (для данного диапазона времени, определяемого длиной интервала времени между последовательными наблюдениями (отсчетами, измерениями)) значения скорости, p_i – доля элементов, чьи скорости примерно равны a_i , в рассматриваемой системе, k – число явно выделяемых типов структур, соответствующих компонентам подогнанной смеси. При этом теорема о свертке смесей позволяет пересчитывать распределение средних скоростей на разных временных масштабах. Как видно из рис. ??, CPC-метод позволяет наглядно выделить ненулевые тенденции развития компонент системы (процесса), хотя, казалось бы, “генеральная” тенденция отсутствует (среднее значение приращений процесса равно нулю). Именно это обстоятельство обусловило выбор эпиграфа к данной книге.

На рис. ?? изображены диффузионные портреты турбулентной плазмы в токамаке. По горизонтальной оси – время (отсчеты), а по вертикальной – значения коэффициентов диффузии компонент смеси нормальных законов, подогнанной к приращениям рассматриваемого процесса. Верхняя картинка содержит больше деталей, но нижняя, более гладкая, дает больше информации об изменениях закономерностей поведения диффузионных компонент волатильности турбулентной плазмы.

Рис. ?? представляет портрет локальных тенденций (распределения средних скоростей) в плазме Л-2М. Обращает на себя внимание отсутствие довольно долгих и гладких локальных тенденций, которые присущи финансовым данным (см. (Korolev and Rey, 2005)). Имеющиеся тенденции визуально более хаотичны сами по себе. В целом картина напоминает поверхность кипящей жидкости.

На рис. ?? представлены графики динамической и диффузионной компонент волатильности турбулентного потока в плазме Л-2М. Видно, что диффузионная компонента в целом превалирует, однако сравнима по величине с динамической компонентой, которая иногда даже превышает диффузионную. Динамическую компоненту можно отождествить с той долей изменчивости, которая обусловлена наличием конвективного переноса, в то время как диффузионная компонента характеризует стохастическую составляющую изменчивости. Это довольно интересный результат. К настоящему времени обнаружено баллистическое (то есть с постоянной скоростью) движение частиц примеси от стенки камеры к центру плазменного шнура (то есть против градиента плотности), полученные в эксперимен-

тах по распространению примеси азота и на токамаке JET (Parail, 2001), и на стеллараторе TJ-II (van Milligen et al., 2002). Данные исследования теснейшим образом связаны с задачами защиты стенки термоядерного реактора от потоков быстрых электронов, из-за которых происходит очень быстрая эрозия стенки. Напуск нейтрального газа за границей плазмы приводит, с одной стороны, к уменьшению потока быстрых электронов, что полезно для защиты стенки камеры. С другой стороны, такой напуск приводит к быстрому (баллистическому) охлаждению центральной области плазменного шнура, что приводит к срыву разряда. (См., например, доклады на XX IAEA (МАГАТЭ) конференции, посвященные таким экспериментам на установках DIII-D (Hollman, 2004) и JT-60U (Bakhtiari, 2004)). Если же конвективные движения частиц в краевой турбулентности превышают диффузионный перенос, то такие методы защиты стенки камеры в будущих больших термоядерных установках (ITER) нуждаются в дополнительных исследованиях.

Разница в поведении приращений потоков в краевой плазме токамака ФТ-2 в L- и H-режимах заметна по рис. ?? и ?. Как видно из рис. ?, распределение приращений амплитуд в L-режиме может быть приближено смесью двух или трех гауссовых распределений.

Распределение приращений амплитуд в H-режиме также может быть приближено смесью двух или трех гауссовых распределений, см. рис. ?. Более того, диффузионные компоненты оказываются более разнесенными и более гладкими, нежели в L-режиме. Эти примеры показывают, что в физике турбулентной плазмы одномерная (скалярная) интерпретация диффузии действительно может привести к неверным представлениям.

Так как диффузия или волатильность тесно связана с выбором масштаба времени, для лучшего понимания структуры плазменной турбулентности очень важно уметь пересчитывать диффузионные портреты, полученные для одних временных масштабов, на другие (большие или меньшие) масштабы. СРС-метод дает такую теоретическую возможность (см. выше). Однако эта теория будет работать, только если соответствующие распределения вероятностей действительно имеют вид конечных смесей нормальных законов. Схожий характер диффузионных портретов прореженного процесса и свертки исходного процесса может свидетельствовать в пользу справедливости последнего условия.

На рис. ?? показаны диффузионные портреты прореженного процесса и свертки исходного процесса (H-режим, ширина окна 300). Заметна аналогия, однако имеются и отличия. Правда, отличия могут быть вызваны не совсем правильным выбором ширины окна СРС-метода.

Список литературы

1. М. Абрамовиц и И. М. Стиган (ред.) *Справочник по специальным функциям с формулами, графиками и таблицами*. “Наука”, Москва, 1979.
2. С. А. Айвазян, И. С. Енюков и Л. Д. Мешалкин. *Прикладная статистика. Основы моделирования и первичная обработка данных*. “Финансы и статистика”, Москва, 1983.
3. С. А. Айвазян, В. М. Бухштабер, И. С. Енюков и Л. Д. Мешалкин. *Прикладная статистика. Классификация и снижение размерности*. “Финансы и статистика”, Москва, 1989.
4. Х. Акаике. Развитие статистических методов. – В кн.: П. Эйкхофф (ред.) *Современные методы идентификации систем*. “Мир”, Москва, 1983, с. 148–176.
5. П. С. Александров (ред.) *Англо-русский словарь математических терминов*. “Мир”, Москва, 1994.
6. В. С. Анищенко, Т. Е. Вадивасова и В. В. Астахов. *Нелинейная динамика хаотических и стохастических систем*. Издательство Саратовского университета, Саратов, 1999, 368 с.
7. В. С. Анищенко, В. В. Астахов, Т. Е. Вадивасова, А. Б. Нейман, Г. И. Стрелкова и Л. Шиманский-Гаер. *Нелинейные эффекты в хаотических системах*. Институт компьютерных исследований, Москва–Ижевск, 2003, 544 с.
8. С. А. Ашманов и А. В. Тимохов. *Теория оптимизации в задачах и упражнениях*. “Наука”, Москва, 1991.
9. С. Н. Антонов и С. Н. Кокшаров. Об асимптотическом поведении хвостов масштабных смесей нормальных распределений. – В сб.: *Статистические методы оценивания и проверки гипотез*. Изд-во Пермского университета. Пермь, 2006, с. 90–105.

10. Э. Б. Багиров. *Метод смесей и его применение к выводу нижних оценок для распределений функций от нормальных случайных величин*. Диссертация на соискание ученой степени кандидата физ.-матем. наук, Математический институт АН СССР, Москва, 1988.
11. Г. М. Батанов, В. Е. Бенинг, В. Ю. Королев и др. Об одном подходе к вероятностно-статистическому анализу процессов турбулентного переноса в плазме. – *Физика плазмы*, 2002, т. 28, № 2, с. 128–143.
12. Г. М. Батанов, В. Е. Бенинг, В. Ю. Королев, А. Е. Петров, А. А. Пшеничников, К. А. Сарксян, В. В. Саенко, Н. Н. Скворцова, В. В. Учайкин, Н. К. Харчев, Ю. В. Хольнов. Структурная плазменная турбулентность и аномальная неброуновская диффузия. – В сб.: *“Стохастические модели структурной плазменной турбулентности”* под ред. В. Ю. Королева и Н. Н. Скворцовой. Макс–Пресс, Москва, 2003 г., с. 148–182.
13. Г. М. Батанов, В. Е. Бенинг, В. Ю. Королев, А. Е. Петров, А. А. Пшеничников, К. А. Сарксян, Н. Н. Скворцова, Н. К. Харчев, Ю. В. Хольнов. Структурная плазменная низкочастотная турбулентность в стеллараторе Л-2М. – *Письма в Журнал экспериментальной и теоретической физики*, 2003, т. 78, вып. 8, с. 974–983.
14. Г. М. Батанов, Л. В. Колик, А. Е. Петров и др. Исследование флуктуаций в высокотемпературной плазме современных стеллараторов методами микроволнового рассеяния. – *Физика плазмы*. 2003, т. 29, № 5, с. 395–413.
15. В. Е. Бенинг и В. Ю. Королев. Асимптотическое поведение обобщенных неординарных процессов Кокса. – *Вестник Московского университета, серия 15 Вычислительная математика и кибернетика*, 1997, вып. 4, с. 3–16.
16. В. Е. Бенинг и В. Ю. Королев. Асимптотические разложения для квантилей обобщенных процессов Кокса и их применение в некоторых задачах финансовой и страховой математики. – *Обозрение промышленной и прикладной математики. Серия “Финансовая и страховая математика”*, 1998, т. 5, вып. 1, с. 23–43.
17. В. Е. Бенинг, Г. Г. Гусаров, В. Н. Колокольников, В. Ю. Королев, Т. А. Сухокурова, В. В. Саенко, В. В. Учайкин. Дробно-устойчивые распределения. – В сб. *“Стохастические модели структурной плазменной турбулентности”* под ред. В. Ю. Королева и Н. Н. Скворцовой. Макс–Пресс, Москва, 2003 г., с. 291–360.
18. П. Бернстайн. *Против богов*. “Олимп–Бизнес”, Москва, 2006.
19. П. Биллингсли. *Сходимость вероятностных мер*. “Наука”, Москва, 1977.

20. К. А. Боровков. *Англо-русский, русско-английский словарь по теории вероятностей, математической статистике и комбинаторике*. Изд-во ТВП / Society for Industrial and Applied Mathematics. Москва–Филадельфия, 1994.
21. М. Вазан. *Стохастическая аппроксимация*. “Мир”, Москва, 1972.
22. Ф. П. Васильев. *Методы оптимизации*. Факториал Пресс, Москва, 2002.
23. Ф. П. Васильев и А. Ю. Иваницкий. *Линейное программирование*. Факториал Пресс, Москва, 2003.
24. С. В. Гавриленко и В. Ю. Королев. Оценки скорости сходимости смешанных пуассоновских случайных сумм – *Системы и средства информатики*, ИПИ РАН, Москва, 2006, специальный выпуск, с. 248–257.
25. Б. В. Гнеденко и А. Н. Колмогоров. *Предельные распределения для сумм независимых случайных величин*. ГИТТЛ, Москва–Ленинград, 1949.
26. Б. В. Гнеденко и Х. Фахим. Об одной теореме переноса. – Доклады АН СССР, 1969, т. 187, № 1, с. 15–17.
27. Ю. Н. Днестровский, А. Ю. Днестровский, С. Е. Лысенко, С. В. Черкасов. Канонические профили в токамаке с произвольным сечением плазменного шнура. – *Физика плазмы*, 2002, т. 28, № 11, с. 963–976.
28. В. М. Золотарев. *Одномерные устойчивые распределения*. “Наука”, Москва, 1983.
29. О. К. Исаенко и В. Ю. Урбах. Разделение смесей распределений вероятностей на их составляющие. – В сб.: *Итоги науки и техники. Теория вероятностей, математическая статистика и теоретическая кибернетика*. Издательство ВИНТИ, Москва, 1976, с. 37–58.
30. Б. А. Каррерас, Д. Ньюман, В. Е. Линч, П. Х. Даймонд. Самоорганизованная критичность как парадигма для процессов переноса в плазме, удерживаемой магнитным полем. – *Физика плазмы*, 1996, т. 22, № 9, с. 819–833.
31. Д. Е. Кащеев. *Моделирование динамики финансовых временных рядов и оценивание производных ценных бумаг*. Диссертация на соискание ученой степени кандидата физ.-матем. наук. Тверской государственный университет, Тверь, 2001.
32. С. Н. Кокшаров. *Асимптотические свойства смесей вероятностных распределений*. Диссертация на соискание ученой степени кандидата физ.-матем. наук. Московский государственный университет, Москва, 2007.

33. А. Н. Колмогоров. Локальная структура турбулентности в несжимаемой вязкой жидкости при очень больших числах Рейнольдса. – *Доклады АН СССР*, 1941, т. 30, с. 9-13. (Также см. А. Н. Колмогоров. *Избранные труды. Книга 1 Математика и механика*. “Наука”, Москва, 1988, с. 281–287.)
34. В. Ю. Королев. Сходимость случайных последовательностей с независимыми случайными индексами. I. – *Теория вероятностей и ее применения*, 1994, т. 39, вып. 2, с. 313–333.
35. В. Ю. Королев. Сходимость случайных последовательностей с независимыми случайными индексами. II. – *Теория вероятностей и ее применения*, 1995, т. 40, вып. 4, с. 907–910.
36. В. Ю. Королев. Построение моделей распределений биржевых цен с помощью методов асимптотической теории случайного суммирования. – *Обозрение промышленной и прикладной математики. Серия “Финансовая и страховая математика”*, 1997, т. 4, вып. 1, с. 86–102.
37. В. Ю. Королев. *Вероятностные модели. Введение в асимптотическую теорию случайного суммирования*. Диалог-МГУ, Москва, 1997.
38. В. Ю. Королев. О сходимости распределений случайных сумм независимых случайных величин к устойчивым законам. – *Теория вероятностей и ее применения*, 1997, т. 42, вып. 4, с. 818–820.
39. В. Ю. Королев. О сходимости распределений обобщенных процессов Кокса к устойчивым законам. – *Теория вероятностей и ее применения*, 1998, т. 43, вып. 4, с. 786–792.
40. В. Ю. Королев. Асимптотические свойства выборочных квантилей, построенных по выборкам случайного объема. – *Теория вероятностей и ее применения*, 1999, т. 44, вып. 2, с. 440–445.
41. В. Ю. Королев. Асимптотические свойства экстремумов обобщенных процессов Кокса и их применение к некоторым задачам финансовой математики. – *Теория вероятностей и ее применения*, 2000, т. 45, вып. 1, с. 182–194.
42. В. Ю. Королев и В. В. Учайкин. Некоторые предельные теоремы для обобщенных процессов восстановления с тяжелыми хвостами. – *Теория вероятностей и ее применения*, 2000, т. 45, вып. 4, с. 809–811.
43. В. Ю. Королев. О стереотипе нормальности и механизмах возникновения распределений с тяжелыми хвостами при математическом моделировании реальных процессов. – В сб. *“Стохастические модели структурной плазменной турбулентности”* под ред. В. Ю. Королева и Н. Н. Скворцовой. Макс-Пресс, Москва, 2003 г., с. 183–273.

