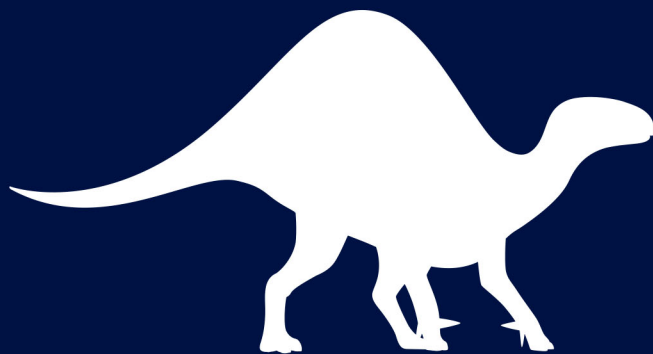


НАССИМ НИКОЛАС ТАЛЕБ

СТАТИСТИЧЕСКИЕ  
ПОСЛЕДСТВИЯ  
ЖИРНЫХ ХВОСТОВ

О НОВЫХ  
ВЫЧИСЛИТЕЛЬНЫХ ПОДХОДАХ  
К ПРИНЯТИЮ РЕШЕНИЙ



НАССИМ НИКОЛАС ТАЛЕБ

# СТАТИСТИЧЕСКИЕ ПОСЛЕДСТВИЯ ЖИРНЫХ ХВОСТОВ

О НОВЫХ  
ВЫЧИСЛИТЕЛЬНЫХ ПОДХОДАХ  
К ПРИНЯТИЮ РЕШЕНИЙ



УДК 336.7/519.2(075.4)

ББК 65.9(2)-97(3)

T16

Nassim Nicholas Taleb

STATISTICAL CONSEQUENCES OF FAT TAILS

Real World Preasymptotics, Epistemology, and Applications

The Technical Incerto Collection

*Перевод с английского Виктора Боруна*

### **Талеб Нассим Николас**

T16 Статистические последствия жирных хвостов : О новых вычислительных подходах к принятию решений / Нассим Николас Талеб ; [пер. с англ. В. Ф. Боруна]. — М. : КоЛибри, Азбука-Аттикус, 2024. — 480 с. ; ил.

ISBN 978-5-389-19584-4

Новая книга всемирно известного мыслителя, автора «Черного лебедя» Нассима Николаса Талеба открывает серию The Technical Incerto Collection и посвящена тем классам статистических распределений, от которых можно ждать экстремальных событий.

Если вы не дружите с графиками и формулами, то из этой книги почерпнете информацию только про скандалы и разоблачение горе-ученых. Если вы учили математическую статистику, эта книга поможет вам переучиться. Если вы студент или ученый, эта книга — бесценный мастер-класс. Впервые под одной обложкой собраны исследовательские статьи Талеба и его учеников, где в неповторимом талебовском стиле живо и ярко прослеживается ход мысли прикладного математика, сталкивающегося с жизненной задачей, не зная, можно ли ее решить аналитически, с чего начать, за что хвататься, — но настроенного пустить в ход, если понадобится, весь арсенал классической и современной математики и всю мощь компьютеров. Автор и его последователи щедро делятся с читателем своими ранними догадками, интуицией и аналогиями, которые помогли им в итоге найти решение.

«Книги серии Incerto посвящены выживанию в реальном мире с его структурой неопределенности, которая слишком сложна для нашего понимания. Цикл ставит целью объединить пять областей знания, связанных с жирными хвостами и экстремальными событиями: в математике, философии, общественных науках, теории контрактов и теории принятия решений, — с опытом профессионалов». (Нассим Николас Талеб)

УДК 336.7/519.2(075.4)

ББК 65.9(2)-97(3)

ISBN 978-5-389-19584-4

© Nassim Nicholas Taleb, 2020

© Боруна В. Ф., перевод на русский язык, 2023

© Издание на русском языке, оформление.

ООО «Издательская Группа «Азбука-Аттикус», 2023

КоЛибри®

## СОАВТОРЫ<sup>1</sup>

Паскуале Чирилло (главы 13, 15 и 16)

Рафаэль Дуади (глава 14)

Андреа Фонтанари (глава 13)

Эльетт Жиман (глава 25)

Дональд Жиман (глава 25)

Эспен Хог (глава 22)

Сотрудники хедж-фонда Universa Investments (глава 23)

Издание подготовлено на основе диссертации Андре Миду и адаптации Лоренцо Пантьери для *Ars Classica*.

С глубочайшей признательностью Андре и Лоренцо.

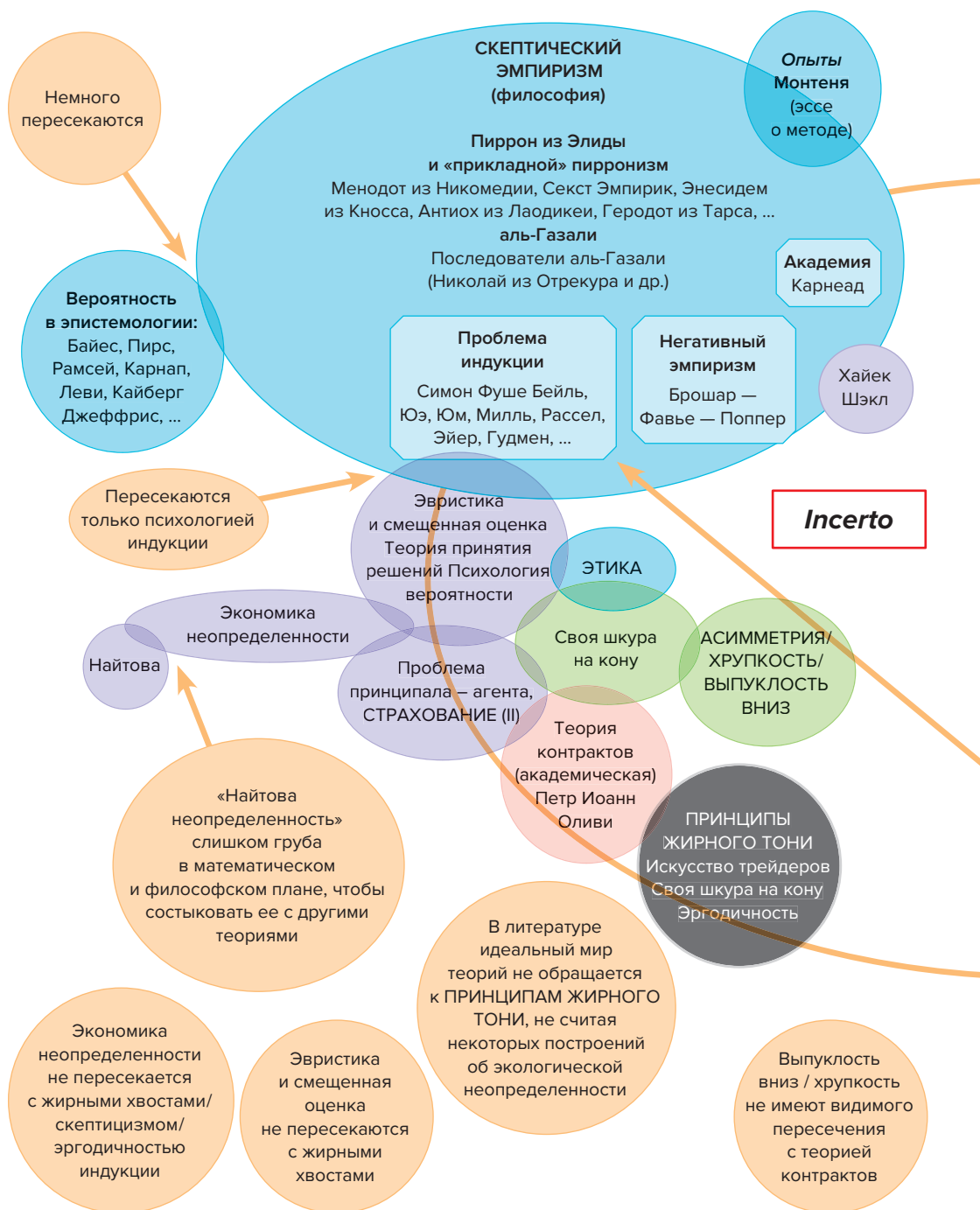
В организации издательской деятельности поддержку оказала компания Scribe Media; особая благодарность Такеру Максу, Элли Коул, Заку Обронту и Эрике Хоффман.

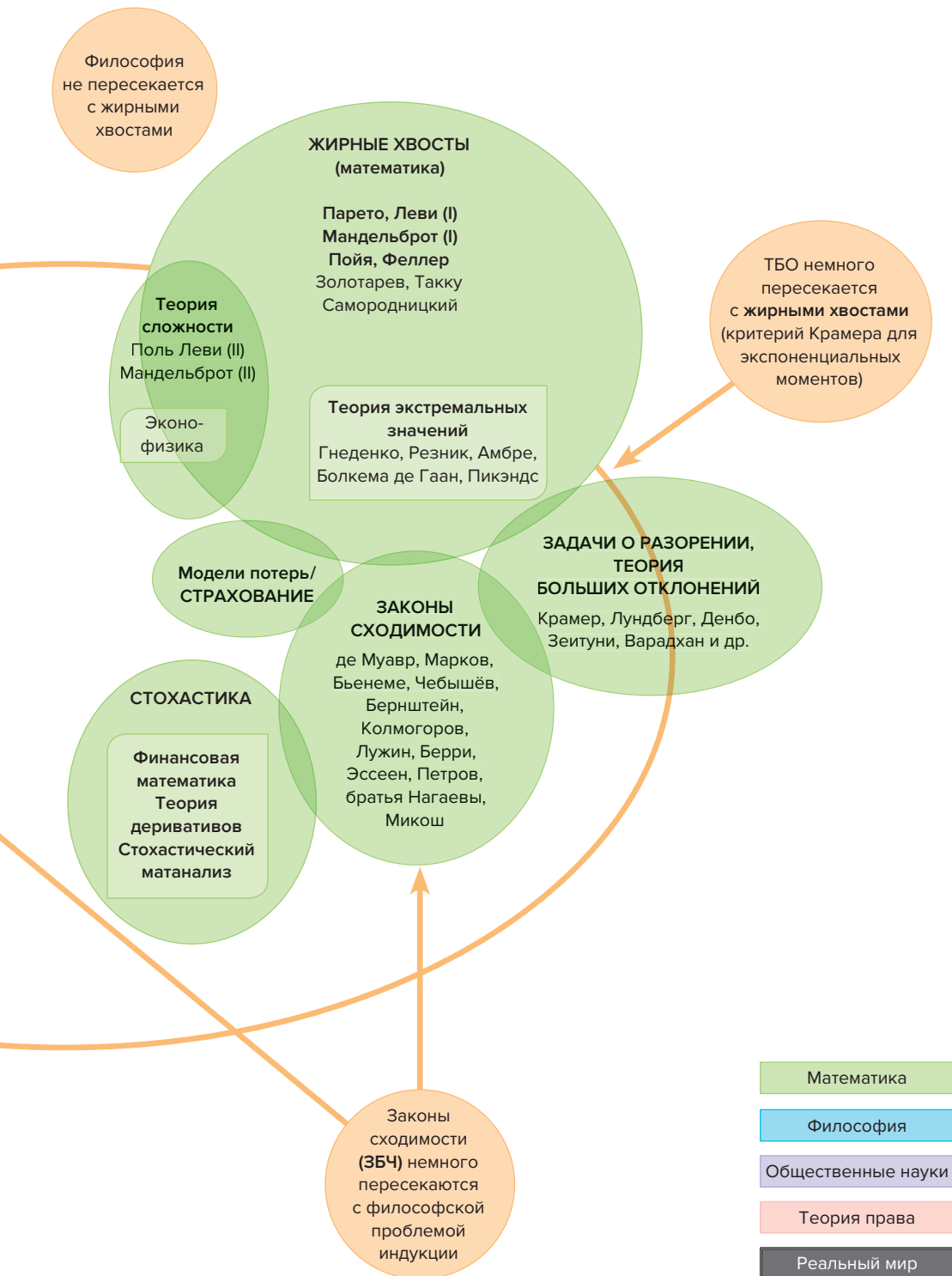
---

<sup>1</sup> Их статьи: [45, 46, 47, 48, 95, 106, 126, 165, 224, 227, 228, 229, 231, 232, 233, 234, 243, 244, 245]. — Прим. автора.



## Генеалогическая карта цикла *Incerto*





Нетехнические главы отмечены звездочкой \*; главы, посвященные дискуссии, отмечены типографским крестиком †; адаптированные версии статей в рецензируемых журналах — двойным типографским крестиком ‡.

Главы нумеруются арабскими цифрами, но вводные и другие короткие главки (отличные от приложений и от полноценных глав) индексируются буквами А, В и т. д.

## СОДЕРЖАНИЕ

1. Пролог** .....	15
ЗАМЕЧАНИЕ О ТЕРМИНАХ .....	17
БЛАГОДАРНОСТИ .....	17
2. ГЛОССАРИЙ, ОПРЕДЕЛЕНИЯ И ОБОЗНАЧЕНИЯ .....	19
2.1. ОБЩИЕ ОБОЗНАЧЕНИЯ И ЧАСТО ИСПОЛЬЗУЕМЫЕ СИМВОЛЫ .....	19
2.2. Систематический каталог общих и идиосинкразических понятий .....	22
2.2.1. Класс степенного закона $\mathfrak{F}$ .....	22
2.2.2. Закон больших чисел (слабый) .....	23
2.2.3. Центральная предельная теорема (ЦПТ) .....	23
2.2.4. Закон средних чисел, или Предасимптотика .....	24
2.2.5. Показатель каппа .....	24
2.2.6. Эллиптическое распределение .....	25
2.2.7. Статистическая независимость .....	25
2.2.8. Устойчивое распределение (устойчивое по Леви) .....	25
2.2.9. Многомерное устойчивое распределение .....	26
2.2.10. Точка Караматы .....	26
2.2.11. Субэкспоненциальность .....	26
2.2.12. t-распределение Стьюдента как прокси .....	27
2.2.13. Круг цитирования .....	27
2.2.14. Погоня за рентой в научном мире .....	28
2.2.15. Псевдоэмпиризм, или Проблема Пинкера .....	28
2.2.16. Предасимптотика .....	29
2.2.17. Стохастизация .....	29
2.2.18. Стоимость под риском, условная стоимость под риском .....	30
2.2.19. Своя шкура на кону .....	30
2.2.20. График MS .....	31
2.2.21. Максимальный аттрактор (MDA) .....	31
2.2.22. Подмена интеграла в литературе для психологов .....	31
2.2.23. Попытка вынести вероятность за скобку (еще одна типичная ошибка) .....	32
2.2.24. Линейка Витгенштейна .....	32
2.2.25. Черные лебеди .....	32
2.2.26. Выборочная функция распределения ненаблюдаема эмпирически .....	33
2.2.27. Скрытый хвост .....	34
2.2.28. Теневой момент .....	35
2.2.29. Зависимость в хвосте .....	35
2.2.30. Метавероятность .....	35
2.2.31. Динамическое хеджирование .....	35

## Часть I. ЖИРНЫЕ ХВОСТЫ И ИХ ПОСЛЕДСТВИЯ, ЗНАКОМСТВО

3. НЕТЕХНИЧЕСКИЙ ОБЗОР — ЛЕКЦИЯ В КОЛЛЕДЖЕ ДАРВИНА*†	39
3.1. О РАЗЛИЧИИ МЕЖДУ ТОНКИМ И ЖИРНЫМ ХВОСТОМ	39
3.2. ХВОСТ, ВЛИЯЮЩИЙ СОБАКАМИ: ИНТУИТИВНО	44
3.3. ДОПОЛНИТЕЛЬНЫЕ КАТЕГОРИИ И ЧТО ИЗ НИХ СЛЕДУЕТ	44
3.4. ОСНОВНЫЕ СЛЕДСТВИЯ И КАК ОНИ СВЯЗАНЫ С ДАННОЙ КНИГОЙ	49
3.4.1. Прогнозирование	60
3.4.2. Закон больших чисел	61
3.5. ЭПИСТЕМОЛОГИЯ И ДЕДУКТИВНАЯ АСИММЕТРИЯ	64
3.6. НАИВНЫЙ ЭМПИРИЗМ: НЕ НАДО СРАВНИВАТЬ ЭБОЛУ И ПАДЕНИЯ СО СТРЕМЯНОК	69
3.6.1. Как изменяется масштаб некоторых мультипликативных рисков	73
3.7. АЗБУКА СТЕПЕННЫХ ЗАКОНОВ (ПОЧТИ БЕЗ МАТЕМАТИКИ)	74
3.8. ГДЕ ПРЯЧУТСЯ СКРЫТЫЕ СВОЙСТВА?	77
3.9. БАЙЕСА-ШМАЙЕСА	81
3.10. $X$ И $F(X)$ : КАК ПУТАЮТ ВОЗДЕЙСТВИЕ ВЕЛИЧИНЫ $X$ С САМОЙ ВЕЛИЧИНОЙ $X$	82
3.11. РАЗОРЕНИЕ И ЗАВИСИМОСТЬ ОТ ПУТИ	86
3.12. ЧТО ДЕЛАТЬ?	89
4. ОДНОМЕРНЫЕ ЖИРНЫЕ ХВОСТЫ УРОВНЯ 1, С КОНЕЧНЫМИ МОМЕНТАМИ†	91
4.1. ПРОСТАЯ ЭВРИСТИКА, КАК СОЗДАВАТЬ СЛЕГКА ЖИРНЫЕ ХВОСТЫ	91
4.1.1. Эвристика, сохраняющая дисперсию	94
4.1.2. Ужирнение хвостов при помощи асимметричной дисперсии	95
4.2. СПОСОБНА ЛИ СТОХАСТИЧЕСКАЯ ВОЛАТИЛЬНОСТЬ ГЕНЕРИРОВАТЬ СТЕПЕННЫЕ ЗАКОНЫ?	98
4.3. ТУЛОВИЩЕ, ПЛЕЧИ И ХВОСТЫ	99
4.3.1. Точки перехода и туннельный эффект	99
4.4. ЖИРНЫЕ ХВОСТЫ, СРЕДНЕЕ ОТКЛОНЕНИЕ И ПОВЫШАЮЩИЕСЯ НОРМЫ	103
4.4.1. Обычные ошибки	103
4.4.2. Немного аналитики	105
4.4.3. Влияние жирных хвостов на «эффективность» STD против MAD	108
4.4.4. Моменты и неравенство о средних разной степени	109
4.4.5. Комментарий: почему среднееквадратическое отклонение нужно отправить в отставку, и немедленно!	112
4.5. ВИЗУАЛИЗИРУЕМ, КАК РОСТ $p$ ВЛИЯЕТ НА ИЗО-НОРМЫ	115
5. УРОВЕНЬ 2: СУБЭКСПОНЕНЦИАЛЬНЫЕ И СТЕПЕННЫЕ ЗАКОНЫ	118
5.0.1. Вернемся к ранжированию	118
5.0.2. Что такое пограничное распределение вероятностей?	120
5.0.3. Давайте выдумаем распределение	121
5.1. УРОВЕНЬ 3: МАСШТАБИРУЕМОСТЬ И СТЕПЕННЫЕ ЗАКОНЫ	123
5.1.1. Масштабируемое и немасштабируемое — более глубокий взгляд на жирные хвосты	123
5.1.2. Серые лебеди	126
5.2. НЕКОТОРЫЕ СВОЙСТВА СТЕПЕННЫХ ЗАКОНОВ	126
5.2.1. Сумма случайных величин	126
5.2.2. Преобразования	127
5.3. КОЛОКОЛООБРАЗНЫЕ И ДРУГИЕ СТЕПЕННЫЕ ЗАКОНЫ	128
5.4. ИНТЕРПОЛЯЦИОННЫЕ ВОЗМОЖНОСТИ СТЕПЕННЫХ ЗАКОНОВ: ПРИМЕР	129
5.5. СВЕРХЖИРНЫЕ ХВОСТЫ: РАСПРЕДЕЛЕНИЕ ЛОГ-ПАРЕТО	130
5.6. ПСЕВДОСТОХАСТИЧЕСКАЯ ВОЛАТИЛЬНОСТЬ: РАССЛЕДОВАНИЕ	131

6. ЖИРНЫЕ ХВОСТЫ В ВЫСШИХ РАЗМЕРНОСТЯХ <sup>+</sup> .....	134
6.1. ТОЛСТЫЕ ХВОСТЫ ПРИ ВЫСОКОЙ РАЗМЕРНОСТИ, КОНЕЧНЫЕ МОМЕНТЫ.....	134
6.2. СОВМЕСТНАЯ ЖИРНОХВОСТОСТЬ И ЭЛЛИПТИЧНОСТЬ РАСПРЕДЕЛЕНИЙ.....	135
6.3. МНОГОМЕРНОЕ t-РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА.....	140
6.3.1. Эллиптичность и независимость при толстых хвостах.....	140
6.4. ЖИРНЫЕ ХВОСТЫ И ВЗАИМНАЯ ИНФОРМАЦИЯ.....	141
6.5. ЖИРНЫЕ ХВОСТЫ И СЛУЧАЙНЫЕ МАТРИЦЫ, КРАТКОЕ ОТСТУПЛЕНИЕ.....	142
6.6. КОРРЕЛЯЦИЯ И НЕОПРЕДЕЛЕННАЯ ДИСПЕРСИЯ.....	143
6.7. ЖИРНОХВОСТЫЕ ОСТАТКИ В МОДЕЛЯХ ЛИНЕЙНОЙ РЕГРЕССИИ.....	144
A. ОСОБЫЕ СЛУЧАИ ТОЛСТЫХ ХВОСТОВ.....	149
A.1. МУЛЬТИМОДАЛЬНОСТЬ И ЖИРНЫЕ ХВОСТЫ, ИЛИ МОДЕЛЬ ВОЙНЫ И МИРА.....	149
A.2. ПЕРЕХОДНЫЕ ВЕРОЯТНОСТИ: ЧТО МОЖЕТ РАЗБИТЬСЯ, РАЗОБЬЕТСЯ.....	152

## Часть II. ЗАКОН СРЕДНИХ ЧИСЕЛ

7. ПРЕДЕЛЬНЫЕ РАСПРЕДЕЛЕНИЯ, КОНСОЛИДАЦИЯ <sup>++</sup> .....	157
7.1. НАПОМИНАНИЕ: СЛАБЫЙ И СИЛЬНЫЙ ЗБЧ.....	157
7.2. ЦЕНТРАЛЬНЫЙ ПРЕДЕЛ В ДЕЙСТВИИ.....	159
7.2.1. Устойчивое распределение.....	160
7.2.2. Закон больших чисел для устойчивого распределения.....	160
7.3. СКОРОСТЬ СХОДИМОСТИ ЦПТ: НАГЛЯДНЫЕ ОПЫТЫ.....	161
7.3.1. Быстрая сходимость: равномерное распределение.....	161
7.3.2. Полузамедленная сходимость: экспоненциальные распределения.....	162
7.3.3. Медленный Парето.....	163
7.3.4. Полукубический Парето и его область сходимости.....	165
7.4. КУМУЛЯНТЫ И СХОДИМОСТЬ.....	166
7.5. ПОВТОРИМ ТЕХНИКУ: ТРАДИЦИОННЫЕ ВЕРСИИ ЦПТ.....	169
7.6. ЗАКОН БОЛЬШИХ ЧИСЕЛ ДЛЯ ВЫСШИХ МОМЕНТОВ.....	170
7.6.1. Высшие моменты.....	170
7.7. СРЕДНЕЕ ОТКЛОНЕНИЕ ДЛЯ УСТОЙЧИВЫХ РАСПРЕДЕЛЕНИЙ.....	173
8. СКОЛЬКО НУЖНО ДАННЫХ? РАБОЧИЙ ПОКАЗАТЕЛЬ ЖИРНОХВОСТОСТИ <sup>+</sup> .....	175
8.1. ВВЕДЕНИЕ И ОПРЕДЕЛЕНИЯ.....	176
8.2. ПОКАЗАТЕЛЬ КАППА.....	178
8.3. УСТОЙЧИВЫЙ БАССЕЙН СХОДИМОСТИ КАК ТОЧКА ОТСЧЕТА.....	179
8.3.1. Эквиваленты устойчивых распределений.....	180
8.3.2. Практическая значимость при достаточной выборке.....	182
8.4. ТЕХНИЧЕСКИЕ СЛЕДСТВИЯ.....	184
8.4.1. Некоторые странности асимметричных распределений.....	184
8.4.2. Скорость сходимости t-распределения Стюдента к гауссову бассейну.....	184
8.4.3. Логнормальный хвост — ни тонкий, ни жирный.....	185
8.4.4. Возможна ли отрицательная каппа?.....	185
8.5. ЗАКЛЮЧЕНИЕ И СЛЕДСТВИЯ.....	185
8.5.1. Портфельная псевдостабильзация.....	186
8.5.2. Другие аспекты статистических выводов.....	187
8.5.3. Последний комментарий.....	187
8.6. ПРИЛОЖЕНИЕ, ВЫВОД ФОРМУЛ, ДОКАЗАТЕЛЬСТВА.....	187
8.6.1. Кубическое t-распределение Стюдента (гауссов бассейн).....	187
8.6.2. Логнормальные суммы.....	190
8.6.3. Экспоненциальное распределение.....	193
8.6.4. Отрицательная каппа, отрицательный эксцесс.....	193

9. ЭКСТРЕМАЛЬНЫЕ ЗНАЧЕНИЯ И СКРЫТЫЕ ХВОСТЫ*†	195
9.1. ПРЕДВАРИТЕЛЬНОЕ ВСТУПЛЕНИЕ К ТЭЗ	197
9.1.1. Любой хвост степенного закона ведет к Фреше	198
9.1.2. Гауссов случай	200
9.1.3. Теорема Пикэндса — Балкемы — де Гаана	202
9.2. НЕВИДИМЫЙ ХВОСТ ПРИ СТЕПЕННОМ ЗАКОНЕ	202
9.2.1. Сравнение с нормальным распределением	205
9.3. ПРИЛОЖЕНИЕ: ЭМПИРИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ НЕ ЭМПИРИЧНО	205
В. СКОРОСТЬ РОСТА И РЕЗУЛЬТАТ ПРИНАДЛЕЖАТ РАЗНЫМ КЛАССАМ РАСПРЕДЕЛЕНИЙ	207
В.1. ЗАГАДКА	207
В.2. У ПАНДЕМИЙ ИСКЛЮЧИТЕЛЬНО ЖИРНЫЕ ХВОСТЫ	209
С. ПРИНЦИП БОЛЬШОГО ОТКЛОНЕНИЯ, ВКРАТЦЕ	211
Простой случай: оценка Чернова	212
Д. КАЛИБРОВКА В СИТУАЦИИ ПАРЕТО	214
Д.1. РАСПРЕДЕЛЕНИЕ ВЫБОРОЧНОЙ ОЦЕНКИ ПОКАЗАТЕЛЯ ХВОСТА	216
10. ПЕЧАЛЬНО, НО ФАКТ: ДИАГНОСТИКА S&P 500 †	219
10.1. ПРИНАДЛЕЖНОСТЬ КЛАССУ ПАРЕТО И МОМЕНТЫ	219
10.2. КРИТЕРИИ СХОДИМОСТИ	221
10.2.1. Критерий 1: эксцесс при агрегации	221
10.2.2. Максимальные падения	222
10.2.3. Эмпирическая каппа	224
10.2.4. Проверка 2: условное математическое ожидание избытка	225
10.2.5. Проверка 3: неустойчивость 4-го момента	227
10.2.6. Проверка 4: график MS	227
10.2.7. Рекорды и экстремальные значения	228
10.2.8. Асимметричность хвостов справа и слева	231
10.3. ЗАКЛЮЧЕНИЕ: ПЕЧАЛЬНО, НО ФАКТ	233
Е. ПРОБЛЕМА С ЭКОНОМЕТРИКОЙ	234
Е.1. ЭФФЕКТИВНОСТЬ СТАНДАРТНЫХ ПАРАМЕТРИЧЕСКИХ ОЦЕНОК РИСКА	235
Е.2. ЭФФЕКТИВНОСТЬ СТАНДАРТНЫХ НЕПАРАМЕТРИЧЕСКИХ ОЦЕНОК РИСКА	238
F. ОСОБЕННОСТИ МАШИННОГО ОБУЧЕНИЯ	242
F.0.1. Калибровка по углам	245

### Часть III. ПРЕДСКАЗАНИЯ, ПРОГНОЗЫ И НЕОПРЕДЕЛЕННОСТЬ

11. КАЛИБРОВКА ВЕРОЯТНОСТИ ПРИ ЖИРНЫХ ХВОСТАХ †	249
11.1. Непрерывные и дискретные платежные функции: определения и комментарии	250
11.1.1. Отходим от словесных формулировок	251
11.1.2. При жирных хвостах нет стандартной величины «коллапса», «катастрофы» или «успеха»	255
11.2. МНИМАЯ ПЕРЕОЦЕНКА ХВОСТОВОЙ ВЕРОЯТНОСТИ В ПСИХОЛОГИИ	257
11.2.1. Тонкие хвосты	258
11.2.2. Жирные хвосты	258
11.2.3. Что с чем путают	259
11.2.4. Неопределенность распределения	263
11.3. КАЛИБРОВКА И МНИМАЯ КАЛИБРОВКА	264
11.4. ПОКАЗАТЕЛИ ДЛЯ КОЛИЧЕСТВЕННОЙ ОЦЕНКИ	264
11.4.1. Вывод распределений	267
11.5. НЕВЕРБАЛЬНЫЕ ПЛАТЕЖНЫЕ ФУНКЦИИ И МАШИННОЕ ОБУЧЕНИЕ	269
11.6. ЗАКЛЮЧЕНИЕ	271
11.7. ПРИЛОЖЕНИЕ: ДОКАЗАТЕЛЬСТВА И ВЫВОД ФОРМУЛ	272

11.7.1. Распределение подсчета пары $P_n^{(P)}$ .....	272
11.7.2. Распределение оценки Брайера.....	272
12. ПРЕДСКАЗАНИЯ ВЫБОРОВ КАК МАРТИНГАЛ: АРБИТРАЖНЫЙ ПОДХОД † .....	276
12.0.1. Основные результаты.....	278
12.0.2. Организация.....	279
12.0.3. Обсуждение нейтральности к риску .....	281
12.1. СТОИМОСТЬ В СТИЛЕ БАШЕЛЬЕ .....	281
12.2. ОГРАНИЧЕННЫЙ ДВОЙСТВЕННЫЙ МАРТИНГАЛЬНЫЙ ПРОЦЕСС.....	283
12.3. СВЯЗЬ С ОЦЕНЩИКОМ ВЕРОЯТНОСТИ ДЕ ФИНЕТТИ.....	285
12.4. ЗАКЛЮЧЕНИЕ И КОММЕНТАРИИ .....	286
ПРИЛОЖЕНИЕ: ВСЕ ДОРОГИ ВЕДУТ К ФИНАНСОВОЙ МАТЕМАТИКЕ .....	287
Некорректные претензии .....	287
Неправильная арбитражная стоимость .....	287
Арбитражные вопросы .....	288
БЛАГОДАРНОСТИ.....	290

#### Часть IV. ОЦЕНОЧНЫЕ НЕРАВЕНСТВА ПРИ ЖИРНЫХ ХВОСТАХ

13. ОЦЕНКА ДЖИНИ ПРИ БЕСКОНЕЧНОЙ ДИСПЕРСИИ † .....	293
13.1. ВВЕДЕНИЕ.....	293
13.2. АСИМПТОТИКИ НЕПАРАМЕТРИЧЕСКОЙ ОЦЕНКИ ПРИ БЕСКОНЕЧНОЙ ДИСПЕРСИИ .....	297
13.2.1. Краткое повторение $\alpha$ -устойчивых случайных величин .....	299
13.2.2. $\alpha$ -устойчивый асимптотический предел индекса Джини.....	300
13.3. ОЦЕНКА МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ.....	301
13.4. ИЛЛЮСТРАЦИЯ С РАСПРЕДЕЛЕНИЕМ ПАРЕТО .....	302
13.5. ПОПРАВКА НА МАЛУЮ ВЫБОРКУ .....	304
13.6. ВЫВОДЫ.....	308
Доказательство Теоремы 1 .....	310
14. СУПЕРАДДИТИВНОСТЬ И СМЕЩЕННЫЕ ОЦЕНКИ ВКЛАДА КВАНТИЛЕЙ † <sup>a</sup> .....	315
14.1. ВВЕДЕНИЕ.....	315
14.2. ОЦЕНКА ДЛЯ НЕСМЕШАННЫХ РАСПРЕДЕЛЕНИЙ С ХВОСТОМ ПАРЕТО .....	317
14.2.1. Смещение и сходимость .....	317
14.3. НЕРАВЕНСТВО АГРЕГАЦИИ НЕРАВЕНСТВ .....	320
14.4. СМЕШАННЫЕ РАСПРЕДЕЛЕНИЯ ПО ПОКАЗАТЕЛЮ ХВОСТА.....	323
14.5. ЧЕМ БОЛЬШЕ СУММА, ТЕМ ВЫШЕ $\hat{\kappa}_q$ .....	326
14.6. ЗАКЛЮЧЕНИЕ И ТОЧНАЯ ОЦЕНКА КОНЦЕНТРАЦИИ .....	326
14.6.1. Робастные методы и использование полных данных .....	327
14.6.2. Как правильно оценивать концентрацию?.....	327

#### Часть V. СТАТЬИ О ТЕНЕВЫХ МОМЕНТАХ

15. ТЕНЕВЫЕ МОМЕНТЫ ЯВЛЕНИЙ С МНИМО БЕСКОНЕЧНЫМ СРЕДНИМ †.....	331
15.1. ВВЕДЕНИЕ.....	331
15.2. ДВОЙСТВЕННОЕ РАСПРЕДЕЛЕНИЕ .....	332
15.3. ВЕРНЕМСЯ К $\mathcal{Y}$ : ТЕНЕВОЕ СРЕДНЕЕ, ИЛИ СРЕДНЕЕ ПО ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ.....	334
15.4. СРАВНЕНИЕ С ДРУГИМИ МЕТОДАМИ .....	337
15.5. ОБЛАСТИ ПРИМЕНЕНИЯ.....	338
16. О ХВОСТОВОМ РИСКЕ ОСТРОГО КОНФЛИКТА (СОВМЕСТНО С П. ЧИРИЛЛО)†.....	340
16.1. ВВЕДЕНИЕ И РЕЗЮМЕ .....	340
16.2. ОБЗОР СТАТИСТИЧЕСКОЙ ДИСКУССИИ.....	343
16.2.1. Результаты .....	343
16.2.2. Заключение .....	345



16.3. ОБСУЖДЕНИЕ МЕТОДОВ .....	345
16.3.1. Метод масштабирования .....	345
16.3.2. Условное математическое ожидание (в нестрогом изложении) .....	347
16.3.3. Надежность данных и влияние на хвостовые оценки .....	347
16.3.4. Определение «события» .....	349
16.3.5. Пропущенные события .....	349
16.3.6. Систематическая ошибка выжившего .....	350
16.4. АНАЛИЗ ДАННЫХ .....	350
16.4.1. Превышения над порогом .....	350
16.4.2. Интервалы во временных рядах и автокорреляция .....	351
16.4.3. Анализ хвоста .....	352
16.4.4. Альтернативный взгляд на максимумы .....	354
16.4.5. Анализ полных данных .....	355
16.5. ДОПОЛНИТЕЛЬНЫЕ ТЕСТЫ РОБАСТНОСТИ И НАДЕЖНОСТИ .....	356
16.5.1. Бутстрэп для GPD .....	356
16.5.2. Внесение возмущений в границы оценок .....	357
16.6. ЗАКЛЮЧЕНИЕ: МИР ОПАСНЕЕ, ЧЕМ КАЖЕТСЯ? .....	358
16.7. БЛАГОДАРНОСТИ .....	359
G. КАКОВА ВЕРОЯТНОСТЬ ТРЕТЬЕЙ МИРОВОЙ ВОЙНЫ?*	360

## Часть VI. СТАТЬИ О МЕТАВЕРОЯТНОСТИ

17. КАК ТОЛСТЫЕ ХВОСТЫ ВОЗНИКАЮТ ИЗ РЕКУРСИВНОЙ ЭПИСТЕМОЛОГИЧЕСКОЙ НЕОПРЕДЕЛЕННОСТИ <sup>†</sup> .....	367
17.1. МЕТОДЫ И ВЫКЛАДКИ .....	368
17.1.1. Уровни неопределенности .....	368
17.1.2. Интегралы высоких порядков в стандартном гауссовом случае .....	369
РЕЖИМ 1 (ВЗРЫВНОЙ): СЛУЧАЙ ПОСТОЯННОГО ПАРАМЕТРА $a$ .....	372
17.1.3. Влияние на малые вероятности .....	374
17.2. РЕЖИМ 2: СЛУЧАИ ЗАТУХАЮЩИХ ПАРАМЕТРОВ $an$ .....	375
17.2.1. Режим 2-a; «потери» погрешности высокого порядка .....	375
17.2.2. Режим 2-b; второй метод, немультимпликативная погрешность .....	376
17.3. ПРЕДЕЛЬНОЕ РАСПРЕДЕЛЕНИЕ .....	376
18. СТОХАСТИЧЕСКИЙ ПОКАЗАТЕЛЬ ХВОСТА ПРИ АСИММЕТРИЧНЫХ СТЕПЕННЫХ ЗАКОНАХ <sup>†</sup> .....	377
18.1. ИСТОРИЯ ВОПРОСА .....	378
18.2. ОДНОХВОСТЫЕ РАСПРЕДЕЛЕНИЯ СО СТОХАСТИЧЕСКОЙ АЛЬФОЙ .....	378
18.2.1. Общие случаи .....	378
18.2.2. Неравенство стохастической альфы .....	379
18.2.3. Аппроксимации для класса $\mathfrak{P}$ .....	381
18.3. СУММЫ СТЕПЕННЫХ ЗАКОНОВ .....	381
18.4. АСИММЕТРИЧНЫЕ УСТОЙЧИВЫЕ РАСПРЕДЕЛЕНИЯ .....	382
18.5. РАСПРЕДЕЛЕНИЕ ПАРЕТО С ЛОГНОРМАЛЬНЫМ РАСПРЕДЕЛЕНИЕМ АЛЬФЫ .....	383
18.6. РАСПРЕДЕЛЕНИЕ ПАРЕТО С ГАММА-РАСПРЕДЕЛЕНИЕМ АЛЬФЫ .....	384
18.7. ОГРАНИЧЕННЫЙ СТЕПЕННОЙ ЗАКОН В РАБОТЕ ЧИРИЛЛО И ТАЛЕБА (2016) .....	385
18.8. ДОПОЛНИТЕЛЬНЫЕ КОММЕНТАРИИ .....	386
18.9. БЛАГОДАРНОСТИ .....	386
19. МЕТАРАСПРЕДЕЛЕНИЕ $p$ -ЗНАЧЕНИЙ И $p$ -ХАКИНГ <sup>†</sup> .....	387
19.1. ДОКАЗАТЕЛЬСТВА И ВЫВОД ФОРМУЛ .....	389
19.2. ОБРАТНАЯ МОЩНОСТЬ ТЕСТА .....	393
19.3. ПРИЛОЖЕНИЕ И ВЫВОДЫ .....	394
БЛАГОДАРНОСТИ .....	395

Н. НЕКОТОРЫЕ НЕДОРАЗУМЕНИЯ В ПОВЕДЕНЧЕСКОЙ ЭКОНОМИКЕ.....	396
Н.1. ПРИМЕР ИССЛЕДОВАНИЯ: ЛОЖНАЯ СПЕЦИФИКАЦИЯ БЛИЗОРУКОЙ БОЯЗНИ ПОТЕРЬ.....	396