44. В. Ю. Королев. *Смешанные гауссовские модели реальных процессов*. МАКС Пресс, Москва, 2004, 124 с.
45. В. Ю. Королев. *Теория вероятностей и математическая статистика*. “Проспект”, Москва, 2006.
46. В. Ю. Королев. Новый подход к определению и анализу компонент волатильности финансовых индексов. – *Актuariй*, 2007, № 1, с. 47–50.
47. В. Ю. Королев. *ЕМ-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений. Теоретический обзор*. Издательство ИПИ РАН, Москва, 2007, 94 с.
48. В. Ю. Королев. Статистическая декомпозиция волатильности. – В сб.: *Статистические методы оценивания и проверки гипотез*. Изд-во Пермского университета. Пермь, 2007, с. 170–206.
49. В. Ю. Королев и Н. Н. Скворцова (ред.). *Стохастические модели структурной плазменной турбулентности*. МАКС Пресс, Москва, 2003.
50. В. Ю. Королев и Н. Н. Скворцова. Новый метод вероятностно-статистического анализа процессов плазменной турбулентности. – *Системы и средства информатики*, ИПИ РАН, Москва, 2005, специальный выпуск, с. 126–179.
51. В. Ю. Королев, В. А. Ломской, Н. Н. Пресняков и М. Рэй. Анализ компонент волатильности с помощью метода скользящего разделения смесей. – *Системы и средства информатики*, ИПИ РАН, Москва, 2005, специальный выпуск, с. 180–206.
52. В. Ю. Королев и И. Г. Шевцова. О точности нормальной аппроксимации. I. – *Теория вероятностей и ее примен.*, 2005, т. 50, вып. 2, с. 353–366.
53. В. Ю. Королев и И. Г. Шевцова. О точности нормальной аппроксимации. II. – *Теория вероятностей и ее примен.*, 2005, т. 50, вып. 3, с. 555–564.
54. В. Ю. Королев, В. Е. Бенинг и С. Я. Шоргин. *Математические основы теории риска*, Физматлит, Москва, 2007.
55. Г. Крамер. *Математические методы статистики*. Государственное издательство иностранной литературы, Москва, 1948.
56. В. М. Круглов. Смеси распределений вероятностей. – *Вестник Московского университета, Серия 15 Вычислительная математика и кибернетика*, 1991, № 2, с. 3–15.
57. В. М. Круглов. Нормальная и пуассоновская сходимости. – *Теория вероятностей и ее применения*, 2003, т. 48, вып. 2, с. 392–398.
58. В. М. Круглов и В. Ю. Королев. *Предельные теоремы для случайных сумм*. Издательство Московского университета, Москва, 1990.

-
59. С. Кульбак. *Теория информации и статистика*. “Наука”, Москва, 1967.
60. П. Леви. *Стохастические процессы и броуновское движение*. “Наука”, Москва, 1972.
61. Р. Дж. А. Литтл и Д. Б. Рубин. *Статистический анализ данных с пропусками*. “Финансы и статистика”, Москва, 1991.
62. М. Лоэв. *Теория вероятностей*. “Мир”, Москва, 1962.
63. Б. Мандельброт. *Фракталы, случай и финансы*. Изд-во “Регулярная и хаотическая динамика”, Москва–Ижевск, 2004, 256 с.
64. В. К. Мацквичюс. О нижней оценке скорости сходимости в центральной предельной теореме. – *Теория вероятностей и ее применения*, 1983, т. 28, вып. 3, с. 565–569.
65. А. Б. Михайловский. *Теория плазменных неустойчивостей*. Т. 1, 2. “Атомиздат”, Москва, 1971.
66. А. Б. Михайловский. *Неустойчивости плазмы в магнитных ловушках*. “Атомиздат”, Москва, 1981.
67. В. К. Мюллер. *Англо-русский словарь*. “Советская энциклопедия”, Москва, 1967.
68. М. Б. Невельсон и Р. З. Хасьминский. *Стохастическая аппроксимация и рекуррентное оценивание*. “Наука”, Москва, 1972.
69. П. В. Новицкий и И. А. Зограф. *Оценка погрешностей результатов измерений*. “Энергоатомиздат”, Ленинград, 1991.
70. А. И. Орлов. Некоторые вероятностные вопросы теории классификации. – В кн. *Прикладная статистика*. “Наука”, Москва, 1983, с. 166–179.
71. В. В. Петров. *Суммы независимых случайных величин*. “Наука”, Москва, 1972.
72. В. В. Петров. *Предельные теоремы для сумм независимых случайных величин*. “Наука”, Москва, 1987.
73. Ю. В. Прохоров. Сходимость случайных процессов и предельные теоремы теории вероятностей. – *Теория вероятностей и ее применения*, 1956, т. 1, вып. 2, с. 177–238.
74. Ю. В. Прохоров, В. Ю. Королев и В. Е. Бенинг. Аналитические методы теории риска, основанные на смешанных гауссовских моделях. – *Вестник московского университета, серия 15 Вычислительная математика и кибернетика*, 2005, вып. 5, с. 95–112.
75. С. Т. Рачев и Л. Рушендорф. Модели и расчеты контрактов с опционами. – *Теория вероятностей и ее применения*, 1994, т. 39, вып. 1, с. 150–190.

76. Дж. Себер. *Линейный регрессионный анализ*. “Мир”, Москва, 1980.
77. Е. Сенета. *Правильно меняющиеся функции*. Наука, Москва, 1985.
78. Н. Н. Скворцова, В. Ю. Королев, Т. А. Маравина, Г. М. Батанов, А. Е. Петров, А. А. Пшеничников, К. А. Сарксян, Н. К. Харчев, Х. Санчес и Ш. Кубо. Новые возможности математического моделирования турбулентных транспортных процессов в плазме. – *Физика плазмы*, 2005, том 31, № 1, с. 64–83.
79. К. Ф. Соколова. *Применение метода скользящего разделения смесей к декомпозиции волатильности финансовых индексов*. Магистерская диссертация. Факультет вычислительной математики и кибернетики МГУ, 2007.
80. М. М. Сущик. Динамика структур в сдвиговых течениях. – В сб. *Нелинейные волны. Структуры и бифуркации*. Под ред. А. В. Гапонова-Грехова и М. И. Рабиновича. “Наука”, Москва, 1987, с. 104–132.
81. С. Уилкс. *Математическая статистика*. “Наука”, Москва, 1967.
82. В. Феллер. *Введение в теорию вероятностей и ее приложения. Том 1*. “Мир”, Москва, 1984.
83. В. Феллер. *Введение в теорию вероятностей и ее приложения. Том 2*. “Мир”, Москва, 1984.
84. У. Фриш. *Турбулентность. Наследие А. Н. Колмогорова*. “Фазис”, Москва, 1998.
85. П. Халмош. *Теория меры*. Изд-во иностранной литературы, Москва, 1953.
86. В. Хенгартнер и Р. Теодореску. *Функции концентрации*. “Наука”, Москва, 1980 г.
87. А. Я. Хинчин. *Предельные законы для сумм независимых случайных величин*. ОНТИ НКТП, Москва–Ленинград, 1938.
88. И. Г. Шевцова. *Уточнение структуры оценок скорости сходимости в центральной предельной теореме для сумм независимых случайных величин*. Диссертация на соискание ученой степени кандидата физ.-матем. наук. Московский государственный университет, Москва, 2006.
89. А. Н. Ширяев. Вероятностно-статистические модели эволюции финансовых индексов. – *Обозрение промышленной и прикладной математики. Серия “Финансовая и страховая математика”*, 1995, т. 2, вып. 4, с. 527–555.
90. А. Н. Ширяев. *Вероятность*. “Наука”, Москва, 1989.
91. А. Н. Ширяев. *Основы стохастической финансовой математики. Том 1. Факты. Модели*. “Фазис”, Москва, 1998.

-
92. А. Н. Ширяев. *Основы стохастической финансовой математики. Том 2. Теория*. “Фазис”, Москва, 1998.
 93. М. И. Шлезингер. О самопроизвольном распознавании образов. – В сб.: *Читающие автоматы*. “Наукова думка”, Киев, 1965, с. 38–45.
 94. М. И. Шлезингер. Взаимосвязь обучения и самообучения в распознавании образов. – *Кибернетика*, 1968, № 2, с. 81–88.
 95. H. Akaike. Information theory and an extension of the maximum likelihood principle. – In: B. N. Petrov and F. Csake (eds.) *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest, 1973, p. 267–281.
 96. H. Akaike. A new look at the statistical model identification. – *IEEE Transactions on Automatic Control*, AC-19, 1974, p. 716–723.
 97. H. Akaike. A Bayesian analysis of the minimum AIC procedure. – *Ann. Inst. Statist. Math.*, 1978, vol. 30A, p. 9–14.
 98. V. Akgiray and G. G. Booth. Compound distribution models of stock returns: an empirical comparison. – *J. of Financial Research.*, 1987, vol. 10, p. 269–280.
 99. S. Alizadeh, M. Brandt and F. Diebold. Range-based estimation of stochastic volatility models. – *Journal of Finance*, 2002, vol. 55, p. 1047–1091.
 100. T. Andersen. Return volatility and trading volume: an information flow interpretation of stochastic volatility. – *Journal of Finance*, 1996, vol. 51, No. 1, p. 169–204.
 101. D. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. – *J. R. Statist. Soc.*, 1974, vol. 36, p. 99–102.
 102. T. Ané and H. Geman. Order flow, transaction clock and normality of asset returns. – *Journal of Finance*, 2000, vol. 55, No. 5, p. 2259–2284.
 103. A. Arneodo, J.-P. Bouchaud, R. Cont, J.-F. Muzy, M. Potters and D. Sornette. *Comment on “Turbulent Cascades in Foreign Exchange Markets*. Eprint arXiv: cond/mat/9607120, 17 July 1996.
 104. L. Bachelier. *Théorie de la spéculation*. Ann. Ecole Norm. Sup., 1900, vol. 17, p. 21–86 (reprinted in: P. H. Cootner (Ed.). *The Random Character of Stock Market Prices*. Cambridge, MA, MIT Press, 1967, p. 517–531).
 105. B. Baeumer, D. A. Benson, M. M. Meerschaert, S. W. Wheatcraft. Subordinated advection–dispersion equation for contaminant transport. *Water Resources Research*, 2001, vol. 37, No. 6, p. 1543–1550.
 106. M. Bakhtiari. Disruption mitigation experiments in the JT-60U tokamak. – In: *20th IAEA Fusion Energy Conference*. Portugal, 2004, report EX/10–6Rb.

107. T. Bali, N. Cakici, X. Yan and Z. Zhang. Does idiosyncratic risk really matter? – *Journal of Finance*, 2005, vol. 60, No. 2, p. 905–929.
108. O. E. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. – *Proc. R. Soc. London*, 1977, vol. A 353, p. 401–419.
109. O. E. Barndorff-Nielsen. Hyperbolic distributions and distributions of hyperbolae. – *Scand. J. Statist.*, 1978, vol. 5, p. 151–157.
110. O. E. Barndorff-Nielsen. Models for non-Gaussian variation, with applications to turbulence. – *Proc. Roy. Soc., ser. A*, 1979, vol. 353, p. 401–419.
111. O. E. Barndorff-Nielsen. *Gaussian\Inverse Gaussian Processes and the Modeling of Stock Returns*. Preprint. Aarhus, Aarhus University, 04 October 1994.
112. O. E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. – *Finance and Stochastics*, 1998, vol. 2, p. 41–68.
113. O. E. Barndorff-Nielsen. Probability and statistics: self-decomposability, finance and turbulence. – In: *Probability Towards 2000 (New York, 1995)*, Lecture Notes in Statistics, vol. 128. Springer, 1998, p. 47–57.
114. O. E. Barndorff-Nielsen and P. Blæsild. Hyperbolic distributions and ramifications: contributions to the theory and application. – In: *Statistical Distributions in Scientific Work. vol. 4*, C. Taillie et al. – Eds. Reidel, Dordrecht, 1981, p. 19–44.
115. O. E. Barndorff-Nielsen, J. Kent and M. Sørensen. Normal variance-mean mixtures and the z-distributions. – *International Statistical Review*, 1982, vol. 50, No.2, p. 145–159.
116. O. E. Barndorff-Nielsen, J. L. Jensen and M. Sørensen. Parametric modeling of turbulence. – *Philosophical Transactions: Physical Sciences and Engineering. Stochastic Processes*, 1990, vol. 332, No. 1627, p. 439–455.
117. O. E. Barndorff-Nielsen, T. Mikosch and S. I. Resnick. *Lévy Processes: Theory and Applications*. Birkhäuser, Boston, 2001.
118. O. E. Barndorff-Nielsen, P. Blæsild and J. Schmiegel. A parsimonious and universal description of turbulent velocity increments. – *European Physical Journal*, 2004, B 41, p. 345–363.
119. O. E. Barndorff-Nielsen and J. Schmiegel. *Ambit Processes; with Application to Turbulence and Tumour Growth*. – Thiele Centre for Applied Mathematics in Natural Science. Research report No. 14, December 2005.
120. O. E. Barndorff-Nielsen and J. Schmiegel. *Time Change and Universality in Turbulence*. – Thiele Centre for Applied Mathematics in Natural Science. Research report No. 15, September 2006.