## Часть VII. ТОРГОВЛЯ ОПЦИОНАМИ И ЦЕНЫ ПРИ ЖИРНЫХ ХВОСТАХ

20. НЕУДАЧИ ФИНАНСОВОЙ ТЕОРИИ КАСАТЕЛЬНО ЦЕН ОПЦИОНОВ <sup>†</sup> .....	403
20.1. БАШЕЛЬЕ, А НЕ БЛЭК — ШОУЛЗ.....	404
20.1.1. Искажения из-за идеализации.....	405
20.1.2. Фактическая процедура репликации.....	406
20.1.3. Провал: погрешность хеджирования может сделать модель непригодной.....	406
21. ЕДИНСТВЕННАЯ МЕРА ДЛЯ ЦЕН ОПЦИОНОВ (БЕЗ ДИНАМИЧЕСКОГО ХЕДЖИРОВАНИЯ ИЛИ ПОЛНОГО РЫНКА) <sup>†</sup> .....	407
21.1. ИСТОРИЯ ВОПРОСА .....	407
21.2. ДОКАЗАТЕЛЬСТВО.....	410
21.2.1. Случай 1: форвард как мера, нейтральная к риску .....	410
21.2.2. Вывод формул.....	411
21.3. СЛУЧАЙ ФОРWARDA БЕЗ НЕЙТРАЛЬНОСТИ К РИСКУ .....	413
21.4. КОММЕНТАРИЙ.....	413
22. ТОРГОВЦЫ ОПЦИОНАМИ НЕ ПОЛЬЗУЮТСЯ ФОРМУЛОЙ БЛЭКА — ШОУЛЗА — МЕРТОНА <sup>*†</sup> .....	415
22.1. ПРЕРЫВАНИЕ ЦЕПИ ПЕРЕДАЧИ .....	415
22.2. ВВЕДЕНИЕ И РЕЗЮМЕ .....	416
22.2.1. Теория Блэка — Шоулза была аргументом в дискуссии .....	416
22.3. МИФ 1: ТРЕЙДЕРЫ НЕ ЗАДАВАЛИ ЦЕНЫ ОПЦИОНАМ ДО БШМ .....	420
22.4. МЕТОДЫ И ВЫВОД ФОРМУЛ .....	421
22.4.1. Формулы опционов и дельта-хеджирование .....	424
22.5. МИФ 2: СОВРЕМЕННЫЕ ТРЕЙДЕРЫ ПОЛЬЗУЮТСЯ БЛЭКОМ — ШОУЛЗОМ.....	425
22.5.1. Когда мы оцениваем стоимость? .....	426
22.6. О МАТЕМАТИЧЕСКОЙ НЕВОЗМОЖНОСТИ ДИНАМИЧЕСКОГО ХЕДЖИРОВАНИЯ .....	426
22.6.1. Сбивающая с толку робастность гауссианы.....	428
22.6.2. Поток заказов и опционы .....	429
22.6.3. Башелье — Торп.....	430
23. ЦЕНООБРАЗОВАНИЕ ОПЦИОНОВ ПРИ СТЕПЕННЫХ ЗАКОНАХ: РОБАСТНАЯ ЭВРИСТИКА <sup>*†</sup> .....	431
23.1. ВВЕДЕНИЕ.....	432
23.2. ЦЕНООБРАЗОВАНИЕ «КОЛЛОВ» ЗА КОНСТАНТОЙ КАРАМАТЫ .....	432
23.2.1. Первый подход, $S$ в классе правильно меняющихся функций .....	432
23.2.2. Второй подход, геометрическая доходность от $S$ в классе правильно меняющихся функций.....	434
23.3. ЦЕНЫ ОПЦИОНОВ «ПУТ» .....	437
23.4. ГРАНИЦЫ АРБИТРАЖА .....	438
23.5. КОММЕНТАРИЙ.....	438
24. ЧЕТЫРЕ ОШИБКИ В ФИНАНСОВОЙ МАТЕМАТИКЕ <sup>*†</sup> .....	439
24.1. ПУТАНИЦА МЕЖДУ ВТОРЫМ И ЧЕТВЕРТЫМ МОМЕНТАМИ.....	439
24.2. НЕУЧЕТ НЕРАВЕНСТВА ЙЕНСЕНА ПРИ АНАЛИЗЕ ДОХОДНОСТИ ОПЦИОНОВ .....	440
24.3. НЕРАЗРЫВНАЯ СВЯЗЬ МЕЖДУ СТРАХОВКОЙ И ПРЕДМЕТОМ СТРАХОВАНИЯ .....	442
24.4. НЕОБХОДИМОСТЬ МАСШТАБА ЦЕН В ФИНАНСОВОЙ МАТЕМАТИКЕ .....	442
24.5. ПРИЛОЖЕНИЕ (СТАВКИ НА ХВОСТЫ РАСПРЕДЕЛЕНИЯ) .....	443
25. ОГРАНИЧЕНИЯ ХВОСТОВОГО РИСКА И МАКСИМАЛЬНАЯ ЭНТРОПИЯ (СОВМЕСТНО С Д. И Э. ЖИМАН) <sup>†</sup> .....	445

25.1. РИСК ЛЕВОГО ХВОСТА КАК ЦЕНТРАЛЬНОЕ ОГРАНИЧЕНИЕ ПОРТФЕЛЯ .....	445
25.1.1. Штанга с точки зрения Э. Т. Джейнса .....	448
25.2. ВЕРНЕМСЯ К ЗАДАНИЮ СРЕДНЕГО И ДИСПЕРСИИ .....	449
25.2.1. Анализ ограничений .....	450
25.3. ВЕРНЕМСЯ К ГАУССОВУ СЛУЧАЮ .....	450
25.3.1. Смесь двух нормальных распределений .....	452
25.4. МАКСИМАЛЬНАЯ ЭНТРОПИЯ .....	453
25.4.1. Случай А: ограничиваем глобальное среднее .....	453
25.4.2. Случай В: ограничиваем абсолютное среднее .....	455
25.4.3. Случай С: степенные законы для правого хвоста .....	455
25.4.4. Расширение на несколько периодов: комментарий .....	457
25.5. КОММЕНТАРИИ И ЗАКЛЮЧЕНИЕ .....	458
25.6. ПРИЛОЖЕНИЕ / ДОКАЗАТЕЛЬСТВА .....	458
ПЕРСОНАЛИИ .....	459
БИБЛИОГРАФИЯ .....	467

# 1

## ПРОЛОГ\*

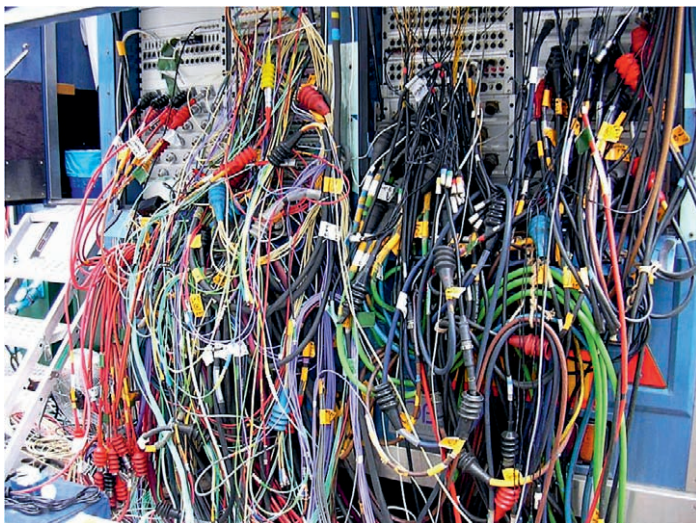
Чем хуже вы понимаете мир,  
тем проще вам принять решение.



**Рисунок 1.1:** Проблема не в том, что люди не слышали о «жирном хвосте», а в том, что не понимают серьезность его последствий. Когда вам встретился «жирный хвост», нельзя выбрать из привычного арсенала статистики соответствующий вариант комплекта инструментов; нужно сменить весь подход к принятию решений. © Stefan Gasic

Главная идея в основе проекта *Incerto* — та, что при всей неопределенности и непроницаемости мира и при нехватке информации и понимания все равно в каждой конкретной ситуации оказывается совершенно ясно, какие действия нужно предпринять на основе того немногого, что известно и понятно.

Эта книга состоит из (1) опубликованных статей и (2) бесцензурного комментария, посвященных тем классам статистических распределений, от которых можно ждать экстремальных событий. Мы изучим, как использовать эти распределения для статистических выводов и принятия решений.



**Рисунок 1.2:** Усложнение из-за непонимания. Что творится в головах профессионалов, когда они применяют статистику и анализ данных, не имея ясного представления об основных понятиях.  
© Wikimedia

«Стандартная» статистика по большей части работает на основе теорем, выведенных для тонких хвостов. Чтобы работать с предасимптотикой<sup>1</sup> жирных хвостов, эти методы придется либо адаптировать нетривиальным образом, либо вовсе исключить из арсенала полезных инструментов.

Автору не раз приходилось слышать фразы вроде «Это и так все знают» и «В жирных хвостах нет ничего нового» — ими пытались защищаться преподаватель или практик, пойманные на совершенно бессмысленной в конкретной ситуации попытке использовать *дисперсию*, *обобщенную авторегрессию*, *коэффициент эксцесса*, *коэффициент Шарпа* или *стоимость под риском* или указать *статистическую значимость* там, где она не значит ничего.

Автор обогатил свой опыт, когда осуществил программу научных исследований и выпустил ряд книг серии *Incerto* [226], посвященных выживанию в реальном мире с его структурой неопределенности, которая слишком сложна для нашего понимания.

Цикл *Incerto* ставит целью объединить пять областей знания, связанных с жирными хвостами и экстремальными событиями: в математике, философии, общественных науках, теории контрактов и теории принятия решений, — с опытом профессионалов. Если вы спросите, при чем здесь теория контрактов и теория принятия решений, то ответ таков: математика опционов основана на идее условной вероятности и объединении контрактов с целью изменить класс воздействия в хвостах распределения; некоторым образом теория опционов — это математическая теория контрактов. Теория принятия решений ставит целью не понять мир, а выбраться из неприятностей и выжить. Этой задаче будет посвящен следующий том *Технического Incerto*, его текущее рабочее название — *Convexity, Risk, and Fragility* («Выпуклость вниз, риск и хрупкость»).

<sup>1</sup> По аналогии с терминами «предыстория» и «доисторический» неологизмы *preasymptotics* и *preasymptotic* можно заимствовать в русский язык как «предасимптотика» и «доасимптотический». — Здесь и далее, если не указано иное, прим. перев.

## ЗАМЕЧАНИЕ О ТЕРМИНАХ

В академическом контексте при описании распределения часто используется термин «толстые хвосты» (*thick tails*). Мы вместо этого будем говорить, что «коэффициент эксцесса выше, чем у гауссианы»; это ближе к профессиональному жаргону финансиста.

Термин «жирные хвосты» (*fat tails*) мы оставим за особо толстыми хвостами, которые характерны для распределений по степенному закону или эквивалентному (жирный хвост и степенной закон, как мы покажем в Главе 8, неотделимы друг от друга). Некоторые авторы придают «жирным хвостам» более узкий смысл, требуя точного степенного закона или хотя бы правильно меняющейся функции. Однако мы, хотя и будем иногда применять степенные законы (в тех случаях, когда известно, что процесс работает именно так), жирными хвостами будем называть все экстремально толстые хвосты.

Во избежание путаницы не будем пользоваться дополнительными терминами вроде «тяжелых хвостов» (*heavy tails*) или «длинных хвостов» (*long tails*).

Термины «толстые хвосты» и «жирные хвосты» будут прояснены в следующих двух главах.



Рисунок 1.3: Классическая реакция, когда «альтернативой» считается только тот анализ, который рекомендует одобрить кредит. © Stefan Gasic

## БЛАГОДАРНОСТИ

Помимо уже названных соавторов, автор благодарен Чжуо Си, Жан-Филипу Бушо, Роберту Фраю, Спиросу Макридакису, Марку Шпицнагелю, Брэндону Ярину, Рафаэлю Дуади, Питеру Карру, Марко Авельянеде, Дидье Сорнетту, Полю Амбре, Бруно Дюпиру, Джамилю Базу, Дамиру Деличу, Яниру Бар-Яму, Диего Цвивовичу, Джозефу Норману, Оле Петерсу, Читпьюниту Манну, Гарри Крейну — и, разумеется, долгим, нескончаемым дискуссиям с великим Бенуа Мандельбротом.

Много опечаток исправили добровольные редакторы в социальных сетях, такие как Максим Бьет, Чао Винчи, Джейсон Торелл и Петри Хэло. Обширный список опечаток и потенциальных нотационных двусмысленностей прислал Кевин Ван Хорн.

Часть статей, ставших главами этой книги, была представлена на конференциях; автор благодарит Лоренца де Гаана, Берта Цварца и других за комментарии по проблемам, связанным с экстремальными значениями. Более точные благодарности сформулированы в конкретных главах. Как обычно, автор хотел бы поблагодарить штат ресторана *Naya* в Нью-Йорке.



Автор представил данную книгу и главные тезисы на ежемесячной конференции Блумберг — Квант<sup>1</sup> в Нью-Йорке в сентябре 2018 года. После лекции ко мне подошел один выдающийся профессор финансовой математики.

— Типичная талебщина, — сказал он. — Вы доказываете, что так-то и так-то нельзя, но взамен не предлагаете альтернатив.

Понятно, что в бизнесе и любой другой сфере, где действует суровая школа реального мира, такой работник долго бы не выжил. Но кто не рискует собственной шкурой [236], до того не доходит, как важно, смотря по обстоятельствам, отложить свои убеждения и как ценны сведения о ненадежности для принятия решений: *не передавай пилоту неточные данные, научись передавать только надежную информацию; сообщая пилоту о неисправности самолета, ты спасаешь жизни*. И до них не доходит, как эффективен подход *via negativa* — когда наука, по Попперу, развивается отсечением неудачных теорий. Покойный Дэвид Фридман предпринял безуспешную попытку укротить маньяков бессмысленного и обманчивого моделирования в статистике, продемонстрировав, как их прогнозы с большим отрывом проигрывают соревнование «ничему», пустой теории.

Между тем в ряде статей и глав этой книги предлагаются решения и альтернативы. Увы, некоторых они не обрадуют, поскольку требуют математических усилий, чтобы построить совершенно другие модели, модели для ситуаций с жирными хвостами.

---

1 Семинар агентства «Блумберг» по финансовой математике (Bloomberg Quantitative Finance Seminars).

## ГЛОССАРИЙ, ОПРЕДЕЛЕНИЯ И ОБОЗНАЧЕНИЯ

Это систематический каталог с пояснениями основных разделов и обозначений. Все обозначения разъясняются и в основном тексте; здесь те же пояснения дублируются для удобства читателя, решившего посмотреть только отдельные отрывки. Некоторые обозначения отличаются в той или иной главе, созданной на основе конкретной статьи; здесь это указывается. Иногда наша терминология расходится с терминологией других исследовательских групп, хотя мы старались не противоречить существующим терминам.

### 2.1. ОБЩИЕ ОБОЗНАЧЕНИЯ И ЧАСТО ИСПОЛЬЗУЕМЫЕ СИМВОЛЫ

$\mathbb{P}$  — вероятность случайного события; обычно в форме  $\mathbb{P}(X > x)$ , где  $X$  — случайная величина, а событием считается, что ее реализация превзошла значение  $x$ . Более формальные определения событий и вероятностей по канонам теории меры и прочих французский встречаются в Главе 11 и других местах, где этот формализм имеет смысл.

$\mathbb{E}$  — оператор математическое ожидание<sup>1</sup>.

$\mathbb{V}$  — оператор дисперсия<sup>2</sup>.

$\mathbb{M}$  — среднее абсолютное отклонение;<sup>3</sup> если центрируется, то относительно среднего (а не медианы).

1 От англ. *expected value* или *mathematical expectation*. В русской литературе обозначается также  $M$ , но в этой книге обозначение  $M(k) = \mu_k$  зарезервировано за статистическим моментом порядка  $k$ . Если случайная величина  $X$  принимает значения на множестве всех действительных чисел  $\mathbb{R}$ , то  $\mathbb{E}X = \int_{\mathbb{R}} x f_X(x) dx$ , где  $f_X(x)$  — плотность вероятности. В этой книге часто используется линейность математического ожидания:  $\mathbb{E}(aX) = a\mathbb{E}X$  и  $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$ , даже если между случайными величинами  $X$  и  $Y$  есть корреляция, — и мультипликативность математического ожидания в случае независимых случайных величин:  $\mathbb{E}XY = \mathbb{E}X \mathbb{E}Y$ .

2 От англ. *variance*. В русской литературе обозначается также  $D$ . Если случайная величина  $X$  принимает значения на  $\mathbb{R}$ , то  $\mathbb{V}X = \int_{\mathbb{R}} (x - \mathbb{E}X)^2 f_X(x) dx$ , где  $f_X(x)$  — плотность вероятности. Другими словами,  $\mathbb{V}X = \mathbb{E}(X - \mathbb{E}X)^2$ . В этой книге часто используется тождество  $\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$ .

3 От англ. *mean absolute deviation*. В литературе встречается также обозначение MAD, но в этой книге MAD зарезервировано за средним абсолютным отклонением от медианы, а не от среднего. Если случайная величина  $X$  принимает значения на  $\mathbb{R}$ , то  $\mathbb{M}X = \int_{\mathbb{R}} |x - \mathbb{E}X| f_X(x) dx$ , где  $f_X(x)$  — плотность вероятности.



$\varphi(\cdot)$  и  $f(\cdot)$  обычно зарезервированы за плотностью вероятности заранее указанного распределения. В некоторых главах делается различие между  $f_X(x)$  и  $f_Y(y)$ , особенно когда случайные величины  $X$  и  $Y$  следуют двум разным распределениям.

$n$  обычно зарезервировано за числом слагаемых.

$p$  обычно зарезервировано за порядком момента.

НСВ — непрерывная случайная величина<sup>1</sup>.

$F(\cdot)$  обычно зарезервировано за функцией распределения, то есть  $F(x) = \mathbb{P}(X < x)$ . Функция выживания  $\mathbb{P}(X > x)$  записывается с чертой сверху,  $\bar{F}(\cdot)$  или обозначается буквой  $S$ <sup>2</sup>.

$\sim$  означает, что случайная величина по одну сторону от тильды распределена согласно закону, указанному по другую сторону от тильды.

$\chi(t) = \mathbb{E} e^{itX_s}$  — характеристическая функция случайной величины  $X_s$ . Иногда для аргумента  $t \in \mathbb{R}$  используется другая буква —  $\omega$ . Сама характеристическая функция иногда обозначается заглавной  $\Psi$ <sup>3</sup>.

$\xrightarrow{D}$  означает сходимость по распределению, то есть следующее. Пусть  $X_1, X_2, \dots$  — последовательность случайных величин; тогда  $X_n \xrightarrow{D} X$  означает, что последовательность соответствующих функций распределения  $F_n$  имеет предел:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

при всяком действительном  $x$ , при котором  $F$  непрерывна.

$\xrightarrow{P}$  означает сходимость по вероятности, то есть что при  $\varepsilon > 0$  для описанной выше последовательности

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

$\xrightarrow{a.s.}$  означает сходимость почти наверное<sup>4</sup>, то есть более сильное требование:

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

$S_n$  обычно обозначает сумму  $n$  слагаемых.

$\alpha$ , а также  $\alpha_p$  и  $\alpha_s$ . Во избежание двусмысленности мы будем прибегать к двум обозначениям:  $\alpha_s \in (0, 2]$  для показателя хвоста платонического (предельного) устойчивого распределения;  $\alpha_p \in (0, \infty)$  для показателя хвоста в распределении Парето (доасимптотическом). В недвусмысленном контексте можем обходиться просто  $\alpha$ .

$\mathcal{N}(\mu_1, \sigma_1^2)$  — нормальное (гауссово) распределение со средним  $\mu_1$  и дисперсией  $\sigma_1^2$ <sup>5</sup>.

Другими словами,  $\mathbb{M}X = \mathbb{E}|X - \mathbb{E}X|$ . Известна теорема:  $\mathbb{M}X \leq \sqrt{\mathbb{V}X}$ ; это следствие неравенства Йенсена. «Центрировать», то есть вычитать  $\mathbb{E}X$ , приходится в тех случаях, где  $\mathbb{E}X \neq 0$ .

1 Это традиционное сокращение в русской литературе. В оригинале традиционное английское сокращение, г. v. от *random variable*.

2 От англ. *survival function*. В русской литературе называется также функцией надежности и обозначается буквой  $R$ , от англ. *reliability function*. Вместо особых обозначений  $\bar{F}(\cdot)$ ,  $S(\cdot)$ ,  $R(\cdot)$  часто просто выписывают разность  $1 - F(\cdot)$ .

3 В литературе можно встретить также обозначение  $\varphi(t)$ .

4 От англ. *almost sure*. Синонимы: сходимость почти всюду (*almost everywhere*), сходимость почти всегда (*almost always*).

5 Случай  $\mathcal{N}(0, 1)$  называется стандартным нормальным распределением, и его функцию распределения иногда обозначают  $\Phi$ , без параметров. Тогда для произвольного нормального распределения  $\mathcal{N}(\mu, \sigma^2)$

$\mathcal{L}(\cdot, \cdot)$  или  $\mathcal{LN}(\cdot, \cdot)$  — логнормальное распределение, с плотностью  $f^{(L)}(\cdot)$ . Здесь обычно параметры указываются как  $\mathcal{L}\left(X_0 - \frac{1}{\sigma^2}, \sigma\right)$ ; тогда математическое ожидание  $X_0$  и дисперсия  $(e^{\sigma^2} - 1)X_0^2$ <sup>1</sup>.

$S(\alpha_s, \beta, \mu, \sigma)$  — устойчивое распределение с показателем хвоста  $\alpha_s \in (0, 2]$ , коэффициентом симметрии  $\beta$  в интервале  $(-1, 1)$ , коэффициентом положения  $\mu \in \mathbb{R}$  и коэффициентом масштаба  $\sigma > 0$ .

$\mathfrak{P}$  — класс степенного закона (см. ниже).

$\mathfrak{S}$  — субэкспоненциальный класс (см. ниже).

$\delta(\cdot)$  — дельта-функция Дирака.

$\vartheta(\cdot)$  — тета-функция Хевисайда<sup>2</sup>.

$\text{erf}(\cdot)$  — функция ошибок, представляющая собой интеграл плотности гауссова распределения<sup>3</sup>

$$\text{erf } z = \frac{2}{\sqrt{\pi}} \int_0^z dt e^{-t^2}.$$

функция распределения  $\Phi_{\mu, \sigma}$  представима через стандартную:  $\Phi_{\mu, \sigma}(z) = \Phi\left(\frac{z - \mu}{\sigma}\right)$ . Автор поступает так в разделе 2.2.3 Центральная предельная теорема (ЦПТ).

Параметр  $\sigma > 0$  называют коэффициентом масштаба или среднеквадратическим отклонением; встречаются также синонимы среднеквадратичное отклонение, стандартное отклонение, STD.

Для гауссова распределения со средним  $\mu$  и масштабом  $\sigma$  плотность вероятности  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right)$ , дисперсия  $\sigma^2$ , коэффициент асимметрии 0, эксцесс 3, четвертый кумулянт 0 и прочие кумулянты 0. Для суммы  $n$  случайных н. о. р. по Гауссу величин последовательность кумулянтов  $\kappa_n^1 = n\mu$ ,  $\kappa_n^2 = n\sigma^2$ ,  $\kappa_n^3 = 0$  и далее только нули; соответственно  $\lim_{n \rightarrow \infty} \kappa_n^p = 0$ .

- 1 Логнормальное распределение — это распределение случайной величины  $e^Y$ , где  $Y$  — гауссова случайная величина. Если гауссова случайная величина  $Y$  имеет среднее  $\mu$  и дисперсию  $\sigma^2$ , то логнормальная случайная величина  $X = e^Y$  имеет среднее  $X_0 = e^{\mu + \sigma^2/2}$  и дисперсию  $(e^{\sigma^2} - 1)X_0^2$ . Величина  $X = e^Y$  принимает только положительные значения; ее распределение одногорбое и несимметричное. Плотность вероятности  $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$ ; медиана  $e^\mu = X_0 e^{-\sigma^2/2}$ , мода  $e^{\mu - \sigma^2} = X_0 e^{-3\sigma^2/2}$ , коэффициент асимметрии  $(e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}$ , коэффициент эксцесса  $e^{4\sigma^2} + 2 e^{3\sigma^2} + 3 e^{2\sigma^2} - 6$ .

- 2 (Heaviside step function) Функция ступенька в нуле,  $\vartheta(x) \triangleq \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$ . Если нужна ступенька в произвольной

точке  $K$ , используют  $\vartheta_K(x) = \begin{cases} 1, & x \geq K \\ 0, & \text{иначе} \end{cases} = \vartheta(x - K)$ ; можно представить и через индексную функцию  $\mathbb{1}_{x \geq K}$ .

Иногда считают, что  $\vartheta(x) \triangleq \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$ ; такое разноточие  $\vartheta(0)$  несущественно при моделировании вероятностных распределений, потому что не влияет на интегралы.

В этой книге функция Хевисайда обозначается некурсивной  $\vartheta$ , чтобы отличать от локального использования  $\vartheta$  для произвольного параметра или функции. Аналогичным образом мнимая единица обозначается некурсивной  $i$ , а число Архимеда обозначается некурсивной  $\pi$ , чтобы отличать от локального использования  $i$  или  $\pi$  для произвольного параметра или функции.

- 3 Связь с нормальным распределением такая: интегрируется функция плотности  $f(t) = \frac{1}{\sqrt{\pi}} e^{-t^2}$  распределения  $\mathcal{N}\left(0, \frac{1}{2}\right)$ , нормального распределения со средним 0 и среднеквадратическим отклонением  $\frac{1}{\sqrt{2}}$ , причем нижний предел интеграла 0 и результат умножен на 2. Можно связать со стандартным нормальным распределением и так:  $\Phi(z) = \frac{1}{2}\left(1 + \text{erf} \frac{z}{\sqrt{2}}\right)$ .

Более удобная связь: в пределы  $\mu \pm \sigma$  попадает  $\text{erf} \frac{\sigma\sqrt{2}}{2}$  всех нормально распределенных наблюдений. Например, в пределы  $\pm \sigma$  попадает  $\text{erf} \frac{\sqrt{2}}{2} \approx 68,2\%$  наблюдений.

$\operatorname{erfc}(\cdot)$  — дополнительная функция ошибок,  $1 - \operatorname{erf}(z)$ .

$\|\cdot\|_p$  — норма; в этой книге<sup>1</sup> применяется к действительному вектору  $\mathbf{X} = (X_1, \dots, X_n)^T$  и определяется как

$$\|\mathbf{X}\|_p \triangleq \left( \frac{1}{n} \sum_{i=1}^n |X_i|^p \right)^{\frac{1}{p}}.$$

Обратите внимание, что компоненты вектора берутся по абсолютной величине.

${}_1F_1(\cdot; \cdot; \cdot)$  — вырожденная гипергеометрическая функция:

$${}_1F_1(a; b; z) = \sum_{k=0}^{\infty} \frac{a_k}{b_k} \frac{z^k}{k!}.$$

${}_2\tilde{F}_2(\cdot, \cdot; \cdot, \cdot; \cdot)$  — регуляризация обобщенной гипергеометрической функции  ${}_2F_2$ :

$${}_2\tilde{F}_2(a_1, a_2; b_1, b_2; z) \triangleq {}_2F_2(a_1, a_2; b_1, b_2; z) \frac{\Gamma(b_1)\Gamma(b_2)}{\Gamma(a_1)\Gamma(a_2)},$$

где обобщенная гипергеометрическая функция  ${}_pF_q(\cdot; \cdot; \cdot)$  раскладывается в ряд

$${}_pF_q(a; b; z) \triangleq \sum_{k=0}^{\infty} \frac{(a_1)_k \dots (a_p)_k}{(b_1)_k \dots (b_q)_k} \frac{z^k}{k!}$$

с использованием символа Похгаммера<sup>2</sup>  $(a)_n = \prod_{i=0}^{n-1} (a+i)$ .

## 2.2. СИСТЕМАТИЧЕСКИЙ КАТАЛОГ ОБЩИХ И ИДИОСИНКРАЗИЧЕСКИХ ПОНЯТИЙ

Ниже дублируются определения из основных разделов.

### 2.2.1. Класс степенного закона $\mathfrak{P}$

Принято определять класс степенного закона по свойству функции выживания следующим образом.

Пусть  $X$  — случайная величина из класса распределений с правым хвостом, подчиняющимся *степенному закону*, то есть:

$$\mathbb{P}(X > x) = L(x) x^{-\alpha}, \quad (2.1)$$

где  $L: [x_{\min}, +\infty) \rightarrow (0, +\infty)$  — медленно меняющаяся функция, определяемая требованием

$$\lim_{x \rightarrow \infty} \frac{L(kx)}{L(x)} = 1$$

для всех  $k > 0$  [22].

<sup>1</sup> В литературе также  $p$ -нормой называют Гёльдерову  $\ell_p$ -норму  $\left( \sum_{i=1}^n |X_i|^p \right)^{\frac{1}{p}}$ , без деления на  $n$ .

<sup>2</sup> Так символ Похгаммера используется в литературе по гипергеометрической функции. В литературе по комбинаторике как  $(a)_n$  обозначают убывающий факториал  $\prod_{i=0}^{n-1} (a-i)$ , тогда как возрастающий факториал  $\prod_{i=0}^{n-1} (a+i)$  там обозначают  $(a)^{(n)}$ .

Тогда говорят, что функция выживания случайной величины  $X$  принадлежит классу *правильно меняющихся на бесконечности функций*  $RV_\alpha^+$ <sup>1</sup>.

Давайте уточним: функция  $f: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  меняется на бесконечности с показателем  $\rho$ , то есть  $f \in RV_\rho^+$ , когда

$$\lim_{x \rightarrow \infty} \frac{f(tx)}{f(t)} = x^{\rho^2}.$$

С практической точки зрения это значит, что рано или поздно  $L(x)$  подходит к своему пределу  $l$  и становится константой, которую мы будем называть *константой Караматы*; рубеж, где достигается константа, будем называть *точкой Караматы*. За этой точкой хвосты степенного закона калибруются стандартными методами, такими как характеристика Хилла. Б. Мандельброт называл распределение в этой области *сильным законом Парето* [162], [75].

То же верно при соответствующих оговорках для левых хвостов.

### 2.2.2. Закон больших чисел (слабый)

Обычно его представляют так. Пусть  $X_1, X_2, \dots, X_n$  — бесконечная последовательность независимых одинаково распределенных случайных величин, интегрируемых по Лебегу, с математическим ожиданием  $\mathbb{E} X_i = \mu$  (вообще говоря, требование н. о. р. можно до некоторой степени ослабить).

Тогда выборочное среднее первых  $n$  величин  $\bar{X}_n = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$  сходится к математическому ожиданию,  $\bar{X}_n \rightarrow \mu$  при  $n \rightarrow \infty$ .

Конечность дисперсии не обязательна (однако весьма желательна: если дисперсия и прочие высшие моменты распределения конечны, то  $\bar{X}_n$  сходится быстрее).

Когда потребуется, рассмотрим и сильный закон больших чисел.

### 2.2.3. Центральная предельная теорема (ЦПТ)

Классический вариант ЦПТ, теорема Линдеберга-Леви, утверждает следующее. Пусть дана последовательность  $X_i$  н. о. р. величин с  $\mathbb{E} X_i = \mu$  и  $\mathbb{V} X_i = \sigma^2 < +\infty$ , и пусть  $\bar{X}_n$  — это среднее по выборке первых  $n$  величин. Тогда по мере приближения  $n$  к бесконечности центрированное и нормированное среднее  $\sqrt{n}(\bar{X}_n - \mu)$  сходится по распределению к гауссову [20] [21]

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma^2).$$

1 От англ. *regularly varying*.

2 Теория строится так. Функция  $f(t)$  правильно меняется на бесконечности, если  $\forall x \in \mathbb{R}^+$  существует и конечен предел  $\lim_{t \rightarrow \infty} \left| \frac{f(tx)}{f(t)} \right|$ . Согласно теореме Караматы, такая функция представима в виде степенного закона  $t^\rho$  с точностью до медленно меняющегося сомножителя  $L(t)$ :  $f(t) = L(t) t^\rho$ .

Очевидным образом верно и обратное: для функции вида  $f(t) = L(t) t^\rho$  и числа  $x \in \mathbb{R}^+$  предел

$$\lim_{t \rightarrow \infty} \left| \frac{f(tx)}{f(t)} \right| = \lim_{t \rightarrow \infty} \left| \frac{L(tx)(tx)^\rho}{L(t)t^\rho} \right| = \lim_{t \rightarrow \infty} \left| \frac{L(tx)}{L(t)} \right| \cdot \lim_{t \rightarrow \infty} \left| \frac{(tx)^\rho}{t^\rho} \right| = 1 \cdot \lim_{t \rightarrow \infty} |x^\rho| = x^\rho.$$

Сходимость по распределению означает, что функция распределения для  $\sqrt{n}(\bar{X}_n - \mu)$  поточечно сходится к  $\mathcal{N}(0, \sigma^2)$ , то есть что для всякого действительного  $z$

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sqrt{n}(\bar{X}_n - \mu) \leq z) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq \frac{z}{\sigma}\right) = \Phi\left(\frac{z}{\sigma}\right), \sigma > 0,$$

где  $\Phi(z)$  — значение стандартного нормального распределения в точке  $z$ .

Есть ряд других вариантов ЦПТ, которые мы представим по мере надобности.

### 2.2.4. Закон средних чисел, или Предасимптотика

Это центральная тема этой книги. Нас интересует поведение случайной величины для умеренно большого  $n$ , или предасимптотика. Вопрос не так актуален для гауссова распределения, поскольку оно сходится быстро (в силу ЗБЧ и ЦПТ); другое дело — негауссовы случайные величины.

Смотрите далее в разделе о показателе каппа.

### 2.2.5. Показатель каппа

Здесь показатель не в алгебраическом смысле, как показатель степени, а в инженерном, как количественный параметр машины<sup>1</sup>. Каппа оценивает доасимптотическое поведение случайной величины. Этот показатель разработан автором, как описано в Главе 8 и статье [235]. Каппа пробегает интервал  $[0, 1]$ ;  $\kappa = 0$  для гауссовой случайной величины и  $\kappa = 1$  для распределения Коши или иной случайной величины, не имеющей математического ожидания<sup>2</sup>.

Пусть  $X_1, \dots, X_n, \dots$  — случайные величины н. о. р. с конечным математическим ожиданием, то есть  $\mathbb{E} X < +\infty$ . Пусть  $S_n = X_1 + X_2 + \dots + X_n$  — частичная сумма. Пусть  $\mathbb{M}(n) = \mathbb{E}|S_n| - \mathbb{E} S_n$  — математическое ожидание абсолютного отклонения частичной суммы  $n$  слагаемых от математического ожидания этой суммы (как мы уже предупреждали, у нас отклонение отсчитывается не от медианы, а от среднего). Определим *скорость сходимости* при увеличении числа слагаемых от  $n_0$  до  $n$ :

$$\frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \left(\frac{n}{n_0}\right)^{1/2 - \kappa_{n_0, n}}, \quad (2.2)$$

где  $n_0, n = 1, 2, \dots$  и  $n > n_0 \geq 1$ ; соответственно

$$\kappa_{n_0, n} = 2 - \frac{\ln n - \ln n_0}{\ln \frac{\mathbb{M}(n)}{\mathbb{M}(n_0)}}. \quad (2.3)$$

1 В оригинале автор делает другую оговорку, из-за иной омонимии: что *Kappa metric* — не метрика в смысле расстояния в той или иной геометрии, а метрика в инженерном смысле.

2 Как и дисперсии.

В дальнейшем мы будем часто пользоваться значениями  $n = n_0 + 1$  и сокращать обозначение до  $\kappa_n$ .

### 2.2.6. Эллиптическое распределение

О случайном векторе  $\mathbf{X}$  размерности  $p \times 1$  говорят, что у него эллиптическое распределение (или распределение с эллиптическим контуром) с параметрами положения  $\mu$ , неотрицательной матрицей  $\Sigma$  и некоторой скалярной функцией  $\Psi$ , если характеристическая функция представима в виде  $\exp(it'\mu)\Psi(t\Sigma t')$ .

С практической точки зрения эллиптическое распределение должно собираться из распределений с одной и той же ковариационной матрицей. Переключение режима или стохастические ковариации (корреляции) мешают распределению быть эллиптическим. И мы покажем в Главе 6, что линейная комбинация случайных величин, следующих распределением с тонким хвостом, способна генерировать взрывные толстохвостые свойства, когда эллиптичность нарушается. Этот эффект, наряду со случаями жирного хвоста, делает несостоятельной значительную часть современной финансовой науки.

### 2.2.7. Статистическая независимость

Независимость между двумя случайными величинами  $X$  и  $Y$  с частными функциями плотности вероятности  $f_X(x)$  и  $f_Y(y)$  и совместной функцией плотности вероятности  $f(x, y)$  определяется тождеством:

$$\frac{f(x, y)}{f_X(x)f_Y(y)} = 1,$$

независимо от коэффициента корреляции. В классе эллиптических распределений, когда совместное гауссово распределение имеет коэффициент корреляции 0, случайные величины независимы, и некоррелированы. Иначе обстоит дело с многомерными формами t-распределения Стюдента или распределения Коши.

### 2.2.8. Устойчивое распределение (устойчивое по Леви)

Это обобщение ЦПТ.

Пусть  $X_1, \dots, X_n$  — независимые одинаково распределенные случайные величины. Рассмотрим их сумму  $S_n$ . Теорема утверждает, что

$$\frac{S_n - a_n}{b_n} \xrightarrow{D} X_S, \quad (2.4)$$

где  $X_S$  следует устойчивому распределению  $S$ ,  $a_n$  и  $b_n$  — нормирующие константы, а  $\xrightarrow{D}$ , как вы помните, означает сходимость по распределению (распределению  $X$  при  $n \rightarrow \infty$ ).

Свойства  $S$  будут должным образом определены и рассмотрены в следующей главе. Пока заметим, что про случайную величину  $X_S$  говорят, что она следует устой-

чивому (или  $\alpha$ -устойчивому) распределению, и пишут  $X_S \sim S(\alpha_S, \beta, \mu, \sigma)$ , если ее характеристическая функция  $\chi(t) = \mathbb{E}e^{itX_S}$  имеет вид:

$$\chi(t) = \exp\left(i\mu t - |t\sigma|^{\alpha_S} \left(1 - i\beta \operatorname{tg} \frac{\pi\alpha_S}{2} \operatorname{sgn} t\right)\right), \text{ где } \alpha_S \neq 1. \quad (2.5)$$

Ограничения:  $-1 \leq \beta \leq 1$  и  $0 < \alpha_S \leq 2$ <sup>1</sup>.

### 2.2.9. Многомерное устойчивое распределение

О случайном векторе  $\mathbf{X} = (X_1, \dots, X_k)^T$  говорят, что он имеет многомерное устойчивое распределение, если каждая линейная комбинация его компонент  $Y = a_1X_1 + \dots + a_kX_k$  имеет устойчивое распределение. То есть каждая векторная константа  $\mathbf{a} \in \mathbb{R}^k$  должна давать устойчивое одномерное распределение для случайной величины  $Y = \mathbf{a}\mathbf{X}$ .

#### 2.2.10. Точка Караматы

См. Класс степенного закона.

#### 2.2.11. Субэкспоненциальность

Естественной границей между Медиокристаном<sup>2</sup> и Экстремистаном служит субэкспоненциальный класс, обладающий следующим свойством.

Пусть  $X_1, \dots, X_n$  — последовательность независимых одинаково распределенных случайных величин с носителем в  $(\mathbb{R}^+)$  и кумулятивной функцией распределения  $F$ .

Субэкспоненциальный класс определяется требованием (см. [248], [196]):

$$\lim_{x \rightarrow \infty} \frac{1 - F^{*2}(x)}{1 - F(x)} = 2, \quad (2.6)$$

где  $F^{*2} = F' * F$  — это кумулятивное распределение  $X_1 + X_2$ , суммы двух независимых копий случайной величины  $X$ . Требование означает, что вероятность того, что сумма  $X_1 + X_2$  превысит значение  $x$ , вдвое выше вероятности того, что значение  $x$  будет пре-

1 Устойчивому распределению, оно же  $\alpha$ -устойчивое по Леви, следует в пределе сумма  $n$  независимых случайных величин при  $n \rightarrow \infty$ ; в случае слагаемых с конечной дисперсией сумма в пределе гауссова, а в более общем случае — нет. Устойчивым это распределение названо потому, что для двух н. о. р.  $X_1$  и  $X_2$  линейная комбинация вида  $aX_1 + bX_2$  следует распределению вида  $cX_1 + d$ .

При  $\alpha_S < 1$  и  $\beta = 1$  устойчивое распределение одностороннее,  $X \in [\mu, +\infty)$ ; при  $\alpha_S < 1$  и  $\beta = -1$  устойчивое распределение одностороннее,  $X \in (-\infty, \mu]$ ; в остальных случаях носитель  $\mathbb{R}$ . При  $\beta = 0$  устойчивое распределение симметричное. При  $\alpha_S = 1$  превращается в распределение Коши, при  $\alpha_S = 2$  превращается в гауссово распределение. При  $1 < \alpha_S < 2$  и  $\beta = 1$  называется распределением Парето.

Среднее равно коэффициенту положения  $\mu$  при  $\alpha_S > 1$ ; иначе не существует. Дисперсия  $2\sigma^2$  при  $\alpha_S = 2$ ; иначе бесконечна. Коэффициент асимметрии 0 при  $\alpha_S = 2$ ; иначе не существует. Экссесс 3 при  $\alpha_S = 2$ ; иначе не существует.

2 Медиокристан и Экстремистан — вымышленные области, от англ. *mediocre* (заурядность) и *extreme* (крайность).

вышено любым отдельно взятым слагаемым. Значит, для больших  $x$  все случаи, когда сумма превышает  $x$ , она обязана этим только одному из слагаемых — большому из двух, — тогда как вклад другого пренебрежимо мал<sup>1</sup>.

Обобщая, можно доказать, что и в сумме  $n$  величин преобладает одна из них, максимальная. Формально следующие два свойства эквивалентны условию субэкспоненциальности [43], [84].

Для данного  $n \geq 2$  пусть  $S_n = \sum_{i=1}^n X_i$  и  $M_n = \max_{1 \leq i \leq n} X_i$ . Тогда

$$\text{a) } \lim_{x \rightarrow \infty} \frac{\mathbb{P}(S_n > x)}{\mathbb{P}(X > x)} = n,$$

$$\text{b) } \lim_{x \rightarrow \infty} \frac{\mathbb{P}(S_n > x)}{\mathbb{P}(M_n > x)} = 1.$$

Таким образом, сумма  $S_n$  сравнима по величине с наибольшим слагаемым  $M_n$ , другими словами — хвосты играют главную роль.

На интуитивном уровне важно понять, что частота событий в хвосте субэкспоненциальных распределений падает медленнее, чем в экспоненциальном распределении, где событиями в далеком хвосте можно пренебречь.

В самом деле, можно доказать, что у субэкспоненциальных распределений нет экспоненциальных моментов:

$$\int_0^{\infty} e^{\varepsilon x} dF(x) = +\infty \quad (2.7)$$

для всех  $\varepsilon$  больше нуля. Однако обратное неверно, поскольку распределения могут не иметь экспоненциальных моментов и все равно не удовлетворять субэкспоненциальному условию.

### 2.2.12. t-распределение Стьюдента как прокси

Мы используем t-распределение Стьюдента с  $\alpha$  степенями свободы как удобное распределение степенного закона с двумя хвостами. При  $\alpha = 1$  оно превращается в распределение Коши, а при  $\alpha \rightarrow \infty$ , естественно, в гауссово.

t-распределение Стьюдента — это главный колоколообразный степенной закон, то есть плотность вероятности непрерывная и гладкая, асимптотически приближается к нулю для больших  $x$ , отрицательных или положительных, и унимодальна, то есть ее максимум — единственный (кроме того, плотность вероятности квазивыпукла вверх, хотя и не выпукла вверх).

### 2.2.13. Круг цитирования

Замкнутый механизм, помогающий академической карьере авторов статей, которые считаются выдающимися, поскольку их цитируют, без фильтрации по внешним критериям; в результате исследователи оседают по уютным углам, сосредоточившись

<sup>1</sup> Имеется в виду, что в хвосте события редкие и при некотором  $x$  в хвосте вероятность превзойти  $x$  увеличится вдвое, если разрешить две попытки:  $\mathbb{P}(X_1 > x \text{ или } X_2 > x) \approx 2\mathbb{P}(X_1 > x)$ . Вообще говоря, если разрешить суммировать результаты двух попыток, шансы превзойти  $x$  еще увеличатся:  $\mathbb{P}(X_1 + X_2 > x) > 2\mathbb{P}(X_1 > x \text{ или } X_2 > x)$ . Но заметной эта прибавка будет только в ситуации, где существенна вероятность совпадения, что в каждой из двух попыток результат сравним с  $x$ .



на несущественных частных вопросах. Этот механизм сложился в условиях функционирования академической системы без контроля извне и без собственной шкуры на кону.

Примеры областей исследований, которые по сути шарлатанство, так как их результаты непереносимы в реальную жизнь и пригодны только как пища для новых статей, которые, в свою очередь, произведут только новые статьи: современная портфельная теория<sup>1</sup>, эконометрия (особенно в части макроэкономических параметров), процедуры GARCH<sup>2</sup>, психометрия, модели стохастического управления финансами, поведенческая экономика и финансирование, принятие решений в условиях неопределенности, макроэкономика и прочее.

#### 2.2.14. Погоня за рентой в научном мире

Существует конфликт интересов<sup>3</sup> между исследователем и его темой. Целью кафедры и ее сотрудников становится цитируемость, награды и прочее в ущерб собственно исследованию; например, чтобы не браться за трудные задачи, многие разбредаются по углам, где исследование полезно для личной карьеры и для кафедры, а не для дела.

#### 2.2.15. Псевдоэмпиризм, или Проблема Пинкера<sup>4</sup>

Привлечение «доказательств», лишенных статистической значимости, или использование показателей, которые неинформативны, поскольку неприменимы к рассматриваемым случайным величинам (например, вывод на основе средних значений или корреляция величин с толстым хвостом). Причина болезни в том, как учат будущих исследователей в общественных науках:

- i) статистике учат на примере гауссовых величин или величин с тонким хвостом,
- ii) не учат теории вероятностей и плохо учат терминам статистики,
- iii) вовсе не учат работать с многомерностью.

1 Англ. *Modern Financial Theory* или *Modern Portfolio Theory*. Идея портфельных инвестиций на основе гипотезы эффективного рынка. Разработана Гарри Марковицем в 1950-е годы.

2 Прогнозирование на основе анализа регрессии; предложено Робертом Энглем в 1982 (ARCH — AutoRegressive Conditional Heteroscedasticity) и развито Тимом Боллерслевом в 1986 (Generalized ARCH — GARCH).

3 Погоня за рентой (*rent-seeking*) — это злоупотребление своим положением или иные нечестные способы конкуренции.

4 Имеется в виду книга канадско-американского социопсихолога и популяризатора Стивена Пинкера *The Better Angels of Our Nature: Why Violence Has Declined* (2011), русское издание «Лучшее в нас. Почему насилия в мире стало меньше» (2021). Нассим Талеб выступил с разоблачением традиционно легкомысленных статистических рассуждений в книге Пинкера; последовала шумная баталия в СМИ. Для тех, кто станет искать подробности, важно знать, что та полемика не имела отношения к недавней атаке на Стивена Пинкера за недостаток политической корректности; в последнем случае Талеб высказался в защиту прав Пинкера.

Вот примеры псевдоэмпиризма: сравнивают число смертей от террористических актов или при эпидемиях вроде эболы (где у случайной величины хвост жирный) и несчастные случаи при падении со стремянок (где хвост тонкий).

Смещение в сторону подтверждения наблюдений — бич современной науки; оно приводит к ложно-положительным выводам как в случаях многомерных случайных величин, так и в случаях жирного хвоста.

На самом деле даже не обязательно знать разницу между величинами с жирным хвостом и гауссовыми, чтобы заметить шаткость таких рассуждений, поскольку они не удовлетворяют простому критерию статистической значимости. Впрочем, рассуждающие обычно вовсе не разбираются в статистической значимости.

### 2.2.16. Предасимптотика

Математическая статистика любит работать с суммой  $n$  слагаемых при  $n = 1$  или  $n = \infty$ . Посередине находится то, что мы называем реальностью и чему главным образом посвящена эта книга. Некоторые суммы (скажем, те, где у слагаемых конечная дисперсия) ведут себя, как гауссова величина, асимптотически, при  $n = \infty$ , однако не делают этого даже при некоторых огромных  $n < \infty$ .

### 2.2.17. Стохастизация

Замена детерминированного параметра на случайный; можно (i) простым способом, можно (ii) с применением более сложного непрерывного или дискретного распределения.

i) Пусть  $s$  — детерминированный параметр; стохастизируем его простейшим способом, заменив на случайную величину  $S$ , следующую распределению<sup>1</sup> Бернулли с двумя значениями:  $s_1$ , которое принимается с вероятностью  $p$ , и  $s_2$ , которое принимается с вероятностью  $1 - p$ . Стохастизация сохранит среднее, если  $ps_1 + (1 - p)s_2 = s$ , то есть когда значение параметра  $s$  сохраняется в виде среднего значения случайной величины  $S$ . Вообще говоря, заменить некоторый параметр  $s$  некоторого распределения  $f$  на случайную величину  $S$  можно так, чтобы сохранить дисперсию распределения или какие-то другие характеристики.

ii) Можно использовать полноценное распределение вероятностей. Обычно берут гауссово, если нужна случайная величина с двумя хвостами, а когда нужен один хвост, берут логнормальное или экспоненциальное распределение, реже степенной закон. Когда  $s$  — это среднееквадратическое отклонение некоторой случайной величины, можно стохастизировать  $s^2$ , создав *стохастическую волатильность*; дисперсию или среднееквадратическое отклонение параметра  $s$  называют *волатильностью волатильности* и обозначают V-vol.

<sup>1</sup> По Бернулли, распределена случайная величина  $X$ , принимающая только два значения: 0 или 1. Если вероятность  $\mathbb{P}(X=1)=p$ , то вероятность  $\mathbb{P}(X=0)=1-p$ ; среднее  $p$ ; дисперсия  $p(1-p)$ ; коэффициент асимметрии  $(1-2p)/\sqrt{p(1-p)}$ ; эксцесс  $(3p^2-3p+1)/p(1-p)$ . Здесь, очевидно, используется случайная величина  $S = s_2 + (s_1 - s_2)X$ ; такая случайная величина принимает значения  $s_1$  и  $s_2$  с вероятностью  $p$  и  $1-p$  соответственно; среднее  $ps_1 + (1-p)s_2$ .

### 2.2.18. Стоимость под риском, условная стоимость под риском

Математическая стоимость под риском (*value at risk*, VaR) при пороге<sup>1</sup>  $\lambda \in [0, 1]$  для случайной величины  $X$  с функцией распределения  $F$  выражается как

$$\text{VaR}_\lambda X = -\inf\{x \in \mathbb{R} \mid F(x) > \lambda\}^2,$$

а соответствующая условная стоимость под риском (*conditional value at risk*, CVaR), она же ожидаемые потери (*expected shortfall*, ES)<sup>3</sup> при пороге  $\lambda$ , как

$$\text{ES}_\lambda X = \mathbb{E}(-X \mid X \leq -\text{VaR}_\lambda X)$$

или, рассматривая положительную величину потерь, работают с положительным хвостом распределения<sup>4</sup>.

Обобщая, ожидаемые потери при пороге  $K$  определяют как  $\mathbb{E}(X \mid X > K)$ <sup>5</sup>.

### 2.2.19. Своя шкура на кону

Фильтрующий механизм, который заставляет повара отвеждать собственное блюдо и пострадать, если оно не удалось; так система избавляется от опасных участников.

Ставят на кон свою шкуру, например, сантехники, дантисты, хирурги, инженеры. Их работа приносит осязаемый результат, или предприятие разоряется.

Где нет своей шкуры на кону: в академических кругах. Участники оценивают друг друга, не ощущая давления со стороны реальности, угрожающего их существованию.

<sup>1</sup> Встречается синоним: доверительный уровень.

<sup>2</sup> Подразумевается задача, в которой  $X$  — величина платежа, положительная в случаях прибыли и отрицательная в случаях убытков, и обсуждаются как раз убытки. Стоимость под риском при заданной пороговой вероятности  $\lambda$ , обычно  $\text{VaR}_{0,05} X$ , противоположна такому платежу  $-\text{VaR}_{0,05} X$  (в рассматриваемой задаче отрицательному), что вероятность этих или худших убытков равна 0,05. Говорят: если реализуются угрозы, имеющие вероятность 5 %, вы потеряете стоимость  $\text{VaR}_{0,05} X$  или хуже того. Иногда тот же термин определяют через квантиль. Для этого рассматривают величину  $-X$ , положительную в случаях убытков. С вероятностью  $\lambda$  величина  $-X \geq \text{VaR}_\lambda X$ , то есть  $\text{VaR}_\lambda X = F_{-X}^{-1}(1 - \lambda)$ . Это квантиль уровня  $1 - \lambda$  случайной величины  $-X$ ,  $Q_{-X}(1 - \lambda)$ ; стоимость под риском 5 % — это квантиль уровня 0,95, или 95-й процентиль величины  $-X$ .

<sup>3</sup> Есть еще несколько синонимов: средняя стоимость под риском (*average value at risk*, AVaR), математическое ожидание хвостовых потерь (*expected tail loss*, ETL), надквантиль (*superquantile*). Часто используется уровень  $\lambda = 0,05$ , то есть среднее по потерям хуже 5-го перцентиля.

<sup>4</sup> В последнем случае, очевидно, рассматриваемая случайная величина  $-X$  положительна в случае убытков. Соответственно формула в определении ожидаемых потерь примет вид  $\text{ES}_\lambda(X) = \mathbb{E}(-X \mid -X \geq \text{VaR}_\lambda X)$ , среднее вычисляется в положительном хвосте  $-X \mid -X \geq \text{VaR}_\lambda X$ .

<sup>5</sup> Здесь величина  $X$  не платеж, как выше, а противоположная величина, «величина потерь», и здесь «порог»  $K$  не по вероятности, как порог  $\lambda$  в приведенных выше определениях  $\text{VaR}_\lambda$  и  $\text{ES}_\lambda$ , а по величине потерь.

### 2.2.20. График MS

График MS (*maximum to sum*, максимум к сумме) показывает проявление ЗБЧ на данный момент, вклад максимального наблюдения в итог и поведение итога по мере роста  $n$ .

Чтобы узнать для НСВ  $X$ , существует ли  $\mathbb{E} X^p$ , нужно пронаблюдать сходимость согласно закону больших чисел или ее отсутствие, изучив поведение высших статистических моментов в данной выборке. Удобно это делать по графику MS, как показано на Рисунке 10.3.

График MS опирается на следствие из закона больших чисел [184], касающегося максимального значения случайной величины. Для последовательности  $X_1, X_2, \dots, X_n, \dots$  неотрицательных случайных величин н. о. р., если  $\mathbb{E} X^p < \infty$  при  $p = 1, 2, 3, \dots$  то отношение

$$R_n^p = \frac{M_n^p}{S_n^p} \xrightarrow{\text{a.s.}} 0 \text{ при } n \rightarrow \infty,$$

где  $S_n^p = \sum_{i=1}^n X_i^p$  — частичная сумма, а  $M_n^p = \max(X_1^p, \dots, X_n^p)$  — частичный максимум. (Заметим, что в качестве  $X$  можно взять абсолютную величину случайной величины, если НСВ может принимать отрицательные значения, и тогда данный подход будет применим и к нечетным статистическим моментам.)

### 2.2.21. Максимальный аттрактор (MDA)

Теория экстремальных значений рассматривает распределение максимума по  $n$  экземплярам НСВ, когда это распределение при  $x \rightarrow x^*$ , где  $x^* = \sup\{x: F(x) < 1\}$  (точка, в которой распределение «кончается»<sup>1</sup>), приближается, с точностью до последовательностей нормирующих коэффициентов, к одному из предельных распределений  $G(x)$  — максимальных аттракторов<sup>2</sup> [116]. Другими словами,

$$\frac{\max(X_1, X_2, \dots, X_n) - b_n}{a_n} \xrightarrow{p} G(x).$$

### 2.2.22. Подмена интеграла в литературе для психологов

В литературе без формул совершается следующее смешение понятий. Пусть  $K \in \mathbb{R}^+$  — порог,  $f(\cdot)$  — плотность вероятности,  $p_K \in [0, 1]$  — соответствующая вероятность превысить порог и  $g(\cdot)$  — функция воздействия. Обозначим как  $I_1$  ожидаемый платеж при превышении  $K$ :

$$I_1 = \int_K^\infty g(x)f(x)dx$$

и как  $I_2$  платеж при  $K$ , умноженный на вероятность превысить  $K$ :

$$I_2 = g(K) \int_K^\infty f(x)dx = g(K)p_K.$$

<sup>1</sup> Может быть и бесконечной.

<sup>2</sup> *Maximum domain of attraction*, MDA.

Подмена происходит из-за смещения  $I_1$  и  $I_2$ , которые на самом деле совпадают только в случае, когда  $g(\cdot)$  — константа в области выше  $K$  (скажем, когда  $g(x) = \mathfrak{g}_K(x)$ , тета-функция Хевисайда).

Если же, напротив,  $g(\cdot)$  изменяется с положительной первой производной, интеграл  $I_1$  будет близок к  $I_2$  только при распределении с тонким хвостом, а в случае жирного хвоста — нет<sup>1</sup>.

### 2.2.23. Попытка вынести вероятность за скобку (еще одна типичная ошибка)

Пусть  $F: \mathcal{A} \rightarrow [0, 1]$  — распределение вероятностей (с производной  $f$ ), а  $g: \mathbb{R} \rightarrow \mathbb{R}$  — измеримая функция «платежа». Понятно, что платеж по области  $\mathcal{A}$  — подмножеству множества  $\mathcal{A}$  составит

$$\int_{\mathcal{A}'} g(x) dF(x) = \int_{\mathcal{A}'} f(x) g(x) dx \neq \int_{\mathcal{A}'} f(x) dx g \left( \int_{\mathcal{A}'} dx \right).$$

В дискретном случае для функции вероятности  $\pi(\cdot)$

$$\sum_{x \in \mathcal{A}'} \pi(x) g(x) \neq \sum_{x \in \mathcal{A}'} \pi(x) g \left( \frac{1}{n} \sum_{x \in \mathcal{A}'} x \right) = \quad (2.8)$$

$\equiv (\text{вероятность события}) \times (\text{платеж при среднем событии}).$

Общая идея в том, что умножение на вероятность не выносится за знак интеграла или суммы; если по ставкам на разные события разная вероятность выигрыша, ожидаемый платеж при некотором наборе ставок разной величины на разные события нельзя посчитать по средней вероятности выигрыша.

### 2.2.24. Линейка Витгенштейна

Имеется в виду следующий парадокс: вы правда измеряете стол линейкой или на самом деле вы измеряете линейку столом?<sup>2</sup> В нашем случае ответ зависит от результата.

Допустим, имелось только две альтернативы: гауссово распределение и степенной закон. Мы докажем, что большое отклонение, скажем на «шесть сигма», — признак степенного закона<sup>3</sup>.

### 2.2.25. Черные лебеди

Черные лебеди появляются из-за неполноты нашего знания и могут быть весьма значительны в области жирных хвостов.

1 В частности, в упомянутом случае  $g'(x) > 0$  ожидаемый платеж при условии превышения  $K$  недооценивается теми, кто умножает  $g(K)$  на  $p_K$ .

2 Бывает, что столяр уверен в линейке и проверяет ею новый стол; бывает, что столяр не уверен в линейке и проверяет ее по старому столу, в котором уверен. Исследователь должен быть готов усомниться и в своей «линейке», и в измеряемом ею «столе».

3 Если по линейке размер стола невероятно большой, разумно в ней усомниться.

По сути, это вещи, которые выпадают из вашего горизонта планирования и моделирования, но могут иметь значительные последствия. Речь не о том, чтобы предсказать их, а только о том, чтобы иметь для них выпуклую вниз (или хотя бы не выпуклую вверх) оценку воздействия: хрупкость по отношению к определенному классу событий поддается обнаружению и даже измерению (путем оценки эффектов второго порядка и асимметрии реакций), даже если статистические параметры этих событий установить не удастся.

Тяжело объяснить разработчикам модели, что им нужно научиться работать с вещами, которых они никогда не видели (и даже не представляли), но это нужно сделать<sup>1</sup>.

**Примечание об эпистемологическом измерении:** черные лебеди зависят от наблюдателя; одно и то же событие может быть черным лебедем для индейки и белым лебедем для торговца мясом. 11 сентября стало черным лебедем для жертв нападения, но не для террористов. Зависимость от наблюдателя — неотъемлемое свойство черного лебедя, и его объективная теоретико-вероятностная модель не просто недоступна, она логически невозможна, ибо разрушит моделируемый объект, нарушив существенную для него неполноту информации и ее распространения.

***Серые лебеди:** так называются большие отклонения со значительными последствиями и низкой частотой, но все же не нарушающие статистических параметров. Разумеется, серость лебедя зависит от наблюдателя: лебедь, который сер для разработчика модели с распределением по степенному закону, окажется черен для наивного статистика, тщетно перебирающего стандартные системы моделирования и репрезентации тонких хвостов.*

Повторим лишний раз: черные лебеди возможны и вне жирных хвостов, но в жирных хвостах они важнее. Связь между жирными хвостами и черными лебедями та, что в области жирных хвостов большие отклонения оказывают более сильное воздействие.

## 2.2.26. Выборочная функция распределения ненаблюдаема эмпирически<sup>2</sup>

Выборочная функция распределения  $\hat{F}(t)$  определяется так.

Пусть  $X_1, X_2, \dots, X_n$  — действительные случайные величины н. о. р. с функцией распределения  $F(t)$ . Тогда

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_i \geq t},$$

где  $\mathbb{1}_{\mathcal{A}}$  — индикаторная функция множества  $\mathcal{A}$ .

<sup>1</sup> Как любит приговаривать Пол Портеси (ссылаясь — кажется, безосновательно — на автора этой книги): «Вы еще не видели другую сторону распределения». — *Прим. автора.*

<sup>2</sup> В оригинале игра слов: по-английски такая функция распределения называется *empirical distribution function*.

Согласно теореме Гливленко — Кантелли, независимо от исходного распределения  $F(t)$  статистика Колмогорова<sup>1</sup> равномерно сходится к распределению<sup>2</sup> Колмогорова — Смирнова, причем

$$\sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F(t)| \xrightarrow{\text{a.s.}} 0. \quad (2.9)$$

Эта «сходимость почти наверное» независимо от распределения гарантирована для вероятности, но не для высших моментов; такой результат автор получил и обобщил для «скрытого момента» выше максимума.

Отметим главный результат<sup>3</sup> (который Донскер в дальнейшем обобщил, выведя теорему броуновского моста, для случая ограничения  $t$  интервалом от 0 до 1):

$$\sqrt{n} (F_n(t) - F(t)) \xrightarrow{D} \mathcal{N}(0, F(t)(1 - F(t))). \quad (2.10)$$

Когда говорят, что «выборочная функция распределения ненаблюдаема эмпирически», имеют в виду, что выборочные распределения неизбежно цензируются на интервале  $[x_{\min}, x_{\max}]$ , и в случае жирного хвоста исследователь попадает в трудное положение: хвост не удастся проанализировать в вероятностном пространстве, только в пространстве платежей.

Смотрите также главку о **скрытом хвосте** (следующую).

### 2.2.27. Скрытый хвост

Рассмотрим  $K_n$  — максимум по выборке из  $n$  независимых одинаково распределенных случайных величин:  $K_n = \max(X_1, X_2, \dots, X_n)$ . Пусть  $\varphi(\cdot)$  — плотность исходного распределения. Разложим его статистический момент порядка  $p$  на два слагаемых, где «скрытый» вклад вносится хвостом выше  $K_n$ :

$$\mathbb{E} X^p = \underbrace{\int_L^{K_n} x^p \varphi(x) dx}_{\mu_{L,p}} + \underbrace{\int_{K_n}^{\infty} x^p \varphi(x) dx}_{\mu_{K,p}}.$$

Здесь  $\mu_L$  — вклад наблюдаемой части распределения и  $\mu_K$  — вклад скрытой части (выше  $K$ ).

Согласно теореме Гливленко — Кантелли, распределение  $\mu_{K,0}$  не может зависеть от исходного распределения  $X$ , но на высшие статистические моменты эта теорема не распространяется, так что у исследователя, полагающегося на критерий<sup>4</sup> Колмогорова — Смирнова, могут быть проблемы.

1 Отклонение выборки от исходного распределения, по формуле  $\sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F(t)|$ .

2 Определяется как распределение случайной величины вида  $K = \sup_{t \in [0,1]} |B(t)|$ , где  $B(t)$  — броуновский мост, вероятность того, что за время  $t$  броуновское движение приведет частицу обратно в точку, где она была в момент 0.

3 Считается вариантом центральной предельной теоремы.

4 Можно ли при заданной доверительной вероятности утверждать, что наблюдаемая выборка порождена процессом с заданным распределением (критерий согласия Колмогорова) или что две наблюдаемые выборки порождены процессами с одним и тем же распределением (критерий однородности Смирнова).

### 2.2.28. Теневой момент

В этой книге рассматривается «дополнительная оценка». Бывает, что недостаточно найти среднее по наблюдаемой выборке: в случае распределений с жирным хвостом оно дает систематическую ошибку. Мы показываем, как оценить методом максимального правдоподобия основные параметры распределения, такие как показатель хвоста  $\alpha$ , и рассчитать соответствующее теневое математическое ожидание или высшие статистические моменты.

### 2.2.29. Зависимость в хвосте

Пусть  $X_1$  и  $X_2$  — две случайные величины, не обязательно из одного класса распределений. Пусть  $F^\leftarrow(q)$  — функция, обратная функции распределения для вероятности  $q$ , то есть  $F^\leftarrow(q) = \inf \{x \in \mathbb{R}: F(x) \geq q\}$ , тогда зависимость в верхнем хвосте  $\lambda_u$  определяется как

$$\lambda_u = \lim_{q \rightarrow 1} \mathbb{P}(X_2 > F_2^\leftarrow(q) \mid X_1 > F_1^\leftarrow(q))^{1/2}. \quad (2.11)$$

Аналогично определяется показатель зависимости в нижнем хвосте.

### 2.2.30. Метавероятность

Вероятность вероятности возникает, когда два вероятностных распределения сравнивают методами, включающими в себя стохастизацию параметров. Или когда параметр стохастизируют, чтобы рассчитать распределение цен опционов «колл» или показателей риска вроде VaR (см. соответствующую главку), CVaR и т. д. и проверить робастность или выпуклость вниз полученного распределения.

### 2.2.31. Динамическое хеджирование

Выигрыш по европейскому опциону «колл»  $C$  с установленным сроком  $T$  при стоимости исходного актива  $S$  следует хеджировать потоком динамических хеджей от настоящего времени  $t$  до  $T$  со следующим пределом:

$$\lim_{\Delta t \rightarrow 0} \sum_{i=1}^{n=T/\Delta t} \frac{\partial C}{\partial S} \bigg|_{t=t+(i-1)\Delta t} \left( S_{t+i\Delta t} - S_{t+(i-1)\Delta t} \right). \quad (2.12)$$

Мы разбиваем интервал на  $n$  частей с инкрементом  $\Delta t$ . Здесь хедж-отношение  $\frac{\partial C}{\partial S}$  вычисляется на промежутке времени  $t + (i-1)\Delta t$ , но мы получаем непредвосхищающую разность между ценой на время инициации хеджа и результирующей ценой на время  $t + i\Delta t$ .

Предполагается, что таким путем выигрыш становится детерминированным в пределе  $\Delta t \rightarrow 0$ . В мире гауссовых распределений пределом будет интеграл Ито — Маккина.

Мы покажем, в каких случаях жирного хвоста специальные доасимптотические свойства делают хеджирование невозможным.

<sup>1</sup> При этом вне хвоста  $q \rightarrow 1$  те же две величины могут вести себя как независимые.





**Часть I**

**ЖИРНЫЕ ХВОСТЫ  
И ИХ ПОСЛЕДСТВИЯ,  
ЗНАКОМСТВО**



## НЕТЕХНИЧЕСКИЙ ОБЗОР — ЛЕКЦИЯ В КОЛЛЕДЖЕ ДАРВИНА<sup>+</sup>

Abyssus abyssum invocat<sup>1</sup>.  
Псалтирь

В этой главе<sup>2</sup> представлены нетехнически, но полно все статистические эффекты проекта жирных хвостов. Все основные идеи кратко изложены в одном месте. Главным образом это список доброй дюжины последствий, которые жирные хвосты имеют для статистических выводов.

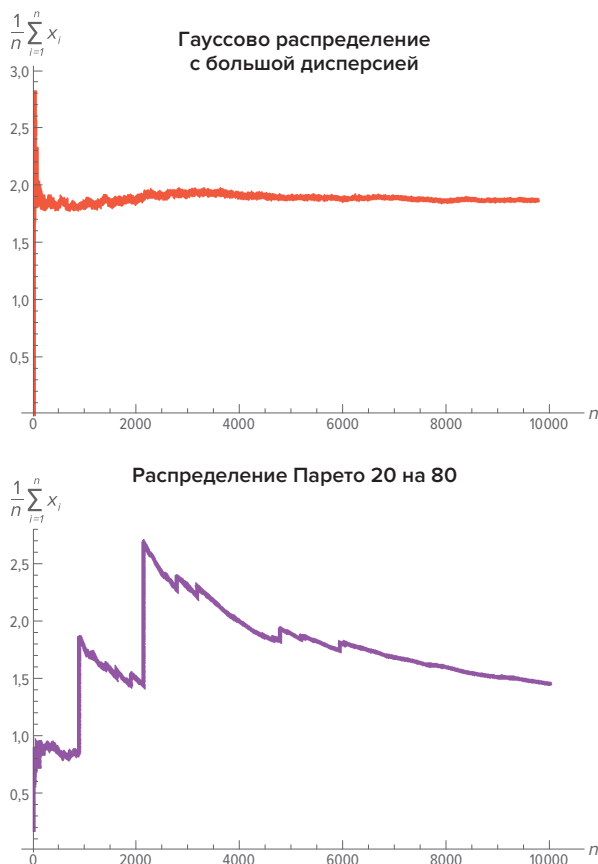
### 3.1. О РАЗЛИЧИИ МЕЖДУ ТОНКИМ И ЖИРНЫМ ХВОСТОМ

Начнем с понятия толстого хвоста и как оно связано с экстремальными значениями. Представим себе два воображаемых царства, Медиокристан (где хвосты тонкие) и Экстремистан (где хвосты толстые).

- В Медиокристани, как только выборка изучаемых событий станет большой, дальнейшие наблюдения перестанут влиять на оценку статистических свойств.
- В Экстремистани на эти свойства непропорционально сильно влияют хвосты (редкие события).

1 Бездна бездну призывает (Псалтирь 41: 7).

2 Краткая версия была представлена Колледжу Дарвина (Кембридж, Великобритания) 27 января 2017 года в цикле *Darwin College Lecture Series on Extremes*. Автор горячо благодарит Ди-Джей Нидема и Юлиуса Вайцдёрфера, а также их невидимых помощников, которые кропотливо переписали лекцию в виде связного текста. Автор также благодарен Сьюзен Пфанненшmidt и Оле Петерсу, исправившим несколько ошибок. Джамил Баз переупрямил меня, уговорив добавить в эту главу комментарии для экономистов и эконометристов, вдруг они со временем примут что-то из изложенного. — *Прим. автора.*



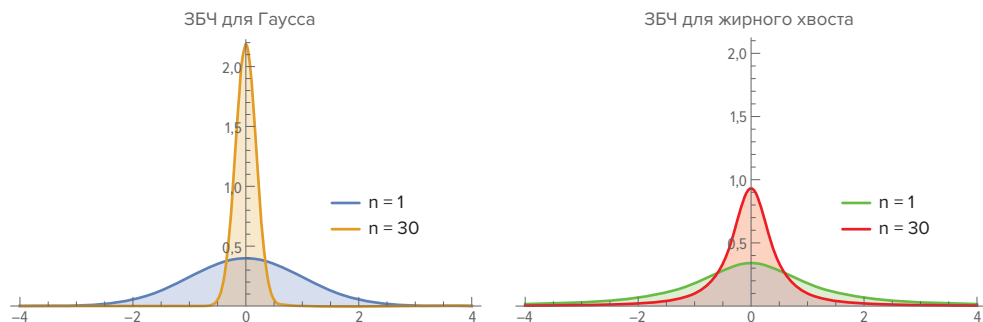
**Рисунок 3.1:** Закон больших чисел, то есть сходимость среднего по выборке к устойчивому значению, медленно работает в Экстремистане (в данном примере взято распределение со степенным законом при показателе хвоста 1,13, обеспечивающем «правило Парето 20 на 80»). В обоих распределениях одно и то же среднеквадратическое отклонение. Заметим, что феномен наблюдается и для других форматов выборки, например, тех, что используются в портфельной теории

Посмотрим с другого бока. Рассмотрим большое отклонение  $X$ .

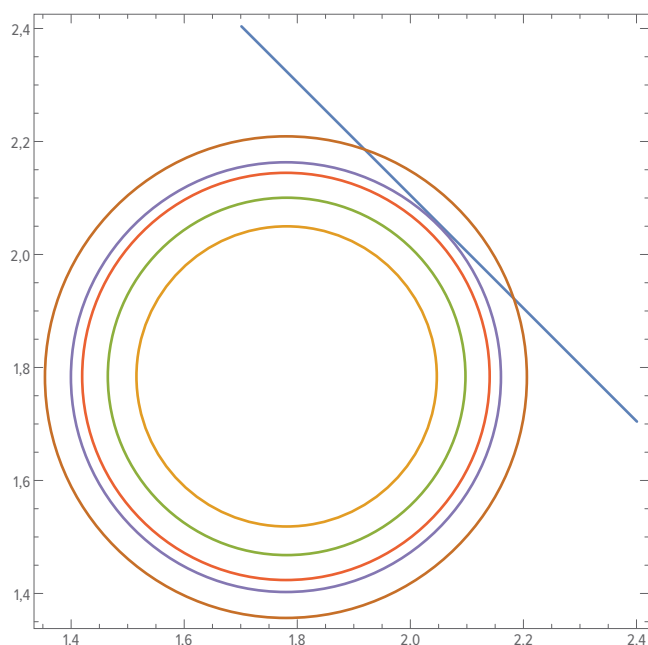
- В Медиокристане вероятнее два раза подряд встретить в выборке отклонение больше  $X$ , чем встретить один раз отклонение больше  $2X$ .
- В Экстремистане скорее попадетсся отклонение выше  $2X$ , чем два раза подряд отклонение выше  $X$ .

Пусть в Медиокристане мы выбрали двоих жителей, и оказалось — редкий, хвостовой случай, — что их суммарный рост 4,1 метра. При гауссовом распределении (точнее, в данном случае речь о его брате с одним хвостом) самое вероятное сочетание ростов окажется 2,05 метра и 2,05 метра. Никак не 10 сантиметров у одного жителя и 4 метра у другого.

Упрощая, можно сказать, что вероятность отклониться дальше чем на 3 сигмы составляет 0,00135, и вероятность отклониться дальше чем на 6 сигм, то есть вдвое сильнее, составляет  $9,86 \times 10^{-10}$ . Значит, вероятность совпадения двух трехсигмовых событий равна  $1,8 \times 10^{-6}$ , и это гораздо вероятнее, чем одно шестисигмовое событие.



**Рисунок 3.2:** Что происходит с распределением среднего по выборке при росте числа наблюдений? Представлен тот же феномен, что на Рисунке 3.1, но в пространстве вероятностных распределений. Сжать распределение с жирным хвостом труднее, чем гауссово. Выборка потребует гораздо большая. Такие дела

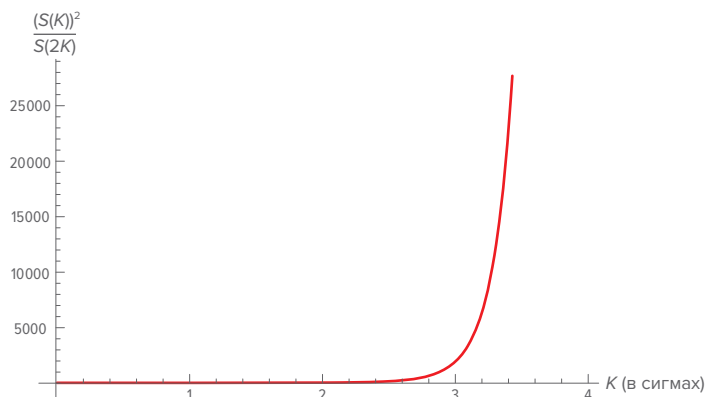


**Рисунок 3.3:** Изолинии плотности вероятности для двух независимых гауссовых распределений. Прямая показывает случаи  $x + y = 4,1$ . Видно, что самый высоковероятный из них — при  $x = y = 2,05$

Так рассуждают, если хвост распределения не толстый.

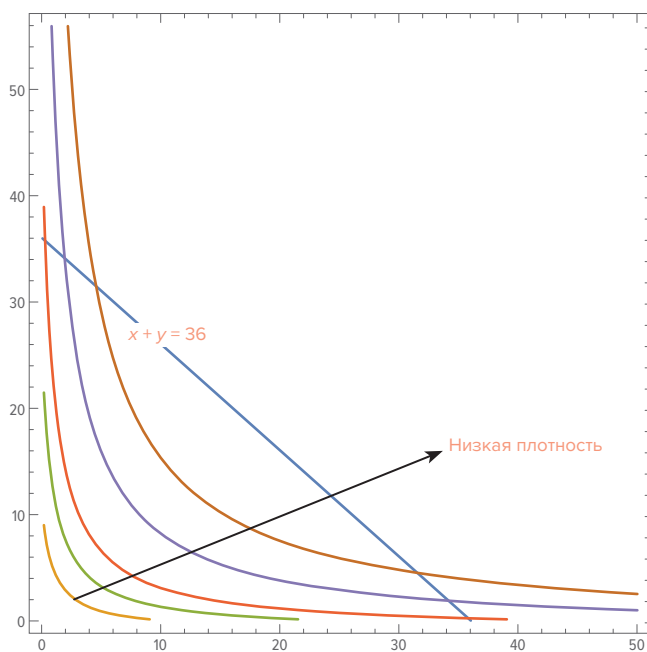
На Рисунке 3.4 показано, что если продвигаться дальше по хвосту, перейти от отношения вероятности совпадения двух 3-сигмовых событий к вероятности одного 6-сигмового к отношению вероятности совпадения двух 4-сигмовых и к вероятности одного 8-сигмового, то окажется, что большее отклонение возможно только при сочетании (сложении) умеренных отклонений; см. правую часть Рисунка 3.4. Другими словами, большая беда может грянуть только в результате серии редких событий, никак не из-за одного-единственного. Такова логика в Медиокристане.