-
121. M. S. Bartlett. Statistical estimation of density function. – *Sankhya, ser. A*, 1963, vol. 25, p. 245–254.
 122. G. M. Batanov *et al.* Statistical properties and radial structure of plasma turbulence in the boundary region of the L2-M stellarator. – *Plasma Phys. Control. Fusion*, 1998, vol. 40, p. 1241–1250.
 123. G. M. Batanov, L. V. Kolik, A. E. Petrov *et al.* Studies of fluctuations in the high-temperature plasma of modern stellarators by the microwave scattering technique. – *Plasma Phys. Reports*, 2003, vol. 29, No. 5, p. 363–369.
 124. G. K. Batchelor. *The Theory of Homogeneous Turbulence*. Cambridge University Press, Cambridge, 1956.
 125. J. T. Bendler and M. F. Shlesinger. Dielectric relaxation via the Montroll–Weiss random walk of defects. – In: M. F. Shlesinger and G. H. Weiss (Eds.) *The Wonderful World of Stochastics*, North-Holland, Amsterdam, 1985, p. 31–46.
 126. V. E. Bening and V. Yu. Korolev. Asymptotic behavior of non-ordinary generalized Cox processes with nonzero means. – *Journal of Mathematical Sciences*, 1998, vol. 92, No. 3, p. 3836–3856.
 127. V. Bening and V. Korolev. *Generalized Poisson Models and their Applications in insurance and finance*. VSP, Utrecht, The Netherlands, 2002.
 128. D. A. Benson, M. M. Meerschaert, R. Schumer and S. W. Wheatcraft. Eulerian derivation of the fractional advection-dispersion equation. *Journal of Contaminant Hydrology*, 2001, vol. 38, p. 69–88.
 129. D. A. Benson, M. M. Meerschaert, R. Schumer and S. W. Wheatcraft. Fractional dispersion, Lévy motion, and the MADE tracer tests. *Transport in Porous Media*, 2001, vol. 42, p. 211–240.
 130. D. A. Benson, M. M. Meerschaert and S. W. Wheatcraft. Application of a fractional advection-dispersion equation. *Water Resources Research*, 2000, vol. 36, No. 6, p. 1403–1412.
 131. J. Bertoin. *Lévy Processes*. Cambridge Tracts in Mathematics, Vol. 121. Cambridge University Press, Cambridge, 1996.
 132. P. Biane, J. Pitman and M. Yor. Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. – *Bulletin of the American Mathematical Society*, 2001, vol. 38, p. 435–465.
 133. J. A. Bilmes. *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*. Technical report TR-97-021. International Computer Science Institute, Berkeley, CA, 1998.

-
134. Z. W. Birnbaum. On random variables with comparable peakedness. – *Ann. Math. Statist.*, 1948, vol. 19, No. 1, p. 76–81.
135. R. Blattberg and N. Gonedes. A comparison of the stable and Student distributions as statistical models of stock prices. – *J. Business*, 1974, vol. 47, p. 244–250.
136. R. P. Boas (Ed.) *A. J. Lohwater's Russian-English Dictionary of the Mathematical Sciences*. American Mathematical Society. Providence, RI, 1990.
137. A. Boness, A. Chen and S. Jatusipitak. Investigations of nonstationary prices. – *J. Business*, 1974, vol. 47, p. 518–537.
138. T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. – *Journal of Econometrics*. 1986, vol. 31, p. 307–327.
139. J. G. Booth and J. P. Hobert. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. – *Journal of the Royal Statistical Society, Series B*, 1999, vol. 61, p. 265–285.
140. J.-P. Bouchaud and A. Georges. Anomalous diffusion in disordered media: Statistical mechanics, models and physical applications. – *Phys. Rep.*, 1990, vol. 145, p. 127–293.
141. R. A. Boyles. On the convergence of the EM-algorithm. – *J. Roy. Statist. Soc.*, 1983, vol. 45, p. 47–50.
142. H. Bozdogan. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. – *Psychometrika*, 1987, vol. 52(3), p. 345–370.
143. H. Bozdogan. Mixture-model cluster analysis using model selection criteria and a new information measure of complexity. – In: H. Bozdogan (Ed.) *Proceedings of the First US-Japan Conference on the Frontiers of Statistical Modeling: An Information Approach*. vol. 2. Kluwer Academic Publishers, Boston, 1994, p. 69–113.
144. J. E. Brada and J. Van Tassel. The distribution of stock-price differences: Gaussian after all? – *Operations Research*, 1966, vol. 14, p. 332–340.
145. M. W. Brandt and J. Kinlay. *Estimating Historical Volatility*. Research article. Investment Analytics. 2005.
146. M. Broniatowski, G. Celeux and J. Diebolt. Reconnaissance de mélanges de densités par un algorithme d'apprentissage probabiliste. – In: E. Diday, M. Jambu, L. Lebart, J.-P. Pagès and R. Tomasone (Eds.) *Data Analysis and Informatics*, III, North Holland, Amsterdam, 1983, p. 359–373.
147. B. S. Caffo, W. S. Jank and G. L. Jones. Ascent-Based Monte Carlo EM. – *Journal of the Royal Statistical Society, Series B*, 2005, vol. 67, p. 235–252.

-
148. J. Y. Campbell. A variance decomposition for stock returns. – *The Economic Journal*, 1991, vol. 101, p. 157–179.
 149. J. Y. Campbell, A. Lo and C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, 1997.
 150. J. Y. Campbell, M. Lettau, B. G. Malkiel and Y. Xu. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. – *Journal of Finance*, 2001, vol. 56, No. 1, p. 1–43.
 151. J. Y. Campbell and R. Shiller. The dividend-price ratio and expectation of future dividends and discount factors. – *The Review of Financial Studies*, 1988, vol. 1, p. 195–228.
 152. P. P. Carr, D. B. Madan and E. C. Chang. The Variance Gamma process and option pricing. – *European Finance Review*, 1998, vol. 2, p. 79–105.
 153. P. P. Carr and D. B. Madan. Option valuation using the Fast Fourier Transform. – *Journal of Computational Finance*, 1999, vol. 2, p. 61–73.
 154. P. P. Carr, H. Geman, D. B. Madan and M. Yor. The fine structure of asset returns: an empirical investigation. – *Journal of Business*, 2002, vol. 75, No. 2, p. 305–332.
 155. P. P. Carr, H. Geman, D. B. Madan and M. Yor. Stochastic volatility for Lévy processes. – *Mathematical Finance*, 2003, vol. 13, No. 3, p. 345–382.
 156. P. P. Carr and L. Wu. Time-changed Lévy processes and option pricing. – *Journal of Financial Economics*, 2004, vol. 71, No. 1, p. 113–141.
 157. B. A. Carreras, B. van Milligen, M. A. Pedrosa et al. Long-range time correlations in plasma edge turbulence. – *Physical Review Letters*, 1998, vol. 80, No. 20, p. 4438–4442.
 158. B. A. Carreras, B. van Milligen, C. Hidalgo et al. Self-similarity properties of the probability distribution function of turbulence-induced particle fluxes at the plasma edge. – *Physical Review Letters*, 1999, vol. 83, No. 18, p. 3653–3657.
 159. B. Castaing, Y. Gagne, E. J. Hopfinger. Velocity probability density functions of high Reynolds number turbulence. – *Physica D*, 1990, vol. 46, p. 177–200.
 160. J. Cavanaugh. Unifying the derivations of the Akaike and corrected Akaike information criteria. – *Statistics and Probability Letters*, 1997, vol. 31, p. 201–208.
 161. G. Celeux and J. Diebolt. *Reconnaissance de mélanges de densité et classification. Un algorithme d'apprentissage probabiliste: l'algorithme SEM*. Rapport de Recherche de l'INRIA RR-0349. Centre de Rocquencourt. 1984.

-
162. G. Celeux and J. Diebolt. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. – *Computational Statistics Quarterly*, 1985, vol. 2, No. 1, p. 73–82.
 163. G. Celeux and G. Govaert. *A Classification EM Algorithm for Clustering and Two Stochastic Versions*. Rapport de Recherche de l'INRIA RR-1364. Centre de Rocquencourt. 1991.
 164. G. Celeux and G. Govaert. A classification EM algorithm for clustering and two stochastic versions. – *Computational Statistics and Data Analysis*, 1992, vol. 14, p. 315–332.
 165. G. Celeux and J. Diebolt. A stochastic approximation type algorithm for the mixture problem. – *Stochastics and Stochastics Reports*, 1992, vol. 41, p. 119–134.
 166. G. Celeux, D. Chauveau and J. Diebolt. *On Stochastic Versions of the EM-algorithm*. Rapport de Recherche de l'INRIA RR-2514. Centre de Rocquencourt. 1995.
 167. J. Chen and J. D. Kalbfleisch. Modified likelihood ratio test in finite mixture models with a structural parameter. – *Journal of Statistical Planning and Inference*, 2005, vol. 129, p. 93–107.
 168. H. Chernoff. On the distribution of the likelihood ratio. – *Annals of Mathematical Statistics*, 1954, vol. 25, p. 573–578.
 169. M. Chernov, R. Gallant, E. Ghysels and G. Tauchen. Alternative models for stock price dynamics. – *Journal of Econometrics*, 2003, vol. 116, p. 225–257.
 170. R. S. Chhikara and J. L. Folks. *Inverse Gaussian Distribution. Theory, Methodology, and Applications*. Marcel Dekker, New York – Basel, 1989.
 171. S. Chrétien and A. Hero. Kullback proximal algorithms for maximum likelihood estimation. – *IEEE Transactions on Information Theory*, 2000, vol. 46, No. 5, p. 1800–1810.
 172. P. K. Clark. *A Subordinated Stochastic Process Model of Cotton Futures Prices*. Ph. D. Thesis, Harvard University, Cambridge, Ma, 1970.
 173. P. K. Clark. A subordinated stochastic process model with finite variance for speculative prices. – *Econometrica*, 1973, vol. 41, p. 135–155.
 174. J. W. Connor, P. Burratti, J. G. Cordey, A. Jacchia, U. Stroth and W. Suttrop. EU-US workshop on transport in fusion plasmas. – *Plasma Physics Control Fusion*, 1999, vol. 41, p. 693–707.
 175. R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman and Hall, Boca Raton, 2004.

-
176. T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.
 177. D. R. Cox. Some statistical methods connected with series of events. – *J. Roy. Statist. Soc., Ser. B*, 1955, vol. 17, p. 129–164.
 178. N. E. Day. Divisive cluster analysis and test for multivariate normality. – *Session of the ISI*, London, 1969.
 179. N. E. Day. Estimating the components of a mixture of normal distributions. – *Biometrika*, 1969, vol. 56, No. 3, p. 463–474.
 180. B. Delyon, M. Lavielle and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. – *The Annals of Statistics*, 1999, vol. 27, p. 94–128.
 181. A. Dempster, N. Laird and D. Rubin. Maximum likelihood estimation from incomplete data. – *Journal of the Royal Statistical Society, Series B*, 1977, vol. 39, p. 1–38.
 182. L. Devroye and L. Györfi. *Nonparametric Density Estimation*. Wiley, New York, 1985.
 183. J. Diebolt and G. Celeux. Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions. *Communications in Statistics B: Stochastic Modelling*, 1993, vol. 9, p. 599–613.
 184. J. Diebolt and E. H.-s. Ip. Stochastic EM: method and application. – In: W. R. Gilks, S. Richardson and D. J. Spiegelhalter (Eds.) *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, 1996, p. 259–273.
 185. Z. Ding and C. Granger. Modeling volatility persistence of speculative returns: A new approach. – *Journal of Econometrics*, 1996, vol. 73, p. 185–215.
 186. J. A. Dutton. Effects of turbulence on aeronautical systems. – In: *Progress in Aerospace Science* (ed.: D. Kucheman), vol. 11, Pergamon Press, New York, 1970, p. 67–109.
 187. J. A. Dutton and D. G. Deaven. Some observed properties of atmospheric turbulence. – In: *Statistical Models and Turbulence. Lecture Notes in Physics* (ed.: M. Rosenblatt and C. W. Van Atta), Springer-Verlag, Heidelberg, 1972, vol. 12, p. 352–383.
 188. E. Eberlein. *Application of generalized hyperbolic Lévy motions to finance*. Preprint No. 64. Freiburg i. Br.: Universität Freiburg, Institut für Mathematische Stochastic, 1999.
 189. E. Eberlein and U. Keller. Hyperbolic Distributions in Finance. – *Bernoulli*, 1995, vol. 1, No. 3, p. 281–299.

-
190. E. Eberlein, U. Keller and K. Prause. New insights into smile, mispricing and value at risk: the hyperbolic model. – *Journal of Business*. 1998, vol. 71, p. 371–405.
 191. E. Eberlein and K. Prause. *The generalized Hyperbolic Model: Financial Derivatives and Risk Measures*. Preprint N 56. Freiburg I. Br.: Universität Freiburg, Institut für Mathematische Stochastic, 1998.
 192. B. Efron and R. Olshen. How broad is the class of normal scale mixtures? – *Ann. Statist.*, 1978, vol. 6, p. 1159–1164.
 193. A. Einstein. On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat. – *Annalen der Physik (Leipzig)*, 1905, vol. 17, p. 549–560.
 194. P. Embrechts. A property of the generalized inverse Gaussian distribution with some applications. – *J. Appl. Probab.*, 1983, vol. 20, p. 537–544.
 195. R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of United Kingdom inflation. – *Econometrica*. 1982, vol. 50, No. 4, p. 987–1008.
 196. R. F. Engle and T. Bollerslev. Modelling the persistence of conditional variance. – *Econometrics Reviews*. 1986, vol. 5, p. 1–50.
 197. R. F. Engle, E. Ghysels and B. Sohn. On the economic sources of stock market volatility. *Discussion paper. New York University and University of North Carolina*, 2006. Available at:
www.unc.edu/depts/econ/workshops/fe/Engle_Ghysels_Sohn_Paper.pdf
 198. R. F. Engle and G. Lee. A permanent and transitory component model of stock return volatility. – In: R. Engle and H. White (Eds.) *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. F. Granger*. Oxford University Press, 1999, p. 475–497.
 199. R. F. Engle and V. K. Ng. Measuring and testing the impact of news on volatility. – *Journal of Finance*. 1993, vol. 48, No. 5, p. 1749–1778.
 200. R. F. Engle, M. E. Sokalska and A. Chanda. *High Frequency Multiplicative Component GARCH*. Stern School of Business working paper. August, 2005, No. Sc-CFE-05-05. Available at:
www.stern.nyu.edu/salomon/docs/CFE/CFE-05-05.pdf
 201. B. S. Everitt and D. J. Hand. *Finite Mixture Distributions*. Chapman and Hall, 1981.
 202. E. Fama. Mandelbrot and stable Paretian hypothesis. – *J. Business*, 1963, vol. 4, p. 420–429.
 203. E. Fama. The behavior of stock market prices. – *J. Business*, 1965, vol. 38, p. 34–105.