Теперь отправимся в Экстремистан и выберем двух жителей. Пусть оказалось, что в сумме их состояние составляет 36 млн долларов. Но вряд ли это будет сочетание

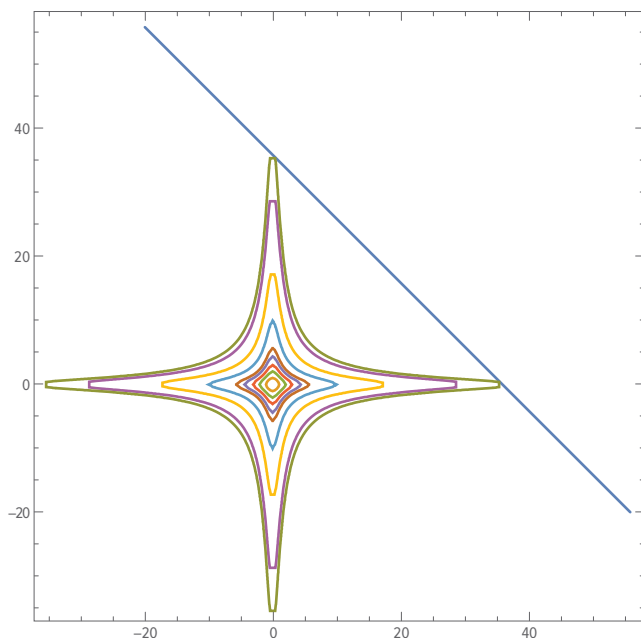


**Рисунок 3.4:** Отношение функций выживания  $S(\cdot)$  для двух аномалий размером  $K$  и одной аномалии размером  $2K$  при гауссовом распределении\*. Чем больше  $K$ , то есть чем дальше мы уходим в хвост, тем решительнее составная аномалия, то есть совпадение двух независимых аномалий размером  $K$ , с функцией выживания, преобладает над единичной аномалией величиной  $2K$

\* Здесь ради наглядности сделано незаконное упрощение. При добросовестном подходе нужно отнести две аномалии размером  $K$  к одной аномалии размером  $(2K \text{ плюс одно обычное отклонение})$ ; впрочем, график в итоге получится тот же самый.



**Рисунок 3.5:** Изолинии плотности для двух независимых распределений с толстым хвостом (в классе степенного закона). Прямая показывает случаи  $x + y = 36$ . Видно, что самые высоковероятные из них — при  $x = 36 - \varepsilon$  или при  $y = 36 - \varepsilon$  и что  $\varepsilon$  будет стремиться к нулю, если рассмотреть большие  $x + y$



**Рисунок 3.6:** Та же ситуация, что на Рисунке 3.5, только показаны оси действительных чисел, включая отрицательные. Видно, что, в отличие от Рисунка 3.3, изолинии степенного закона напоминают крест, особенно при низких плотностях, где эллиптичность совершенно утрачивается

18 млн долларов и 18 млн долларов. Скорее окажется, что это 35 999 000 долларов и 1000 долларов.

Мы подчеркнули четкое различие между двумя областями. В классе субэкспоненциальных распределений катастрофа скорее наступит из-за одного редкого события, чем из-за серии аварий. Эта логика лежит в основе классической теории рисков, намеченной страховым статистиком Филипом Лундбергом в начале XX века [155] и формализованной в 1930-е Харальдом Крамером [51], но забытая современными экономистами. Чтобы имело смысл страховать убытки, у них должно быть много ожидаемых причин, а не одна-единственная; только при большом числе ожидаемых причин возможна диверсификация.

Это показывает, что страховой бизнес работает только в Медиокристане; не выпишите страховки без верхнего предела возмещаемого ущерба, если рискуете разориться на одной-единственной катастрофе. Это правило называется принципом катастрофы<sup>1</sup>.

Как мы видели ранее, при распределениях с толстым хвостом экстремальные события, далекие от центра распределения, играют весьма важную роль. Не то чтобы черные лебеди здесь встречаются *чаще*, это недоразумение, просто их последствия существеннее. Самый жирный хвост — это когда в распределении всего одно отклонение, зато огромное, а не многочисленные умеренные аномалии. На Рисунке 4.4 показано, что, если взять распределение вроде гауссова и начать ужирнять его хвосты, число событий за пределами одного стандартного отклонения падает. При гауссовом

<sup>1</sup> Принцип катастрофы (*catastrophe principle*), он же принцип большого скачка (*single big jump*), эквивалентен следующему критерию субэкспоненциальности хвоста распределения: если такому распределению следуют независимые случайные величины  $X_1, X_2, \dots, X_n$ , то  $\mathbb{P}(X_1 + X_2 + \dots + X_n > x) \rightarrow \mathbb{P}(\max\{X_1, X_2, \dots, X_n\} > x)$  при  $x \rightarrow \infty$ . Достаточным признаком субэкспоненциальности хвоста оказывается такой:  $\mathbb{P}(X_1 + X_2 > x) \rightarrow \mathbb{P}(\max\{X_1, X_2\} > x)$  при  $x \rightarrow \infty$ , что эквивалентно требованию  $1 - \int_0^x F(x-y)dF(y) \sim 2(1-F(x))$  при  $x \rightarrow \infty$ , где  $F(x) = \mathbb{P}(X > x)$  — кумулятивная функция распределения.



распределении вероятность того, что случайное событие попадет в интервал плюс-минус одного стандартного отклонения от математического ожидания, составляет 68 процентов. По мере ужирнения хвостов, скажем, до уровней, типичных для финансовых рынков, вероятность того, что событие останется в пределах одного стандартного отклонения от математического ожидания, возрастает до 75–95 %. Чем жирнее хвосты, тем выше и уже пик и вместе с тем сильнее эффект очень больших отклонений. Поскольку сумма всех вероятностей дает 1 (даже во Франции), при добавлении жира в хвосты хвосты худеют склоны пика.

### 3.2. ХВОСТ, ВЛИЯЮЩИЙ СОБАКАМИ: ИНТУИТИВНО

#### Эффект хвоста, влияющего собакой

Центральная идея та, что чем толще хвост распределения, тем больше хвост влияет собакой, то есть важная информация сосредотачивается в хвосте, покидая «туловище» (центральную часть) распределения. В случае очень жирного хвоста все отклонения, кроме больших, делаются информационно стерильными.

Центр становится просто шумом. Хотя «доказательная наука» еще не вполне осознала этот феномен, но есть обстоятельства, когда основной корпус данных ни о чем не свидетельствует.

Это свойство также объясняет, почему закон больших чисел медленно срабатывает при наблюдении таких областей, ведь хвостовые наблюдения, где сосредоточена главная информация, по определению хвоста случаются редко.

Это свойство объясняет, например, почему наблюдение миллиона белых лебедей не доказывает несуществование черных лебедей или почему миллион подтверждающих наблюдений стоит меньше, чем одно опровергающее. Мы свяжем этот феномен с асимметрией в духе Поппера позже в этой главе.

Оно также объясняет, почему нельзя сопоставлять случайные величины, определяемые своим хвостом (например, число жертв пандемии), со случайными величинами, определяемыми своим телом (например, число утонувших в своем бассейне). О следствиях системных рисков для политики смотрите статью Чирилло и Талеба 2020 года [48].

### 3.3. ДОПОЛНИТЕЛЬНЫЕ КАТЕГОРИИ И ЧТО ИЗ НИХ СЛЕДУЕТ

Рассмотрим разные степени толстохвостости, для начала бегло; позже в этой книге нам еще предстоит углубиться, и не раз. Упорядочим их по серьезности.

Распределения бывают:

Толстохвостые  $\supset$  Субэкспоненциальные  $\supset$   
 $\supset$  По степенному закону (по Парето)

Во-первых, есть толстохвостые начального уровня. Чтобы попасть в «толстые», хвосту достаточно всего лишь быть толще, чем у гауссова распределения; это значит, что в пределы  $\pm$  среднеквадратическое отклонение должно попасть больше чем  $\operatorname{erf} \frac{1}{\sqrt{2}} \approx 68,2\%$  наблюдений, или что эксцесс (представляющий собой функцию четвертого центрального момента<sup>1,2</sup>) превышает  $3^3$ .

Во-вторых, есть субэкспоненциальные распределения, которые соответствуют приведенному выше мысленному эксперименту — тому, где иллюстрируется принцип катастрофы. У чисто субэкспоненциального распределения, не вошедшего в класс степенного закона, хвост лишь умеренно толст и не ставит распределение под чудовищное, непропорциональное воздействие со стороны редких событий. А именно у чисто субэкспоненциального распределения в наличии все статистические моменты.

Распределения третьего уровня называют по-всякому — степенной закон, правильно меняющийся класс, класс «с хвостом Парето»; речь о самых толстых, жирных хвостах, но и среди таких различают разную жирность. Определения различий внутри этого класса зависят от выбираемой системы параметров; не углубляясь пока в подробности того или иного показателя хвоста, заметим, что у всякого жирнохвостого распределения некоторый статистический момент обращается в бесконечность, и все моменты более высоких порядков также бесконечны.

Рассмотрим пирамиду на Рисунке 3.7, снизу вверх. Слева внизу вырожденное распределение, когда случайная величина может принимать только одно значение, то есть никакой случайности и никакой изменчивости нет. Этажом выше распределение Бернулли, при котором возможны ровно два исхода. Еще выше два гауссовых распределения. Есть естественное гауссово распределение (с носителем от минус бесконечности до плюс бесконечности) и есть приближения, полученные сложением случайного блуждания (с более-менее компактным<sup>4</sup> носителем, если не допустить бесконечно большого числа слагаемых). Это две разные вещи, ведь первое разрешает бесконечно большие значения,

1 Момент порядка  $p$  случайной величины  $X$  — это математическое ожидание  $p$ -той степени случайной величины  $X$ , т. е.  $\mathbb{E}(X^p)$ . — Прим. автора.

2 Однако центральный момент порядка  $p$  — это математическое ожидание  $p$ -той степени не самой случайной величины  $X$ , а разности  $X - \mathbb{E}X$ , т. е.  $\mu_p = \mathbb{E}(X - \mathbb{E}X)^p$ .

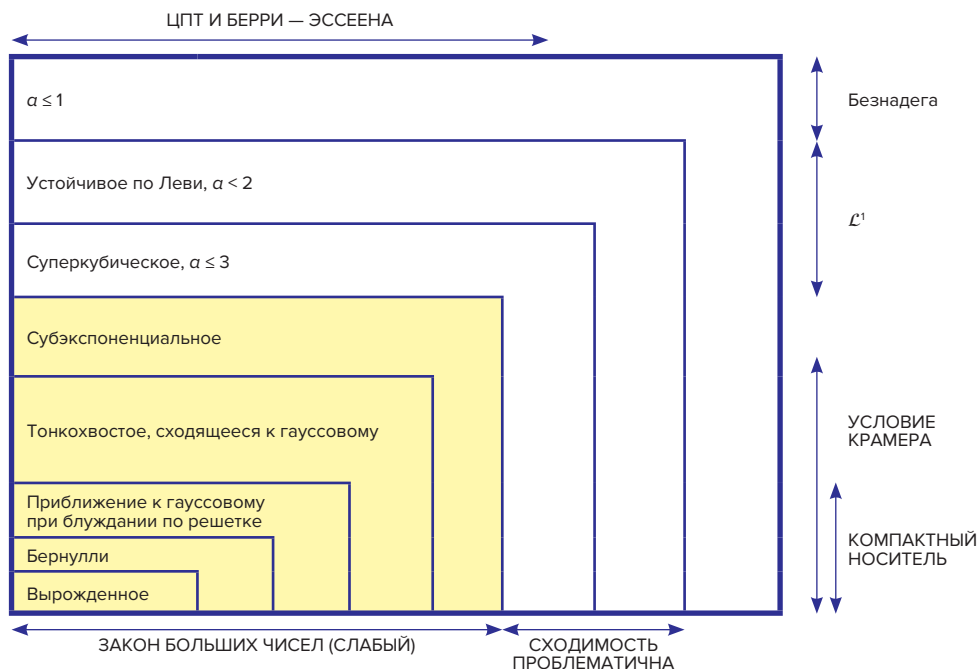
3 Знакомые с коэффициентом эксцесса, принятым в русской математической литературе:  $\gamma_2 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{\sigma^4} - 3$ , где  $\mathbb{E}(X - \mathbb{E}X)^4$  — 4-й центральный момент и  $\sigma$  — среднеквадратическое отклонение случайной величины  $X$ , — привыкли к другому: если  $X$  имеет стандартное нормальное распределение, то  $\gamma_2 = 0$ . Однако в английской литературе эксцессом (*kurtosis*) называют  $\operatorname{Kurt}(X) = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{\sigma^4} - 3$ ; если  $X$  следует нормальному распределению, то  $\operatorname{Kurt}(X) = 3$ . Коэффициент эксцесса  $\gamma_2$  в английской литературе называют «избыточным эксцессом» (*excess kurtosis*).

Смысл эксцесса поясняют так. Распределению с самым малым эксцессом, 1, следует случайная величина, принимающая только два значения, причем с равной вероятностью. Представляя эти значения как  $\mu \pm \sigma$ , получаем среднее  $\mu$  и среднеквадратическое отклонение  $\sigma$ .

Для произвольной случайной величины  $X$  со средним  $\mu$  и среднеквадратическим отклонением  $\sigma$  эксцесс оценивает рассеяние случайной величины  $X$  вокруг двух значений,  $\mu \pm \sigma$ . Распределение с одним горбом может иметь большой эксцесс в двух случаях: (1) если большая масса сосредоточена в узком пике вблизи  $\mu$  или (2) если большая масса находится в хвостах  $x \ll \mu - \sigma$  и  $x \gg \mu + \sigma$ . Строгая формула  $\operatorname{Kurt}(X) = \mu \frac{\mathbb{V}(X - \mu)^2}{\sigma^4}$ ; здесь величина  $(X - \mu)^2$  имеет среднее  $\sigma^2$ , и дисперсия этой величины оценивает ее рассеяние вокруг  $\sigma^2$  и, тем самым, рассеяние величины  $X$  вокруг двух значений  $\mu \pm \sigma$ .

4 Компактный носитель означает, что действительная случайная величина  $X$  ограничена интервалом вида  $[a, b]$ ,  $(a, b)$ ,  $[a, b)$  и т. д. Хвост гауссианы убывает по экспоненциальному закону  $\exp(-x^2)$ , то есть чем

а второе — нет (не считая асимптотического приближения к бесконечным значениям). Над гауссовыми распределениями расположен класс субэкспоненциальных, не принадлежащих классу степенного закона. В субэкспоненциальном классе у распределения существуют все моменты. К этому классу относятся логнормальные распределения, и это самые коварные звери во всей статистике, способные ввести исследователя в заблуждение. При низкой дисперсии они тонкохвостые; при высокой дисперсии обнаруживают ярко выраженное толстохвостое поведение. Некоторые спешат обрадоваться, когда данные оказываются распределены не по Парето, а логнормально, однако иногда радоваться не стоит. В причудливые свойства логнормальных распределений мы углубимся в Главе 8.



**Рисунок 3.7:** *Пирамида толстых хвостов и ряд классификаций по сходимостям (сходимость по закону больших чисел и др.) и по серьезности проблем для выводов. Распределения по степенному закону на белом фоне, остальные на желтом. См. Амбре и др. [82]*

Принадлежность субэкспоненциальному классу не обеспечивает условия Крамера, разрешающего заниматься страхованием, как мы видели в мысленном эксперименте в начале главы и проиллюстрировали на Рисунке 3.4. Говоря технически, условие теоремы Крамера означает существование математического ожидания для экспоненты случайной переменной<sup>1</sup>.

больше отклонение, тем быстрее; поэтому некоторые авторы, например Адриен Дуади, считают действующий носитель гауссианы компактным. — Прим. автора.

1 Вот строгая формулировка условия Крамера. Пусть  $X$  — случайная величина. Для всех  $r > 0$

$$\mathbb{E}e^{rX} < +\infty,$$

где  $\mathbb{E}$  — оператор математическое ожидание. — Прим. автора.

Покинув желтую зону, где закон больших чисел (ЗБЧ) более-менее работает<sup>1</sup>, мы перейдем в классы, где не будет работать центральная предельная теорема (ЦПТ)<sup>2</sup> и начнутся проблемы со сходимостью. Это зона степенных законов. Мы ранжируем их по показателю хвоста  $\alpha$ , о котором поговорим позже; пока примем, что чем ниже показатель хвоста, тем жирнее хвост. При  $\alpha \leq 3$  распределение называется субкубическим и при  $\alpha = 3$  — кубическим. Эта часть жирнохвостой зоны неформально приграничная: у распределений есть моменты первого и второго порядка, а значит, закон больших чисел и центральная предельная теорема применимы... теоретически.

Следом идет класс с  $\alpha \leq 2$ , который мы для простоты называем классом устойчивости по Леви, хотя распределение степенного закона с показателем степени меньше 2 не является устойчивым по Леви; мы, однако, обращаем внимание на то, что по мере сложения все большего числа случайных величин распределение суммы приблизится, хотя бы теоретически, именно к устойчивости по Леви, а не гауссовому; это гарантируется так называемой обобщенной центральной предельной теоремой (ОЦПТ).

Еще выше по пирамиде жирность хвоста усиливается и пропадает дисперсия. При  $1 \leq \alpha \leq 2$  дисперсии нет, но еще существует абсолютное среднее отклонение (то есть математическое ожидание абсолютной величины отклонения).

Еще выше, в верхней секции, пропадает даже математическое ожидание. Мы назвали этот класс Безнадегой<sup>3</sup>. Если вы что-то увидели в потоке данных этой категории, вернитесь домой и никому не рассказывайте.

У статистиков сложилась традиция в отношении толстых хвостов: пообещать, что будут использоваться особые распределения, а потом как ни в чем не бывало опять использовать старые показатели, критерии и оценки значимости. Но после выхода из желтой зоны, для которой раньше разрабатывались статистические методы, ничто не работает по плану. В следующем разделе представлен целый ворох проблем, и почти все безнадежны. С этого места развернем технические подробности и начнем использовать математический жаргон.

### Обзор проблемы со злоупотреблением стандартной статистикой

Статистическая оценка основана на двух элементах: на центральной предельной теореме (предполагающей работу с «большими» суммами, когда

1 Примем пока следующую грубую формулировку закона больших чисел. Если у распределения существует конечное среднее, то с ростом выборки — т. е. накоплением все новых наблюдений данной случайной величины — рано или поздно наблюдаемое среднее сходится к среднему распределения. Другой вопрос, сколько потребуются наблюдений; этот вопрос в данной книге исследуется. — *Прим. автора.*

2 Мы еще напомним центральную предельную теорему *ad nauseam*, а пока вот первый интуитивный взгляд. Она утверждает, что сумма  $n$  независимых случайных величин, имеющих конечные статистические моменты второго порядка, рано или поздно начинает напоминать гауссово распределение. При каких  $n$  произойдет такое чудо? Для случайных слагаемых из класса степенного закона теория требует бесконечно большого числа слагаемых, т. е. предсказывает, что сумма так и не станет гауссовой. В Главе 7 рассматриваются предельные распределения и ищется ответ на центральный вопрос: *сколько ждать?* — для обеих теорем, ЦПТ и ЗБЧ. Вопрос важен в реальном мире, который отличается от случая  $n = \infty$ . — *Прим. автора.*

3 В оригинале *Fuhgetaboudit*, кличка персонажа из гангстерской комедии «Клан Сопрано», искаженное *forget about it*.

почти всякое явление в мире становится уютным и нормальным) и на законе больших чисел, согласно которому дисперсия оценки падает по мере роста выборки. К сожалению, не все так просто; есть ограничения. В Главе 8 мы покажем, что нужная выборка сильно зависит от того, какому распределению следует исходный процесс, и драматически различается даже в одном классе. Как показали Бушо и Поттерс в [27] и Сорнетт в [214], хвосты с конечной дисперсией даже при бесконечных высших моментах могут сходиться, с ростом числа слагаемых  $n$ , к гауссовому в области  $\pm \sqrt{n \ln n}$ , то есть центральная часть распределения в этой полосе становится гауссовой, тогда как далекие части, то есть хвосты, этого не делают; надо понимать, что многие свойства определяются как раз хвостами.

Жизнь протекает в доасимптотических областях.

К сожалению, в статье про статистические оценки в монументальной «Энциклопедии статистических наук» [147] Василий Хёфдинг пишет:

«Обычно точное распределение статистического параметра оказывается сложным и неудобным. Отсюда потребность приблизить его более простой формулой, имеющей более прозрачные свойства. Важный инструмент для таких приближений дают предельные теоремы теории вероятностей. В частности, классические центральные предельные теоремы утверждают, что сумма большого числа независимых случайных величин распределена приблизительно по нормальному закону при весьма общих условиях. Собственно, нормальное распределение преобладает среди возможных предельных распределений. Процитируем из текста Гнеденко и Колмогорова [[111], Гл. 5]:

...если для сходимости функций распределения сумм независимых слагаемых к нормальному закону на слагаемые помимо требования их бесконечной малости (или предельного постоянства) приходится налагать лишь ограничения весьма общего характера, то для сходимости к другим предельным законам от слагаемых требуется наличие некоторых весьма специальных свойств<sup>1</sup>.

Более того, многие статистические распределения асимптотически ведут себя как суммы независимых случайных величин. Все это помогает объяснить важную роль нормального распределения как приближающего распределения».

<sup>1</sup> Цит. по: Б. В. Гнеденко и А. Н. Колмогоров. Предельные распределения для сумм независимых случайных величин. Государственное издательство технико-теоретической литературы. М.; Л., 1949.

Но что, если мы не успели достичь нормального распределения, если жизнь настигла нас до выхода на асимптоту? Вот о чем рассказывается в данной книге<sup>а</sup>.

<sup>а</sup> Приглашаем читателя поискать статью про «статистическую оценку» в любом учебнике или онлайн-энциклопедии. Вполне вероятно, что вопрос «а что, если мы не достигли асимптоты» так и не будет поднят, как этого не случилось на 9500 страницах монументальной *Энциклопедии статистики*. Еще спросите обычного пользователя статистических инструментов, какой объем данных нужен для использования того или иного распределения, и постарайтесь не показать удивления, когда услышите ответ. Проблема в том, что головы забиты готовыми комплектами статистических инструментов, и никому не приходится самостоятельно выводить формулы для своей задачи. В данной же книге девиз такой: статистика никогда не бывает стандартной.

### 3.4. ОСНОВНЫЕ СЛЕДСТВИЯ И КАК ОНИ СВЯЗАНЫ С ДАННОЙ КНИГОЙ

Вот некоторые следствия выхода за пределы желтой части пирамиды, зоны статистического комфорта:

#### Следствие 1

*Закон больших чисел в реальном мире даже если работает, то слишком медленно.*

Это невообразимый шок — большинство статистических оценок отменяется. Иллюстрацию смотрите на Рисунке 3.1 в этой главе. Тема рассматривается в Главе 8, где проводится соответствующая классификация распределений<sup>1</sup>.

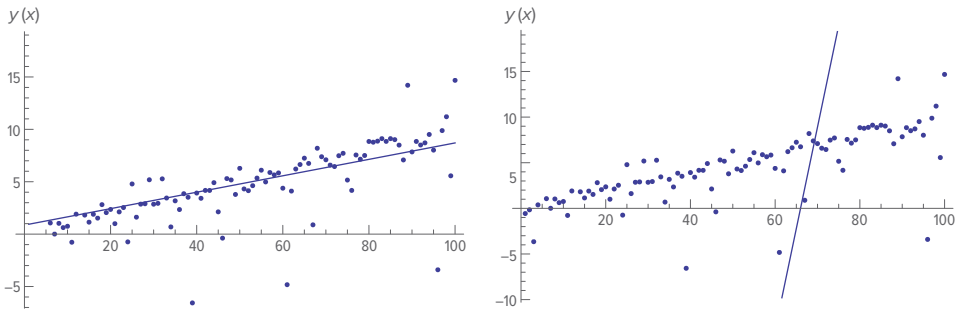
#### Следствие 2

*Среднее по выборкам редко концентрируется вокруг среднего, присущего генерирующему процессу; чаще проявляется стойкое смещение малой выборки в одну сторону, особенно сильное, когда распределение генерирующего процесса асимметрично (или однохвостое).*

Это еще одна проблема из-за недостаточности выборки. Собственно, ни при каком толстохвостом или однохвостом распределении невозможно правильно оценить среднее генеральной совокупности по среднему выборки — первое зависит от редких

<sup>1</sup> Поведение суммы или последовательности, когда  $n$  велико, но не бесконечно, мы называем предасимптотикой. Это одна из центральных тем данной книги. — Прим. автора.

событий, а такие события становятся видны только при большом объеме данных<sup>1</sup>. Возьмем некоторые степенные законы. При законе, дающем пресловутые «20 к 80», в 92 % наблюдений по выборке среднее занижается. Чтобы на основании среднего по выборкам можно было как-то судить о распределении, требуются объемы данных на порядки больше, чем практически доступный объем (исследователи в области экономики до сих пор этого не понимают, хотя трейдеры инстинктивно чувствуют). Проблема кратко обсуждается ниже в главке 3.8 и с более детальным формализмом в главах о *теневого среднем* — 15 и 16. Кроме того, в главке 3.8 мы представим концепцию скрытых свойств. Понятно, что дисперсия при малой выборке недооценивается.



**Рисунок 3.8:** При наличии толстого хвоста можно подогнать весьма различные линейные регрессии к одним и тем же данным (а теорема Гаусса — Маркова<sup>2</sup>, на которую опирается метод линейной регрессии, неприменима). Слева: обычная (наивная) регрессия. Справа: линейная регрессия, которая пытается адаптироваться к большой дисперсии — так сказать, домножить на хеджирующий коэффициент, то есть защитить агента от большого отклонения, идя на худшее приближение слабых отклонений. Иногда фатально именно пропустить большое отклонение. Заметим, что выборка не содержит критических наблюдений, об их существовании лишь делаются допущения методами теневого среднего.

### Следствие 3

Такие показатели, как среднеквадратическое отклонение и дисперсия, неприменимы.

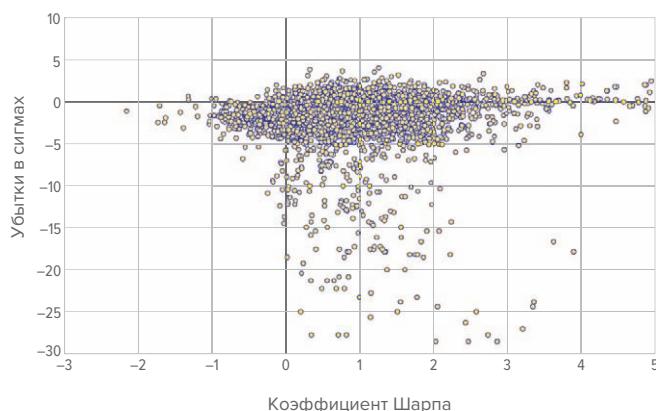
- 1 Среднее генеральной совокупности — это среднее выборки, если бы мы могли выбрать всю генеральную совокупность целиком. Среднее по выборке — это то, чем мы располагаем на практике. Даже в тех случаях собраны данные по всей существующей на текущий момент генеральной совокупности — например, о благосостоянии населения или военных потерях, — среднее по этим данным может не совпасть со средним, отвечающим процессу или механизму генерирования данных. Последнее называют «теневым средним». — *Прим. автора.*
- 2 Теорема Гаусса — Маркова утверждает, что если корреляцию между двумя случайными величинами  $X$  и  $Y$  представлять линейной регрессией  $y_i = a + bx_i + \varepsilon_i$ , то оценка коэффициентов  $a$  и  $b$  по методу наименьших квадратов оптимальна. Посылка этой теоремы: (1) корреляция представима линейной зависимостью, (2) все наблюдения  $X_i$  известны, (3) ошибки не содержат систематической ошибки,  $E\varepsilon = 0$ , (4) дисперсия ошибок конечна и постоянна, (5) ошибки не коррелированы между собой,  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$  при всех  $i, j$ .

Выборка не отражает их, даже если они существуют, даже если статистические моменты существуют. Всяческие подробности обсуждаются в Главе 4. Научный предвзвешенный догм, будто концепция среднеквадратического отклонения (которую пользователи вдобавок путают со средним отклонением) — универсально полезный показатель вариативности; на самом деле этот параметр в лучшем случае работает в предписанной ему узкой области.

#### Следствие 4

*Бета-коэффициент, коэффициент Шарпа и прочие расхожие финансовые показатели неинформативны.*

Это простое следствие предыдущего пункта<sup>1</sup>. Для этих показателей либо требуется слишком много данных, на много порядков величины больше имеющегося объема, либо исходная модель нужна не та, что используется, а другая, которую еще не изобрели. На Рисунке 3.6 было показано, как коэффициент Шарпа, разработанный, чтобы предсказывать эффективность, провалился по выборке и даже сработал обратно своему назначению. То, что такой показатель по-прежнему используется, демонстрирует, как легко люди ведутся на цифирь.



**Рисунок 3.9:** Данные о хедж-фондах: по горизонтали коэффициент Шарпа накануне кризиса 2008 года, по вертикали потери в период кризиса, в среднеквадратических отклонениях. Коэффициент Шарпа не только не предсказал эффективность хедж-фонда в выборке, он скорее показал себя как слабый предиктор неудачи. С решения Рафаэля Дуади

Практически каждая экономическая величина и стоимость ценных бумаг имеет толстый хвост. Из 40 000 изученных ценных бумаг ни одна не оказалась тонкохвостой. В этом главная причина неудач в финансах и экономике.

<sup>1</sup> Грубо говоря, бета-коэффициент показывает, насколько изменится актив  $A$  в ответ на изменение на рынке в целом (оговаривается конкретный вид рыночного индекса); вычисляется как отношение ковариации между  $A$  и рынком к дисперсии рынка.

Коэффициент Шарпа показывает среднюю или избыточную доходность (*average return, excess return*) актива или стратегии, поделенную на ее среднеквадратическое отклонение. — Прим. автора.



Финансовые теоретики делают вопиюще безосновательные заявления вроде «если у распределения толстый хвост, но существуют математическое ожидание и дисперсия, то портфельная теория на основе среднего и дисперсии работает»; иногда делают оговорку насчет эллиптичности, которую мы обсудим позже. Беда в том, что *даже когда дисперсия существует, мы не имеем сколько-нибудь точного представления о ее величине; будучи вторым статистическим моментом, дисперсия подчиняется закону больших чисел еще медленнее, чем среднее, потому что имеет еще более толстый хвост, чем исходная случайная величина*. Хуже того, стохастические корреляции или ковариации проявляют толстые хвосты своего рода (или теряют эллиптичность), тем самым обесценивая такие показатели.

Практически всякая статья по экономике, где используются ковариационные матрицы, — подозрительна.

Подробности приведены в Главе 4 для одномерной ситуации и в Главе 6 — для многомерной.

### Следствие 5

*Робастная статистика не робастна, а выборочное распределение не эмпирично.*

Старая история. Название советской газеты *Правда* стало восприниматься как издательство; робастная статистика врет не меньше, но в среде профессиональных статистиков прозрение пока не наступило.

Во-первых, робастная статистика гонится за параметрами, которые слабо реагируют на хвостовые события, наблюдения больших значений. Такое понимание робастности порочно, потому что отсутствие реакции показателя на хвостовое событие вполне может быть следствием неинформативности этого показателя. Более того, такие параметры не помогут оценить ожидаемый платеж.

Во-вторых, типичная робастная статистика строится в рамках так называемой «непараметрической» ветви статистической науки, где исследователи мнят, будто без параметров анализ будет меньше зависеть от распределения. На протяжении всей этой книги будет демонстрироваться, что зависимость от распределения только обострится.

Винсоризация данных, устраняя выбросы, уродует процесс сходимости к математическому ожиданию и, по сути, сокращает доступные данные, хотя бывает полезно перепроверить достоверность выброса — вдруг это случайная ошибка, такая как опечатка или компьютерный глюк; в финансовом мире мы называем такие выбросы *bad print* («непринт»).

В непараметрической статистике популярно *выборочное распределение*<sup>1</sup>, которое не работает эмпирически, поскольку, как мы покажем в Главе 10, оно неправильно

<sup>1</sup> В английской литературе — *empirical distribution*.

представляет ожидаемые платежи в хвостах, во всяком случае, та версия, которая используется в управлении финансами и рисками. Пока поясним только, что будущие максимумы плохо отслеживаются по прошлым данным, если не применить разумную экстраполяцию.

Представим себе, что планируется построить систему дамб для защиты от наводнения. Данные по уровню воды покажут наихудшее в истории наводнение, и этот уровень будет историческим максимумом. Наивно построенное выборочное распределение предскажет, что вероятность более страшного наводнения ноль (или около того). Но исторический максимум, по определению, являет пример, дискредитирующий оценку по историческому максимуму: в том году, когда случился исторический максимум, он превзошел предыдущий исторический максимум. И если бы к тому году мы столь же наивно построили эмпирическое распределение по имевшимся данным, мы не предвидели бы явление нового исторического максимума. При толстом хвосте различие между историческим максимумом и ожидаемым максимумом много драматичнее, чем при тонком хвосте.

#### **Следствие 6**

*Линейная регрессия по методу наименьших квадратов не работает (провал теоремы Гаусса — Маркова).*

Посмотрите на Рисунок 3.8 и подпись к нему. Логика за методом наименьших квадратов та, что по теореме Гаусса — Маркова наилучшая прямая по точкам данных оказывается единственной, когда у распределения тонкий хвост. Вывод: когда теорема Гаусса — Маркова применима и минимизировать квадраты отклонений можно, данных потребуется гораздо больше, чем бывает в реальном, доасимптотическом мире конечных данных; если же второй статистический момент не существует, то минимизация квадратов отклонений невозможна ни при каком объеме данных. В последнем случае можно попробовать минимизировать средние абсолютные отклонения (*mean absolute deviations*, MAD), но тогда, как мы увидим в 4.1, можно не только столкнуться с недостаточностью данных для хорошей сходимости, но и угловой коэффициент линейной регрессии может оказаться не единственным.

Эту проблему мы обсудим подробнее в Главе 6.7 и покажем, как из-за эффекта малой выборки при толстом хвосте завышается коэффициент детерминации ( $R^2$ ). В случае бесконечной дисперсии  $R$ -квадрат должен быть нулем. Но по конечной выборке получаются ложные значения выше 0. Вывод: при толстом хвосте  $R$ -квадрат бесполезен, неинформативен, и им злоупотребляют шарлатаны (например, в исследованиях IQ).

#### **Следствие 7**

*Методы максимального правдоподобия могут работать в отношении некоторых параметров распределения (это хорошая новость).*

Возьмем степенной закон. Форму хвоста по степенному закону можно оценить параметром — показателем хвоста (который мы в этой книге обозначаем греческой  $\alpha^1$ ), и при добавлении еще одного параметра (масштаба) мы получаем лучшую связь со средним, чем непосредственное взятие среднего по выборке.

**Пример:** для простого распределения Парето с минимальным значением  $L$ , показателем хвоста  $\alpha$  и плотностью вероятности  $\alpha L^\alpha x^{-\alpha-1}$  среднее зависит от  $\alpha$  как  $L \frac{\alpha}{\alpha-1}$ . Есть смысл оценивать среднее именно по этой формуле, особенно если один из двух параметров уже известен (этот метод мы и назвали выше дополнительной статистической оценкой). Можно оценить  $\alpha$  с небольшой погрешностью, пользуясь визуальным инструментом, а можно использовать методы максимального правдоподобия и получить оценку с низкой дисперсией — у нее обратное гамма-распределение; найдя  $\alpha$ , можно получить среднее. И выйдет лучше, чем просто среднее по выборке.

Логика этого метода стоит того, чтобы ее выделить:

Показатель хвоста экстраполирует низковероятные отклонения, которые не встретились в собранных данных, но которые вносят непропорционально большой вклад в среднее.

Этот общий подход к статистическим оценкам применим, в частности, к индексу Джини и другим оценкам социального неравенства.

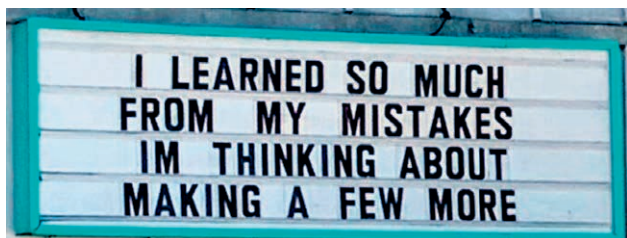


**Рисунок 3.10:** Показатели неравенства, такие как коэффициент Джини, требуют других методов оценки при толстом хвосте, как мы увидим в Части III. Наука дается трудно

Итак, возможно получить надежные (или хотя бы менее шаткие) оценки для, так сказать, функции показателя хвоста в некоторых ситуациях. Но никоим образом не во всех.

<sup>1</sup> Поясним термины. Разные авторы разные параметры хвоста обозначают буквой  $\alpha$ . В этой книге это предел отношения логарифма функции выживания после  $K$  к  $\ln K$ ; такой параметр достигает 1 в случае распределения Коши. Некоторые исследователи используют  $\alpha - 1$ , где  $\alpha$  — показатель степени в функции плотности. — Прим. автора. [У такого распределения бесконечны моменты, порядок которых выше  $\alpha - 1$ .]

Здесь неизбежен вопрос из реального мира: а что делать, когда у нас нет надежной статистической оценки? Сидеть дома. Не следует подставяться при наличии хрупкости; можно тем не менее принимать рискованные решения, если нас несет к максимальным потерям (Рисунок 3.11).



**Рисунок 3.11:** «Я УЧИЛСЯ НА СВОИХ ОШИБКАХ, Я СТАЛ МУДРЕЕ. ХОЧУ СДЕЛАТЬ ЕЩЕ ПАРОЧКУ». Дело в том, что учиться на собственных ошибках хорошо, только пока хвосты тонкие. При жирных хвостах ошибки фатальны. Источник: сетевой фольклор с хештегом #youhadjob (У тебя была одна задача)

### Следствие 8

*В отличие от обычной статистики, в случае жирных хвостов возникает пропасть между эмпиризмом опровержений и конфирматорным эмпиризмом; отсутствие доказательств здесь не служит доказательством отсутствия. (Так называемая доказательная наука, если не ограничивается строго одними опровержениями, обычно оказывается интерполяторской, бездоказательной и ненаучной.)*

Из полемики с когнитивным лингвистом и популяризатором Стивеном Пинкером автор вынес следующее: нельзя делать выводы и строить теории на основе последних отклонений в данных, если не следовать стандартам значимости, а по ним в ситуации толстого хвоста требуется больше данных (эта логика аналогична рассуждениям касательно медленного ЗБЧ).

Утверждение Пинкера, что «уровень насилия упал», поскольку уменьшилось число убитых в войнах по сравнению с предыдущим годом или десятилетием, — ненаучно; научное суждение основывается не на анекдотическом свидетельстве, а на выборке, и учитывает такую вещь, как статистическая значимость.

Повторим лишний раз: утверждения, лишённые статистической значимости, не принадлежат миру науки. Практика вычитывать описательную статистику бывает оправданной в случае тонкого хвоста (где не требуется огромного размера выборки); в случае толстого хвоста так делать не надо — за исключением, повторимся, ситуации, когда наблюдалось большое отклонение; так, утверждение о росте числа острых конфликтов может вполне научным образом базироваться на одном-единственном наблюдении.

### Следствие 9

*Метод главных компонент и факторный анализ, скорее всего, нагенерируют кучу паразитных факторов и насчитают иллюзорные нагрузки.*

Это техническая область; понятие недостаточной выборки применяется к так называемому методу главных компонент (*principal component analysis*, PCA), который призван сократить размерность больших случайных векторов. Проблемы с законом больших чисел имеют свою версию для высоких размерностей. Лучше всего эту историю разъясняет Рисунок 3.26, демонстрирующий предвзятое восприятие так называемого «эффекта Вигнера» из-за недостаточности данных для PCA. Кроме того, говоря о технике дела, заметим, что распределение случайных матриц Марченко — Пастура неприменимо, когда не существует конечный четвертый момент (или, как показано в [23], когда показатель хвоста превышает 4)<sup>1</sup>.

### Следствие 10

*Метод моментов (method of moments, MoM) не работает. Высшие моменты неинформативны или не существуют.*

То же относится к ОММ, обобщенному методу моментов, увенчанному премией Банка Швеции по прозвищу Нобелевка. Это долгая история, а пока ограничимся тем, что оценка данного распределения путем подбора моментов проваливается, если моменты не конечны, и от выборки к выборке оценка момента скачет — как мы скоро увидим на примере 4-го момента S&P 500<sup>2</sup>.

Попросту говоря, высшие моменты в толстохвостых распределениях вообще склонны взрываться, а в экономике и подавно.

### Следствие 11

*Такой вещи, как типичное большое отклонение, не существует.*

При условии «большого» изменения величина изменений расходится, особенно при серьезной толщине хвоста (класс степенного закона). Ситуация похожа на принцип катастроф, который мы описали выше. В Гауссовом мире математическое ожидание изменения при условии, что величина изменения больше 4 среднеквадратических отклонений, примерно 4 среднеквадратических отклонения. При степенном законе получается в разы больше. Мы называем это свойством Линди; оно обсуждается отчасти в Главе 5, а главным образом в Главе 11.

1 Еще дальше углубляясь в технику дела, укажем, что главные компоненты независимы при корреляциях 0. Однако для распределений с толстым хвостом, как мы увидим в технической главе 6.3.1, отсутствие корреляции еще не означает независимости. — Прим. автора.

2 Фондовый индекс компании *Standard & Poor's*. В корзину включено 505 акций пятисот публичных компаний, торгуемых на фондовых биржах США и имеющих наибольшую капитализацию.

**Следствие 12**

*Коэффициент Джини теряет аддитивность.*

Данные выборки для коэффициента Джини измеряют интерполяционно — по существу, с ними та же проблема, что мы видели выше с недооценкой среднего, если использовать выборочное среднее. Теперь дополнительное осложнение возникает из-за сверхаддитивности коэффициента Джини при толстом хвосте. По мере роста объема выборки условное измерение коэффициента Джини создает иллюзию суперконцентрации богатств. (Например, для Европы в целом можно насчитать неравенство доходов жителей, которое превысит средневзвешенное среднее по странам, составляющим Европу.)

Дело не просто в коэффициенте Джини, та же ситуация и с другими показателями концентрации, такими как процент национального богатства в собственности 1 % самых богатых и т. д. Формулы выводятся в Главах 13 и 14.

**Следствие 13**

*Теория больших отклонений к толстым хвостам неприменима.*

Я не оговорился, именно так: неприменима<sup>1</sup>. Методы, связанные с принципом большого отклонения ([260], [59] и др.), очень полезны в мире тонких хвостов. И только там. Обсуждение и вывод формул смотрите в Приложении С, а также в главах, затрагивающих центральную предельную теорему, особенно в Главе 7.

**Следствие 14**

*Динамическое хеджирование не снижает риски опционов.*

Возможно, эта тема технически слишком сложна и при этом недостаточно интересна для тех, кто не занимается финансами, но сама основа хеджирования по модели<sup>2</sup> ценообразования опционов Блэка — Шоулза покоится на возможности и необходимости динамического хеджирования, и оба предположения несостоятельны, как показано в Главах 20, 21 и 22. Для этой модели требуется экспоненциальное падение частоты отклонений вдали от центра, то есть исключается распределение вероятности

1 Не путаем теорию больших отклонений (*large deviation theory*, LDT) с теорией экстремальных значений (*extreme value theory*, EVT); последняя охватывает все основные классы распределений. — Прим. автора.

2 Связывает цены опционов с ожидаемой рынком волатильностью базового актива. Статья с такой формулой была опубликована Блэком и Шоулзом в 1973 году.

в субэкспоненциальном классе. Здесь мы опять говорим о вещах, связанных с условием Крамера, то есть в сухом остатке — с экспоненциальным моментом.

Вспомните, что автор занимался торговлей опционами, а среди трейдеров цена на опцион устанавливается отнюдь не путем динамического хеджирования, с которым, как показали Хог и автор, в прошлые века биржа не имела ничего общего.

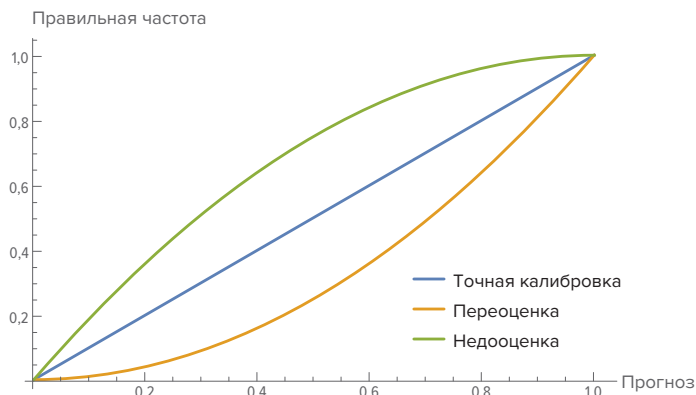
### Следствие 15

*Прогнозирование в пространстве частот расходится с ожидаемым платежом.*

А также:

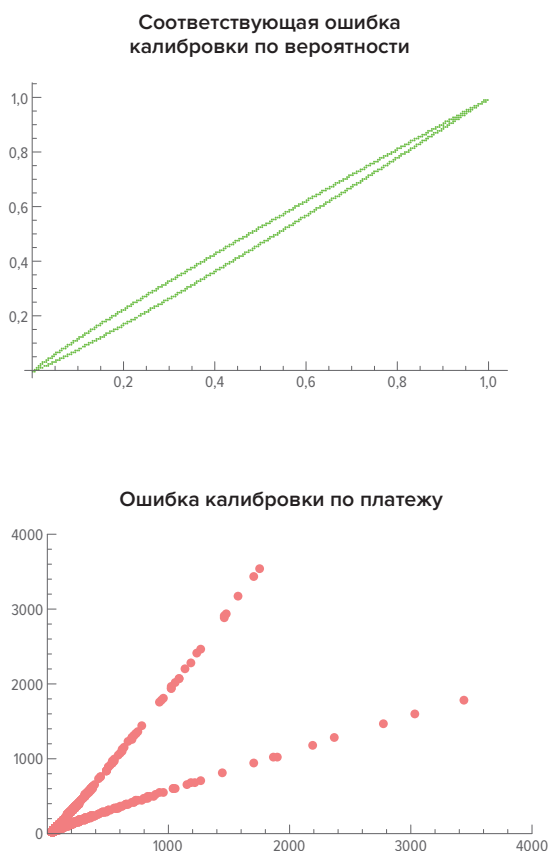
### Следствие 16

*Львиная доля утверждений в литературе по психологии и по принятию решений, где говорится о якобы переоценке вероятностей в хвосте и якобы иррациональном поведении из-за редких событий, делается из-за непонимания исследователями природы хвостового риска, смешения вероятности и ожидаемого платежа, злоупотребления распределениями вероятностей и незнания с теорией больших отклонений.*



**Рисунок 3.12:** Калибровка вероятностей, встречаемая в литературе по психологии. На оси  $x$  показана оценка вероятности, данная прогнозистом, а на  $y$  — что происходит в действительности; например, если после того, как некий метеоролог объявил вероятность дождя 30 %, дождь происходит именно в 30 % случаев, оценка этого метеоролога объявляется «откалиброванной». Мы утверждаем, что калибровка в пространстве частот (вероятностей) — академическая тема (в плохом смысле), создающая неверное представление о реальной жизни за пределами узкой области пари о двух возможных исходах. Данный подход особенно обманчив при толстых хвостах. Подробно вопрос разбирается в Главе 11

Этот вопрос рассматривается в следующем разделе и в Главе 11: дурацкая идея сфокусироваться на частотах вместо математического ожидания может дать небольшой эффект, но только при тонком хвосте, никак не при толстом. Эффект показан на Рисунках 3.12 и 3.13.



**Рисунок 3.13:** Как ошибка калибровки по вероятности отвечает ошибке калибровки по платежу при степенном законе. Рассматривается распределение Парето с показателем хвоста  $\alpha = 1,15$ . Подробнее этот вопрос будет рассмотрен опять-таки в Главе 11

### Следствие 17

*Проблемы с разорением обостряются, и эргодичность становится неизбежной при толстых хвостах.*

Здесь без технических деталей не обойтись, но мы разберемся с этой темой к концу данной главы.

Обсудим несколько вопросов.



### 3.4.1. Прогнозирование

В «Одураченных случайностью»<sup>1</sup> одного персонажа спрашивают про некоторый рынок, что *вероятнее* к концу месяца — подъем или падение. «Подъем, — отвечает персонаж, — намного вероятнее». После чего оказывается, что он заключил сделки, выгодные в случае *падения*. Разумеется, эта ситуация парадоксальна только на взгляд тех, кто незнаком с теорией вероятностей; для трейдеров это обычное дело, особенно когда распределение отличается от нормального: да, у подъема рынка вероятность была выше, но при этом у падения, если оно случится, ожидалась много бо́льшая глубина.



**Рисунок 3.14:** В жизни важен ожидаемый выигрыш, а не бинарный прогноз, и различие особенно велико в Экстремистане. (Почему «габииш», а не «капиш»? Воспроизведен сицилийско-калабрезский акцент: «п» звучит как «б», а «к» звучит как семитский коф, твердое карфагенское «К». Примерно так же «капиколи» превращается в «габагул».) © Stefan Gasic

Это иллюстрирует распространенное смешение бинарного прогноза, который сообщает вероятность одного из двух исходов, с прогнозом эффекта, который учитывает больше нюансов и зависит от распределения вероятностей. Это самая элементарная, хотя и распространенная ошибка: называть в прогнозе одну-единственную вероятность, когда в действительности возможен целый диапазон различных исходов и ему отвечает распределение вероятностей. По мере углубления в тему мы встречаемся с менее очевидными парадоксами, которые не так известны. Но суть в том, что, по мнению автора, ненаучно говорить о вероятности как некой итоговой оценке, по которой принимают решение.

В реальном мире платеж представляет собой не проценты вероятности, а доллары, выживание и т. д. Чем жирнее хвост, тем важнее пространство платежей; как говорят в Экстремистане, «платеж поглотил вероятность» (см. текст в рамке в начале следующей главки). Можно позволить себе то и дело ошибаться и проигрывать, если потери невелики и выпуклы вниз по платежу (благодаря большому доходу при выигрышах). И наоборот, можно делать верные ставки в 99,99 % случаев и обанкротиться (собственно, такого рода правота даже делает банкротство

1 См.: Талеб Н. Н. Одураченные случайностью: О скрытой роли шанса в бизнесе и в жизни. М.: КоЛибри, Азбука-Аттикус, 2022. Первая книга философских эссе в *Incerto*, четырехтомном цикле Нассима Талеба. Этот цикл продолжается данной книгой — первой в техническом цикле *Incerto*.

2 Итал. *capisce*, «понимаешь?».

более вероятным: именно фонды с безупречным послужным списком разорились во время краха 2008–2009)<sup>1</sup>. Прозвучит немного сложно для незнакомых с финансовой математикой, но дело в том, что есть разница между «ванильным»<sup>2</sup> опционом и бинарным<sup>3</sup> опционом с той же страйк-ценой, как описано в «Динамическом хеджировании» [225]: против интуиции, толстый хвост снижает стоимость бинарного опциона и повышает стоимость ванильного. Отсюда авторская поговорка: «В жизни не видел ни одного богатого прогнозиста». Мы глубже изучим вопрос в главке 4.3.1, где покажем, что ужирнение хвоста снижает вероятность событий с отклонением больше 1 среднеквадратического, однако усиливает их последствия (в терминах вклада в статистический момент — повышает влияние на среднее и прочие показатели).

На Рисунке 3.12 показан размах этой проблемы.

### Замечание 1

*Неточный прогноз вероятности («ошибка калибровки») лежит в другом классе распределений вероятности, нежели колебания прибылей и убытков в реальном мире (или чистых платежей).*

*«Калибровка» (мера точности) при прогнозе вероятности принадлежит пространству вероятностей, то есть диапазону от 0 до 1. Всякая стандартная мера этого рода обязательно имеет тонкий хвост (скорее даже сверхтонкий хвост, поскольку эта величина ограниченная) — даже когда случайная величина, в отношении которой делается прогноз, имеет толстый хвост.*

*Платежи в реальном мире, напротив, могут иметь толстый хвост, и у их «калибровки» свойства распределения должны следовать свойствам распределения случайной величины.*

Вывод всех формул и доказательства утверждений мы покажем в Главе 11.

### 3.4.2. Закон больших чисел

Обсудим теперь закон больших чисел, основу основ статистики. Он утверждает, что по мере накопления наблюдений среднее значение становится все более устойчивым,

1 Р. Дуади, по данным об оценке рисков для фондов, рухнувших в кризис 2008. — Прим. автора.

2 «Ванильный» (не экзотический) опцион охватывает два наиболее распространенных вида: американский опцион, который владелец имеет право погасить (то есть продать по страйк-цене оговоренное количество базового актива продавцу опциона, если это опцион «пут», *put*, или купить, если это опцион «колл», *call*) в любой день до истечения срока опциона; европейский опцион, который владелец имеет право погасить только в последний день. Страйк-цена базового актива прописана в опционе, и рыночная цена того же базового актива в день погашения опциона может отличаться; владелец воспользуется правом погасить опцион, если разность будет в его пользу (опцион окажется «в деньгах», *in the money*), и не воспользуется, если разность будет не в его пользу (опцион окажется «вне денег», *out of the money*) или нулевой (опцион окажется «на деньгах», *at the money*).

3 Бинарный опцион считается экзотическим. В зависимости от выполнения оговоренного условия в оговоренное время он либо обеспечивает фиксированную премию, либо не приносит ничего.

и скорость этой стабилизации около  $\sqrt{2}$ . На Рисунке 3.1 показано, что в случае распределения с жирным хвостом (нижний график) для выхода на устойчивое среднее требуется гораздо больше наблюдений.

В таком случае теоретическая *эквивалентность* наблюдаемого среднего и среднего, которое на самом деле присуще порождающему процессу, может вас подвести.

**В Экстремистане платеж поглотил вероятность:** чтобы увидеть главное различие между Медиокристаном и Экстремистаном, рассмотрим такое событие, как авиакатастрофа. Последствия тяжелые, много погибших; допустим, от 100 до 400. Даже одно такое событие — трагедия. В рамках прогнозирования и управления рисками мы стараемся минимизировать такую вероятность, сделать ее пренебрежимо малой.

А теперь представим себе катастрофу, которая убивает всех, кто когда-либо летал самолетом, даже в далеком прошлом, всех. Считать ли ее событием того же типа? Такого рода события известны в Экстремистане, и, работая с ними, фокусируются не на снижении вероятности события, а на снижении его величины.

- Для первого типа управление рисками состоит в том, чтобы снизить вероятность, то есть частоту, происшествий. Мы подсчитываем число событий и стараемся его уменьшить.
- Для второго типа мы стараемся уменьшить масштаб катастрофы, если она все-таки разразится. Мы заняты не числом событий, а ущербом от одного события.

Если вам наш мысленный пример показался странным, примите во внимание, что центральные банки потеряли в 1982 году больше денег, чем заработали за всю свою историю, и что в 1991 году та же участь постигла в США ссудо-сберегательную отрасль (ныне не существующую), а в 2008–2009 годах вся американская банковская система потеряла все нажитое до последнего пенни. На финансовом рынке сплошь и рядом видишь, как люди лишаются всех накоплений в рамках одного-единственного события. То же относится ко многим другим отраслям (например, производителям автомобилей и самолетов).

Ладно банки, они теряют только деньги; примем во внимание войны — уж для них-то мы точно не можем сфокусироваться на частоте в ущерб масштабу, как сделал научно-популярный писатель Стивен Пинкер в [194]; см. обсуждение в Главе 16. А ведь мы еще даже не коснулись проблемы разорения (и неэргодичности), которую приберегаем на конец данного раздела.

Говоря техничнее, только при выполнении условия Крамера, исключающего субэкспоненциальность последовательных событий (даже если их

величины просто складываются), можно придать какой-то смысл работе просто с вероятностью.

Аналогия с авиакатастрофами была предложена изобретательным Рассом Робертсом во время передачи EconTalk с участием автора данной книги.

Один из самых известных феноменов статистики — это открытое Парето распределение 20 на 80; например, 20 % итальянцев владеет 80 % земли. В Таблице 3.1 показано, что при гауссовом распределении устойчивость достигается за 30 наблюдений, а при распределении Парето требуется 10 наблюдений, чтобы достичь той же выборочной ошибки для такого же среднего (если оно существует).

Хотя построить и обсчитать это все нетрудно, мало кто заморачивается. Что достойно сожаления: на таком примере убеждаешься, что про распределение с толстым хвостом нельзя утверждать, что накопленная выборка достигла устойчивого среднего. Способ узнать среднее этой случайной величины существует, но это ни в коем случае не простое отслеживание выборочного среднего.

**Таблица 3.1:** Разные альфы и соответствующие  $n_\alpha$  или Сколько требуется наблюдений, чтобы разброс выборочного среднего стал как при  $\alpha$ -устойчивом распределении (параметр подробно обсуждается в Главе 8). В гауссовом случае  $\alpha = 2$ . Если рассматривать распределения с хвостом 20 на 80, то требуемое число наблюдений при распределении Парето больше, чем при гауссовом, в  $10^{11}$  раз

$\alpha$	$n_\alpha$ Симметричное	$(n_\alpha)^{\beta=\pm\frac{1}{2}}$ Скошенное	$(n_\alpha)^{\beta=\pm 1}$ Однохвостое
1	Безнадега	—	—
$\frac{9}{8}$	$6,09 \times 10^{12}$	$2,8 \times 10^{13}$	$1,86 \times 10^{14}$
$\frac{5}{4}$	574 634	895 952	$1,88 \times 10^6$
$\frac{11}{8}$	5027	6002	8632
$\frac{3}{2}$	567	613	737
$\frac{13}{8}$	165	171	186

Продолжение таблицы 3.1

$\frac{7}{4}$	75	77	79
$\frac{15}{8}$	44	44	44
2	30	30	30

3.5. ЭПИСТЕМОЛОГИЯ И ДЕДУКТИВНАЯ АСИММЕТРИЯ

Определение 3.1 (Асимметрия в распределениях)

Скорее преступник постарается выдать себя за честного человека, чем честный человек — выдать себя за преступника. Аналогичным образом легче принять жирнохвостое распределение за тонкохвостое, чем тонкий хвост за жирный.

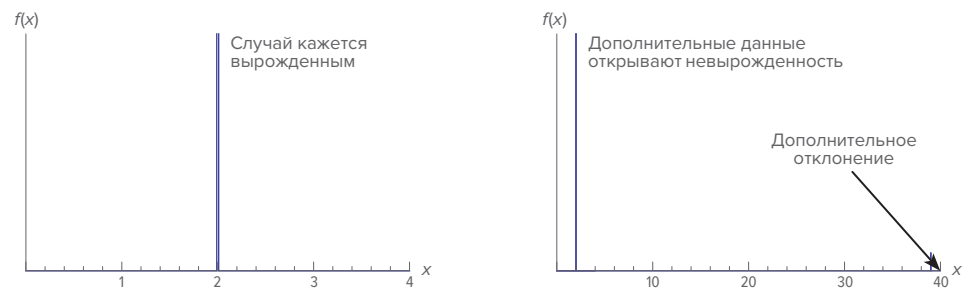


Рисунок 3.15: Проблема маскарада (фундаментальная асимметрия статистического вывода). Слева вырожденная случайная величина все время принимала одно и то же значение, и гистограмма наблюдений похожа на шест Дирака<sup>1</sup>. Однако исключить невырожденность невозможно. На графике справа случайная величина принимает более одного значения. Здесь можно исключить вырожденность. Такую фундаментальную асимметрию можно обобщить и сформулировать соответствующие принципы. Например, можно придать строгий смысл словам «не удалось отвергнуть». Асимметрию можно учесть в строгой системе правил вывода

**Принцип 3.1 (Эпистемология: невидимость генератора)**

- Наблюдаемо не вероятностное распределение, а только его реализация.
- Зная свойства некоторого вероятностного распределения, невозможно утверждать, что ему принадлежит известная реализация.
- Для обсуждения хвостовых событий требуется метавероятностное распределение (т. е. распределение условных вероятностей того, что случайная величина следует тому или иному вероятностному распределению).

<sup>1</sup> Имеется в виду  $\delta$ -функция Дирака — спектральная функция для спектра, состоящего из одной-единственной спектральной линии. При любом экспериментальном наблюдении  $\delta$ -функция Дирака в какой-то степени размывается; площадь наблюдаемого пика при любой степени размывания остается равной единице, но толщина кажется больше нуля, а высота кажется конечной.

Изучим эпистемологические следствия. Рисунок 3.15 иллюстрирует Проблему маскарада (она же Фундаментальная асимметрия статистического вывода). Слева вырожденная случайная переменная принимает постоянные на вид значения, порождая гистограмму вида Шест Дирака.

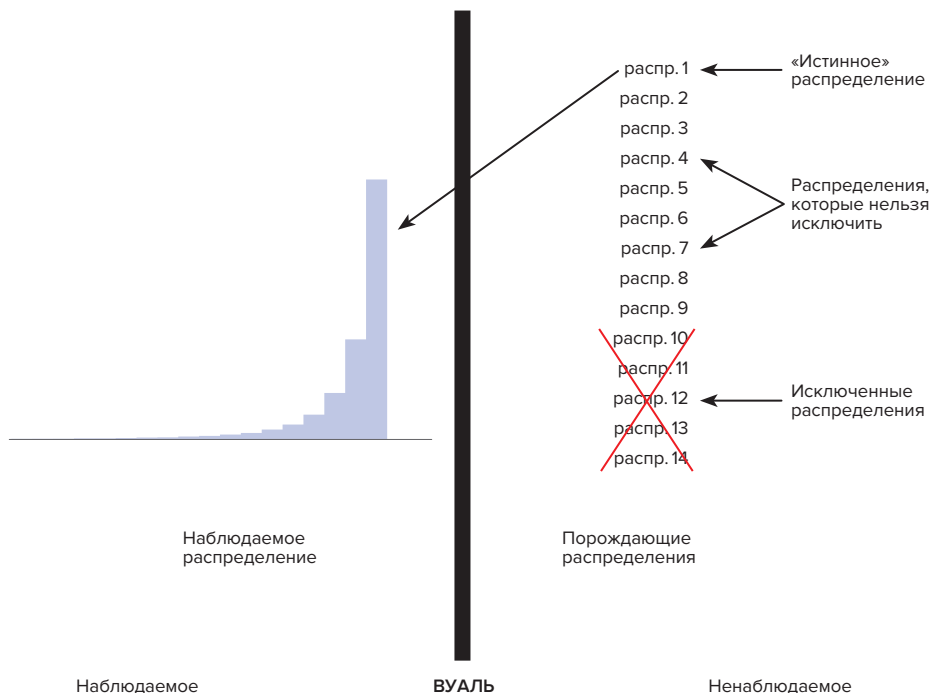
Со времен самого позднего Секста Эмпирика<sup>1</sup> мы знаем, что не можем вывести заключение о вырожденности, однако в некоторых ситуациях можем вывести заключение о невырожденности. Если я вижу распределение, в котором нет случайности, я не имею права заявить, что величина неслучайная. То есть не могу заявить, что черных лебедей не существует. Если же появится хотя бы одно наблюдение в стороне, я увижу, что величина случайная, и могу исключить вырожденность. Могу утверждать, что величина «не неслучайная».

На рисунке справа мы видим черного лебедя и заключаем, что утверждение, которое отрицало существование черных лебедей, оказалось ложным. Так работает негативный эмпиризм, легший в основу западной науки. По мере сбора информации удастся исключить те или иные вещи. Невырожденное распределение могло до поры скрываться под видом распределения слева, но вырожденное распределение никогда не сможет выдать себя за распределение справа. Тем самым мы получаем некоторые возможности для работы со случайностью. Рисунок 3.16 обобщает процедуру исключения тех или иных распределений.

Увидев событие в 20 сигм, мы можем исключить тонкохвостость распределения. Не увидев ни одного большого отклонения, мы не можем исключить толстый хвост, если не знаем порождающего принципа. Исходя из этого, ранжируем возможные распределения. Возвращаясь к Рисунку 3.7, начнем снизу и по мере наблюдения отклонений будем исключать распределения одно за другим. Ранг отвечает способности распределения проявлять хвостовые события. Принцип оказался прост: порядок отвечает приоритету при вынесении суждений. Логика здесь та, что, когда кажется, что произошло десятисигмовое событие, скорее всего, это не событие десятисигмовое, это распределение выбрано не то и сигма посчитана не та (мы уточним этот довод позже в этой главе). Дальше в том же духе: как мы уже видели, толстохвостые распределения выдают мало отклонений от среднего, зато порой случается большое отклонение. Встретив такое, можно исключить  $\mathcal{E}$ -распределение и понять, что мы не в Медиокристане. Можно объявить данное распределение толстохвостым, в силу устранения других возможностей. Другое дело, что никакое наблюдаемое распределение нельзя объявить тонкохвостым. Такова принципиальная проблема черного лебедя.

**Применение Проблемы маскарада: аргентинский фондовый рынок до и после 12 августа 2019.** В качестве иллюстрации дедуктивной асимметрии применительно к параметрам распределения и того, как распределение может маскироваться под относительно тонкохвостое, рассмотрим наши сведения об аргентинском рынке до и после обвала 12 августа 2019 (как показано на Рисунке 3.21). Рассуждая, как описано выше, нам иногда приходится со временем пересматривать хвосты в сторону ужирнения, но не в сторону утоньшения. Рафал Верон показал [264], что при подгонке устойчивого распределения мы рискуем преувеличить показатель хвоста и тем самым преуменьшить толщину хвоста.

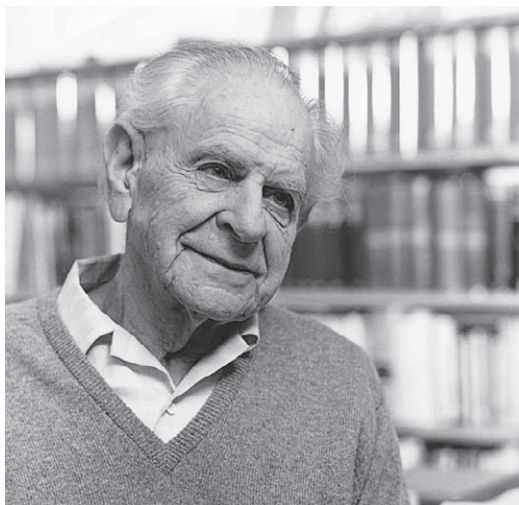
<sup>1</sup> Скептицизм (а философ Секст Эмпирик был главой школы скептиков) признает, что иногда можно опровергнуть теорию (указав противоречия), но не признает, что теорию можно доказать.



**Рисунок 3.16: «Теоретико-вероятностная вуаль неведения»<sup>1</sup>.** Талеб и Пильпель в [245] обсуждают вопрос с эпистемологической точки зрения, и в мысленном эксперименте с вуалью наблюдатель получает данные, которые производятся владельцем идеальной статистической информации при помощи генератора временного ряда. Задача наблюдателя: видя только поступающие данные и не зная ничего о порождающем процессе, оценить статистические свойства (вероятностное распределение, среднее, дисперсию, стоимость под риском и т. д.). Понятно, что наблюдатель без доступа к полной информации о генераторе и без надежной теории о принципе, порождающем данные, неизбежно будет ошибаться, но в его ошибках будет некоторая система. Исследовать эту систему — центральная задача в управлении рисками

<sup>1</sup> Вуаль неведения (*veil of ignorance*) в американской философии появилась в книге Джона Роллза «Теория справедливости» (*A Theory of Justice*, 1971) в мысленном эксперименте, предложенном Уильямом Викри и Яношем Харшаньи. Представьте, что вам разрешено выбрать общественную систему, при которой вам предстоит жить, но вы не будете знать заранее, в каком окажетесь исходном положении (*original position*): какая у вас будет расовая принадлежность, социальный статус и гендер; та же вуаль неведения скрывает, каковы будут представления о счастливой жизни у вас и у других.





**Рисунок 3.17:** *Поппер дал проблеме индукции асимметричное решение: полагаясь на подтверждающий эмпиризм, действовать в негативном стиле — то есть исключать то, что не сработало. Мы распространяем этот подход на статистические выводы с учетом вероятностной вуали и последовательно класс за классом исключаем опровергнутые вероятностные распределения*

### Научная строгость и асимметричные законы согласно российско-советской школе теории вероятностей

Можно доверять математической строгости утверждений о вероятностях, не попадаясь в ловушку наивных расчетов, уязвимых к ошибкам моделирования. Чудесный пример осознания асимметрии мы видим в работах российско-советской школы теории вероятностей — эта асимметрия составляет математическую аналогию идее Поппера.

Школа насчитывала три поколения: П. Л. Чебышёв, А. А. Марков, А. М. Ляпунов, С. Н. Бернштейн, Е. Е. Слуцкий, Н. В. Смирнов, Л. Н. Большев, В. И. Романовский, А. Н. Колмогоров, Ю. В. Линник, а также новое поколение: В. В. Петров, С. В. Нагаев, А. Н. Ширяев и ряд других.

Они сделали большое дело в истории научной мысли: стали работать не с равенствами, а с неравенствами (самые знаменитые — Маркова<sup>1</sup>, Чебышёва<sup>2</sup>, Бернштейна<sup>3</sup>, Ляпунова<sup>4</sup>). Вместо оценок они нашли границы. Даже для центральной предельной теоремы они построили версии, в которых делалось утверждение о границах, и мы воспользуемся ими позже и посмотрим,

1 Неравенство Маркова:  $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}$ .

2 Неравенство Чебышёва:  $\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{\mathbb{V}X}{a^2}$ .

3 Неравенство Бернштейна: пусть  $X_1, \dots, X_n$  — независимые случайные величины, принимающие значения 1 и  $-1$  с вероятностью  $\frac{1}{2}$ ; тогда для всякого  $\varepsilon > 0$

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n X_i > \varepsilon\right) \leq 2 \exp\left(-\frac{n\varepsilon^2}{2\left(1 + \frac{\varepsilon}{3}\right)}\right).$$

4 Неравенство Ляпунова: если  $0 < r < s < \infty$ , то  $(\mathbb{E}|X|^r)^{\frac{1}{r}} \leq (\mathbb{E}|X|^s)^{\frac{1}{s}}$ .



что делается за этими границами. Они ушли далеко вперед по сравнению с нынешним поколением пользователей, мыслящим в терминах точной вероятности — или, того хуже, механистических общественных наук. Их метод учитывает скептицизм, рассуждение в одну сторону. Они рассматривали условия не вида  $A = x$ , а вида  $A > x$  или  $AO(x)$  — заглавное О перед  $x$  означает «порядка  $x$ ».

Это великий источник для тех, кто работает над интеграцией математической строгости в теорию рисков. Мы всегда знаем только одну сторону. Мы знаем минимальную сумму, которую готовы заплатить за страховку, и не знаем верхней границы (или наоборот)<sup>a</sup>.

<sup>a</sup> Вот связь между асимметрией и робастностью. Робастный параметр не перескакивает в другой класс при изменении параметров вероятностного распределения. Если же класс изменяется, но асимметрично, т. е. проявляя выпуклость вниз или вверх при этих возмущениях, то этот класс соответственно хрупкий или антихрупкий; смотрите [223].



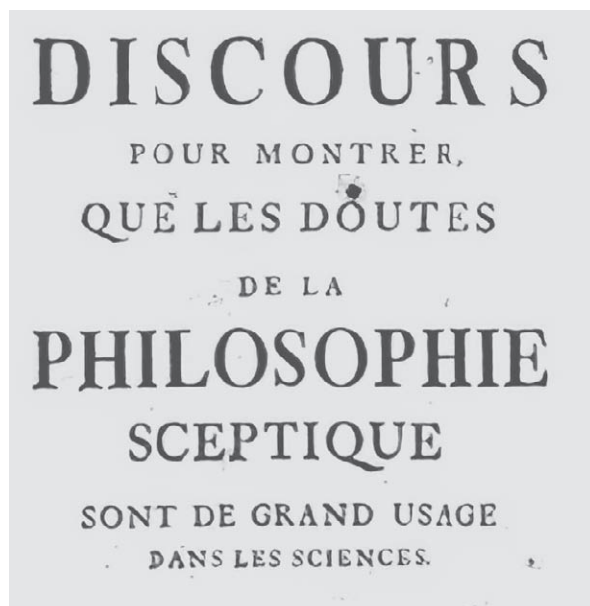
**Рисунок 3.18:** Проблема индукции. Философскую проблему перечислительной индукции можно сформулировать как вопрос:

— **Сколько белых лебедей нужно насчитать, прежде чем исключить появление черного лебедя в будущем?**

Он на удивление точно ложится на вопрос к работе закона больших чисел:

— **Сколько нужно собрать данных, чтобы сделать утверждение с приемлемой частотой ошибок?**

Оказывается, что статистический вывод по самой своей природе основывается на ясном определении и количественном измерении механизма индукции, и в случае толстого хвоста требуется гораздо больше данных. Как мы увидим в Главах 7 и 8, есть способ измерить относительную скорость индуктивного механизма, несмотря на то, что сама проблема индукции так и не получает идеального решения. Постановку проблемы индукции часто и ошибочно приписывают Юму [227]



**Рисунок 3.19:** Тракат о пользе скептической философии для науки<sup>1</sup>, автор Франсуа де ла Мот ле Ваје (1588–1672). По-видимому, послужил источником для епископа Юэ. Всякий раз, натываясь на очередного мыслителя, открывшего скептическое решение проблемы черного лебеда, я обнаруживаю, что бедолага повторил предшественников — не плагиатом, а по неумению отыскать корни. Мы настаиваем, что «проблема Юма» лишь в малой степени поставлена Юмом, возившим с собой по всей Европе многотомный Словарь<sup>2</sup>, изданный его предшественником Пьером Бейлем. Некоторое время я полагал, что первым был Юэ, но историки еще раз нашли более раннего предшественника

### 3.6. НАИВНЫЙ ЭМПИРИЗМ: НЕ НАДО СРАВНИВАТЬ ЭБОЛУ И ПАДЕНИЯ СО СТРЕМЯНОК

Давайте проиллюстрируем проблему тонкохвостого мышления в области толстых хвостов примером из реального мира. Некоторые цитируют так называемые «эмпирические» данные и объявляют, что мы зря тревожимся из-за эболы, от которой за 2016 год умерло всего двое американцев. Объявляют, что больше надо тревожиться из-за смертей от диабета и несчастных случаев, когда человек запутался в простынях. Давайте подумаем в терминах хвостов. Если однажды вы прочитаете в газетах о внезапной смерти 2 млрд человек, что вероятнее: что их убьет эбола или что они погибнут от курения, диабета и опутавших простыней?

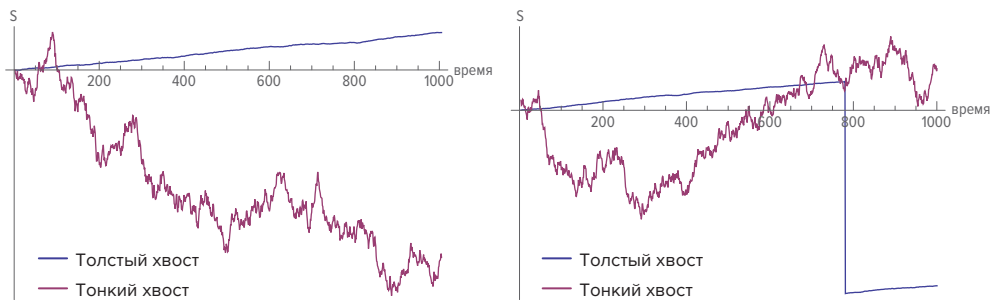
#### Принцип 3.2

*Не сравнивай толстохвостый процесс субэкспоненциального класса в Экстремистане, оказывающий мультипликативное воздействие, и тонкохвостый процесс в Медиокристане, особенно если он ограничен оценками Чернова<sup>3</sup>...*

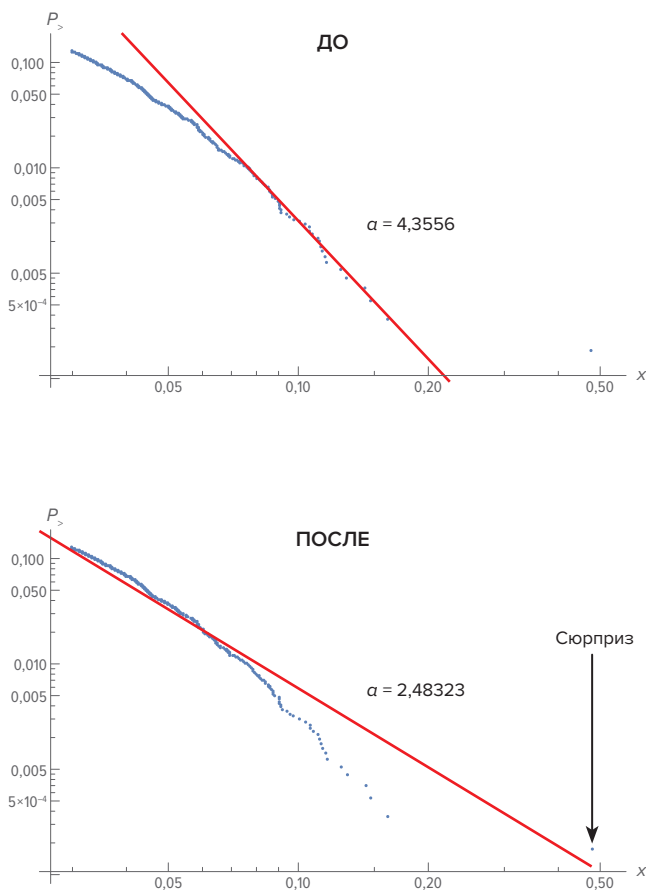
1 *Discours pour montrer que les doutes de la philosophie sceptique sont de grand usage dans les sciences*; издан в 1669.

2 Исторический и критический словарь (*Dictionnaire historique et critique*) в последнем, 16-томном издании выходил в 1820–1824 годах.

3 Оценки Чернова — серия теорем, полученных Германом Черновым для оценки вероятности больших отклонений сумм независимых случайных величин.



**Рисунок 3.20:** Невозможно «подтвердить» тонкохвостость, но опровергнуть ее удастся — как только случатся резкие скачки. Спокойные дни не позволяют исключить возможность скачков



**Рисунок 3.21:** Один день, открывший глаза на истинную природу распределения. Фондовый рынок Аргентины до и после 12 августа 2019. Иногда приходится резко пересмотреть хвосты в сторону утолщения (снизить показатель  $\alpha$ ); пересмотр в обратную сторону займет долгое время. Данные получены благодаря Диего Цвиговичу



**Рисунок 3.22:** *Наивный эмпиризм: не сравнивайте толстохвостые величины с тонкохвостыми, поскольку их выборочные средние из разных классов распределений. Данная ошибка генерализации сделана журналом The Economist, но она вообще обычна в так называемом ученом дискурсе. Однажды на это купилось само Королевское статистическое общество, предложив руководящий пост по «коммуникации риска» человеку с опытом работы в социологии и журналистике*

Это простое следствие принципа катастрофы, который мы уже упоминали и иллюстрировали на Рисунке 3.4.

К сожалению, немногие из заикленных на *доказательной науке* осознают (когда пишут статью), как работает эффект хвоста, вливающего собакой.

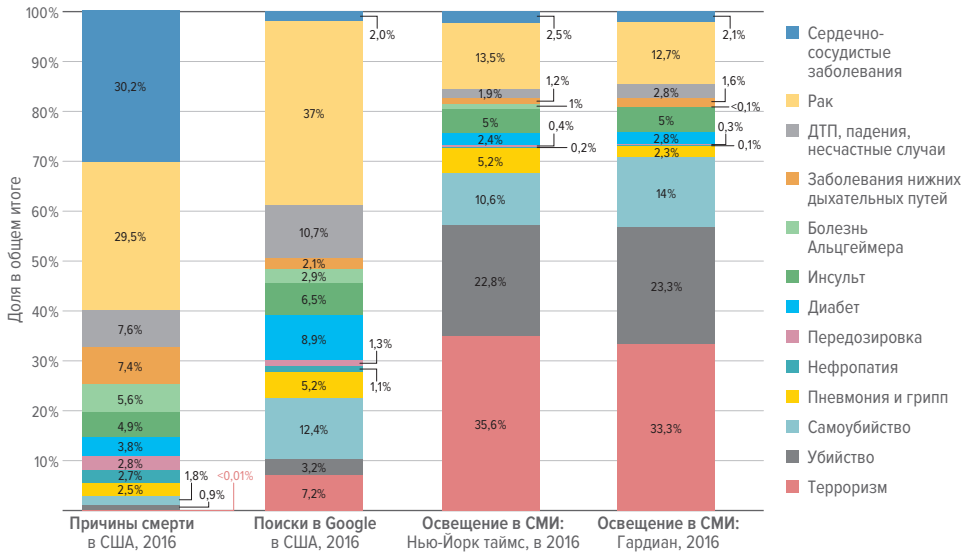
Это наивный эмпиризм — сравнить такие процессы и заявить, что мы слишком беспокоимся об эболе (эпидемиях, пандемиях) и недостаточно о диабете. На самом деле все наоборот. Мы слишком беспокоимся о диабете и недостаточно об эболе и других болезнях с мультипликативным воздействием. Ошибочное суждение делается из-за непонимания толстых хвостов, которое прискорбным образом ширится. Хуже того, подобные ошибки в рассуждениях продвигаются **эмпирической психологией**, которая не отличается особой эмпиричностью. Кроме того, они используются пропагандистами промышленности, которые под видом «коммуникаторов риска» впаривают нам пестициды и заверяют, что нет причин для беспокойства, потому что в прошлом ущерб был ничтожным.