-
204. M. A. T. Figueiredo. *Lecture Notes on the EM Algorithm*, 2004. Available at: <http://www.stat.duke.edu/courses/Spring06/sta376/Support/EM/EM.Mixtures.Figueiredo.2004.pdf>
205. R. Fletcher and M. J. D. Powell. A rapidly convergent descent method for minimization. – *Computer Journal*, 1963, vol. 6, p. 163–168.
206. C. Fraley and A. E. Raftery. *How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis*. Technical Report No. 329, Department of Statistics, University of Washington, Seattle, 1998. Available at <http://www.ics.uci.edu/~smyth/courses/ics274/fraleyraftery.pdf>
207. C. Fraley and A. E. Raftery. How many clusters? Which clustering method? Answers via model-based cluster analysis. – *Computer Journal*, 1998, vol. 41, p. 578–588.
208. A. Frazzini and I. Marsh. *Idiosyncratic Volatility in the US and UK Equity Markets*. Working paper, Yale University, SSRN-id: 571067, 2003.
209. M. C. Fu, R. A. Jarrow, Ju-Yi. J. Yen, R. J. Elliott (Eds.) *Advances in Mathematical Finance*. Birkhäuser, Boston, 2007.
210. A. R. Gallant, C.-T. Hsu and G. Tauchen. Using daily range data to calibrate volatility diffusion and extract the forward integrated variance. – *Review of Economic Statistics*, 1999, vol. 81, p. 617–631.
211. M. Garman and M. Klass. On the estimation of security price volatilities from historical data. – *Journal of Business*, 1980, vol. 53, p. 67–68.
212. J.-M. Gaspar and M. Massa. *Idiosyncratic Volatility and Product Market Competition*. Discussion Paper 4812, Centre for Economic Policy Research, London, 2004
213. H. Geman. Pure jump Lévy processes for asset price modelling. – *Journal of Banking and Finance*, 2002, vol. 26, No. 7, p. 1297–1316.
214. H. Geman and T. Ané. Stochastic subordination. – *Risk*, 1996, vol. 9, p. 146–149.
215. S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner and Y. Dodge. Turbulent cascades in foreign exchange markets. – *Nature*, 1996, vol. 381, p. 767–770.
216. B. V. Gnedenko and V. Yu. Korolev. *Random Summation: Limit Theorems and Applications*. CRC Press, Boca Raton, FL, 1996.
217. D. Goldfarb. A family of variable metric updates derived by variational means. – *Mathematics of Computing*, 1970, vol. 24, p. 23–26.
218. A. Goyal and P. Santa-Clara. Idiosyncratic risk matters! – *Journal of Finance*. 2003, vol. 58, p. 975–1007.

-
219. J. Grandell. *Doubly Stochastic Poisson Processes*. – Lecture Notes Math., vol. 529, 1976.
220. J. Grandell. *Mixed Poisson Processes*. Chapman and Hall, London, 1997.
221. C. W. J. Grander and O. Morgenstern. *Predictability of Stock Market Prices*. D. C. Heath and Co., Lexington, Ma, 1970.
222. B. Grigelionis. Processes of Meixner type. – *Lithuanian Mathematical Journal*, 1999, vol. 39, No. 1, p. 33–41.
223. B. Grigelionis. *Generalized z -Distributions and Related Stochastic Processes*. Preprint No 2000–22, Institute of Mathematics and Informatics, Vilnius, 2000.
224. H. Guo and R. Savickas. *Aggregate Idiosyncratic Volatility in G7 Countries*. Working Paper, Federal Reserve Bank of St. Louis and George Washington University. SSRN-id: 615242, 2005.
225. P. Hall. On measures of the distance of a mixture from its parent distribution. – *Stochastic Processes Appl.*, 1979, vol. 8, p. 357–365.
226. E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. – *Journal of the Royal Statistical Society*, 1979, vol. B41, p. 190–195.
227. V. Hasselblad. Estimation of parameters for a mixture of normal distributions. – *Technometrics*, 1966, vol. 8, No. 3, p. 431–446.
228. M. J. R. Healy and M. H. Westmacott. Missing values in experiments analyzed on automatic computers. – *Applied Statistics*, 1956, vol. 5, p. 203–206.
229. C. C. Heyde and J. R. Leslie. On moment measures of departure from the normal and exponential laws. – *Stochastic Processes Appl.*, 1976, vol. 4, p. 317–328.
230. B. Holdom. From turbulence to financial time series. – *Physica A*, 1998, vol. 254, p. 569–576.
231. E. M. Hollman. Disruption thermal quench mitigation by noble gas injection in DIII-D. – In: *20th IAEA Fusion Energy Conference*. Portugal, 2004, report EX/10–6Ra.
232. I. Hosokawa, C. W. Van Atta and S. T. Thoroddsen. Experimental study of Kolmogorov refined similarity variable. – *Fluid Dynamics Research*, 1994, vol. 13, p. 329–333.
233. J. C. Hull. *Options, Futures and Other Derivative Securities*. 5th ed. Prentice–Hall, New Jersey, 2003.
234. J. Hull and A. White. The pricing of options on assets with stochastic volatilities. – *Journal of Finance*, 1987, vol. 42, p. 281–308.

-
235. C. M. Hurvich and C.-L. Tsai. Regression and time series model selection in small samples. – *Biometrika*, 1989, vol. 76, p. 297–307.
236. C. M. Hurvich and C.-L. Tsai. Model selection for extended quasi-likelihood models in small samples. – *Biometrics*, 1995, vol. 51, p. 1077–1084.
237. E. H.-s. Ip. *A Stochastic EM Estimator in the Presence of Missing Data – Theory and Practice*. PhD Dissertation, Stanford University, 1994, 127 p.
238. M. B. Isichenko. Percolation, statistical topography, and transport in random media. – *Rev. Mod. Phys.*, 1992, vol. 64, p. 961–1043.
239. M. Ishiguro, Y. Sakamoto and G. Kitagawa. Bootstrapping log likelihood and EIC, an extension of AIC. – *Annals of the Institute of Statistical Mathematics*, 1997, vol. 49, p. 411–434.
240. J. Jang. Jump diffusion processes and their applications in insurance and finance. – *Insurance: Mathematics and Economics*, 2007, vol. 41, p. 62–70.
241. W. Jank. *Implementing and Diagnosing the Stochastic Approximation EM Algorithm*. Technical Report, University of Maryland, 2004. Also available at: <http://www.smith.umd.edu/faculty/wjank/monitorSAEM.pdf>
242. D. Jansen and C. de Vries. On the frequency of large stock returns: Putting booms and busts into perspective. – *The Review of Economics and Statistics*, 1991, vol. 73, p. 18–24.
243. M. Jordan and L. Xu. Convergence results for the EM approach to mixtures of experts architectures. – *Neural Networks*, 1996, vol. 8, p. 1409–1431.
244. B. Jørgensen. *Statistical properties of the generalized inverse Gaussian distribution*. Lecture Notes in Statistics, vol. 9. Springer, Berlin, 1982.
245. G. G. Judge and T. Takayama. Inequality restrictions in regression analysis. – *Journal of American Statistical Association*, 1966, vol. 61, No. 1, p. 166–181.
246. J. N. Kapur. *Maximum-Entropy Models in Science and Engineering*. – Wiley, New York, 1989.
247. R. L. Karandikar and S. T. Rachev. A generalized binomial model and option pricing formulae for subordinated stock-price processes. – *Probability and Mathematical Statistics (Wrocław)*, 1995, vol. 15, p. 427–447.
248. D. E. Kashcheev. Compound Cox processes and option pricing. – *Journal of Mathematical Sciences*, 2001, vol. 106, No. 1, p. 2682–2697.
249. R. E. Kass and A. E. Raftery. Bayes factors. – *Journal of the American Statistical Society*, 1995, vol. 90, p. 773–795.
250. J. Kazakos. Recursive estimation of prior probabilities using a mixture. – *IEEE Transactions on Information Theory*, 1977, vol. 23, No. 2, p. 203–210.

-
251. J. Keilson and F. W. Steutel. Mixtures of distributions, moment inequalities and measures of exponentiality and normality. – *Annals of Probability*, 1974, vol. 2, No. 1, p. 112–130.
252. M. G. Kendall. The analysis of economic time series: Part I, Prices. – *J. Royal Statist. Soc.*, 1953, vol. 96, p. 11–25.
253. J. Klafter, A. Blumen and M. F. Shlesinger. Stochastic pathway to anomalous diffusion. – *Physical Review, A*, 1987, vol. 35, p. 3081–3085.
254. J. Klafter, G. Zumofen and M. F. Shlesinger. Lévy walks in dynamical systems. – *Physica, A*, 1993, vol. 200, p. 222–230.
255. A. N. Kolmogorov. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. – *Journal of Fluid Mechanics*. 1962, vol. 13, p. 82–85.
256. V. Kolokoltsov, V. Korolev and V. Uchaikin. *Fractional Stable Distributions*. Research report No. 23/00, Nottingham Trent University, 2000, 12 p.
257. V. Kolokoltsov, V. Korolev and V. Uchaikin. Fractional Stable Distributions. – *Journal of Mathematical Sciences*, 2001, vol. 105, No. 6, p. 2569–2576.
258. V. Kolokoltsov, V. Bening, V. Korolev and T. Sukhorukova. *Convergence rate estimates for superpositions of independent stochastic processes with applications to estimation of the accuracy of approximation of the distributions of continuous-time random walks by fractional stable laws*. Preprint. The Nottingham Trent University, 2003.
259. S. Kon. Models of stock returns – a comparison. – *J. Finance*, 1984, (March), p. 147–165.
260. V. Yu. Korolev. Limit behavior of centered random sums of independent identically distributed random variables. – *Journal of Mathematical Sciences*, 1995, vol. 76, p. 2153–2162.
261. V. Yu. Korolev. A general theorem on the limit behavior of superpositions of independent random processes with applications to Cox processes. – *Journal of Mathematical Sciences*, 1996, vol. 81, No. 5, p. 2951–2956.
262. V. Yu. Korolev and V. V. Uchaikin. Limit distributions of sums of random numbers of random variables expressible in terms of stable distributions. – In: *XX International Seminar on Stability Problems for Stochastic Models. Lublin–Nałęczów, Poland, 5-11 September, 1999*. Abstracts of Communications. Maria Curie-Skłodowska University Publishing House, Lublin, 1999, p. 94.
263. V. Yu. Korolev. On distributions whose symmetrization is a scale mixture of normal laws. – *Journal of Mathematical Sciences*, 2004, vol. 119, No. 3, p. 377–380.

-
264. V. Yu. Korolev and N. N. Skvortsova (Eds.) *Stochastic Models of Plasma Turbulence* VSP, Utrecht, 2005.
265. V. Yu. Korolev and M. Rey. Statistical analysis of volatility of financial time series and turbulent plasmas by the method of moving separation of mixtures. – In: V. Yu. Korolev and N. N. Skvortsova (Eds.) *Stochastic Models of Plasma Turbulence* VSP, Utrecht, 2005.
266. M. Kotulski. Asymptotic distributions of the continuous-time random walks: a probabilistic approach. – *J. Stat. Phys.*, 1995, vol. 81, p. 777–779.
267. S. G. Kou. A jump–diffusion model for option pricing. – *Management Science*, 2002, vol. 48, p. 1086–1101.
268. S. G. Kou and H. Wang. Option pricing under a double exponential jump–diffusion model. – *Management Science*, 2004, vol. 50, No. 9, p. 1178–1192.
269. U. Kühler and S. Tappe. Bilateral gamma distributions and processes in financial mathematics. – *Stochastic Processes and their Applications*, 2007 (to appear).
270. U. Kühler and S. Tappe. On the shapes of bilateral gamma densities. – 2007. Available at:
www.mathematik.uni-muenchen.de/~tappe/Bilateral_Shapes.pdf.
271. S. Kullback and R. A. Leibler. On information and sufficiency. – *Annals of Mathematical Statistics*, 1951, vol. 22, p. 79–86.
272. S. I. Lashkul, A. B. Altukhov, A. D. Gurchenko et al. Poloidal inhomogeneity of the particle fluctuation induced fluxes near the LCFS at lower hybrid heating and improved confinement transition at the FT-2 tokamak. – *Czechoslovak Journal of Physics*, 2005, vol. 55, No. 3, p. 341–348.
273. R. A. Levine and G. Casella. Implementations of the Monte Carlo EM algorithm. – *Journal of Computational and Graphical Statistics*, 2001, vol. 10, p. 422–439.
274. R. F. Levine and J. Fan. An automated (Markov Chain) Monte Carlo EM algorithm. – *Journal of Statistical Computation and Simulation*, 2004, vol. 74, p. 349–359.
275. P. Lévy. *Processus Stochastiques et Mouvement Brownien*. Gauthier–Villars, Paris, 1948.
276. Z. Liang, R. J. Jaszczak, R. E. Coleman. Parameter estimation of finite mixtures using the EM algorithm and information criteria with applications to medical image processing. – *IEEE Transactions Nuclear Science*, 1992, vol. 39, No. 4, p. 901–913.

-
277. A. W. Lo, F. C. MacKinlay and J. Zhang. Econometric models of limit order execution. – *Journal of Financial Economics*, 2002, vol. 65, p. 31–71.
278. Y. Lo, N. R. Mendell and D. B. Rubin. Testing the number of components in a normal mixture. – *Biometrika*, 2001, vol. 88, No. 3, p. 767–778.
279. Y. Lo. Likelihood ratio tests of the number of components in a normal mixture with unequal variances. – *Statistics and Probability Letters*, 2005, vol. 71, p. 225–235.
280. O. Lundberg. *On Random Processes and and their Application to Sickness and Accident Statistics*. Almqvist & Wiksell, Uppsala, 1964 (1st ed. 1940).
281. T. Lux. The stable Paretian hypothesis and the frequency of large returns: an examination of major German stocks. – *Applied Financial Economics*, 1996, vol. 6, p. 463–475.
282. D. B. Madan and E. Seneta. The variance gamma (V.G.) model for share market return. – *Journal of Business*, 1990, vol. 63, p. 511–524.
283. J. M. Maheu and T. H. McCurdy. News arrival, jump dynamics and volatility components for individual stock returns. – *Journal of Finance*, 2004, vol. 59, No. 2 (April), p. 755–793.
284. D. G. Malkiel and Y. Xu. Investigating the behavior of idiosyncratic volatility. – *Journal of business*, 2002, vol. 76, p. 613–644.
285. B. B. Mandelbrot. The variation of certain speculative prices. – *J. Business*, 1963, vol. 36, p. 394–419.
286. B. B. Mandelbrot. The variation of some other speculative prices. – *J. Business*, 1967, vol. 40, p. 393–413.
287. B. B. Mandelbrot. *Fractals and Scaling in Finance. Discontinuity, Concentration, Risk*. Springer, Berlin–Heidelberg, 1997.
288. R. Mantegna and H. E. Stanley. Turbulence and financial markets. – *Nature*, 1996, vol. 383, p. 587–588.
289. R. Mantegna and H. E. Stanley. Stock market dynamics and turbulence: parallel analysis of fluctuation phenomena. – *Physica A*, 1997, vol. 239, p. 255–266.
290. B. Martinet. Regularisation d'inéquations variationnelles par approximations successives. – *Revue Francaise d'Informatique et le Recherche Operationelle*, 1970, vol. 3, p. 154–170.
291. A. G. McKendrick. Applications of mathematics to medical problems. – *Proceedings of the Edinburgh Mathematical Society*, 1926, vol. 44, p. 98–130.