Корректные рассуждения не приняты в кругах, занимающихся теорией решений и управления рисками; исключение составляют работающие над теорией экстремальных значений и отдел «Адаптивное поведение и познание»<sup>1</sup> в берлинском Обществе Макса Планка, руководимый Гердом Гигеренцером [108], — эти люди скажут вам, что не стоит игнорировать инстинкты и наставления вашей бабушки; когда ее советы

<sup>1</sup> *Adaptives Verhalten und Kognition*, отдел в Институте человеческого развития Общества Макса Планка (Max-Planck-Institut für Bildungsforschung). В англоязычной литературе известен как *The Center for Adaptive Behavior and Cognition (ABC)*.

## Причины смерти в США

От чего американцы умирают, что ищут в Google и о чем сообщают СМИ



**Рисунок 3.23:** Наивный (не статистический) эмпиризм Билла Гейтса: основатель Майкрософт\* продвигает и финансирует эту диаграмму, притом что он же заявляет об экзистенциальном риске из-за климата, не осознавая противоречивость двух терминов: экзистенциальные риски по определению отсутствуют в исторических данных. Далее, внимательное изучение диаграмм показывает, что на них присутствуют возрастные болезни: рак, сердечно-сосудистые заболевания и Альцгеймер, — которые волнуют молодежь и население среднего возраста меньше, чем терроризм и эпидемия.

Другой логический порок в том, что число смертей от рук террористов низкое как раз потому, что оно в центре внимания. Если снизить бдительность, оно может выйти из-под контроля. То же относится и к убийствам: страхи укрепляют безопасность. Так что скорее эта диаграмма показывает рациональность обычных людей с лучшим детектором риска хвостовых событий, чем у невежественных «экспертов». Восприятие последствий и свойств распределений откалибровано у людей лучше, чем утверждают психологи

\* Microsoft — технологическая компания; еще существовала на момент написания книги



**Рисунок 3.24:** Из-за медленной работы закона больших чисел при толстом хвосте в прошлом случилось, что тогдашнее будущее не было похоже на тогдашнее прошлое. Точно так же нынешнее прошлое не будет похоже на нынешнее будущее. Здесь вам не то что при тонком хвосте.

© Stefan Gasic





**Рисунок 3.25:** Опасайтесь лоббистов, использующих псевдоэмпирические аргументы. «Специалистов по коммуникации рисков» вроде вот этой личности, пришедшей из журналистики, нанимают такие фирмы, как Monsanto (или автомобильные концерны, или производители табака), для черного пиара со ссылкой на науку, эмпирические аргументы и опыт, чтобы дискредитировать «иррациональные» общественные опасения. Лоббистские организации инфильтровались в такие центры, как Harvard Center for Risk Analysis (звучит солидно для человека со стороны). Типичные

темы пиарщиков: «опыты доказали безвредность» и «будем рациональными». Их подхватывают журналисты, рассчитывая увлечь статистически наивную аудиторию. Распространение ложных концепций научно и, главное, статистически непросвещенными умами нанесло урон грамотности в вопросах вероятности и рисков статистической науке и журналистике

противоречат психологам и теоретикам принятия решений, обычно бабушкины советы ближе к научной строгости. Достаточно взглянуть на резюме книги моего «любимого» автора (Барух Фишхофф, Риск. Очень краткое введение [93]), чтобы увидеть полное нежелание различать два класса распределения вероятностей. Проблема психологов, не разобравшихся в «калибровке рисков» и «калибровке вероятностей», обсуждается с техническими подробностями в Главе 11, где рассматривается экспертная калибровка при толстом хвосте<sup>1</sup>.

### 3.6.1. Как изменяется масштаб некоторых мультипликативных рисков

«Доказательный подход» слишком примитивен, чтобы работать с эффектами второго порядка (и управлять рисками). Он безусловно нанес огромный ущерб в пандемию COVID-19 и должен быть решительно отвергнут за пределами управления рисками одного пациента. Одна из проблем состоит в переносе индивидуального риска на коллективный (еще одна — неправильная характеристика наблюдений и смешение их с отсутствием наблюдений).

В начале пандемии COVID-19 ряд эпидемиологов, невежественных по части вероятностей, уподобляли риск умереть от этой инфекции риску утонуть в бассейне. Положим, на том этапе это было верно для индивидуума (хотя очень скоро COVID-19 превратился во многих частях страны в главную причину смерти, и в Нью-Йорке достигал уровня 80 % всех смертей).

<sup>1</sup> Школа Гигеренцера тоже не имеет иммунитета от ошибок, как показывает недооценка ими рисков COVID-19 в начале 2020 г. Различия между Меднокристаном и Экстремистаном доходят тугу, но автор оптимистически надеется, что со временем дойдут. — Прим. автора.

Но и с самого начала можно было заметить, что если рассмотреть событие 1000 одновременных смертей, то вероятность приписать их утонувшим в бассейне окажется ничтожной.

Причина в том, что COVID у вашего соседа повышает ваши шансы заразиться, тогда как если ваш сосед утонет в бассейне, для вас вероятность утонуть в бассейне не вырастет (и если изменится, то в сторону уменьшения. Это как с авиакатастрофами: один утонувший уменьшает риск утонуть для остальных).

Технические детали этой проблемы накопления наряду с эллиптичностью смотрите в разделе 6.8 — совместное распределение теряет эллиптичность, и сумма оказывается жирнохвостой, даже когда отдельные слагаемые тонкохвостые.

Кроме того, рассматривается соответствующая проблема в этике [247]: заразившись инфекцией, вы можете стать причиной большего числа смертей, чем одна ваша. Даже когда риск смерти от некоторой заразной болезни меньше, чем, например, от автокатастрофы, «психология рациональности» (с ее моделями рациональности первого порядка) здесь будет психопатией, поскольку вы в итоге наносите вред системе и, естественно, самому себе тоже.

### 3.7. АЗБУКА СТЕПЕННЫХ ЗАКОНОВ (ПОЧТИ БЕЗ МАТЕМАТИКИ)

Обсудим интуицию, стоящую за законом Парето. По-простому он определяется так. Пусть  $X$  — случайная величина, для которой значение  $x$  — большое. Если вероятность того, что  $X$  превысит  $2x$ , разделить на вероятность превысить  $x$ , полученная величина «не слишком сильно отличается» от вероятности превысить  $4x$ , деленной на вероятность превысить  $2x$  и так далее. Это свойство называется «масштабируемость»<sup>1</sup>.

Таким образом, при распределении Парето или ему подобном отношение числа людей, владеющих 16 млн долларов, к числу людей, владеющих 8 млн долларов, такое же, как отношение числа двухмиллионеров к числу миллионеров. Неравенство в богатстве постоянное. У такого распределения нет характерного масштаба, и по-своему оно устроено очень просто. Хотя при таком распределении у случайной величины часто нет среднего и среднеквадратического отклонения, мы неплохо понимаем это распределение, даже лучше, чем более стандартные статистические распределения. Однако ввиду отсутствия среднего мы не можем обратиться за помощью к учебникам по статистике, и нам придется совершить более надежное, строгое исследование, которое на первый взгляд покажется менее математическим.

<sup>1</sup> Добавим минимум математики. Случайная величина  $X$  принадлежит классу распределений с правым хвостом по «степенному закону» с показателем  $\alpha$ , если

$$\mathbb{P}(X > x) = L(x) x^{-\alpha}, \quad (3.1)$$

где  $L: [x_{\min}, +\infty) \rightarrow (0, +\infty)$  — медленно меняющаяся функция, определяемая критерием  $\lim_{x \rightarrow \infty} \frac{L(kx)}{L(x)} = 1$  при всех  $k > 0$ . С соответствующим преобразованием применимо и к отрицательной области. — Прим. автора.

Таблица 3.2: Пример степенного закона

Богаче 1 млн	1 из 62,5
Богаче 2 млн	1 из 250
Богаче 4 млн	1 из 1000
Богаче 8 млн	1 из 4000
Богаче 16 млн	1 из 16 000
Богаче 32 млн	1 из ?

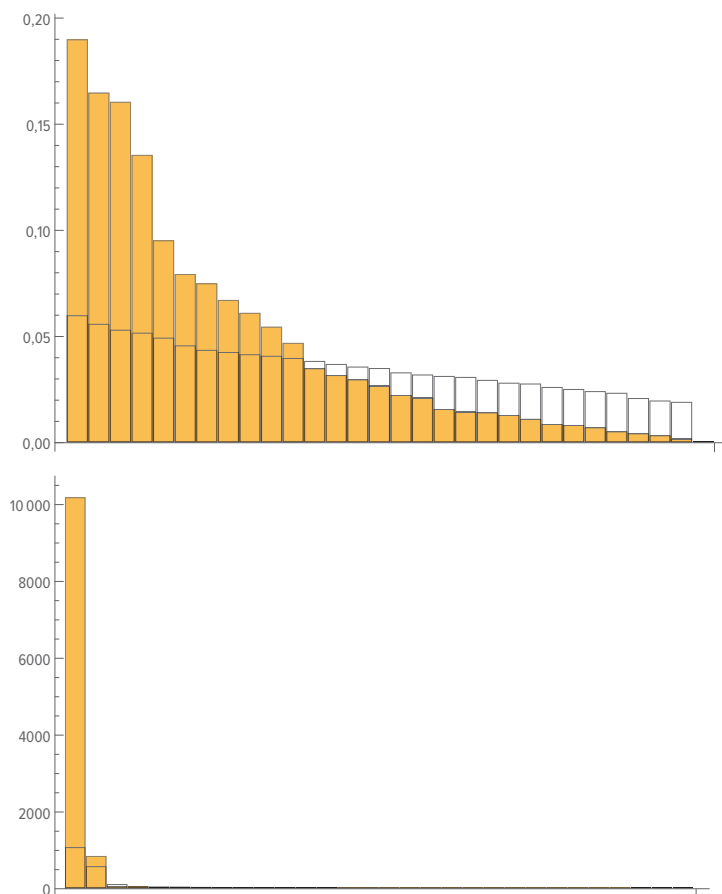
У распределения Парето нет высших моментов: они либо не существуют, либо статистически неустойчивы. Поэтому займемся теперь проблемой экономики и эконометрики. В 2009-м я взял данные за 55 лет и попробовал оценить эксцесс (функция четвертого момента) по самым обширным наблюдениям — смотрите Таблицу 3.3. Для гауссова распределения максимальный вклад за такой период времени должен быть в диапазоне  $0,008 \pm 0,0028$ . Но для портфеля S&P 500 получилось 80 %. Это говорит о том, что мы понятия не имеем, каков эксцесс для биржевой стоимости этих ценных бумаг. Ошибка определения значения по выборке огромная; возможно, параметр вообще не существует, и оттого наша оценка критически зависит от выборки. Раз мы ничего не знаем о четвертом моменте, мы ничего не знаем об устойчивости второго момента. Значит, мы не в том классе распределений, чтобы работать с дисперсией, даже если она существует. Гранит науки тверд, и финансовая математика — тоже трудное дело.

Таблица 3.3: Вклад одного наблюдения в эксцесс в финансовых данных,  $\frac{\max_{i=0}^n (X_{t-\Delta t i}^4)}{\sum_{i=0}^n X_{t-\Delta t i}^4}$

Ценные бумаги	Max Q	Годы
Серебро	0,94	46
Индекс S&P 500	0,79	56
Сырая нефть	0,79	26
Процентный фьючерс	0,75	17
Топочный мазут	0,74	31
Индекс Никкей	0,72	23
Индекс FTSE	0,54	25
Японские гособлигации	0,48	24
1-месячный евродоллар	0,31	19
Сахар	0,3	48
Иена	0,27	38
Индекс Bovespa	0,27	16
3-месячный евродоллар	0,25	28
СТ	0,25	48
Индекс DAX	0,2	18

Для серебра за 46 лет 94 % эксцесса имело источником одно-единственное наблюдение. Для финансовых данных не годятся стандартные статистические методы. GARCH (метод, популярный в академических кругах) не работает, потому что мы имеем дело с квадратами. Дисперсия в квадратах аналогична четвертому моменту.





**Рисунок 3.26: Ложные главные компоненты при толстых хвостах:** эксперимент Монте-Карло показывает, насколько острее проблема ложных корреляций и ковариаций при толстых хвостах. На верхней гистограмме главные компоненты ранжированы по дисперсии для 30 гауссовых некоррелированных величин, когда число точек данных  $n = 100$  (закрашено) и 1000 (не закрашено). На нижней гистограмме главные компоненты ранжированы по дисперсии для 30 величин с устойчивым распределением (показатель хвоста  $\alpha = \frac{3}{2}$ , коэффициент симметрии  $\beta = 1$ , коэффициент положения  $\mu = 0$ , коэффициент масштаба  $\sigma = 1$ ) при том же числе точек данных  $n = 100$  (закрашено) и  $n = 1000$  (не закрашено). На обеих диаграммах представлены случайные величины «некоррелированные» и одинаково распределенные. Мы видим уплощение спектра главных компонент при росте числа  $n$  гауссовых величин — различие между главными компонентами тает. При жирном хвосте подобное уплощение за разумное число наблюдений не достигается

Мы не знаем дисперсию. Зато с распределениями Парето можем работать очень даже легко. Они дают меньше информации, но более строги, если данные неограниченные или если есть открытые переменные.

Таблица 3.3, с финансовыми данными, опровергает все нынешние вузовские учебники. Куча эконометрики, работающей с квадратами, идет на помойку. Становится

понятно, почему экономисты не могут предсказывать события — они пользуются не теми методами и строят неверные доверительные интервалы. Что работает на выборке, не будет работать за пределами выборки — а выборки по определению всегда конечны и всегда имеют конечные статистические моменты. Как объявить дисперсию или эксцесс бесконечными, когда мы не наблюдаем ничего бесконечного в выборке?

Метод главных компонент (см. Рисунок 3.26) сокращает размерность больших данных и нормально работает при тонких хвостах (по крайней мере иногда). Но при недостаточности данных предлагает иллюзорную структуру  $n$  переменных, которая уплощается с ростом объема данных — это так называемый эффект Вигнера для случайных матриц (в честь Юджина Вигнера. Не путать с открытиями Вигнера о дислокациях атомов в материале под действием радиации). В этой симуляции данные не имеют никакой структуры и главные компоненты должны оказаться все одной величины (асимптотически, при большом объеме данных); однако эффект малой выборки придает главным компонентам уклон вниз.

Для главных компонент нужно получить матрицу с нулевой корреляцией. При распределении с толстым хвостом (нижняя часть рисунка) требуется гораздо больше данных для устранения ложных корреляций, т. е. при толстых хвостах сократить размерности не удастся.

### 3.8. ГДЕ ПРЯЧУТСЯ СКРЫТЫЕ СВОЙСТВА?

Следующее резюмирует все, что я написал в «Черном лебеде» (почему-то неискаженная версия этой идеи пробивалась больше десяти лет). Распределения бывают с одним хвостом (левым или правым) или с двумя. Толстохвостое распределение бывает однохвостым или двуххвостым. Толстохвостое однохвостое бывает с хвостом слева или с хвостом справа.

Интуитивную подсказку смотрите на Рисунке 3.28: при толстом хвосте редкие события скрылись от наблюдения, и наивный наблюдатель неверно оценивает среднее однохвостого распределения. Хвостовые события еще внесут свой вклад, но, по определению, они редки.

Если судить об эффективности инвестиций по временному ряду с краткосрочной волатильностью (скошенность влево, угроза чувствительных потерь), скорее окажешься *одураченным случайностью*, чем если судить по ряду с долгосрочной хвостовой волатильностью (скошенность вправо, возможность крупных выигрышей). Говоря по-простому, краткосрочная волатильность завышает оценку эффективности, а долгосрочная — занижает (см. Рисунок 3.28). Это еще один вариант несимметричности по Попперу, о которой говорилось выше в этой главе.

Фокус в том, что нужно каким-то образом оценить распределение и потом уже выводиться среднее (опираясь на экстраполяцию). В этой книге метод называется «дополнительной оценкой»; смотрите Таблицу 3.4. Не стоит полагаться просто на среднее по выборке наблюдений; оно при распределениях с толстым хвостом имеет

систематическую ошибку. По этой причине кажется, пока не грянул кризис, что банки приносят большие прибыли. А потом оказывается, что они потеряли все, что имели, и даже больше, и что их приходится выкупать на деньги налогоплательщиков. Чтобы не попадаться в ловушку, мы отличаем истинное среднее (которое я называю *теневым*) от реализовавшегося среднего, что и показано в Таблице 3.4.

**Таблица 3.4:** Выборочное среднее, теневое среднее и их отношение при различном минимальном пороге  $L$ . Теневое среднее получено с дополнительными оценками по методу максимального правдоподобия. Масштаб изменен. По Чирилло и Талебу [46]. Подробности в Главах 16 и 13

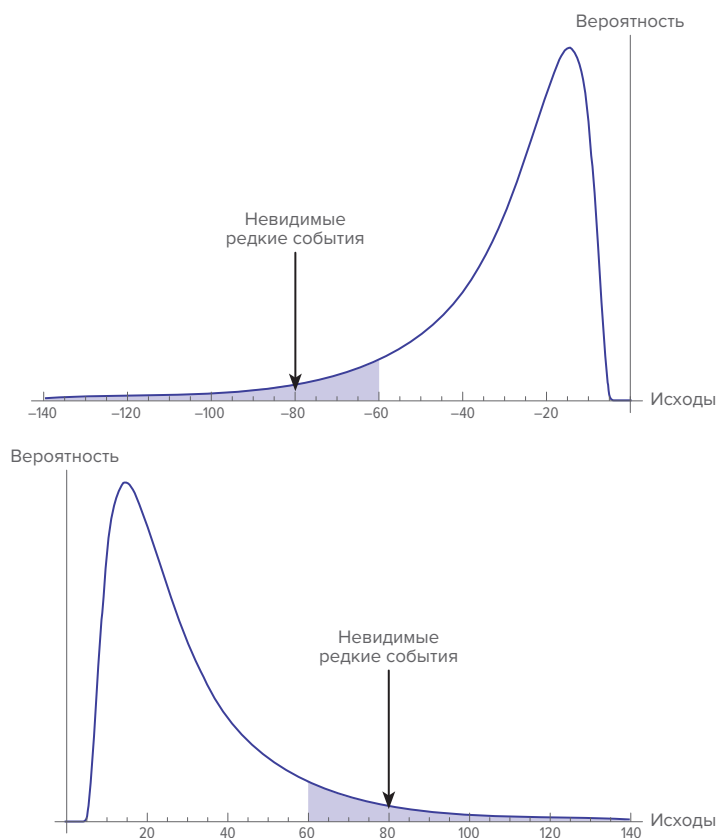
$L$	Выборочное среднее	Среднее по ММП	Отношение
10 тыс.	$9,079 \times 10^6$	$3,11 \times 10^7$	3,43
25 тыс.	$9,82 \times 10^6$	$3,62 \times 10^7$	3,69
50 тыс.	$1,12 \times 10^7$	$4,11 \times 10^7$	3,67
100 тыс.	$1,34 \times 10^7$	$4,74 \times 10^7$	3,53
200 тыс.	$1,66 \times 10^7$	$6,31 \times 10^7$	3,79
500 тыс.	$2,48 \times 10^7$	$8,26 \times 10^7$	3,31

Аналогичным образом можно обработать коэффициент Джини, чтобы оценить «теневое», а не наивно наблюдаемое социальное неравенство.

Вот что мы имеем в виду, когда говорим, что выборочное распределение «не эмпирично». Другими словами: (1) есть разница между атрибутами генеральной совокупности и выборки и (2) даже на исчерпывающие исторические данные следует смотреть как на выборку из более широкого диапазона (прошлые



**Рисунок 3.27: Фундаментальная асимметрия** — различие между отсутствием свидетельств и свидетельством отсутствия — усложняет исследование толстых хвостов. Требуется более продвинутое понимание случайных событий — или более реалистичное. (Прошу не понимать здесь баллы IQ буквально, как в обычной психометрике. Есть подозрение, что как раз лица с высоким IQ не разберутся с данной асимметрией. Считайте, что здесь речь об идеальном IQ, отвечающем «настоящей» рациональности, а не результатам тестирования.) © Stefan Gasic



**Рисунок 3.28:** Теневое среднее за работой. Нижний график: обратная проблема индейки — невидимое редкое событие позитивное. Наблюдая позитивно скошенный (антихрупкий) временной ряд и делая выводы о невидимом без учета параметров, вы упускаете из виду хорошее и недооцениваете выгоды. Верхний график: противоположная проблема. Закрашенная область соответствует тому, что мы обычно упускаем в малых выборках, из-за недостаточного числа точек данных. Интересно, что закрашенная площадь растет с ростом ошибки модели (благодаря выпуклой вниз зависимости хвостовых вероятностей от неопределенности)

является выборкой; выводы из него являются статистическими суждениями на основе выборки).

Разобравшись в типе распределения, можно оценить математико-статистическое среднее. Оценка будет намного точнее, чем простое измерение среднего по выборке. Так, при распределении Парето 98 % наблюдений ниже среднего; наблюдения недооценивают среднее. Но после того, как мы поняли, что имеем дело с распределением Парето, мы можем игнорировать выборочное среднее и использовать другие методы. Техника дела обсуждается в Главах 13 и 15.

Заметим, что исследователи теории экстремальных значений [115] [82] [116] фокусируются на свойствах хвоста, а не на поиске среднего или статистических выводов.

### ЛИНЕЙКА ВИТГЕНШТЕЙНА: ЧТО, ЭТО ПРАВДА БЫЛО 10-СИГМОВОЕ СОБЫТИЕ?

Летом 1998 хедж-фонд с громким названием *Long Term Capital Management* («Долгосрочное управление капиталом») приказал долго жить, разорившись из-за «неожиданных» отклонений на рынках. Это была чувствительная утрата: двое партнеров успели получить премию шведского госбанка, рекламируемую как «Нобелевка в экономике». Фонд привлек многочисленных профессоров финансовой математики и имел подражателей, также привлечших профессоров (и по самому скромному счету, 60 докторов наук разорились в тот период благодаря трейдерам, похожим на LTCM и использующим идентичные методы управления рисками). Минимум двое партнеров выступили с заявлением, что произошло *10-сигмовое событие* (отклонение в 10 раз больше среднеквадратического) и что это снимает с них обвинения в некомпетентности (я лично был свидетелем двух таких заявлений).

Давайте применим в этой ситуации подход, который автор называет *витгенштейновской линейкой*: спросим себя, правда ли мы измеряем линейкой стол, как обычно, или на этот раз мы измеряем столом линейку?

Примем для простоты, что есть только две возможности: гауссово распределение и распределение по степенному закону. При гауссовом вероятностность, которую мы определим как функцию выживания для события величиной 10 среднеквадратических отклонений, составит 1 из  $1,31 \times 10^{23}$ . Для степенного закона в том же масштабе — *t*-распределения Стьюдента с показателем хвоста 2, — функция выживания составит 1 из 203.

Ну и какова вероятность, что распределение гауссово при условии наблюдения 10-сигмового события и выбора из наших двух альтернатив?

Начнем с правила Байеса.  $\mathbb{P}(A | B) = \frac{\mathbb{P}(A)\mathbb{P}(B | A)}{\mathbb{P}(B)}$ . Заменим  $\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B | A) + \mathbb{P}(\bar{A})\mathbb{P}(B | \bar{A})$  и применим к нашему случаю.

$\mathbb{P}(\text{Гаусс} | \text{Событие}) =$

$$= \frac{\mathbb{P}(\text{Гаусс})\mathbb{P}(\text{Событие} | \text{Гаусс})}{(1 - \mathbb{P}(\text{Гаусс}))\mathbb{P}(\text{Событие} | \text{не-Гаусс}) + \mathbb{P}(\text{Гаусс})\mathbb{P}(\text{Событие} | \text{Гаусс})}$$

$\mathbb{P}(\text{Гаусс})^1$

$\mathbb{P}(\text{Гаусс} | \text{Событие})$

0,5

$2 \times 10^{-21}$

0,999

$2 \times 10^{-18}$

0,9999

$2 \times 10^{-17}$

0,99999

$2 \times 10^{-16}$

0,999999

$2 \times 10^{-15}$

1

1

<sup>1</sup> Очевидно, имеется в виду априорная вероятность того, что распределение гауссово (до известия о 10-сигмовом событии).

Мораль: если до наблюдения была хоть крошечная,  $< 10^{-10}$ , вероятность того, что распределение не гауссово, после случившегося можно сделать твердый выбор в пользу альтернативы — распределения с толстым хвостом.

Простое эвристическое правило: смело отвергайте гипотезу о гауссовости, если случилось хотя бы одно событие  $> 4$  или  $5$  среднеквадратических отклонений. Никакие заплатки вроде условной дисперсии, как мы увидим в этой книге, не дают адекватного результата и должны быть признаны мошенничеством<sup>a</sup>.

<sup>a</sup> Великий человек Бенуа Мандельброт крайне критически относился к методам, дополняющим гауссовость разрывами или иными трюками *ad hoc*, чтобы объяснить, что не так с данными (например, процесс диффузионных скачков Мертон [173]). Задним числом всегда можно подогнать нужные разрывы. Мандельброт цитировал слова, приписываемые Джону фон Нейману: «С четырьмя параметрами я могу подогнать кривую под слона, а с пятью — заставить его вилять хоботом».

### 3.9. БАЙЕСА-ШМАЙЕСА

Если нет надежных сведений, от Байесовых методов мало помощи. Автора этой книги не раз спрашивали после выхода «Черного лебедя», нельзя ли решить проблемы с неизвестными толстыми хвостами при помощи чего-нибудь такого байесовского. Но невозможно заменить отсутствующие наблюдения сфабрикованной информацией, и никакие байесовские-шмайесовские ухищрения не помогут. Прежде всего, Байесовы методы срабатывают при наличии надежной априорной вероятности, но она в обычных наблюдательных данных отсутствует (смотрите у Диакониса и Фридмана [66] о том, как трудно агенту сформулировать априорную вероятность).

Проблема в скорости обновления, которая, как мы увидим в Главе 7, сильно зависит от распределения. В литературе по теории рациональных ожиданий делается ошибочное предположение, что два наблюдателя по одной и той же информации обязательно придут к одному и тому же мнению. К сожалению, требуются весьма специальные условия, чтобы это случилось в реальном времени или вообще случилось.

Разумеется, можно пользоваться методами Байеса (при наличии адекватных априорных оценок) для оценки параметров — если (1) есть ясное представление о диапазоне значений (например, по классу универсальности или по иным бассейнам устойчивости) и (2) эти параметры следуют поддающемуся анализу распределению с низкой дисперсией — например, для распределения Парето наблюдаемый показатель степени хвоста следует обратному гамма-распределению [11].

**МОРАЛЬНЫЙ РИСК<sup>1</sup> И ПОГОНЯ ЗА РЕНТОЙ** в финансовом образовании: одно из самых депрессивных воспоминаний в жизни автора — как он читал курс о жирных хвостах в недолгую пору работы в Школе бизнеса Массачусетского университета в Амхерсте. Один аспирант по финансам прямо за-

<sup>1</sup> От англ. *moral hazard*, риск обмана со стороны участника договора.

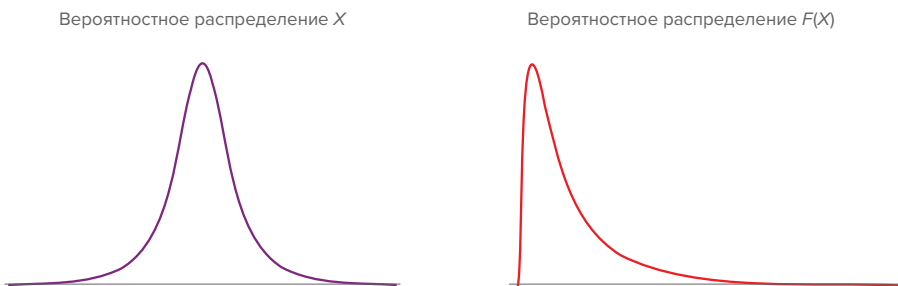
явил, что идеи ему нравятся, но карьера в финансовом образовании требует «добиться самого высокого жалования в стране» (в академической сфере). Он выбрал специализацию по методам Марковица (зная, что они не работают в области толстого хвоста), поскольку их применяли остальные профессора, и такой выбор открывал дорогу к публикации статей и получению высокооплачиваемой должности.

Я был разочарован, но предсказал ему в будущем успешную карьеру автора нестатей. Он ее добился.

### 3.10. $X$ И $F(X)$ : КАК ПУТАЮТ ВОЗДЕЙСТВИЕ ВЕЛИЧИНЫ $X$ С САМОЙ ВЕЛИЧИНОЙ $X$

Рассмотрим случайную или неслучайную величину  $X$  и ее воздействие на вас  $F(X)$ <sup>1</sup>, например платеж, ущерб, итоговый баланс. (Нередко величина  $X$  многомерная, но примем для простоты, что у нас она одномерная.)

Практики и венчурные инвесторы часто замечают такое недоразумение: не-практики начинают рассказывать практикам об  $X$ , считая, что практикам это интересно, тогда как те думают об  $F(X)$  и ни о чем другом. Смешение  $X$  с ее последствиями  $F(X)$  — хроническая болезнь со времен Аристотеля; она обсуждается в *Антихрупкости* [230] как главная тема книги. Иные делают шаг от  $X$  к функции полезности, но не доходят до итогового платежа. Путаница происходит и на бытовом уровне, и в литературе по теории принятия решений; в теории различие признается, но не осознается, что действовать удобнее с оглядкой на  $F(X)$ , а не на  $X$ .



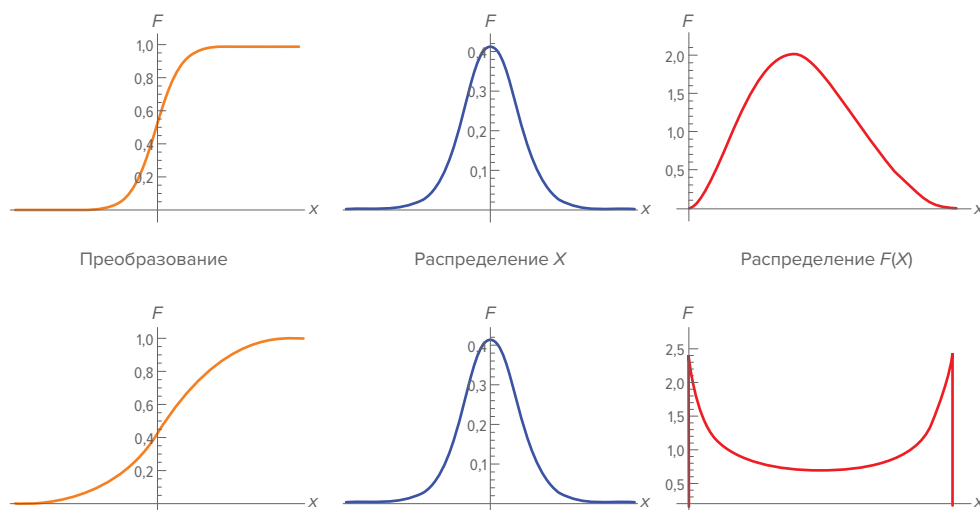
**Рисунок 3.29:** Проблема смешения случайной величины  $X$  и ее платежа,  $F(X)$ . Когда знаешь, что функция  $F$  выпукла вниз, подробностями  $X$  можно не интересоваться. Когда стоит задача изменить ситуацию, надежнее изменить  $F$ , а не  $X$

- Величина  $X$  может быть показателем безработицы в Сенегале,  $F_1(X)$  — ее воздействием на итоговый отчет МВФ, а  $F_2(X)$  — ее воздействием на финансы вашей бабушки (надо полагать, минимальным).

<sup>1</sup> Тем самым в данной главе  $F(x)$  обозначает не функцию распределения случайной величины  $X$ ,  $F(x) = \mathbb{P}(X < x)$ , как в других местах книги, а функцию платежа, или последствий того, что случайная величина  $X$  примет значение  $x$ .

- $X$  может быть курсом акций, причем вы владеете опционом, и  $F(X)$  — воздействие  $X$  на стоимость вашего опциона или даже еще хитрее — функция полезности этого воздействия на стоимость опциона.
- $X$  может быть изменением вашего богатства, а  $F(X)$  — его выпукло-вогнутым воздействием на ваше благополучие. Можно заметить, что  $F(X)$  намного устойчивее, или робастнее — то есть имеет более тонкие хвосты, — чем  $X$ .

**Выпуклые и линейные функции случайной величины  $X$ .** Рассмотрим Рис. 3.30: чем меньше зависимость  $F$  (по вертикали) от  $X$  (по горизонтали) похожа на линейную<sup>1</sup>, тем меньше остается общего между вероятностным распределением  $X$  и  $F(X)$ . Даже при небольшой выпуклости  $F$  статистические и прочие свойства  $F(X)$  отличаются от свойств  $X$ . Например, математическое ожидание  $F(X)$  отличается от  $F(\mathbb{E} X)$  согласно неравенству Йенсена<sup>2</sup>. При резкой нелинейности  $F$  влияние вероятностного распределения  $X$  на вероятностное распределение  $F(X)$  тает. Мораль: чтобы что-то изменить, фокусируйтесь на  $F$  и не гонитесь за измерением трудноуловимых свойств  $X$ .



**Рисунок 3.30:** Проблема смещения: выпуклая вниз-вверх S-образная функция  $F(x)$  превращает случайную величину  $X$  с толстохвостым распределением на бесконечном промежутке  $(-\infty, +\infty)$  в случайную величину  $F(X)$  с бесхвостым распределением на промежутке  $(0, 1)$ . Сильная выпуклость (нижний пример) способна превратить пик в яму с вертикальными стенками на краях, аналогичную арксинусу

**Ограниченные знания.** Важно, что ограниченность наших знаний касательно распределения  $X$  не обязательно транслируется в неопределенность распределения  $F(X)$ . Бывает, что мы не управляем  $X$ , но можем повлиять на  $F$ . Иногда это дает нам весьма существенное влияние на  $F(X)$ .

<sup>1</sup> Платежная функция на графиках слева. В верхнем примере она близка к линейной, в нижнем примере далека от линейной.

<sup>2</sup> Неравенство Йенсена гласит: если функция  $F(x)$  выпуклая вниз, то  $F(\mathbb{E} X) \leq \mathbb{E} F(X)$ .



Опасное заблуждение относительно проблемы черного лебедя — фокусироваться на  $X$ , пытаться предсказать  $X$ . Я пытаюсь всем объяснить, что, даже не понимая  $X$ , можно работать с понятным нам  $F$ ; другие же упорно бьются над предсказанием  $X$  — бесплодно, потому что маловероятные события не поддаются расчету, особенно в области толстого хвоста. Между тем итоговое воздействие на нас оказывает как раз  $F(X)$ .

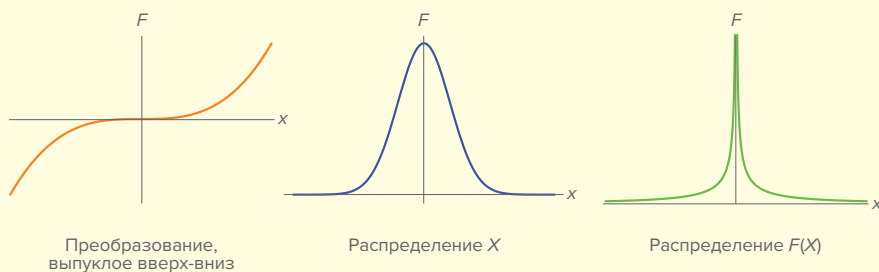
Вероятностное распределение у  $F(X)$  не такое, как у  $X$ , когда функция  $F$  нелинейная. Чтобы получить  $F(X)$  из  $X$ , нужно выполнить нелинейное преобразование. Только в 1964 началось обсуждение «выпуклого преобразования случайной величины», см. ван Звет [259], так как до того тема не считалась важной.

**Повсеместная распространенность S-образных кривых.** Функция  $F$  практически всегда нелинейная (лично я не знаю ни одного исключения) и часто «S-образная», то есть выпуклая вниз-вверх возрастающая. Более подробное обсуждение см. в Приложении F к Главе 10.

**Хрупкость и антихрупкость.** Когда  $F$  выпуклая вверх (хрупкая), ошибки относительно  $X$  могут преобразоваться в огромные отрицательные отклонения  $F(X)$ . Когда  $F$  выпуклая вниз, мы защищены от суровых отрицательных отклонений. Если мы действуем методом проб и ошибок или выбираем опцию или опцион, нам важно не столько понимать про случайную величину  $X$ , сколько про создаваемые риски. Грубо говоря, статистические свойства  $X$  тонут в свойствах  $F$ . Книга *Антихрупкость* посвящена тому, что воздействие  $X$  важнее и наивна установка «хочу все знать», то есть понимать  $X$ .

Чем сильнее нелинейность  $F$ , тем меньше вероятностное распределение  $X$  влияет на вероятностное распределение итогового пакета,  $F(X)$ .

Многие авторы путают вероятности тех или иных значений  $X$  с вероятностями  $F(X)$ . Я не шучу: чуть ли не вся литература базируется на этой ошибке. Ради Ваала, фокусируйтесь на  $F$ , а не на  $X$ .



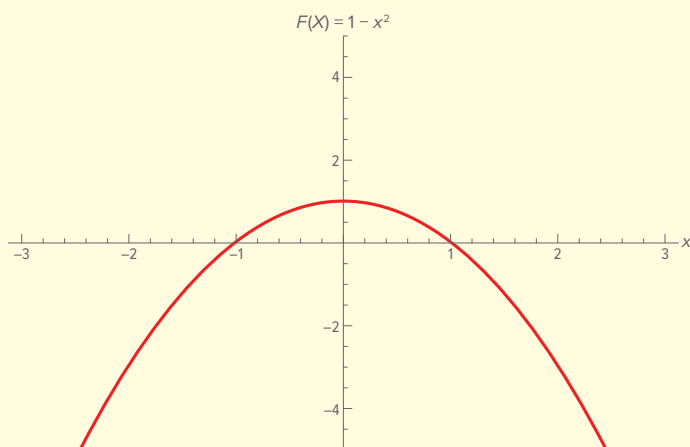
**Рисунок 3.31:** Преобразование выпуклое вверх-вниз (такое, как пробит-функция<sup>1</sup> — обратная к интегральной функции стандартного нормального распределения; другой пример — логит-функция<sup>2</sup>) делает хвосты распределения  $F(X)$  толще

1  $\text{probit}(p) = \Phi^{-1}(p) = \sqrt{2} \operatorname{erf}^{-1}(2p - 1)$ , где  $p \in (0, 1)$ .

2  $\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$ , где  $p \in (0, 1)$ .

**Будь хоть сто раз прав, нужна выпуклость вниз.** Осенью 2017 одна фирма обанкротилась, делая ставки по поводу волатильности. Эти люди прогнозировали дисперсию ниже той, что ожидалась рынком... и были правы. Почему же они обанкротились? Потому что использовали выпуклую вверх функцию платежа. Вспомним, что случайная величина  $X$  распределена не так, как  $F(X)$ , и что в реальном мире практически все функции  $F$  нелинейны.