-
292. G. McLachlan and K. Basford. *Mixture Models: Inference and Application to Clustering*. Marcel Dekker, New York, 1988.
293. G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley and Sons, New York, 1997.
294. G. McLachlan and D. Peel. *Finite Mixture Models*. John Wiley and Sons, New York, 2000.
295. P. Medgyessy. *Decomposition of Superpositions of Distribution Functions*. Publishing House of the Hungarian Academy of Sciences, Budapest, 1961.
296. M. M. Meerschaert and H.-P. Scheffler. *Limit Theorems for Continuous Time Random Walks*. Preprint. 18 October 2001. Available at <http://unr.edu/homepage/mcubed/LimitCTRW.pdf>
297. M. M. Meerschaert and H.-P. Scheffler. Stochastic solution of space-time fractional diffusion equations. *Physical Review E*, 2002, v. 65, No. 4, p. 1103–1106.
298. C. Meneveau and K. R. Sreenivasan. The multifractal nature of turbulent energy dissipation. – *Journal of Fluid Mechanics*, 1991, vol. 224, p. 429–484.
299. R. C. Merton. Theory of rational pricing. – *Bell J. Econom. Management Sci.*, 1973, vol. 4, p. 141–183.
300. R. C. Merton. Option pricing when underlying stock returns are discontinuous. – *Journal of Financial Economics*, 1976, vol. 3, p. 125–144.
301. R. Metzler, E. Barkai and J. Klafter. Anomalous transport in disordered systems under the influence of external fields, *Physica A*, 1999, vol. 266, p. 343–350.
302. B. B. van Milligen, E. de la Luna, F. L. Tabares et al. Ballistic transport in TJ-II. – *Nuclear Fusion*, 2002, vol. 42, p. 787–795.
303. S. Mittnik and S. Rachev. Stable distributions for asset returns. – *Appl. Math. Lett.*, 1989, vol. 2, p. 301–304.
304. S. Mittnik and S. Rachev. *Modeling Financial Assets with Alternative Stable Models*. J. Wiley and Sons, New York–Chichester, 1994.
305. I. Monroe. Processes that can be imbedded in Brownian motion. – *Annals of Probability*, 1978, vol. 6, No. 1, p. 42–56.
306. E. W. Montroll and G. H. Weiss. Random walks on lattices. – *Journal of Mathematical Physics*, 1965, vol. 6, p. 167–181.
307. O. Motojima, K. Akaishi, H. Chikaraishi et al. Progress summary of LHD engineering design and construction. – *Nuclear Fusion*, 2000, vol. 40, No. 3, p. 599–610.

-
308. M. M. Navier. Memoire sur les loi du mouvement des fluides. – *Mem. Acad. Roy. Sci.*, 1823, vol. 6, p. 389–440.
309. R. Neal and G. Hinton. A view of the EM algorithm that justifies incremental, sparse and other variants. – In: M. I. Jordan (Ed.). *Learning in Graphical Models*. Kluwer Academic Publishers, 1998, p. 355–368.
310. S. Newcomb. Discussion and results of observations on transits of Mercury from 1677 to 1881. – *Astronomic Papers*, 1882, vol. 1, p. 363–487. Published by the US Nautical Almanac Office.
311. S. F. Nielsen. The stochastic EM algorithm: estimation and asymptotic results. – *Bernoulli*, 2000, vol. 6, No. 3, p. 457–489.
312. A. M. Obukhov. Some specific features of atmospheric turbulence. – *Journal of Fluid Mechanics*, 1962, vol. 13, p. 77–81.
313. M. O’Connell. *Business of Options: Time-Tested Principles and Practice*. Wiley, New York, 2001.
314. G. Oldfield, R. Rogalski and R. Jarrow. An autoregressive jump process for common stock returns. – *J. Financial Economics*, 1977, vol. 5, p. 289–418.
315. V. V. Parail. Heat transport experiments in JET: stiffness and non-locality. – In: *12th International Toki Conf. on Plasma Phys. and Contr. Nucl. Fusion*. Toki, 2001, report O-7.
316. M. Parkinson. The extreme value method for estimating the variance of the rate of return. – *Journal of Business*, 1980, vol. 53, p. 61–68.
317. E. Parzen. On estimation of a probability density and mode. – *Ann. Math. Statist.*, 1962, vol. 33, p. 1065–1076.
318. K. Pearson. Contribution to the mathematical theory of evolution. – *Phil. Trans. Roy. Soc. London*, 1897, vol. A185, p. 71–100 (см. также *Karl Pearson’s Early Statistical Papers*. Cambridge, 1948, p. 1–40).
319. M. A. Pedrosa, M. A. Ochando, J. A. Jimenez et al. Magnetic configuration effects on the TJ-IU torsatron plasma edge turbulence – *Plasma Physics Control Fusion*, 1996, vol. 38, p. 365–373.
320. A. Piryatinska, A. I. Saichev and W. A. Woyczynski. Models of anomalous diffusion: the subdiffusive case. – *Physica A: Statistical and Theoretical Physics*, 2005, vol. 349, No. 3–4, p. 375–420.
321. B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. – *SIAM Journal of Control and Optimization*, 1992, vol. 30, p. 838–855.
322. P. Praetz. The distribution of share prices changes. – *J. Business*, 1972, vol. 45, p. 49–55.

-
323. B. L. S. Prakasa Rao. *Identifiability in Stochastic Models*. Academic Press, Boston–San Diego–New York–London–Sydney–Tokyo–Toronto, 1992.
324. K. Prause. *Modeling Financial Data Using Generalized Hyperbolic Distributions*. Preprint N 48. Freiburg i. Br.: Universität Freiburg, Institut für Mathematische Stochastic, 1997.
325. S. J. Press. A compound events model for security prices. – *J. Business*, 1967, vol. 40, p. 317–335.
326. S. T. Rachev. *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester – New York, 1991.
327. S. T. Rachev and G. Samorodnitsky. Option pricing formulae for speculative prices modelled by subordinated stochastic processes. – *Serdica – Bulgaricae Mathematicae publicationes*, 1993, vol. 19, p. 175–190.
328. *Random House Webster's Concise Dictionary*. Random House, New York, 1993.
329. R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM algorithm. – *SIAM Review*, 1984, vol. 26, No. 2, p. 195–239.
330. J. Rissanen. Minimax entropy estimation of models for vector processes. – In: R. K. Mehra and D. G. Lainiotis (Eds.) *System Identification. Advances and Case Studies*. Academic Press, New York, 1976, p. 97–119.
331. H. Robbins and S. Monro. A stochastic approximation method. – *The Annals of Mathematical Statistics*, 1951, vol. 22, p. 400–407.
332. R. Rockafellar. Monotone operators and the proximal point algorithm. – *SIAM Journal on Control and Optimization*, 1976, vol. 14, p. 877–898.
333. L. C. G. Rogers and S. E. Satchell. Estimating variance from high, low and closing prices. – *Annals of Applied Probability*, 1991, vol. 1, p. 504–512.
334. M. Romanowski. *Random Errors in Observations and the Influence of Modulation of Their Distribution*. Verlag Konrad Wittwer. Stuttgart, 1979.
335. M. Rosenblatt. Remarks on some nonparametric estimates of a density function. – *Ann. Math. Statist.*, 1956, vol. 27, p. 832–837.
336. S. A. Ross. Information and volatility: the no–arbitrage martingale approach to timing and resolution irrelevancy. – *Journal of Finance*, 1989, vol. 44, No. 1, p. 1–17.
337. A. I. Saichev and G. M. Zaslavsky. Fractional kinetic equations: solutions and applications. – *Chaos*, 1997, vol. 7, p. 753–764.
338. G. Samorodnitsky and M. Taqqu. *Stable Non–Gaussian Random Processes*. Chapman and Hall, London, 1994.

-
339. K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge, 1999.
340. H. Scher and M. Lax. Stochastic transport in a disordered solid. – *Physical Review, B*, 1973, vol. 7, p. 4491–4502.
341. W. R. Schneider. Stable distributions: Fox function representation and generalization. – In: *Stochastic Processes in Classical and Quantum Systems*, S. Albeverio, G. Casati, D. Merlini, Eds. Springer, Berlin, 1986, p. 497–511.
342. W. R. Schneider. Generalized one-sided stable distributions. – In: *Proc. Second BiBos Symp.*, Springer, Berlin, 1987, p. 269–287.
343. W. R. Schneider and W. Wyss. Fractional diffusion and wave equations. – *J. Math. Phys.*, 1989, vol. 30, p. 134–144.
344. W. Schoutens and J. L. Teugels. Lévy processes, polynomials and martingales. – *Communications in Statistics – Stochastic Models*, 1998, vol. 14, No 1–2, p. 335–349.
345. W. Schoutens. *Stochastic Processes and Orthogonal Polynomials*. Lecture Notes in Statistics, vol. 146. Springer, 2000.
346. W. Schoutens. *The Meixner Processes in Finance*. EURANDOM Report 2001–002, EURANDOM, Eindhoven, 2000.
347. W. Schoutens. *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley, New York, 2003.
348. V. Seshadri. *The Inverse Gaussian Distribution: Statistical Theory and Applications*. Springer, Berlin–Heidelberg–New York, 1999.
349. G. Schwartz. Estimating the dimension of a model. – *The Annals of Statistics*, 1978, vol. 6, p. 461–464.
350. M. Shaked. Bounds on the distance of a mixture from its parent distribution. – *Journal of Applied Probability*. 1981, vol. 18, No. 4, p. 853–863.
351. A. N. Shiryaev. *Kolmogorov and the Turbulence*. – Thiele Centre for Applied Mathematics in Natural Science. Research report No. 04, February 2006.
352. M. F. Shlesinger. Williams–Watts dielectric relaxation: a fractal time stochastic process. – *J. Stat. Phys.*, 1984, vol. 36, p. 639–648.
353. M. F. Shlesinger, B. J. West and J. Klafter. Lévy dynamics of enhanced diffusion. – *Physical Review Letters*, 1987, vol. 58, p. 1110–1103.
354. N. N. Skvortsova, G. M. Batanov, L. V. Kolik et al. – *J. Plasma Fusion Rev. Series*, 2002, vol. 5, p. 328–332.

-
355. N. N. Skvortsova, V. Yu. Korolev, G. M. Batanov et al. Statistical analysis and modeling of turbulent fluxes in plasma of L-2M stellarator and FT-2 tokamak – *J. Plasma Phys. Contr. Fusion*, 2006.
356. N. N. Skvortsova, G. M. Batanov, L. V. Kolik et al. Recent results of studies of fluctuations in high-temperature plasma of modern stellarators by microwave scattering techniques. – In: *VI Workshop “Strong Microwaves in Plasma”*, Nizhnii Novgorod, 2005.
357. A. F. M. Smith. Present position and potential developments: some personal views of Bayesian statistics. – *Journal of the Royal Statistical Association*, 1984, vol. 147(3), p. 245–259.
358. K. R. Sreenivasan and R. A. Antonia. The phenomenology of small-scale turbulence. – *Ann. Rev. Fluid Mech.*, 1997, vol. 29, p. 435–472.
359. S. V. Stigler. Simon Newcomb, Percy Daniell, and the history of robust estimation 1885–1920. *J. Amer. Statist. Assoc.*, 1973, vol. 68, p. 872–879.
360. G. Stolovitzky, P. Kailasnath and K. R. Sreenivasan. Kolmogorov’s refined similarity hypothesis. – *Phys. Rev. Letters*, 1992, vol. 69, p. 1178–1181.
361. G. Stolovitzky and K. R. Sreenivasan. Kolmogorov’s refined similarity hypotheses for turbulence and general stochastic processes. – *Rev. Modern Phys.*, 1994, vol. 66, p. 229–239.
362. N. Sugiura. Further analysis of the data by Akaike’s information criterion and the finite corrections. – *Communications in Statistics. Theory and Methods*, 1978, vol. A7, p. 13–26.
363. K. Takeuchi. Distribution of informational statistics and a criterion of model fitting. – *Suri-Kagaku (Mathematical Sciences)*, 1976, vol. 153, p. 12–18 (in Japanese).
364. G. M. Tallis. The identifiability of mixtures of distributions. – *Journal of Applied Probability*, 1969, vol. 6, No. 2, p. 389–398.
365. W. Y. Tan and W. C. Chang. Some comparisons of the method of moments and the method of maximum likelihood in estimating parameters of a mixture of normal densities. – *Biometrics*, 1971, vol. 27, No. 2, p. 489–501.
366. M. Tanner. *Tools for Statistical Inference*. Springer-Verlag, New York, 1993.
367. H. Teicher. Identifiability of mixtures. – *Ann. Math. Stat.*, 1961, vol. 32, p. 244–248.
368. H. Teicher. Identifiability of finite mixtures. – *Ann. Math. Stat.*, 1963, vol. 34, No. 4, p. 1265–1269.