Разобраться нам поможет следующий пример. Рассмотрим платежную функцию на рисунке ниже<sup>1</sup>. Это функция  $F(x) = 1 - x^2$ , где  $x$  — отклонение цены закрытия; таков ваш платеж за этот день, и если  $x$  не превышает единицу (допустим, единица — это среднее квадратическое отклонение цены закрытия), то вы в плюсе, а иначе несете убытки. Типичный договор типа «вариационный своп»<sup>2</sup>.



Теперь рассмотрим следующие две последовательности отклонения  $x$  в течение 7 дней (в тех же единицах, среднее квадратических отклонениях).

Последовательность 1 (тонкохвостая):  $\{1, 1, 1, 1, 1, 0, 0\}$ . Среднее отклонение 0,71. Ваша чистая прибыль 2.

Последовательность 2 (толстохвостая):  $\{0, 0, 0, 0, 0, 0, 5\}$ . Среднее отклонение 0,71 (то же самое). Ваша чистая прибыль = -18, прямо разорение.

В обоих примерах прогноз среднего 0,71 сбывается, но одна и та же стратегия в первом примере принесет прибыль, а во втором — убытки; все испортит неоднородность в волатильности — жирный хвост вероятностного распределения.

Вот почему в реальном мире из «плохих» прогнозистов получаются успешные трейдеры и ответственные руководители. Все профессионалы об этом знают, но литература по «прогностике», математически беспомощная и практически наивная, до сих пор не разобралась, в чем дело, после столетий кипучей активности.

<sup>1</sup> Эта функция выпуклая вверх.

<sup>2</sup> Современный финансовый жаргон; от англ. *variance swap*.

### 3.11. РАЗОРЕНИЕ И ЗАВИСИМОСТЬ ОТ ПУТИ

Напоследок поговорим о зависимости от пути и о вероятности по времени. Наши прабабушки разбирались в толстых хвостах. Не так страшен черт, как его малюют; мы разобрались, как выживать, принимая рациональные решения с учетом фундаментальных статистических свойств.

Зависимость от пути означает следующее. Если я сначала поглажу рубашки, а потом постираю, результат будет не такой, как если сначала постирать, а потом погладить. В своей первой книге, *Динамическое хеджирование* [225], я писал про то, как трейдеры стараются не «врезаться в барьер»: если разоришься, то выйдешь из игры; если стратегия окончилась банкротством, потеряв *весь* выигрыш.

Физики Оле Питерс и Мюррей Гелл-Манн [186] пролили новый свет на этот вопрос и революционизировали теорию принятия решений, показав ложность принципа, на котором базировалась традиционная прикладная теория вероятностей в экономике. Физики обнаружили, что все учебники экономики делают общую ошибку; избежали ее только специалисты по теории информации, такие как Келли и Торп.

Разберемся сначала, что такое вероятности для статистического ансамбля.

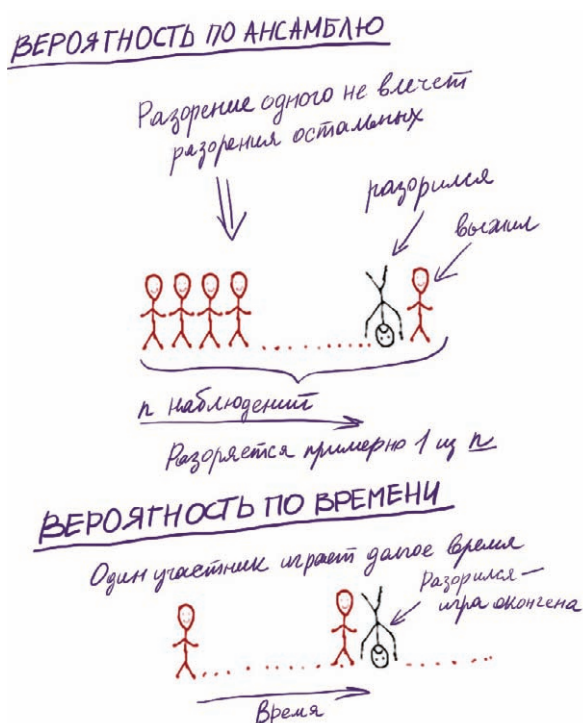
Допустим, что мы — это 100 случайно выбранных человек и что мы отправились в казино и играем там в азартные игры. Если 28-й игрок разорился, это не помешало 29-му играть дальше. Значит, можно посчитать доход от похода в казино, сложив наши 100 результатов. Закон больших чисел позволяет нам повторить опыт 2–3 раза и точно измерить так называемое *преимущество* данного казино.

Проблема появится при попытке применить это ансамблевое свойство к 100 визитам одного человека. Прежний подход не сработает, потому что если кто-то на 28-й день разорится, то на 29-й день уже не придет. По этой причине Крамеру пришлось сформулировать *условие Крамера*, за пределами которого страхование не работает; по этому условию недопустим риск разорения из-за одного потрясения. Аналогичным образом отдельный инвестор не может достигнуть доходности, равной среднему показателю фондового рынка, потому что ресурсы отдельного инвестора конечны (или, как заметил Оле Питерс, отдельный инвестор не может много раз прогнать свою жизнь по множеству ветвящихся вселенных).

Добиться дохода на фондовом рынке можно лишь при выполнении строгих условий.

Вероятность по времени и вероятность по ансамблю — не одно и то же. Активному инвестору удастся приблизиться к вероятности по ансамблю, только если этот инвестор использует политику распределения ресурсов, отвечающую критерию Келли [142], [250] с диаграммами. Питерс написал три статьи о вероятности по времени (одну в соавторстве с Мюрреем Гелл-Манном) и показал, что ряд парадоксов устраняется.

Посмотрим, как работать с этими вероятностями и что не так в литературе. Для наглядности подействуем на трейдера небольшим, но часто повторяющимся риском разорения; со временем вероятность разорения приблизится к единице.



**Рисунок 3.32:** Вероятность по ансамблю и вероятность по времени. Трейдеры на рынке опционов рискуют врезаться в барьер. Я все время указываю на это, и в «Динамическом хеджировании» [225], и в «Антихрупкости» [223], как, например, смещения случайной величины  $X$  с ее действием  $F(X)$ , когда последнее содержит барьер



**Рисунок 3.33:** Харальд Крамер, автор условия Крамера и анализа разорения

Если мы решим кататься на мотоцикле с небольшим риском катастрофы, но много раз, мы заметно сократим ожидаемую продолжительность жизни. Измерить это можно так.

### Принцип 3.3 (Повторение воздействия)

Фокусируйтесь на сокращении ожидаемого срока службы того модуля, который подвергается многократному воздействию с некоторой плотностью, или частотой.

Финансовые теоретики-бихевиористы пока что выводят свои суждения из статистики, а не динамики, и поэтому данную картину не видят. Они вырывают компромиссы из контекста и приходят к консенсусу, будто человеку свойственно иррационально преувеличивать хвостовые риски (и его надо *подталкивать* к более смелым действиям). Но катастрофическое событие является барьером. Смелые действия нельзя анализировать по отдельности: риски накапливаются. Если мы катаемся на мотоцикле, курим, водим винтовой самолет и вступаем в мафию, перечисленные риски складываются в весьма вероятную преждевременную гибель. Хвостовые риски — не то же, что риск потерять заменимый ресурс.

Каждый выживший венчурный инвестор это понимает. Уоррен Баффетт это понимает. Группа «Голдман Сакс»<sup>1</sup> это понимает. Их не удовлетворяют низкие риски, им требуются нулевые риски, и этим фирма, выжившая на протяжении двадцати, тридцати, ста лет, отличается от разорившейся. Такой подход к рискам объясняет, почему группе «Голдман Сакс» исполнилось 149 лет. Она управлялась как партнерство с неограниченной ответственностью первые примерно 130 лет, но в 2009 получила финансовую помощь от правительства, после того как стала банком. Понимание хвостовых рисков не отражено в литературе по принятию решений, но для практиков, рискующих своей шкурой, это повседневная работа. Рассматриваем уязвимый модуль, оцениваем, какой желателен для него срок службы и насколько ожидаемый срок службы сократится при многократном воздействии риска.

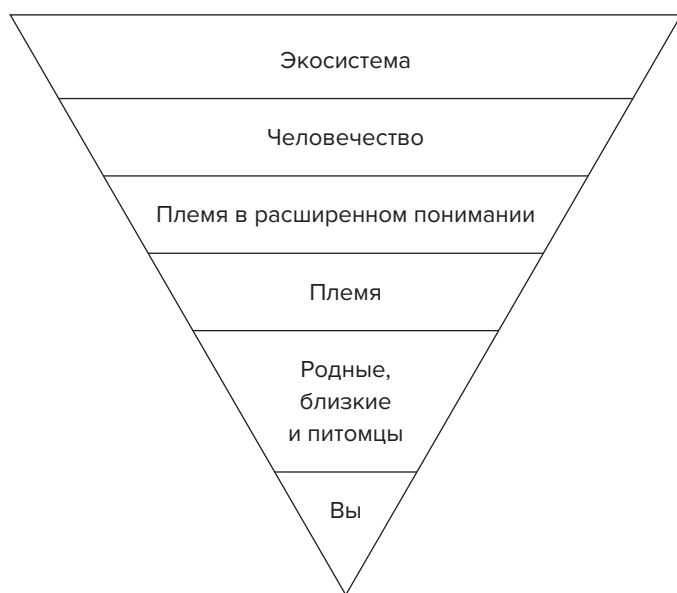
### Замечание 2: Психология принятия решений

Психологическая литература фокусируется на воздействии одного эпизода и опирается на узкое определение анализа «затраты — выгоды». Иногда анализ объявляет людей параноиками, преувеличивающими небольшие риски; авторам такого анализа не приходит в голову, что при малейшей терпимости к набиранию хвостовых рисков мы не выжили бы на протяжении миллионов лет<sup>2</sup>.

1 *Goldman Sachs* — инвестиционный конгломерат со штаб-квартирой в Нью-Йорке.

2 По молекулярно-генетическим данным, эволюционные пути человека и шимпанзе разошлись 5–12 млн лет назад.

Рассмотрим теперь разные уровни и почему системный риск попадает в другую категорию по сравнению с риском идиосинкразическим (индивидуальным). Посмотрим на перевернутую пирамиду на Рисунке 3.34: худший сценарий не тот, где индивидуум погибает. Хуже, если погибают ваши родные, близкие и питомцы. Хуже, если вы погибаете, а ваш заклятый враг выживает. В этих сценариях набирается большая потеря ожидаемой продолжительности жизни из-за одного хвостового терминального события.



**Рисунок 3.34:** Иерархия выживания. Объекты на верхних уровнях содержат больший объем ожидаемой продолжительности жизни, и хвостовые риски для этих объектов важнее. Объекты на нижних уровнях, вроде нас с вами, заменимы

Отсюда разные уровни. Самая большая угроза — гибель всей экосистемы. Принцип предосторожности строит структуру на идее риска для тех или иных модулей с учетом их ожидаемого срока службы.

Эргодичность в этом контексте означает перенос результатов анализа вероятностей по ансамблю на вероятности по времени. Если перенос невозможен, игнорируйте вероятность по ансамблю.

### 3.12. ЧТО ДЕЛАТЬ?

Резюмируя, отметим как первую задачу необходимость различать Медиокристан и Экстремистан, две области, которые практически не перекрываются. Если мы не научимся видеть разницу, наши попытки анализа будут бессмысленны. Во-вторых, если мы не будем отличать вероятность по времени (зависящую от пути) и по ансамблю (независимую от пути), наши попытки анализа будут бессмысленны.

Следующий этап нашего проекта *Incerto* — познакомиться с хрупкостью, робастностью и, наконец, антихрупкостью. Узнав о жирном хвосте, можно использовать эвристику и исследовать, как на отдачу от вложений в этой области влияют хвостовые события — какой ущерб они наносят одному модулю. Намного эффективнее изолировать ущерб от случайных событий, чем пытаться установить с достаточной точностью параметры их распределения (поскольку, как мы видели, ошибки вывода при толстых

хвостах огромны). Надежнее, мудрее, этичнее и эффективнее сфокусироваться на эвристике и политиках обнаружения, чем производить статистические параметры.

Хорошая новость — наше открытие того, что все хрупкое обязательно представлено воздействием, выпуклым вверх [223], аналогичным и равным отдаче от вложений в шорты, то есть отрицательному воздействию волатильности. Оно должно быть нелинейным. Ущерб должен расти с величиной события, вплоть до полного краха. Если я спрыгну с высоты 10 м, я получу повреждения серьезнее, чем если 10 раз спрыгну с высоты один метр. Это обязательное свойство хрупкости. Нам достаточно посмотреть, как ускоряется ущерб с уходом в хвост. Основываясь на этом свойстве, аналогичном опциону, мы построили эффективную эвристику для стресс-тестов [240].

В реальном мире нужны простые и работающие решения [109]; нам нужно произвести впечатление на нашего бухгалтера, а не на коллег-рецензентов. (Мой довод в последней книге серии *Incerto* — *Своя шкура на кону* — был тот, что в системах, где успехом считается суждение коллег, а не эволюционный отбор, происходит переусложнение и загнивание.) Чтобы выжить, нужны понятные методы, отвечающие нашей методологической интуиции.

Перенесем фокус на то, как обнаружить и измерить выпуклость вниз и вверх. Это узнать намного проще, чем вероятностное распределение.

## ДАЛЕЕ

В следующих трех главах в дискуссионной форме и без лишних формальностей изучается методическая интуиция в основе толстых хвостов. Вывод формул и строгие доказательства отложим на главы, излагающие, в адаптированной версии, статьи из научных журналов.

## ОДНОМЕРНЫЕ ЖИРНЫЕ ХВОСТЫ УРОВНЯ 1, С КОНЕЧНЫМИ МОМЕНТАМИ<sup>\*</sup>

Следующие две главы устроены так. Мы посмотрим на три уровня жирных хвостов, делая ударение скорее на интуиции и эвристике, чем на математических различиях, которые покажем позже, во время обсуждения предельных теорем. Эти три уровня называются так:

- Жирные хвосты начального уровня, с конечными моментами
- Субэкспоненциальный класс
- Класс степенного закона

О первом уровне мы будем говорить дольше всего, поскольку на его примере построим интуитивные понятия. Хотя этот уровень менее всего фигурирует в математических статьях (обычно под жирными хвостами понимают степенной закон и сходящееся к нему предельное поведение сумм), на этот нижний уровень больше всего полагаются в аналитических моделях и на практике. Мы без особого труда извлекаем непосредственные следствия жирнохвостости — например, производная функции дает хорошее представление о локальной чувствительности. Как трейдер, автор сумел получить максимальный эффект от жирнохвостости, пользуясь простенькой эвристикой при усреднении цен опциона по двум волатильностям; этой эвристики при всей ее простоте оказалось достаточно.

### 4.1. ПРОСТАЯ ЭВРИСТИКА, КАК СОЗДАВАТЬ СЛЕГКА ЖИРНЫЕ ХВОСТЫ

Пара напоминаний о выпуклости вниз и неравенстве Йенсена.

Пусть  $\mathcal{A}$  — выпуклое множество в векторном пространстве над  $\mathbb{R}$ , и пусть  $\varphi: \mathcal{A} \rightarrow \mathbb{R}$  — функция;  $\varphi$  называется выпуклой вниз, если  $\forall x_1, x_2 \in \mathcal{A}, \forall t \in [0, 1]$ :

$$\varphi(tx_1 + (1-t)x_2) \leq t\varphi(x_1) + (1-t)\varphi(x_2)^1.$$

<sup>1</sup> Это определение можно прочесть так. Пусть  $\varphi(x)$  — высота рельефа в каждой точке  $x$  на физической карте. Эта функция называется выпуклой вниз на участке  $\mathcal{A}$ , если автострада — прямолинейная и, возможно, наклонная — между любыми точками рельефа  $(x_1, \varphi(x_1))$  и  $(x_2, \varphi(x_2))$  в пределах участка  $\mathcal{A}$ , прой-



Для произвольной случайной величины  $X$  и выпуклой вниз функции  $\varphi(\cdot)$  согласно неравенству Йенсена [135]:

$$\varphi(\mathbb{E} X) \leq \mathbb{E} \varphi(X).$$

### Замечание 3: Жирные хвосты и неравенство Йенсена

Для гауссова распределения (и членов его семейства распределений с коэффициентами сдвига и масштаба<sup>1</sup>) хвосты вероятностного распределения выпуклы вниз «в масштабе распределения», в данном случае — при отклонении, равном среднеквадратическому отклонению распределения  $\sigma$  (оно же корень из дисперсии). Это дает возможность построить более жирные хвосты так: «стохастизировать» среднеквадратическое отклонение или дисперсию и посмотреть, изменилось ли вероятностное распределение согласно неравенству Йенсена.

Гетероскедастичность — это общий технический термин, которым при анализе временного ряда характеризуют процесс с непостоянным масштабом<sup>2</sup>. Согласно нашему методу нужно «стохастизировать», то есть пробовать изменять дисперсию или среднеквадратическое отклонение<sup>3</sup> распределения так, чтобы сохранять среднее.

Заметим, что, вообще говоря, и для процесса со сколь угодно жирным хвостом, даже с хвостом по степенному закону, наблюдаемую выборку (то есть конечное число дискретных наблюдений) можно описать как гауссов процесс с непостоянной дисперсией, процесс с переключением режима или сочетание гауссова процесса и нескольких скачков случайной величины; исключения составляют процессы со скачками равной величины, смотрите резюме в [174]<sup>4</sup>.

---

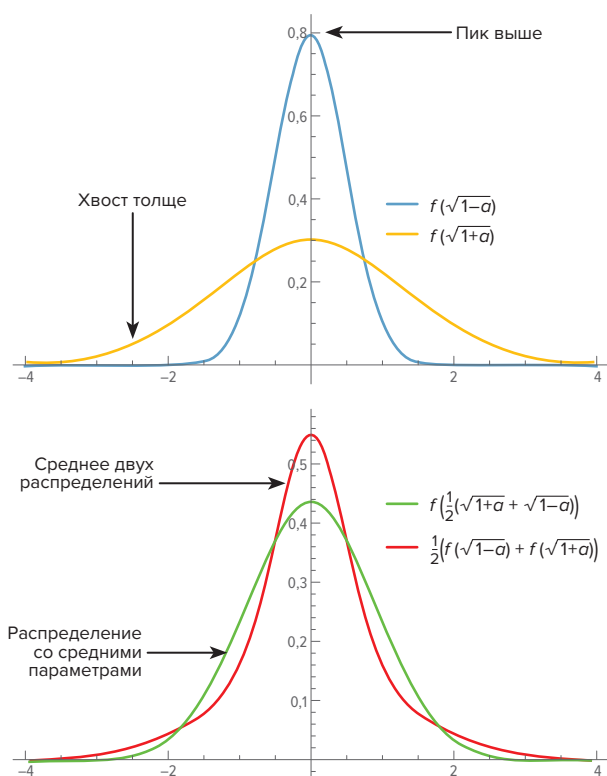
дет не ниже рельефа; пусть с мостами, но без туннелей. Пусть автомобиль проезжает по автострате за промежуток времени  $t \in [0, 1]$  на постоянной скорости; тогда положение автомобиля на карте изменяется от  $x_1$  до  $x_2$  по закону  $x(t) = tx_1 + (1-t)x_2$ . Высота рельефа в точке  $x(t)$  изменяется от  $\varphi(x_1)$  до  $\varphi(x_2)$  по закону  $\varphi(tx_1 + (1-t)x_2)$  и во все моменты времени  $t \in [0, 1]$  остается не выше автомобиля, который на прямой автострате находится на высоте, изменяющейся от  $\varphi(x_1)$  до  $\varphi(x_2)$  по закону  $t\varphi(x_1) + (1-t)\varphi(x_2)$ . Далее в этой книге критерий выпуклости вниз чаще применяется в одномерном случае, когда вместо карты рассматривается профиль рельефа,  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ , и участок  $\mathcal{A}$  представляет собой отрезок числовой прямой  $\mathbb{R}$ , на котором профиль рельефа изображается графиком функции и автострада — хордой этого графика.

- 1 Семейство распределений с коэффициентами сдвига и масштаба (*location-scale family*) — множество распределений всех случайных величин, полученных из данной случайной величины  $X$  линейным преобразованием; другими словами, множество распределений всех случайных величин вида  $aX + b$ , где  $a > 0$  (это коэффициент масштаба, изменяющий среднеквадратическое отклонение) и  $b$  любое (это коэффициент положения, изменяющий среднее).
- 2 Другими словами, с непостоянным среднеквадратическим отклонением.
- 3 На жаргоне финансовых математиков «волатильность» приписывается среднеквадратическому отклонению, хотя обычно «стохастической волатильностью» называют стохастическую дисперсию. — *Прим. автора.*
- 4 Но и для такого процесса скачки можно смоделировать как переключение гауссова процесса в режим с низкой дисперсией, большим средним и низкой вероятностью переключения в этот режим. Таким образом, формально даже скачки распределения Пуассона представимы как смесь гауссовых распределений. — *Прим. автора.*

Описанный ниже метод позволит смоделировать процессы со слегка ужирненными хвостами и заодно, в главке 4.3, ответить на великий вопрос: где начинаются хвосты?

Обозначим  $f(\sqrt{\sigma^2}, x)$  плотность нормального распределения в точке  $x$  при математическом ожидании 0 и дисперсии  $\sigma^2$ .

Сравним распределение со средним корнем из дисперсии  $f\left(\frac{\sqrt{1-a} + \sqrt{1+a}}{2}, x\right)$  и среднее от распределений с дисперсиями  $1 \pm a$ ,  $\frac{f(\sqrt{1-a}, x) + f(\sqrt{1+a}, x)}{2}$ ; различия будут отражать выпуклость согласно неравенству Йенсена. Мы сохранили постоянной среднюю дисперсию  $\sigma^2 = 1$ , хотя метод сработал бы и в том случае, если мы заботились вместо этого о постоянстве среднеквадратического отклонения  $\sigma$ . Среди математиков давно идет борьба традиций добиваться постоянной дисперсии или добиваться постоянного среднеквадратического отклонения, хотя (1) оба способа работают, лишь бы придерживаться чего-то одного, и (2) для целей нашей иллюстрации разницы не будет никакой.



**Рисунок 4.1:** Как внесение случайной волатильности<sup>1</sup> создает более жирные хвосты благодаря выпуклости вниз тех частей распределения плотности вероятности, где отклонение от нуля примерно равно масштабу распределения

<sup>1</sup> На всех графиках использовано  $a = 0,75$ , то есть сверху сравниваются две гауссианы со среднеквадратическим отклонением  $\sqrt{1-a} = 0,5$  и  $\sqrt{1+a} \approx 1,32$ , а внизу сравнивается гауссиана с усредненным среднеквадратическим отклонением  $\frac{\sqrt{1-a} + \sqrt{1+a}}{2} \approx 0,91$  (зеленая кривая) и стохастизированное распределение, возникающее при усреднении двух гауссиан сверху (красная кривая).

Поскольку при жирных хвостах моменты высокого порядка растут, даже когда низшие моменты остаются те же, мы сможем повысить жирнохвостость (судя по четвертому моменту), сохранив неизменными два или три низших момента<sup>1</sup>.

#### 4.1.1. Эвристика, сохраняющая дисперсию

Попробуем так «стохастизировать» дисперсию распределения, чтобы сохранять постоянное  $\mathbb{E}X^2$ , но добиться большего  $\mathbb{E}X^4$ . За величиной  $\mathbb{E}X^4$  проще следить, чем за дисперсией случайной величины  $\mathbb{E}X^2$ , то есть центральной величиной,  $\mathbb{E}(X^2 - \mathbb{E}X^2)^2$ , и мы сфокусируемся на простой величине, пока центральность не важна. Позже мы проследим за «стохастизацией» подробнее, в дальнейших разделах этой главы.

Эффективная эвристика, поддерживающая интуицию при попытках ужирнить хвосты, состоит в том, чтобы воспроизвести случайную величину опять со средним значением 0 и с той же дисперсией, но применив следующий прием ужирнения хвоста: рассмотреть случайную величину, которая с вероятностью  $p = \frac{1}{2}$  следует распределению  $\mathcal{N}(0, \sigma\sqrt{1-a})$  и с вероятностью  $p = \frac{1}{2}$  следует распределению  $\mathcal{N}(0, \sigma\sqrt{1+a})$ , где  $0 \leq a < 1$ <sup>2</sup>.

Характеристическая функция<sup>3</sup> будет такая:

$$\varphi(t, a) = \frac{1}{2} e^{-\frac{1+a}{2} a \sigma^2 t^2} (1 + e^{a \sigma^2 t^2}). \quad (4.1)$$

Нечетные моменты нулевые. Второй момент сохранен, поскольку

$$M(2) = (-i)^2 \frac{\partial^2 \varphi(t)}{\partial t^2} \Big|_{t=0} = \sigma^2, \quad (4.2)$$

а четвертый момент оказывается

$$M(4) = (-i)^4 \frac{\partial^4 \varphi(t)}{\partial t^4} \Big|_{t=0} = 3(a^2 + 1)\sigma^4, \quad (4.3)$$

что дает традиционный эксцесс (без вычитания 3 ради сравнения с гауссовым) равный  $3(a^2 + 1)$ . Таким образом, всякий эксцесс можно представить через «предположительное» значение  $a$ . Грубо говоря,  $a$  — это среднее отклонение стохастизированного параметра дисперсии, «волатильность волатильности»; в формулах будем обозначать  $V_{vol}$ .

**Ограничения простой эвристики.** Данный эвристический прием полезен для первого интуитивного понимания, но его возможности ограничены: поднять экс-

1 Повторим сказанное в прошлой главе: в литературе иногда тщатся различать «жирный хвост» и «тяжелый хвост», резервируя первый исключительно за степенным законом, а второй — за субэкспоненциальным распределением (о котором поговорим позже). Безнадега. Мы называем «жирным хвостом» просто всякое распределение, эксцесс которого больше, чем гауссов, или не определен. Такое употребление слов работает для практиков, имеющих дело с жирными хвостами, то есть трейдеров, торгующих опционами, и удобно при обсуждении «ужирнения хвостов», как мы увидим далее в этом разделе. — *Прим. автора.*

2 Здесь в  $\mathcal{N}(., .)$  указаны параметры среднее и среднеквадратическое отклонение, а не так, как в справочном Разделе 2.1, где параметры нормального распределения — среднее и дисперсия.

3 Заметим, при среднем 0 нет никакой разницы между характеристической функцией и производящей функцией моментов; это свойство пригодится нам потом, в более технических главах. — *Прим. автора.*

цесс удастся максимум вдвое<sup>1</sup> по сравнению с гауссовым. Тем самым ценность этого метода в основном педагогическая, познакомиться с тем, как работает выпуклость вниз. В Разделе 4.1.2 будет представлена более развитая техника.

#### Замечание 4: Пики

*Как показывает Рисунок 4.4, жирные хвосты проявляются более высокими пиками — наблюдения концентрируются в центре распределения.*

Многие этого не понимают.

#### 4.1.2. Ужирнение хвостов при помощи асимметричной дисперсии

Можно следующим образом усовершенствовать эвристику получения жирных хвостов, описанную в главке 4.1 (ограниченную эксцессом только вдвое больше, чем гауссов). Будем переключаться между гауссовыми распределениями с дисперсией:

$$\begin{cases} \sigma^2(1+a), & \text{с вероятностью } p \\ \sigma^2(1+b), & \text{с вероятностью } 1-p, \end{cases} \quad (4.4)$$

где  $p \in [0, 1)$  и  $b = -a \frac{p}{1-p}$ ; это даст характеристическую функцию

$$\varphi(t, a) = pe^{-\frac{1+a}{2}\sigma^2 t^2} - (p-1)e^{-\frac{ap+p-1}{2(p-1)}\sigma^2 t^2},$$

с эксцессом  $\frac{3((1-a^2)p-1)}{p-1}$ , что позволяет регулировать контраст между состояниями

и повышать эксцесс, сохраняя дисперсию<sup>2</sup>.

Так, при  $p = \frac{1}{1000}$  и соответствующем максимально допустимом<sup>3</sup>  $a = 999$  можно добиться эксцесса около 3000.

Эта эвристика отлично приближает действие, которое оказывает на вероятности логнормальное взвешивание с характеристической функцией

$$\varphi(t, V) = \int_0^\infty e^{-\frac{t^2 v}{2} - \frac{\left(\ln v - \ln V + \frac{V}{2v}\right)^2}{2V_v^2}} \frac{dv}{\sqrt{2\pi v} V_v}, \quad (4.5)$$

где  $v$  — дисперсия и  $V_v$  — дисперсия второго порядка, так называемая волатильность волатильности. Благодаря интегрированию по частям преобразование Фурье позво-

<sup>1</sup> Чтобы смоделировать более жирный хвост, как при степенном законе  $f_X(x) \sim x^{-\alpha-1}$  с  $\alpha \leq 4$ , нужно добиться бесконечно большого эксцесса.

<sup>2</sup> Средневзвешенная дисперсия получится  $p\sigma^2(1+a) + (1-p)\sigma^2\left(1 - \frac{ap}{1-p}\right) = \sigma^2$ .

<sup>3</sup> Подразумевается, что в формуле (4.4) во втором состоянии, с вероятностью  $1-p$ , дисперсия  $\sigma^2(1+b)$  должна быть неотрицательной. С учетом  $b = -a \frac{p}{1-p}$  требуется  $a \leq \frac{1-p}{p}$ , и при  $p = \frac{1}{1000}$  требуется  $a \leq 999$ .

ляет получить все разновидности функции платежа (см. Гезерал [102]). С другой стороны, можно поправить отсутствие аналитической формулы и следующим способом, при помощи более удобных для анализа распределений дисперсии.

**Гамма-дисперсия.** Гамма-распределение<sup>1</sup> выгодно применить к дисперсии гауссова распределения, чтобы быстро получить полное распределение дисперсии и выйти за узкие рамки эвристического нащупывания [36]. Его удобнее анализировать, чем логнормальное.

Примем, что дисперсия гауссова распределения следует гамма-распределению

$$\Gamma_a(v) = \frac{v^{a-1} \left(\frac{V}{a}\right)^{-a} e^{-\frac{av}{V}}}{\Gamma(a)}$$

со средним  $V$  и среднеквадратическим отклонением  $\frac{V}{\sqrt{a}}$ . На Рисунке 4.2 показано соответствие логнормальному с теми же первыми двумя моментами, где мы откалибровывали логнормальное распределение под математическое ожидание  $\frac{1}{2} \ln \frac{aV^3}{aV+1}$  и среднеквадратическое отклонение  $\sqrt{-\ln \frac{aV}{aV+1}}$ .

Итоговое распределение получается такое (снова предполагая, что средняя стохастизированная дисперсия та же, что была фиксированная):

$$f_{a,V}(x) = \int_0^\infty \frac{e^{-\frac{(x-\mu)^2}{2v}}}{\sqrt{2\pi}\sqrt{v}} \Gamma_a(v) dv, \quad (4.6)$$

откуда удастся получить явный вид

<sup>1</sup> Гамма-распределение с коэффициентом формы  $a > 0$  и коэффициентом масштаба  $b > 0$  (или со средним  $V = \frac{a}{b}$ ) в частном случае  $a \in \mathbb{N}$  присуще случайному времени  $X > 0$  ожидания  $a$  событий, когда события происходят случайно, независимо друг от друга и со средней частотой  $b^{-1} = \frac{V}{a}$ . В этом частном случае гамма-распределение называется также распределением Эрланга; плотность вероятности  $f_X(X) = \frac{x^{a-1} e^{-bx} b^a}{(a-1)!} = \frac{x^{a-1} e^{-\frac{ax}{V}} \left(\frac{V}{a}\right)^{-a}}{(a-1)!}$ ,

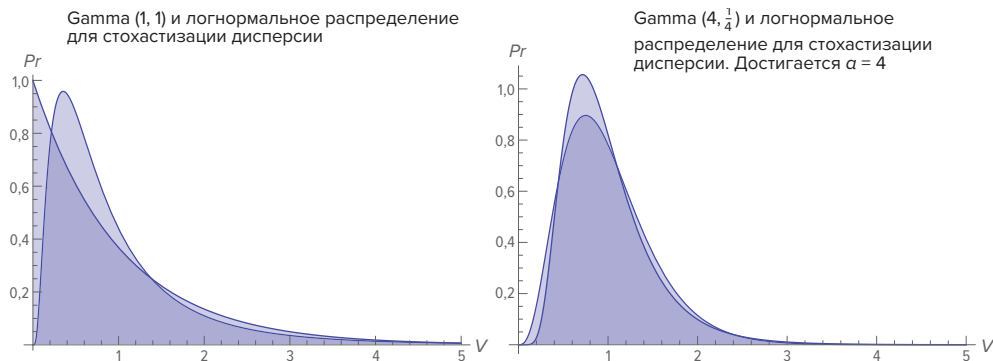
$x > 0$ . В общем случае  $a \in \mathbb{R}^+$ ; плотность вероятности  $f_X(X) = \frac{x^{a-1} e^{-bx} b^a}{\Gamma(a)} = \frac{x^{a-1} e^{-\frac{ax}{V}} \left(\frac{V}{a}\right)^{-a}}{\Gamma(a)}$ ,  $x > 0$ . Как в частном

случае  $a \in \mathbb{N}$ , так и в общем случае  $a \in \mathbb{R}_{>0}$  среднее  $\frac{a}{b} = V$ , дисперсия  $\frac{a}{b^2} = \frac{V^2}{a}$ , коэффициент асимметрии  $\frac{2}{\sqrt{a}}$ , эксцесс  $3 + \frac{6}{a}$ . Кумулянты  $\kappa_1^1 = \frac{a}{b} = V$ ,  $\kappa_2^2 = \frac{a}{b^2} = \frac{V^2}{a}$ ,  $\kappa_3^3 = \frac{2a}{b^3} = \frac{2V^3}{a^2}$ ,  $\kappa_4^4 = \frac{6a}{b^4} = \frac{6V^4}{a^3}$ , ...,  $\kappa_p^p = \frac{a(p-1)!}{b^p} = \frac{V^p(p-1)!}{a^{p-1}}$ .

Для суммы  $n$  н. о. р.  $X_1 + \dots + X_n$  кумулянты  $\kappa_n^1 = \frac{an}{b} = nV$ ,  $\kappa_n^2 = \frac{an}{b^2} = \frac{nV^2}{a}$ ,  $\kappa_n^3 = \frac{2an}{b^3} = \frac{2nV^3}{a^2}$ ,  $\kappa_n^4 = \frac{6an}{b^4} = \frac{6nV^4}{a^3}$ , ...,  $\kappa_n^p = \frac{an(p-1)!}{b^p} = \frac{nV^p(p-1)!}{a^{p-1}}$ ,  $\lim_{p \rightarrow \infty} \kappa_n^p = \infty$ .

Характеристическая функция  $\psi_X(t) = \left(1 - \frac{it}{b}\right)^{-a}$ .

Случайную величину  $X \in \mathbb{R}^+$ , следующую гамма-распределению с коэффициентом формы  $a > 0$  и коэффициентом масштаба  $b > 0$ , принято обозначать как  $\text{Gamma}(a, b^{-1})$ ; этой традиции следуют надписи на графиках Рисунка 4.2.

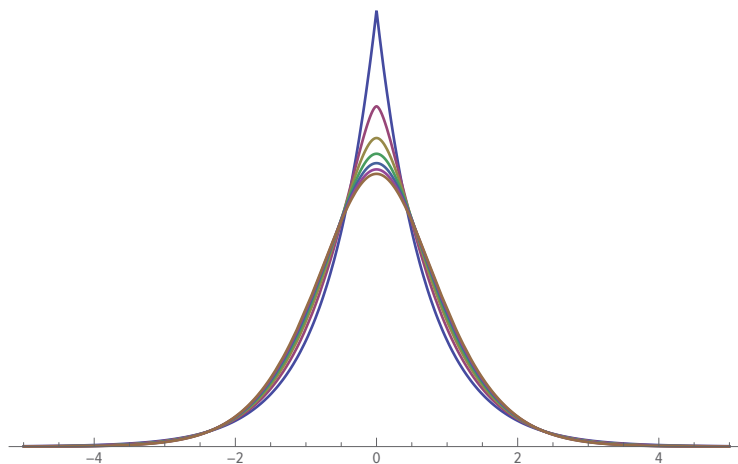


**Рисунок 4.2:** Чем можно стохастизировать дисперсию: гамма-распределением и логнормальным распределением с теми же математическим ожиданием и дисперсией

$$f_{a,V}(x) = \frac{2^{\frac{3}{4}-\frac{a}{2}} a^{\frac{a}{2}+\frac{1}{4}} |x-\mu|^{a-\frac{1}{2}} K_{a-\frac{1}{2}}\left(\frac{\sqrt{2}\sqrt{a}|x-\mu|}{\sqrt{V}}\right)}{\sqrt{\pi}\Gamma(a)}, \quad (4.7)$$

где  $K_a(z)$  —  $K$ -функция Бесселя<sup>1</sup>, удовлетворяющая дифференциальному уравнению  $-y(\alpha^2 + z^2) + z^2 y'' + zy' = 0$ <sup>2</sup>.

Стохастизация гауссова распределения  
гамма-распределением дисперсии



**Рисунок 4.3:** Стохастизация дисперсии гамма-распределением при различных  $\alpha$  в формуле 4.7

Углубимся теперь в различные формы стохастической волатильности.

1 В русской литературе называется также модифицированной функцией Бесселя второго рода, или функцией Макдональда.  $K$ -функции Бесселя экспоненциально убывают, и чем больше параметр  $\alpha$ , тем позже начинается хвост функции  $K_a(z) = \int_0^\infty e^{-z \operatorname{ch} t} \operatorname{ch} \alpha t \, dt$ .

2 Как одно из двух решений этого дифференциального уравнения (это т. н. модифицированное уравнение Бесселя). Еще одно решение обозначают  $I_a(z)$ ; в русской литературе называют модифицированной функцией Бесселя первого рода, или функцией Инфельда.  $I$ -функции Бесселя экспоненциально возрастают.

## 4.2. СПОСОБНА ЛИ СТОХАСТИЧЕСКАЯ ВОЛАТИЛЬНОСТЬ ГЕНЕРИРОВАТЬ СТЕПЕННЫЕ ЗАКОНЫ?

Мы еще не дали определение степенным законам; примем пока как критерий то, что хотя бы один из статистических моментов должен оказаться бесконечным.

И вот ответ на вопрос: нет, не способна, до тех пор, пока мы стохастизируем  $\sigma$  или  $\sigma^2$ , но будет способна, когда стохастизируем  $\frac{1}{\sigma}$  или  $\frac{1}{\sigma^2}$ .

Пусть исходное распределение гауссово, то есть случайная величина  $X \sim \mathcal{N}(\mu, \sigma)$ . Есть несколько способов стохастизировать масштаб  $\sigma$ . Заметим, что  $\sigma$  может быть только положительной; значит, для ее стохастизации требуется однохвостое распределение.

- Можем заставить  $\sigma^2$  (вариант:  $\sigma$ ) следовать логнормальному распределению. Из него не получить аналитические решения, но можно узнать статистические моменты сгенерированного распределения и убедиться, что его хвост не следует степенному закону.
- Можем заставить  $\sigma^2$  (вариант:  $\sigma$ ) следовать гамма-распределению. Можно получить аналитические решения, как мы видели в примере выше, формула 4.7.
- Можем заставить  $\frac{1}{\sigma^2}$  — параметр точности — следовать гамма-распределению.
- Можем заставить  $\frac{1}{\sigma^2}$  следовать логнормальному распределению.