-
369. D. M. Titterington, A. F. Smith and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley and Sons, Chichester–New York–Brisbane–Toronto–Singapore, 1987.
370. J. W. Tukey. A survey of sampling from contaminated distributions. – In: I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow and H. B. Mann (Eds.) *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling*. Stanford University Press, Stanford, CA, 1960, p. 448–485.
371. V. V. Uchaikin and V. M. Zolotarev. *Chance and Stability. Stable Distributions and their Applications*. VSP, Utrecht, 1999.
372. N. G. Ushakov. *Selected Topics in Characteristic Functions*. VSP, Utrecht, 1999.
373. A. Vincent and M. Maneguzzi. The spatial structure and statistical properties of homogeneous turbulence. – *Journal of Fluid Mechanics*, 1991, vol. 225, p. 1–25.
374. M. S. Waterman. A restricted least squares problem. – *Technometrics*, 1974, vol. 16, No. 1, p. 135–136.
375. G. C. G. Wei and M. A. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. – *Journal of the American Statistical Association*, 1990, vol. 85, p. 699–704.
376. S. Wei and C. Zhang. Idiosyncratic risk does not matter: A re-examination of the relationship between average returns and average volatilities. – *Journal of Banking and Finance*, 2005, vol. 29, p. 603–621.
377. S. Wei and C. Zhang. Why did individual stocks become more volatile? – *Journal of Business*, 2006, vol. 79, p. 259–292.
378. G. H. Weis and R. J. Rubin. Random walks – theory and selected applications. – *Advances in Chemical Physics*, 1983, vol. 52, p. 363–505.
379. B. J. West and W. Deering. Fractal physiology for physicists: Lévy statistics. – *Phys. Rep.*, 1994, vol. 246, p. 1–100.
380. S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. – *The Annals of Mathematical Statistics*, 1938, vol. 9, p. 60–62.
381. J. H. Wolfe. Pattern clustering by multivariate mixture analysis. – *Multivariate Behavioral Research*, 1970, vol. 5, p. 329–350.
382. C. F. Wu. On the convergence properties of the EM-algorithm. – *The Annals of Statistics*, 1983, vol. 11, No. 1, p. 95–103.
383. L. Xu and M. I. Jordan. On convergence properties of the EM algorithm for Gaussian mixtures. – *Neural Computation*, 1996, vol. 8, p. 129–151.

- 384. S. J. Yakowitz and J. D. Spragins. On the identifiability of finite mixtures. – *Ann. Math. Statistics*, 1968, vol. 39, p. 209–214.
- 385. D. Yang and Q. Zhang. Drift independent volatility estimation based on high, low and close prices. – *Journal of Business*, 2000, vol. 73, p. 477–491.
- 386. G. M. Zaslavsky. Chaos, Fractional Kinetics and Anomalous Transport. *Physics Reports. A Review Section of Physics Letters*. North Holland, Amsterdam, 2002. Also available at www.sciencedirect.com
- 387. G.M. Zaslavsky, M.Edelman, H.Weiltzner et al. Large-scale behavior of tokamak density fluctuations. – *Plasma Physics*, 2000, vol. 7, p. 3691–3698.
- 388. Y. Zhu, R. A. Antonia and I. Hosokawa. Refined similarity hypotheses for turbulent velocity and temperature fields. – *Phys. Fluids*, 1995, vol. 7, p. 1637–1648.
- 389. G. Zumofen, J. Klafter and A. Blumen. Lévy walks and propagators in intermittent chaotic systems. – *Physical Review, E*, 1993, vol. 47, p. 2183–2186.

Иллюстрации

Иллюстрации

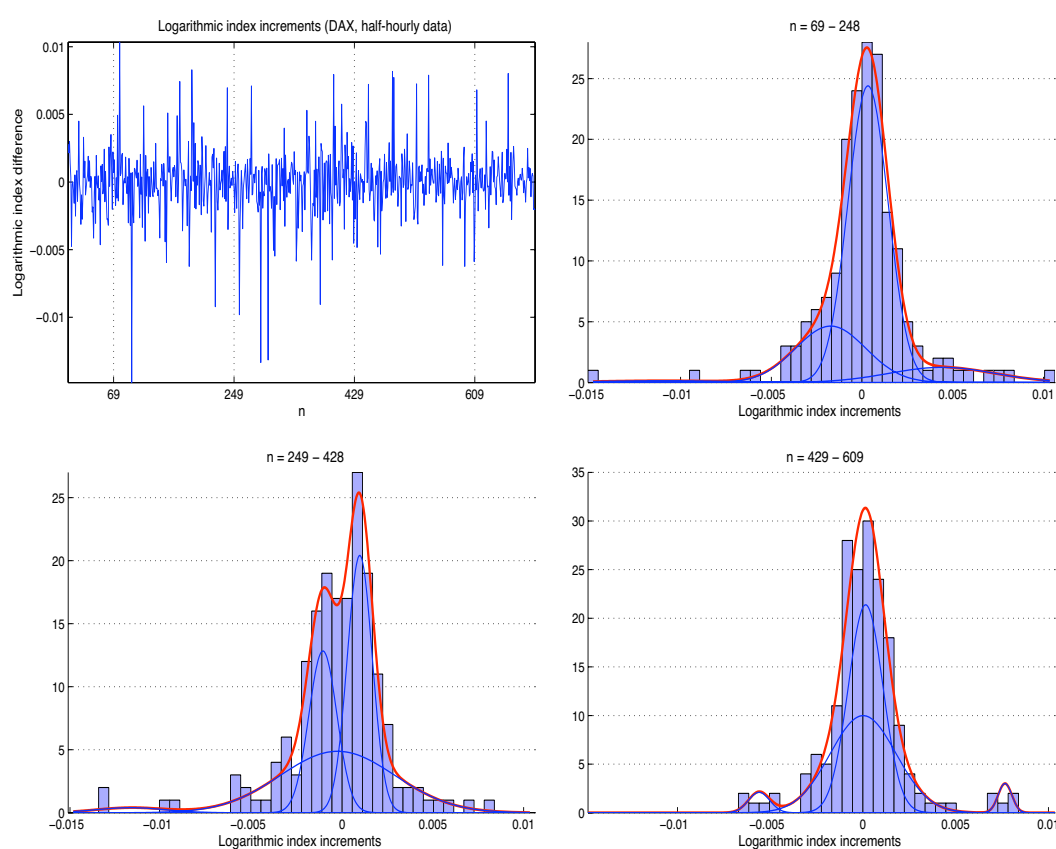


Рис. 1: Пример подгонки четырехкомпонентной смеси нормальных распределений к приращениям одного и того же временного ряда (индекс DAX) на непересекающихся отрезках временного ряда одинаковой длины (по 180 наблюдений). На всех трех интервалах степень согласия очень высока, но все три подогнанные смеси имеют разные параметры, так что ‘универсальной’ четырехкомпонентной смеси, обеспечивающей столь же высокое согласие, нет

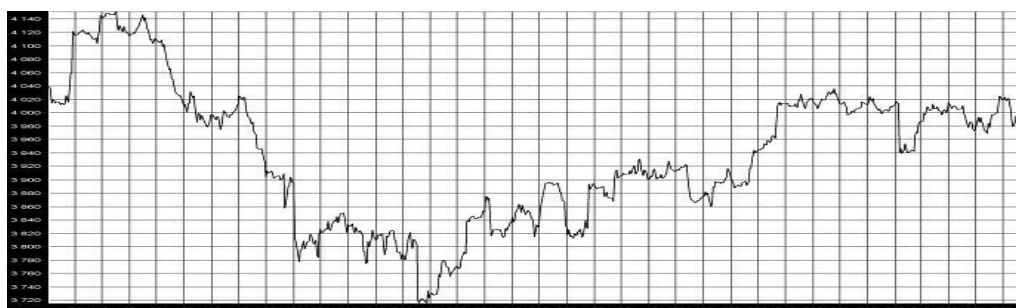


Рис. 2: Индекс немецкой фондовой биржи XETRA DAX. Получасовые данные

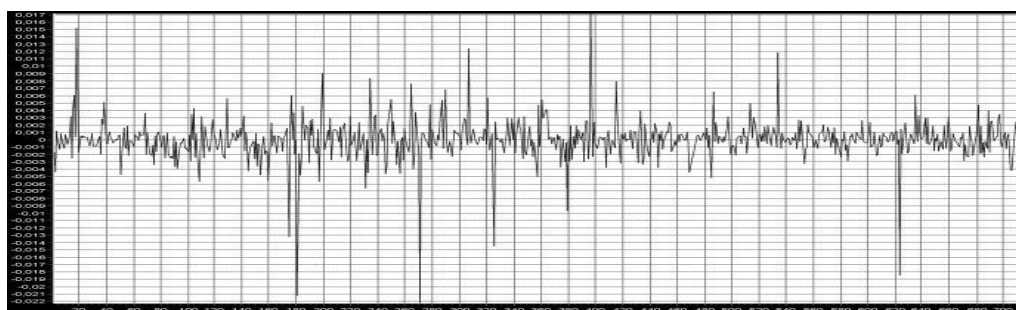


Рис. 3: Получасовые приращения логарифмов индекса DAX

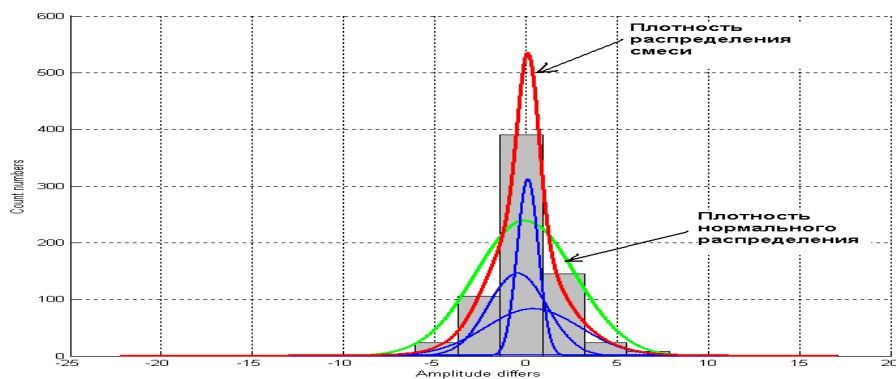


Рис. 4: Графики смеси нормальных законов, подогнанной с помощью ЕМ-алгоритма к получасовым приращениям логарифмов индекса XETRA DAX, а также компонент смеси (с учетом весовых коэффициентов), плотности нормального распределения с соответствующими параметрами и гистограмма приращений логарифмов исследуемого ряда.

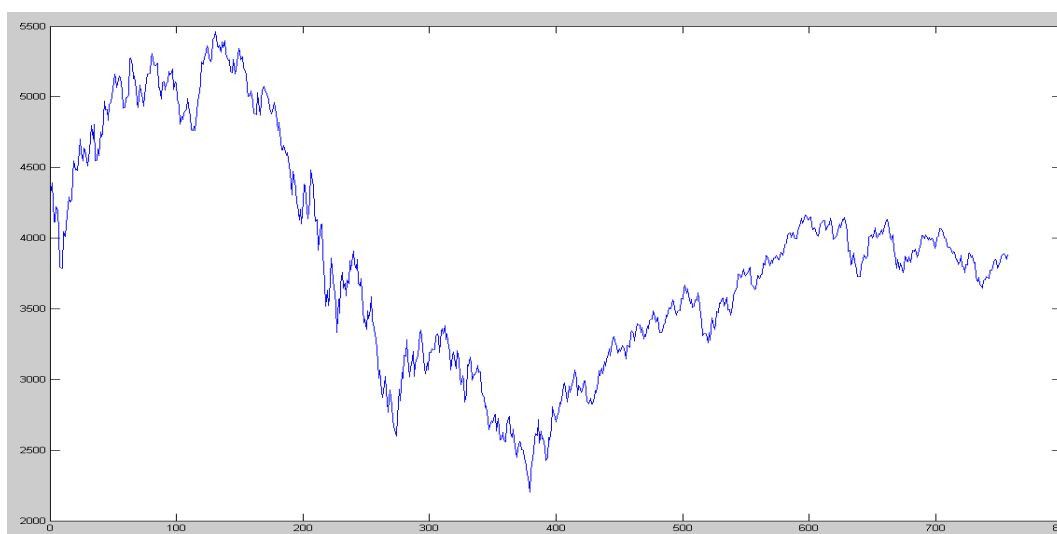


Рис. 5: Индекс немецкой фондовой биржи XETRA DAX. Дневные данные

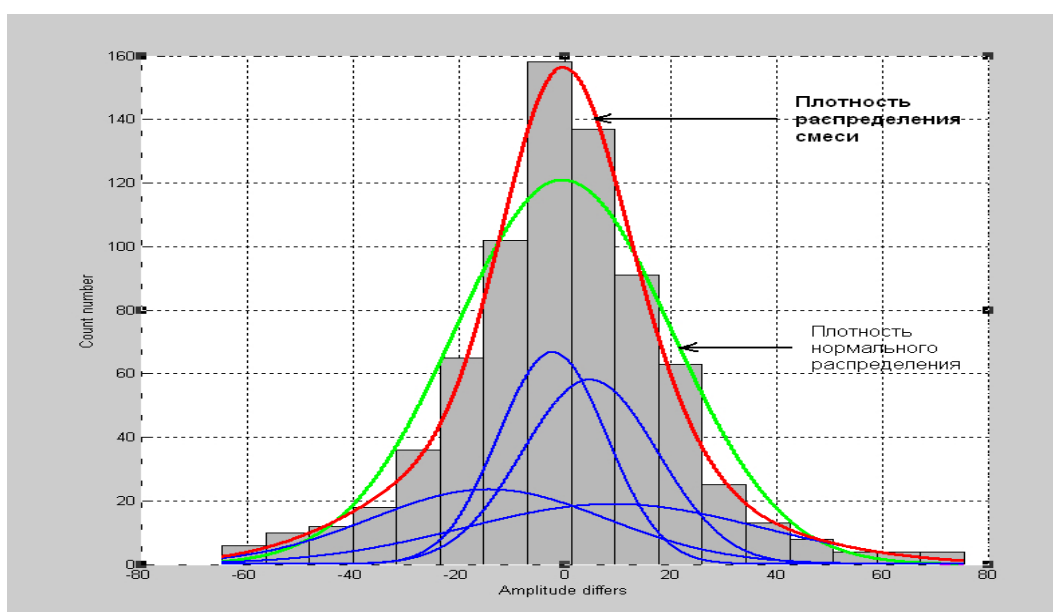


Рис. 6: Графики смеси нормальных законов, подогнанной с помощью ЕМ-алгоритма к дневным приращениям логарифмов индекса XETRA DAX, а также компонент смеси (с учетом весовых коэффициентов), плотности нормального распределения с соответствующими параметрами и гистограмма приращений логарифмов исследуемого ряда.

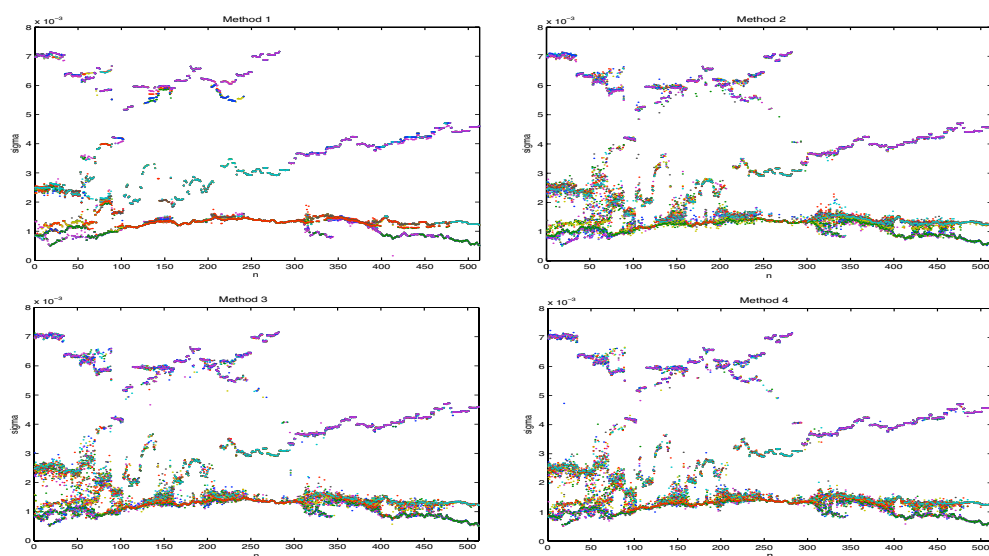


Рис. 7: Портреты волатильности временного ряда приращений логарифмов индекса DAX (6.1.1), вычисленные в соответствии с четырьмя способами организации выбора начальных приближений

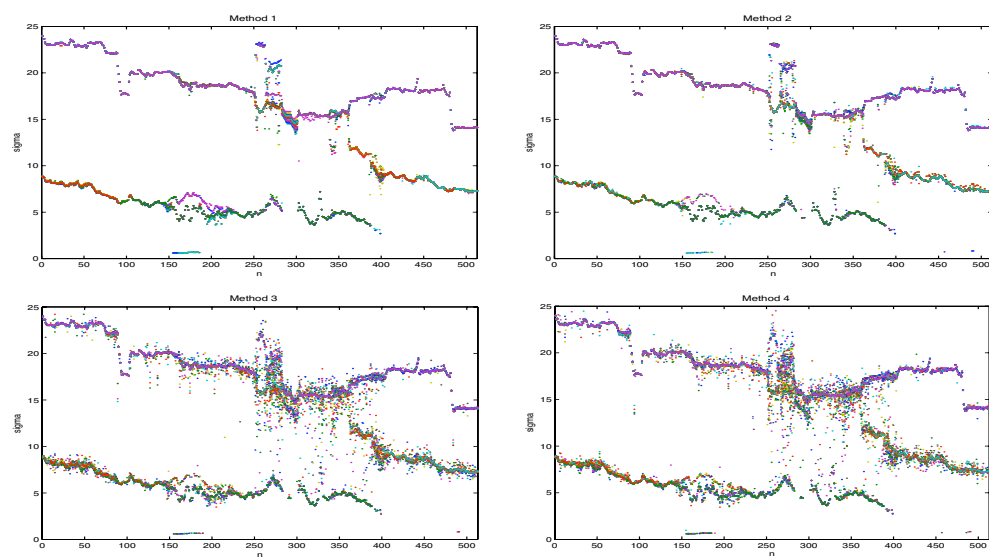


Рис. 8: Портреты волатильности временного ряда, полученного вычитанием из исходного ряда индекса DAX скользящего среднего (см. (6.1.2)), вычисленные в соответствии с четырьмя способами организации выбора начальных приближений

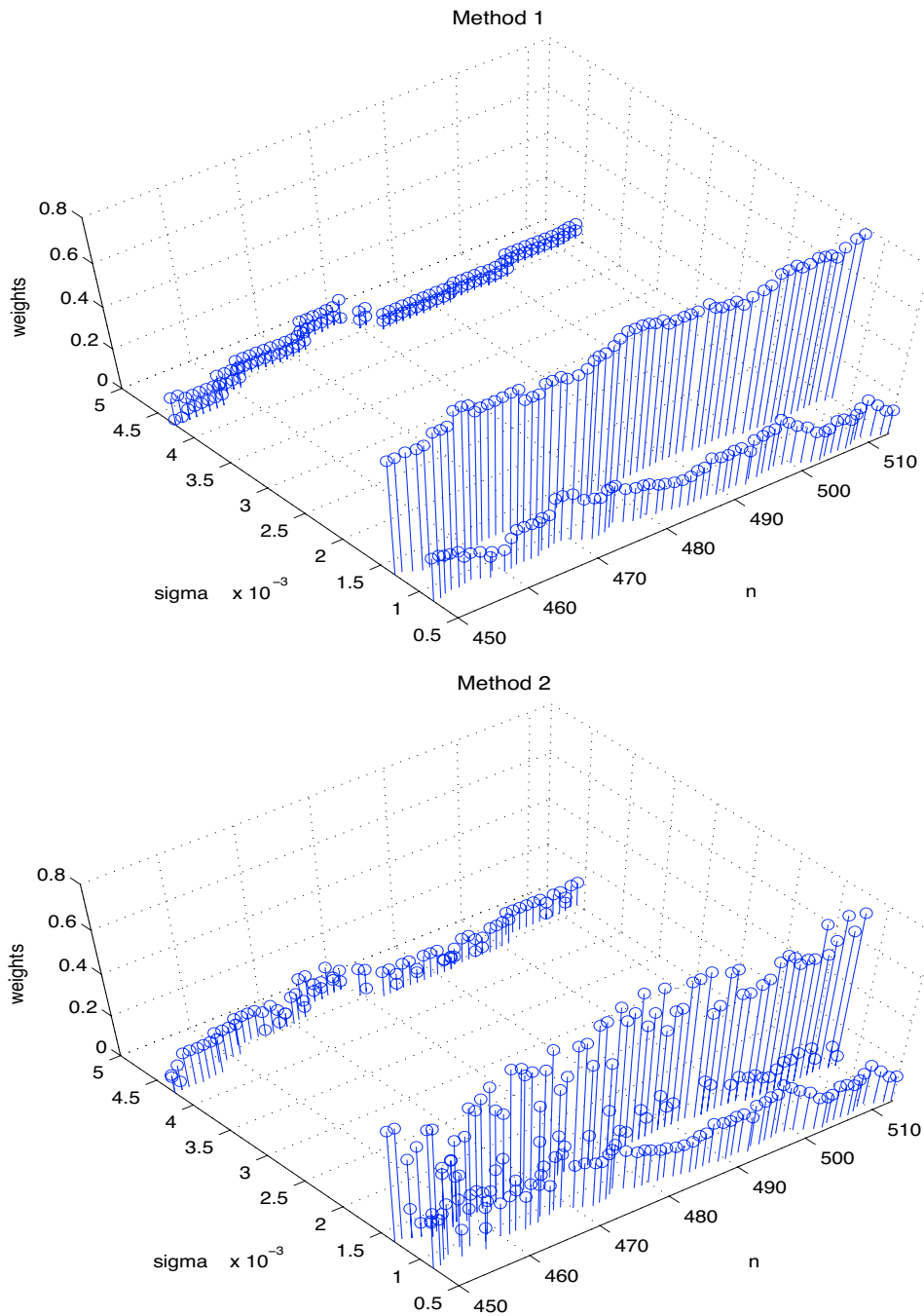


Рис. 9: Визуализация влияния начального выбора весов на оценку компонент волатильности, получаемую СРС-методом. На верхнем графике начальные приближения весов одни и те же для всех положений скользящего окна. На нижнем графике при каждом новом сдвиге начальные приближения для весов выбираются заново. (Процесс (6.1.1) приращений логарифма индекса.)

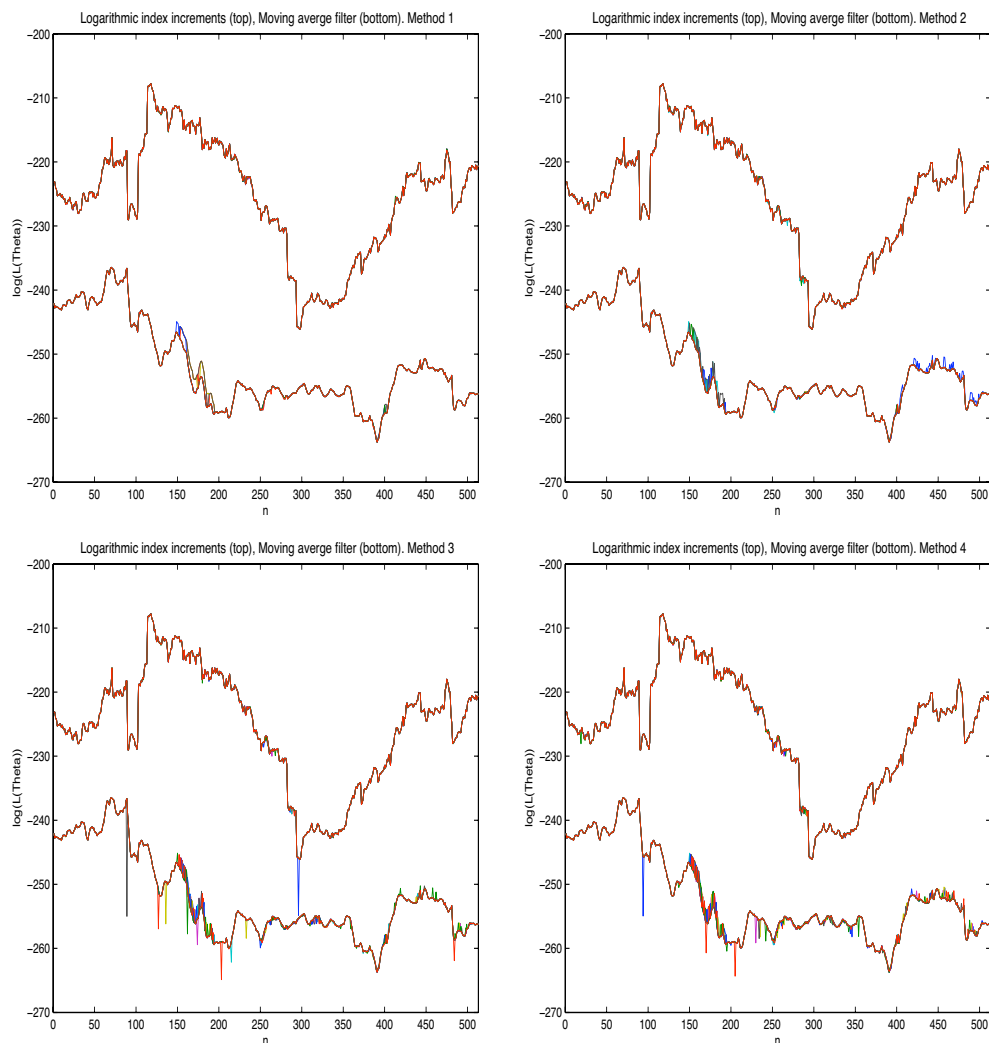


Рис. 10: Логарифмы функций правдоподобия, соответствующих смесям нормальных законов с параметрами, оцененными СРС-методом. Смеси нормальных законов глобально более правдоподобны в качестве моделей распределений приращений логарифмов индекса, чем в качестве моделей распределений остатков временного ряда после отфильтровывания скользящего среднего. Заметны различия между четырьмя методами, подчеркивающие выводы, сделанные в тексте

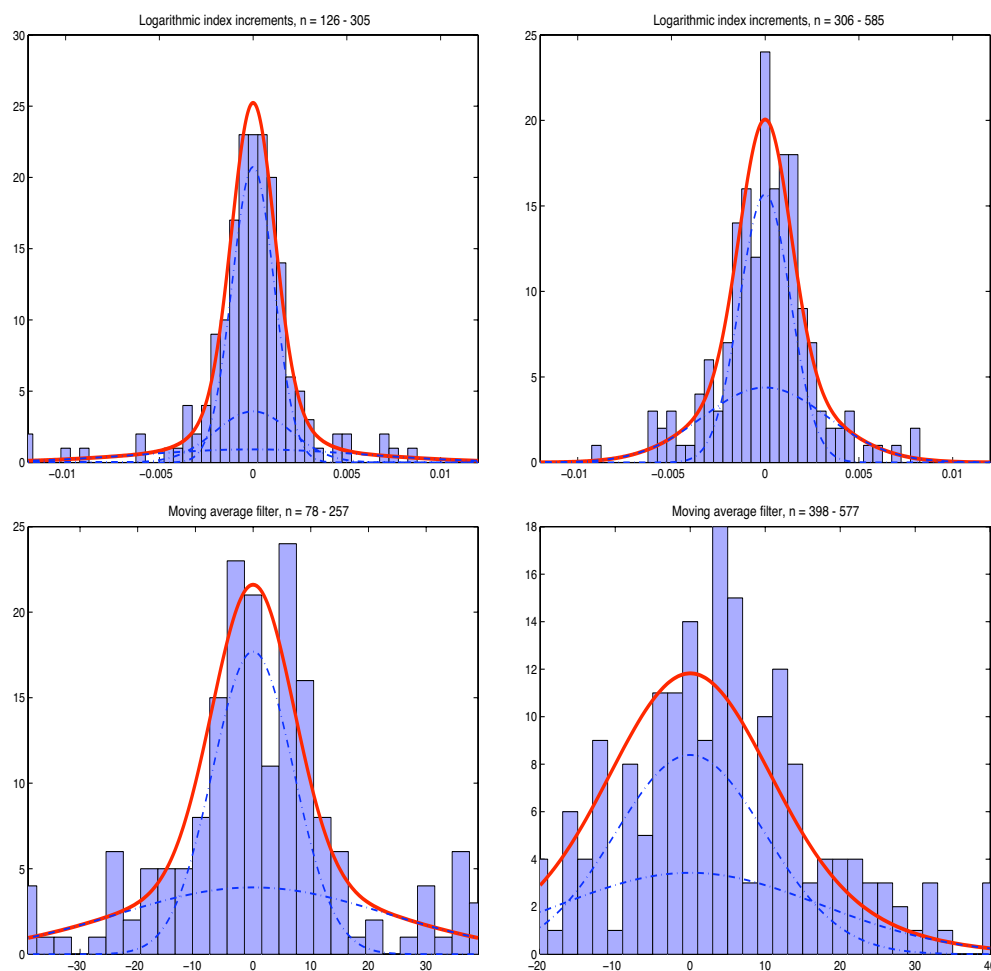


Рис. 11: Четыре примера, иллюстрирующие качество модели, имеющей вид конечной смеси нормальных законов. Для $n = 126 \div 305$, интервала максимума (по n) функции правдоподобия, гистограмма приращений логарифмов индекса и график подогнанной смеси нормальных плотностей очень похожи. Точно так же для $n = 306 \div 585$, интервала минимума (по n) функции правдоподобия различия между гистограммой приращений логарифмов индекса и графиком подогнанной смеси нормальных плотностей очень невелики. Для данных типа остатков после отфильтровывания скользящего среднего картина несколько иная: для интервала максимума (по n) функции правдоподобия ($n = 78 \div 257$) графики схожи, однако для интервала минимума (по n) функции правдоподобия ($n = 398 \div 577$) различия между гистограммой и подогнанной смесью очевидны

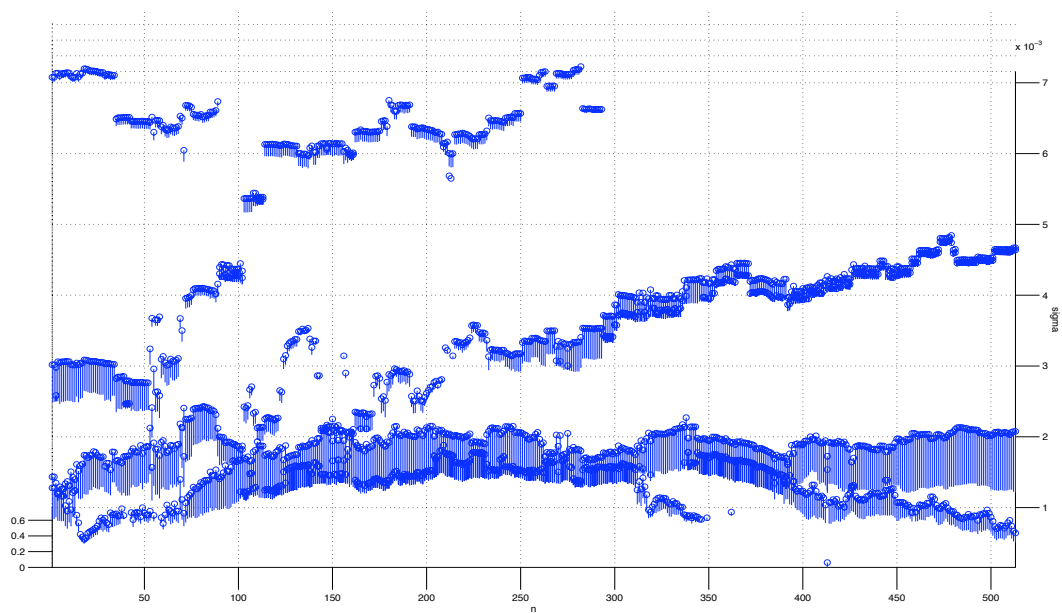


Рис. 12: Трехмерный ‘стерженьковый’ график диффузионных компонент волатильности и соответствующих весов

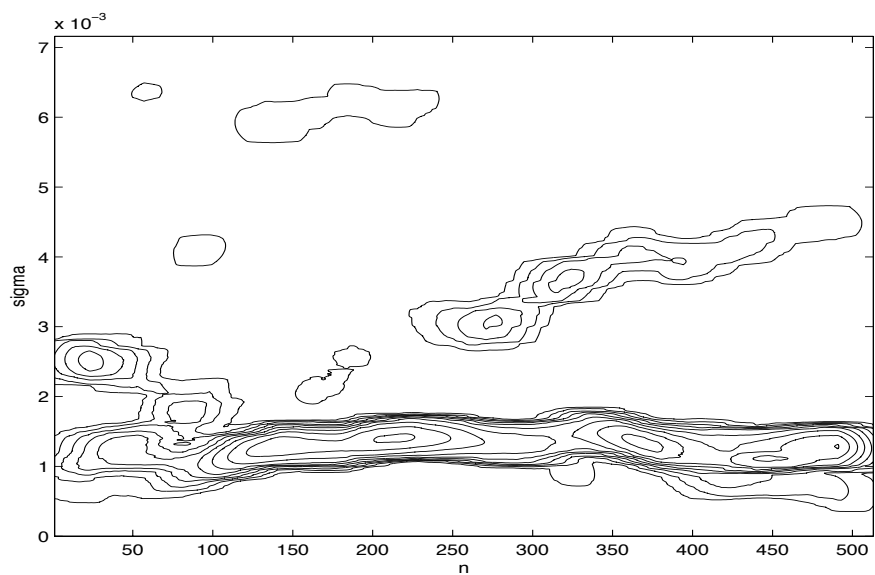


Рис. 13: Контурный (топографический) график диффузионных компонент волатильности и соответствующих весов

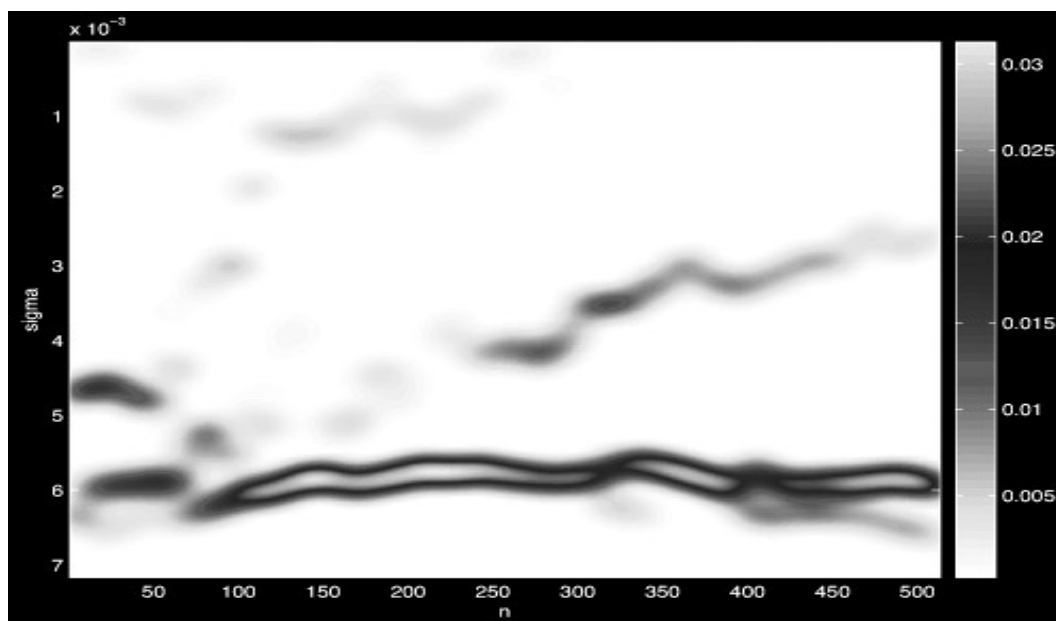


Рис. 14: Двумерный ‘температурный’ график диффузионных компонент волатильности и соответствующих весов

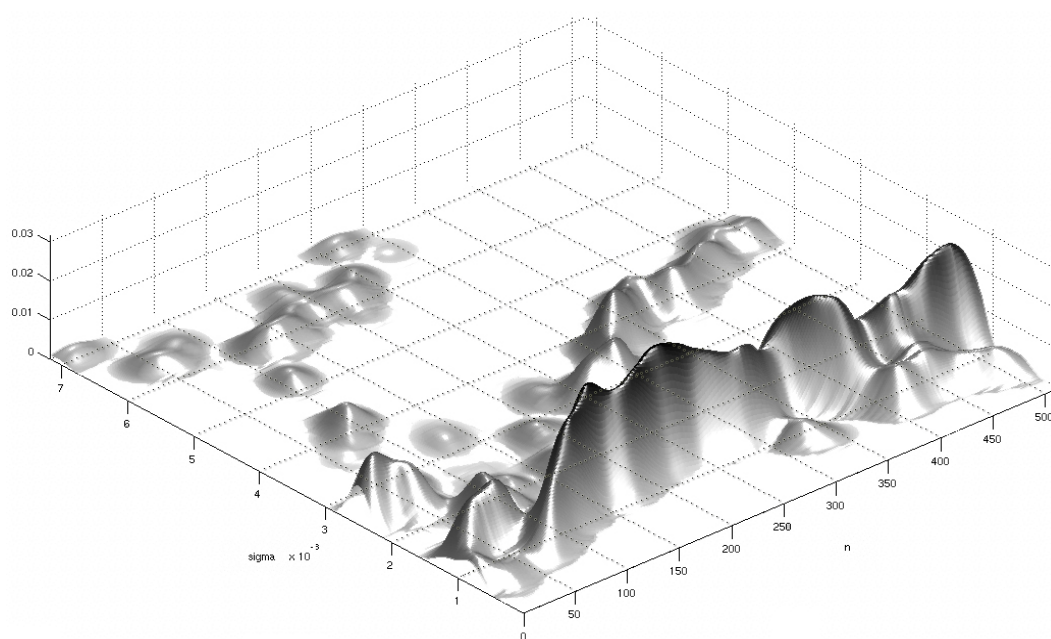


Рис. 15: Трехмерный ‘температурный’ график (‘температурная’ поверхность) диффузионных компонент волатильности и соответствующих весов

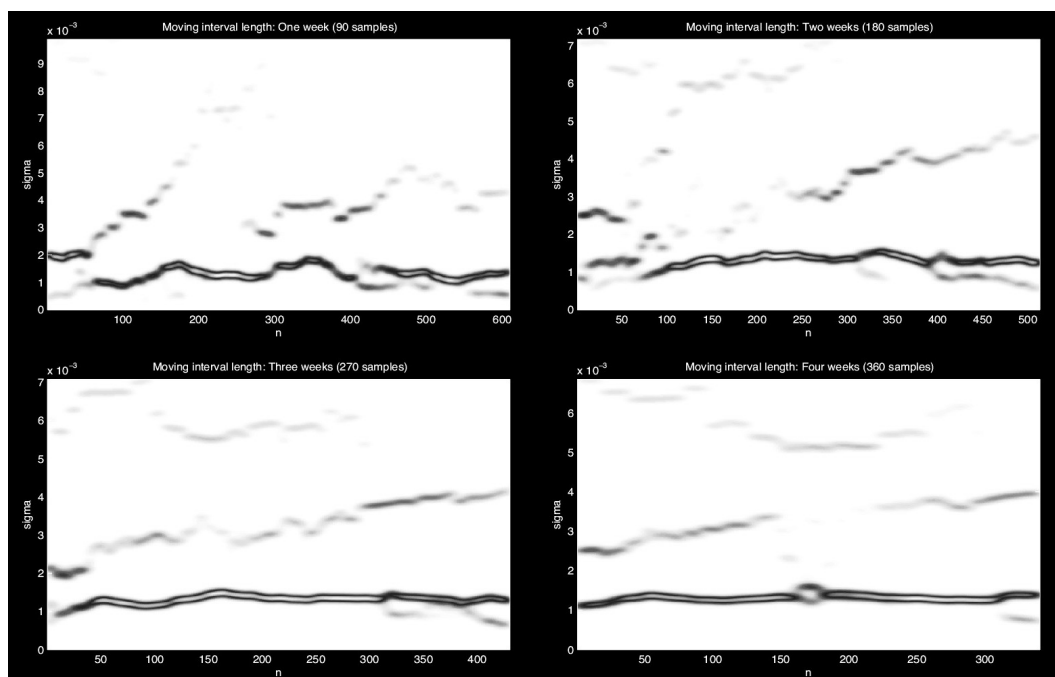


Рис. 16: ‘Температурные’ графики, иллюстрирующие влияние длины скользящего интервала (ширины окна) на результат, получаемый СРС-методом

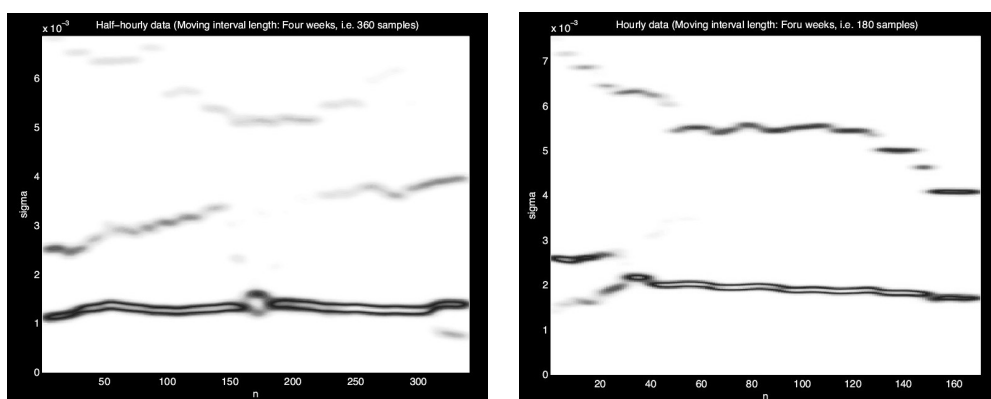


Рис. 17: ‘Температурные’ графики, иллюстрирующие влияние длины интервала времени между последовательными наблюдениями на результат, получаемый СРС-методом. Полукасовые и часовые данные