Результаты, приведенные в Таблице 4.1, происходят из следующих простых свойств функций плотности и операторов математического ожидания. Пусть  $X$  — случайная величина с функцией плотности вероятности  $f(\cdot)$  в семействе распределений с коэффициентами сдвига и масштаба,  $\lambda$  — случайная величина с функцией плотности вероятности  $g(\cdot)$ ;  $X$  и  $\lambda$  предполагаются независимыми. В силу стандартных теорем (о преобразованиях Меллина) моменты порядка  $p$  для произведения  $X\lambda$  и для отношения  $\frac{X}{\lambda}$  представимы в виде:

$$\mathbb{E} (X\lambda)^p = \mathbb{E} X^p \mathbb{E} \lambda^p$$

и

$$\mathbb{E} \left( \frac{X}{\lambda} \right)^p = \mathbb{E} \frac{1}{\lambda^p} \mathbb{E} X^p.$$

Заметим, что в силу свойств семейства распределений с коэффициентами положения и масштаба

$$\frac{1}{\lambda} f_X(x) = f_X\left(\frac{x}{\lambda}\right),$$

поэтому, в частности, если случайная величина  $X$  следует нормальному распределению  $X \sim \mathcal{N}(0, \sigma)$ , то  $\frac{X}{\sigma} \sim \mathcal{N}(0, 1)$ .

**Таблица 4.1:** Преобразования, стохастизирующие волатильность. По плотности после преобразования  $\frac{1}{x}$  или  $\frac{1}{\sqrt{x}}$  можно понять, удалось ли получить степенной закон.  $\mathcal{LN}$ ,  $\mathcal{N}$ ,  $\mathcal{G}$  и  $\mathcal{P}$  — соответственно логнормальное, нормальное, гамма-распределение и распределение Парето.

Распределение	$f(x)$	$f\left(\frac{1}{x}\right)$	$f\left(\frac{1}{\sqrt{x}}\right)$
$\mathcal{LN}(\mu, \sigma)$	$\frac{e^{-\frac{(\mu - \ln x)^2}{2\sigma^2}}}{\sqrt{2\pi} \sigma x}$	$\frac{e^{-\frac{(\mu + \ln x)^2}{2\sigma^2}}}{\sqrt{2\pi} \sigma x}$	$\frac{\sqrt{2}}{\pi} e^{-\frac{(\mu + 2 \ln x)^2}{2\sigma^2}}}{\sigma x}$
$\mathcal{N}(\mu, \sigma)$	$\frac{e^{-\frac{(\mu - x)^2}{2\sigma^2}}}{\sqrt{2\pi} \sigma}$	$\frac{e^{-\frac{\left(\mu - \frac{1}{x}\right)^2}{2\sigma^2}}}{\sqrt{2\pi} \sigma x^2}$	$\frac{\sqrt{2}}{\pi} e^{-\frac{\left(\mu - \frac{1}{x^2}\right)^2}{2\sigma^2}}}{\sigma x^3}$
$\mathcal{G}(a, b)$	$\frac{b^{-a} x^{a-1} e^{-\frac{x}{b}}}{\Gamma(a)}$	$\frac{b^{-a} x^{-a-1} e^{-\frac{1}{bx}}}{\Gamma(a)}$	$\frac{2b^{-a} x^{-2a-1} e^{-\frac{1}{bx^2}}}{\Gamma(a)}$
$\mathcal{P}(1, \alpha)$	$\alpha x^{-\alpha-1}$	$\alpha x^{\alpha-1}$	$2\alpha x^{2\alpha-1}$

**Таблица 4.2:**  $p$ -тые моменты возможных распределений для дисперсии

Распределение	$\mathbb{E} X^p$	$\mathbb{E} \frac{1}{X_p}$	$\mathbb{E} \left( \frac{1}{\sqrt{X}} \right)^p$
$\mathcal{LN}(\mu, \sigma)$	$e^{\mu p + \frac{p^2 \sigma^2}{2}}$	$e^{\frac{p(p\sigma^2 - 2\mu)}{2}}$	$e^{\frac{p(p\sigma^2 - 4\mu)}{8}}$
$\mathcal{G}(a, b)$	$b^p(a)_p$	$\frac{(-1)^p b^{-p}}{(1-a)_p}, p < a$	Безнадега
$\mathcal{P}(1, \alpha)$	$\frac{\alpha}{\alpha - p}, p < \alpha$	$\frac{\alpha}{\alpha + p}$	$\frac{2\alpha}{2\alpha + p}$

### 4.3. ТУЛОВИЩЕ, ПЛЕЧИ И ХВОСТЫ

Где начинаются хвосты?

Примем, что хвосты начинаются на уровне выпуклости вниз сегмента вероятностного распределения, отвечающем масштабу распределения, — другими словами, в сегменте, на который влияет стохастизация.

#### 4.3.1. Точки перехода и туннельный эффект

На Рисунке 4.4 обратите внимание на переходные зоны, не зависящие от  $a$ . У так называемых колоколообразных распределений выпуклость направлена вниз-вверх-вниз, это то, что называется «квазивыпуклость вверх».

Пусть  $X$  — случайная величина, у которой плотность вероятности  $p(x)$  принадлежит некоторому общему классу  $p_\sigma$  всех унимодальных<sup>1</sup> непрерывных функций с одним параметром, определенных на носителе  $D \subseteq \mathbb{R}$  параметром масштаба  $\sigma$ . Пусть  $p(\cdot)$

<sup>1</sup> В русской литературе унимодальность называется также одновершинностью (*unimodality*). Имеется в виду функция одного аргумента, возрастающая до точки максимума и убывающая после точки макси-



квазивыпукла вверх на области определения, но не является выпуклой вниз или вверх. Плотность вероятности  $p(x)$  удовлетворяет для всех  $\varepsilon > 0$  неравенству  $p(x) \geq p(x + \varepsilon)$  при  $x > x^*$  и неравенству  $p(x) \geq p(x - \varepsilon)$  при  $x < x^*$ , где  $x^* = \operatorname{argmax}_x p(x)$ <sup>1</sup>.

Квазивыпуклость вверх функции  $p$  означает, что для любых двух точек  $x, y$  и любой точки между ними  $\omega x + (1 - \omega)y$ , где  $\omega \in [0, 1]$ ,

$$p(\omega x + (1 - \omega)y) \geq \min(p(x), p(y)).$$

Рассмотрим две возможности.

**А.** Случайная величина двуххвостая, то есть определена на множестве  $D = (-\infty, \infty)$ ; тогда, обозначив  $p^\delta(x) \triangleq \frac{p(x, \sqrt{\sigma^2 + \delta}) + p(x, \sqrt{\sigma^2 - \delta})}{2} \frac{p_{\sigma+\delta}(x) + p_{\sigma-\delta}(x)}{2}$ , находим, что

1) существует «туннель под вершиной», такой интервал  $A_T = (a_2, a_3)$ , что возмущенное на величину  $\delta$  вероятностное распределение  $p^\delta(x) \geq p(x)$  при  $x \in (a_2, a_3)$ ;

2) существуют два «периферийных туннеля», такие «хвосты», что  $p^\delta(x) \geq p(x)$  при  $x \in (-\infty, a_1)$  или  $x \in (a_4, \infty)$ ;

3) существуют два промежуточных моста, такие «плечи»<sup>2</sup>, где  $p^\delta(x) \leq p(x)$  при  $x \in (a_1, a_2)$  или  $x \in (a_3, a_4)$ .

**Проблема черного лебедя.** Как мы видели, дело не только в том, что события в хвостах распределения имеют значение, случаются, играют важную роль и т. п. Суть в том, что эти события играют главную роль, а их вероятности не то что трудно вычислить, их даже оценить не удастся. Вывод тот, что не обязательно винить за черных лебедей жирные хвосты; проблема может заключаться в неполной оценке хвостовых событий.

Найдем  $A = \{a_i\}$  как множество решений уравнения,  $A = \left\{ x \left\{ \frac{\partial^2 p(x, \sigma)}{\partial \sigma^2} = 0 \right\} \right\}$ .

Для гауссова распределения с параметрами  $(\mu, \sigma)$  уравнение примет вид:<sup>3</sup>

$$\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}} (2\sigma^4 - 5\sigma^2(x-\mu)^2 + (x-\mu)^4)}{\sqrt{2\pi}\sigma^7} = 0,$$

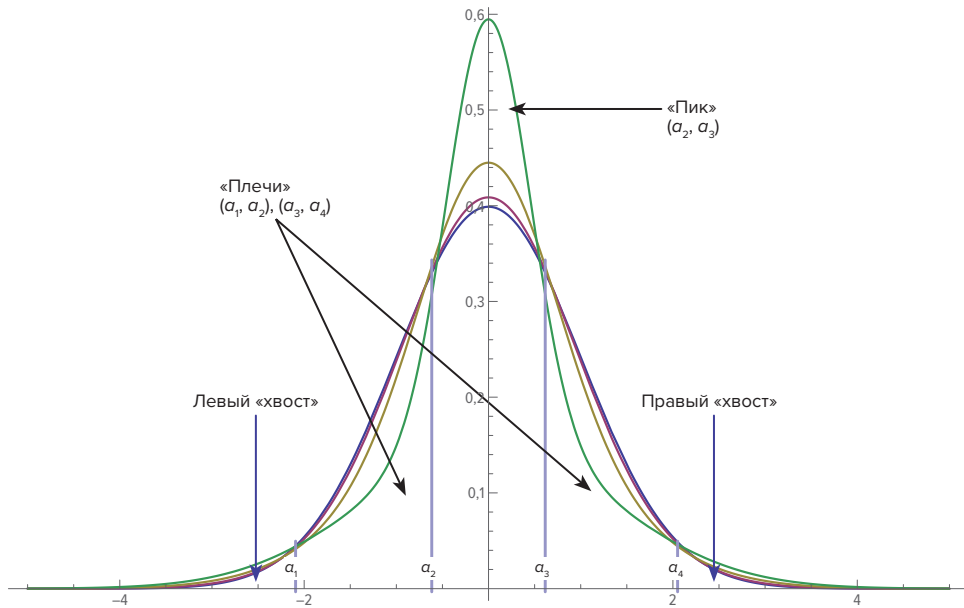
откуда точки перехода

мума; другие, локальные максимумы запрещены. Для функции одного аргумента унимодальность равносильна квазивыпуклости вверх.

1 Обозначение  $\operatorname{argmax}_x f(x)$  (от *arguments of the maxima*, аргументы максимизации) в русской литературе встречается также в виде  $x_{\max}$  — точки максимума функции  $f(x)$ . В данном случае  $x^*$  — положение горба.

2 В русской литературе плечи (*shoulders*) называются также склонами. Антитуннели можно также назвать мостами.

3 В самом деле, если допустить  $\mu \neq 0$ , то плотность  $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ .



**Рисунок 4.4:** Где начинаются хвосты? Чтобы сделать хвосты жирнее, стохастизируем гауссово распределение, сделав переменным параметр масштаба  $\sigma$ . В одних местах плотность вероятности повысится, в других понизится. Средние по величине отклонения от нуля станут менее вероятными, а хвостовые и малые — более. Обратим внимание на точки пересечения, от  $a_1$  до  $a_4$ . Собственно «хвосты» простираются от  $a_4$  вправо и от  $a_1$  влево\*

\* Изображено семейство  $\frac{p(x, \sqrt{1+\delta}) + p(x, \sqrt{1-\delta})}{2}$  при  $\delta = 0, 0,2, 0,5$  и  $0,8$ ; синяя кривая — гауссиана, без стохастизации; зеленая кривая — при  $\delta = 0,8$ . — Прим. перев.

$$\begin{aligned} \{a_1, a_2, a_3, a_4\} = \\ = \left\{ \mu - \sqrt{\frac{5+\sqrt{17}}{2}}\sigma, \mu - \sqrt{\frac{5-\sqrt{17}}{2}}\sigma, \mu + \sqrt{\frac{5-\sqrt{17}}{2}}\sigma, \mu + \sqrt{\frac{5+\sqrt{17}}{2}}\sigma \right\}. \end{aligned} \quad (4.8)$$

На Рисунке 4.4 эти точки<sup>1</sup> показаны численно, как  $\{-2,13\sigma, -0,66\sigma, 0,66\sigma, 2,13\sigma\}$ .

Для симметричного степенного закона (как мы увидим ниже) можно использовать  $t$ -распределение Стьюдента с масштабом  $s$  и показателем хвоста  $\alpha$ :

$$p(x) \triangleq \frac{\left( \frac{\alpha}{\alpha + \frac{x^2}{s^2}} \right)^{\frac{\alpha+1}{2}}}{\sqrt{\alpha} s B\left(\frac{\alpha}{2}, \frac{1}{2}\right)},$$

где бета-функция  $B(., .)$  определяется как  $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 dt \, t^{a-1} (1-t)^{b-1}$ .

<sup>1</sup> При  $\mu = 0$ . Этим точкам отвечают пересечения гауссианы (синяя кривая) со слабо стохастизированным распределением (лиловая кривая,  $\delta = 0,2$ ).

Отсюда точки перехода следующие:

$$\{a_1, a_2, a_3, a_4\} = \left\{ -\sqrt{\frac{5\alpha + \sqrt{(\alpha+1)(17\alpha+1)} + 1}{2(\alpha-1)}} s, -\sqrt{\frac{5\alpha - \sqrt{(\alpha+1)(17\alpha+1)} + 1}{2(\alpha-1)}} s, \sqrt{\frac{5\alpha - \sqrt{(\alpha+1)(17\alpha+1)} + 1}{2(\alpha-1)}} s, \sqrt{\frac{5\alpha + \sqrt{(\alpha+1)(17\alpha+1)} + 1}{2(\alpha-1)}} s \right\}.$$

При «кубическом» t-распределении Стьюдента, т. е. при  $\alpha = 3$ :

$$\{a_1, a_2, a_3, a_4\} = \{-\sqrt{4+\sqrt{13}} s, -\sqrt{4-\sqrt{13}} s, \sqrt{4-\sqrt{13}} s, \sqrt{4+\sqrt{13}} s\}.$$

### Итак, где начинается хвост?

Для общего класса симметричных распределений по степенным законам хвост начинается в точке  $\pm \sqrt{\frac{5\alpha + \sqrt{(\alpha+1)(17\alpha+1)} + 1}{2(\alpha-1)}} s$ . В случае стохастизированного гауссова распределения  $\alpha$  бесконечно, а  $s$  — среднее квадратическое отклонение; хвост располагается между 2–3 среднее квадратических отклонениями. Это вытекает из нашего определения хвоста как промежутка, где распределение выпукло вниз к изменениям коэффициента масштаба.

Впрочем, на практике из-за эффектов малой выборки историческая оценка среднее квадратического отклонения занижается (как мы любим повторять, жирные хвосты усиливают эффекты малой выборки), так что хвостовые отклонения будут больше, чем от 2 до 3 среднее квадратических отклонений по выборке.

Мы можем доказать, что при  $\alpha \rightarrow \infty$  точки перехода превращаются в гауссовы. Например, точка  $a_1$ :

$$\lim_{\alpha \rightarrow \infty} \left( -\sqrt{\frac{5\alpha + \sqrt{(\alpha+1)(17\alpha+1)} + 1}{2(\alpha-1)}} s \right) = -\sqrt{\frac{5-\sqrt{17}}{2}} s.$$

**В.** Некоторые однохвостые распределения тоже колоколообразны, то есть выпуклы вниз-вверх-вниз, и для них по-прежнему существуют 4 точки перехода. Частный случай — логнормальное распределение:

$$\{a_1, a_2, a_3, a_4\} = \left\{ e^{\frac{1}{2}(2\mu - \sqrt{2}\sqrt{5+\sqrt{17}}\sigma)}, e^{\frac{1}{2}(2\mu - \sqrt{2}\sqrt{5-\sqrt{17}}\sigma)}, e^{\frac{1}{2}(2\mu + \sqrt{2}\sqrt{5-\sqrt{17}}\sigma)}, e^{\frac{1}{2}(2\mu + \sqrt{2}\sqrt{5+\sqrt{17}}\sigma)} \right\}.$$

**Стохастические параметры.** Проблема с эллиптическими распределениями в том, что они не отображаются в доходность ценных бумаг, поскольку ни в какой момент времени не имеют одной дисперсии; смотрите у Бушо и Шишпортика (2010) [42]. Когда масштабы распределений индивидуальных участников изменяются не в тандеме, резуль-

тирующее распределение уже не эллиптическое. На Рисунках 6.2–6.5 показан эффект от применения аналога методов стохастической волатильности, создающего более сильное возмущение, — метода стохастической корреляции. Вместо того чтобы вносить возмущение в корреляционную матрицу  $S$  как единое целое, подобно тому, что мы делаем в Разделе 6, начнем вносить возмущения в корреляции — и получим удивительный эффект.

#### 4.4. ЖИРНЫЕ ХВОСТЫ, СРЕДНЕЕ ОТКЛОНЕНИЕ И ПОВЫШАЮЩИЕСЯ НОРМЫ

Далее рассмотрим использование среднеквадратического отклонения и то, как его интерпретировать.

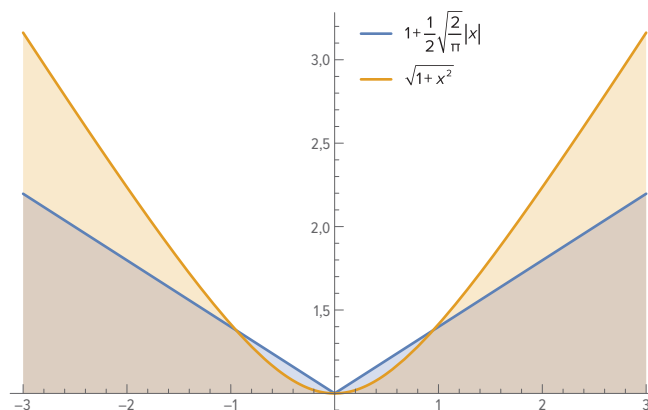
##### 4.4.1. Обычные ошибки

Сначала рассмотрим среднеквадратическое отклонение и дисперсию как свойства высших моментов. Но придется пояснить, что такое среднеквадратическое отклонение. Потому что распространившаяся путаница с жирными хвостами отравила и наше понимание среднеквадратического отклонения тоже.

Есть различие между среднеквадратическим отклонением и средним абсолютным отклонением, и это различие усиливается в случае жирных хвостов.

Примем для простоты, что среднее и медиана равны 0; тогда среднеквадратическое отклонение  $\sigma = \sqrt{\frac{1}{n} \sum x_i^2}$ , а среднее абсолютное отклонение  $\text{MAD} = \frac{1}{n} \sum |x_i|$ .

Рисунок 4.5 дает общее представление о том, как растет различие между этими статистическими моментами при больших средних отклонениях.



**Рисунок 4.5:** Сравним поведение взвешивающих функций  $\sqrt{K + x^2}$  и  $K + |x|$ . Оно расходится при больших значениях случайной величины  $x$ , показывая, почему при жирном хвосте среднеквадратическое отклонение расходится со средним абсолютным отклонением; это обобщается и на моменты более высоких порядков

Дэн Голдстейн и автор [113] ставили такой вопрос перед профессиональными инвесторами и студентами выпускных курсов по финансовому инжинирингу — то есть людьми, которые постоянно работают с рисками и отклонениями.

Акция (или фонд) имеет среднюю доходность 0 %. Она изменяется в абсолютном выражении в среднем на 1 % за день; рост в среднем составляет 1 %, падение в среднем составляет 1 %.

Это не значит, что рост всегда на 1 % — в некоторые дни рост 0,6 %, в другие дни 1,45 % и пр.

Допустим, что мы живем в гауссовом мире, где доходность (или процент изменения за день) надежно моделируется нормальным распределением. Допустим, что в году 256 рабочих дней. Каково среднеквадратичное отклонение доходности (то есть процента изменения), параметр  $\sigma_{Adj}$ , характеризующий волатильность в финансовых прикладных программах?

Каково среднеквадратическое отклонение за день?

Каково среднеквадратическое отклонение за год?

Как может видеть читатель, в условии задачи было описано среднее абсолютное отклонение, а найти требовалось отклонение среднеквадратическое. Ответы на задачу в подавляющем большинстве оказались ошибочными. Почти все указали отклонение за день 1 %. На самом деле гауссова переменная величина, которая за день изменяется в абсолютном выражении в среднем на 1 %, имеет среднеквадратическое отклонение выше процента, около 1,25 %. В выборочных распределениях должно быть еще больше, до 1,7 %. Про отклонение за год больше всего ответов было 16 % или около того, что составляет около 80 % от правильного ответа.

Профессиональные инвесторы для перехода от волатильности за день к волатильности за год умножали на  $\sqrt{256}$ , и это дало бы правильный результат, если бы применялось к правильной оценке волатильности за день.

Итак, испытуемые были склонны сообщать MAD, когда их спрашивали о среднеквадратическом отклонении. Когда профессионалы финансовых рынков, постоянно слышащие, как обсуждают волатильность, говорят о «стандартном отклонении», они имеют в виду не среднеквадратическое отклонение, а среднее абсолютное, из-за чего недооценивают отклонения на 20–40 %; для некоторых рынков недооценка может достигать до 90 %. И когда им указывали на ошибку, мало кто из них понимал, что ошибся. При всем при том на просьбу выписать формулу среднеквадратического отклонения они успешно рисовали квадратный корень из среднего квадрата отклонения. Некоторые удивлялись, первый раз услышав от нас о существовании MAD.

Почему это важно: сложилась ситуация, где руководители, принимающие решения, рассуждают о «волатильности», имея о ней смутное представление. В финансовой печати отмечаем, что в ряде мест ту же ошибку делает и журналист, пытаясь объяснить термин VIX (от *volatility index*, индекс волатильности). Даже на официальном сайте Министерства торговли США волатильность определена неверно.

Ошибка ведет к недооценке отклонений, потому что MAD, согласно неравенству Йенсена, меньше среднеквадратического отклонения (или равно ему).

**Как растет отношение этих параметров.** Для гауссовой случайной величины отношение  $\sim 1,25$ , а для жирных хвостов — больше.

**Пример.** Возьмем распределение с чрезвычайно жирным хвостом, в котором из  $n = 10^6$  наблюдений одно-единственное показало миллион, а все остальные — минус единицу:

$$X = \{-1, -1, \dots, -1, 10^6\}.$$

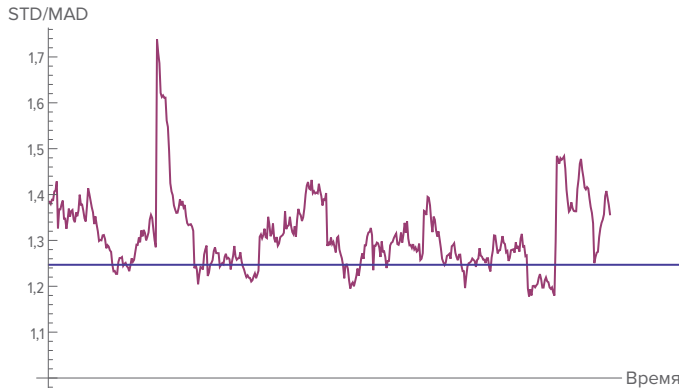
Среднее абсолютное отклонение

$$\text{MAD}(X) = 2.$$

Среднеквадратическое отклонение

$$\text{STD}(X) = 1000.$$

Отношение среднеквадратического отклонения к среднему абсолютному отклонению составило 500.



**Рисунок 4.6:** Отношение  $\text{STD}/\text{MAD}$  для ежедневной доходности S&P 500 за последние 47 лет по скользящему временному окну 1 месяц. На уровень  $\sqrt{\frac{\pi}{2}} \approx 1,253$  (как приближенное значение для гауссовых отклонений) можно смотреть как на точку отсечения жирных хвостов

#### 4.4.2. Немного аналитики

**Отношение при тонких хвостах.** Как полезную эвристику рассмотрим отношение  $h$ :

$$h = \frac{\sqrt{\mathbb{E} X^2}}{\mathbb{E} |X|},$$

где  $\mathbb{E}$  — оператор математического ожидания (применяемый к исследуемому вероятностному параметру), а  $X$  — центральная случайная величина, то есть  $\mathbb{E} X = 0$ ; это отношение растет с ростом жирности хвостов распределения. Пока речь об отношении для статистического момента второго порядка; в общем случае можно рассмотреть  $\frac{\sqrt{\mathbb{E} X^{1/p}}}{\mathbb{E} |X|}$ , где  $p > 1$  и для  $X$  существуют конечные статистические моменты до

$p$ -того, а наш случай будет при  $p = 2$ <sup>1</sup>.

Упрощая,  $X^p$  можно считать взвешивающим оператором, умножающим каждое слагаемое  $x_i$  на вес  $x_i^{p-1}$ ; этот вес велик при больших значениях  $x$  и мал при малых.

<sup>1</sup> Термин «бесконечный» несколько двусмыслен; лучше представить проблему как «неопределенный» момент, в том смысле, что он зависит от выборки и не воспроизводится за ее пределами. Заметим, что, например, у двухвостого распределения, то есть определенного на всей оси действительных чисел, может быть бесконечно большой момент четвертого порядка, но не третьего. — *Прим. автора.*

Эффект недооценки волатильности по среднему абсолютному отклонению порождается выпуклостью вниз разности между двумя функциями. Вторая функция,  $|X|$ , кусочно-линейная и не имеет выпуклости за пределами области вокруг нуля.

### Среднее абсолютное отклонение против среднеквадратического, подробнее

Почему в статистике предпочли STD, а не MAD? Рассказываю, добавляя аналитические выкладки, которых не вижу в доступной литературе. Читаем у Хьюбера [131]:

Был спор между Эддингтоном и Фишером, примерно в 1920 году, об относительных достоинствах  $dn$  (среднее абсолютное отклонение) и  $Sn$  (среднеквадратическое отклонение). Потом Фишер указал, что для **в точности нормальных** наблюдений  $Sn$  на 12 % эффективнее, чем  $dn$ , и, видимо, это решило исход спора (выделение мое. — *Авт*).

Давайте реконструируем выкладки Фишера и поймем, что он имел в виду.

Пусть  $n$  — число слагаемых.

Асимптотическую относительную эффективность (*Asymptotic Relative Efficiency*, ARE) можно определить как

$$\text{ARE} = \lim_{n \rightarrow \infty} \left( \frac{\mathbb{V} \text{STD}}{\mathbb{E} \text{STD}^2} \bigg/ \frac{\mathbb{V} \text{MAD}}{\mathbb{E} \text{MAD}^2} \right)^1.$$

Примем, что  $X_i$ , компоненты выборки, следуют гауссову распределению, нормализованному к математическому ожиданию 0 и среднеквадратическому отклонению 1.

### Относительная погрешность среднеквадратического отклонения

Характеристическая функция  $\Psi_1(t)$  для распределения квадрата отклонения,  $X^2$ :

$$\Psi_1(t) = \int_{-\infty}^{+\infty} e^{\frac{-x^2}{2} + itx^2} dx = \frac{1}{\sqrt{1-2it}}.$$

Обозначая сумму квадратов отклонения  $Z$ , выпишем плотность вероятностного распределения при  $n$  слагаемых

$$f_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx^2} \left( \frac{1}{\sqrt{1-2it}} \right)^n dt = \frac{2^{-\frac{n}{2}} e^{-\frac{z}{2}} z^{\frac{n}{2}-1}}{\Gamma \frac{n}{2}},$$

где  $z > 0$ <sup>2</sup>. Тогда для  $Y = \sqrt{Z}$  получаем<sup>3</sup> распределение

1 Здесь STD обозначает  $\text{STD}(X_1, \dots, X_n)$  и MAD обозначает  $\text{MAD}(X_1, \dots, X_n)$ .

2 Это распределение известно как хи-квадрат с  $n$  степенями свободы; таково распределение суммы квадратов  $n$  независимых случайных величин, которые все гауссовы, со средним 0 и дисперсией 1.

3 Пользуясь тем, что если для случайной величины  $Z$  с распределением плотности вероятностей по закону  $f_Z(z)$  рассматривается гладкое возрастающее отображение  $y = \tau(z)$ , то существует обратное отображение  $z = \tau^{-1}(y)$ , существует производная  $(\tau^{-1}(y))' > 0$  и  $Y = \tau(Z)$  является случайной величиной с распределением плотности вероятностей по закону  $f_Y(y) = f_Z(\tau^{-1}(y))(\tau^{-1}(y))'$ .

$$f_y(y) = \frac{2^{\frac{1-n}{2}} e^{-\frac{y^2}{2}} y^{n-1}}{\Gamma \frac{n}{2}},$$

где  $y > 0$ , что соответствует хи-распределению с  $n$  степенями свободы<sup>1</sup>. Интегрируя,

находим дисперсию:  $\mathbb{V} \text{STD}(n) = n - \frac{2 \left( \Gamma \frac{n+1}{2} \right)^2}{\left( \Gamma \frac{n}{2} \right)^2}$ . С учетом того, что математическое

ожидание равно  $\frac{\sqrt{2} \Gamma \frac{n+1}{2}}{\Gamma \frac{n}{2}}$ , получаем отношение  $\frac{\mathbb{V} \text{STD}}{\mathbb{E} \text{STD}^2} = \frac{n \left( \Gamma \frac{n}{2} \right)^2}{2 \left( \Gamma \frac{n+1}{2} \right)^2} - 1$ .

### Относительная погрешность абсолютного отклонения

Характеристическая функция для  $|x|$  — просто сложенное нормальное распределение, но давайте повторим выкладки:

$$\Psi_2(t) = \int_0^{+\infty} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2} + itx} dx = e^{-\frac{t^2}{2}} \left( 1 + i \operatorname{erfi} \frac{t}{\sqrt{2}} \right),$$

где т. н. мнимая функция ошибок  $\operatorname{erfi} z \triangleq \frac{\operatorname{erf} iz}{i}$ .

Первый момент

$$M_1 = -i \frac{\partial}{\partial t} \left( e^{-\frac{t^2}{2n^2}} \left( 1 + i \operatorname{erfi} \frac{t}{\sqrt{2}} \right) \right) \Bigg|_{t=0} = \sqrt{\frac{2}{\pi}}.$$

Второй момент

$$M_2 = (-i)^2 \frac{\partial^2}{\partial t^2} \left( e^{-\frac{t^2}{2n^2}} \left( 1 + i \operatorname{erfi} \frac{t}{\sqrt{2}} \right) \right) \Bigg|_{t=0} = \frac{2n + \pi - 2}{\pi n}.$$

Значит<sup>2</sup>,

$$\frac{\mathbb{V} \text{MAD}}{\mathbb{E} \text{MAD}^2} = \frac{M_2 - M_1^2}{M_1^2} = \frac{\pi - 2}{2n}.$$

В самом деле, здесь рассматривается случайная величина  $Z$  с распределением  $f_z(z) = \frac{2^{-\frac{n}{2}} e^{-\frac{z^2}{2}} z^{\frac{n}{2}-1}}{\Gamma \frac{n}{2}}$  в области  $z > 0$  и возрастающее отображение  $y = \sqrt{z}$ . Поэтому в рассматриваемой области существует обратное отображение  $z = y^2$ , при  $y > 0$ , и положительная производная обратного отображения,  $(y^2)' = 2y > 0$  при  $y > 0$ , и

$Y = \sqrt{Z}$  является случайной величиной с распределением  $f_y(y) = f_z(y^2)(y^2)' = \frac{2^{-\frac{n}{2}} e^{-\frac{y^2}{2}} (y^2)^{\frac{n}{2}-1}}{\Gamma \frac{n}{2}} \cdot 2y = \frac{2^{\frac{1-n}{2}} e^{-\frac{y^2}{2}} y^{n-1}}{\Gamma \frac{n}{2}}$ .

1 Распределение квадратного корня суммы квадратов  $n$  независимых случайных величин, которые все гауссовы, со средним 0 и дисперсией 1.

2 Использовано тождество  $\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$  применительно к  $X = \text{MAD}$ .



### Вуаля, асимптотическая относительная эффективность для гауссианы

$$\text{ARE} = \lim_{n \rightarrow \infty} \frac{n}{\pi - 2} \left( \frac{n \left( \Gamma \frac{n}{2} \right)^2}{\left( \Gamma \frac{n+1}{2} \right)^2} - 2 \right) = \frac{1}{\pi - 2} \approx 0,875 ,$$

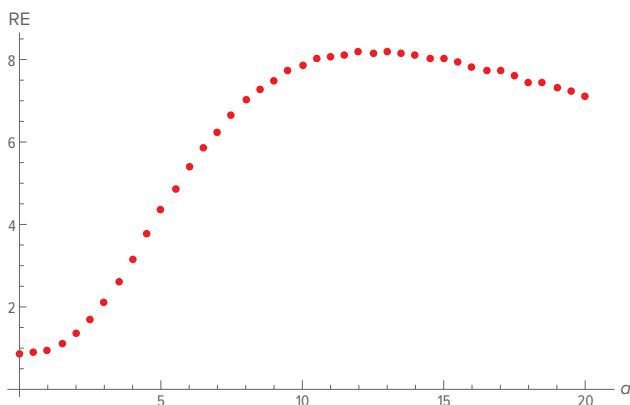
то есть среднеквадратическое отклонение на 12,5 % «эффективнее» среднего абсолютного отклонения, когда данные Гауссовы, и наши герои приняли такой довод. Хотя от этого отношения ничего не останется при малейшем загрязнении данных. Мы покажем далее, почему норма  $\ell^2$  не подходит почти нигде, а пока просто бросим взгляд на то, как хрупко среднеквадратическое отклонение STD.

#### 4.4.3. Влияние жирных хвостов на «эффективность» STD против MAD

Рассмотрим волатильность в обычной модели смешения, где с вероятностью  $p$  происходит случайное переключение на одно из двух распределений. Пусть они оба будут гауссовы и с математическим ожиданием 0; тогда и у модели математическое ожидание будет такое же. Дисперсию будем смешивать:

$$\mathbb{V} = \begin{cases} \sigma^2(1+a) & \text{с вероятностью } p \\ \sigma^2 & \text{с вероятностью } (1-p). \end{cases}$$

Как простой способ исследовать модель используем метод Монте-Карло. Пусть загрязнение выбросами будет скромное, с вероятностью  $p = 0,01$ . На Рисунке 4.7 показано, что уже начиная с  $a = 2$  ничего не остается от преимущества в эффективности у среднеквадратического отклонения по отношению к среднему абсолютному. С появлением выбросов MAD становится эффективнее, чем STD. Даже такие скромные «выбросы», как 5 сигм, делают MAD впятеро эффективнее, чем STD<sup>1</sup>.



**Рисунок 4.7:** Моделирование эффективности (Relative Efficiency, RE) среднеквадратического отклонения сравнительно со средним абсолютным отклонением, если подмешать в выборку выбросы величиной  $\sqrt{1+a} \sigma$ , где  $\sigma$  — среднеквадратичное отклонение

<sup>1</sup> Обычно MAD центрируют вокруг медианы, но мы находим, что для многих задач в этой книге (и вообще в теории принятия решений) важнее видеть отклонения от среднего. Делая соответствующую оговорку, мы не раз будем центрировать MAD относительно среднего. — Прим. автора.

#### 4.4.4. Моменты и неравенство о средних разной степени

Определим  $n$ -мерный вектор:<sup>1</sup>

$$\mathbf{X} \triangleq (x_i)_{i=1}^n$$

и норму степени  $p$ :<sup>2</sup>

$$\|\mathbf{X}\|_p \triangleq \left( \frac{1}{n} \sum_{i=1}^n |x_i|^p \right)^{1/p}.$$

Для любых  $1 \leq p < q$  выполняется неравенство:

$$\sqrt[p]{\sum_{i=1}^n w_i |x_i|^p} \leq \sqrt[q]{\sum_{i=1}^n w_i |x_i|^q}, \quad (4.9)$$

где веса  $w_i$  в сумме дают 1<sup>3</sup>. (Заметим, что избегаем случая  $p < 1$ , поскольку в нем нарушается неравенство треугольника.)

*Доказательство.* Пользуясь тем, что  $p$  и  $q$  положительные, определим функцию

$$f: \mathbb{R}^+ \rightarrow \mathbb{R}^+; f(x) = x^{\frac{q}{p}}.$$

Поскольку  $f$  — степенная функция, ее вторая производная тоже степенная:

$$f''(x) = \frac{q}{p} \left( \frac{q}{p} - 1 \right) x^{\frac{q}{p} - 2}$$

и строго положительна на всей области определения функции  $f$ , так как  $p < q$ ; значит,  $f$  выпукла вниз. Следовательно, в силу неравенства Йенсена,

$$f\left(\sum_{i=1}^n w_i x_i^p\right) \leq \sum_{i=1}^n w_i f(x_i^p),$$

поэтому

$$\sqrt[p]{\sum_{i=1}^n w_i x_i^p} \leq \sqrt[q]{\sum_{i=1}^n w_i x_i^q}.$$

Возведя обе части в степень  $1/q$  (это возрастающее преобразование, поскольку  $1/q$  положительно), получаем требуемое неравенство.

В нашей задаче, как и вообще в изучении влияния жирных хвостов, важно, что для нормы данной степени с ростом дисперсии исследуемых данных растет норма.

Начнем с примера плоского<sup>4</sup> распределения  $\mathbf{X} = \{1, 1\}$ <sup>5</sup>. Нормы всех степеней

$$\|\mathbf{X}\|_1 = \|\mathbf{X}\|_2 = \dots = \|\mathbf{X}\|_n = 1.$$

1 Обозначен жирным шрифтом; в литературе встречается также обозначение  $\bar{x}$ .

2 То же самое, что среднее степени  $p$  данных  $n$  чисел  $x_1, x_2, \dots, x_n$ ; в литературе встречается также обозначение  $A_p(x_1, x_2, \dots, x_n)$ .

3 Оба участника неравенства — так называемое среднее взвешенное степени  $p$  или  $q$  данных  $n$  чисел  $x_1, x_2, \dots, x_n$  с весами  $w_1, w_2, \dots, w_n$ ; в литературе встречается также обозначение  $\bar{x}$ .

4 «Плоская» последовательность значит константа. В русской литературе обычно эта метафора относится только к области постоянства функции: участок плато на графике.

5 Цепочка чисел выписана в фигурных скобках; в литературе встречается также запись с другими скобками — круглыми:  $(1, 1)$  или угловыми:  $\langle 1, 1 \rangle$ .

Внесем возмущение, сохранив  $\|\mathbf{X}\|_1$ . Для  $\mathbf{X} = \left\{ \frac{1}{2}, \frac{3}{2} \right\}$  получим рост нормы с ростом степени:

$$\left\{ \|\mathbf{X}\|_n \right\}_{n=1}^5 = \left\{ 1, \frac{\sqrt{5}}{2}, \frac{\sqrt[3]{7}}{2^{2/3}}, \frac{\sqrt[4]{41}}{2}, \frac{\sqrt[5]{61}}{2^{4/5}} \right\}. \quad (4.10)$$

Если усилить возмущение, дальше отклонившись от постоянного значения, и взять  $\mathbf{X} = \left\{ \frac{1}{4}, \frac{7}{4} \right\}$ , получим еще большие нормы:

$$\left\{ \|\mathbf{X}\|_n \right\}_{n=1}^5 = \left\{ 1, \frac{5}{4}, \frac{\sqrt[3]{43}}{2}, \frac{\sqrt[4]{1201}}{4}, \frac{\sqrt[5]{2101}}{2 \times 2^{3/5}} \right\}. \quad (4.11)$$

Как можно видеть, высшие моменты демонстрируют взрывной рост.

Полезно знать свойство степенных законов с бесконечными моментами:

$$\|\mathbf{X}\|_\infty = \sup_{i=1}^n |x_i|. \quad (4.12)$$

**Гауссов случай.** Рассмотрим гауссову случайную величину  $X \sim \mathcal{N}(0, \sigma^2)$ ; мы полагаем математическое ожидание 0, не теряя общности.

Пусть  $\mathbb{E}$  — оператор математического ожидания. Тогда<sup>1</sup>

$$\frac{\mathbb{E} X^p}{\mathbb{E} |X|} = 2^{\frac{p-3}{2}} \left( (-1)^p + 1 \right) \sigma^{p-1} \Gamma \frac{p+1}{2}$$

или, как удобнее для работы с четными  $p$ ,

$$\frac{\mathbb{E} X^p}{\mathbb{E} |X|} = 2^{\frac{1}{2}(p-3)} \left( 1 + (-1)^p \right) \left( \frac{1}{\sigma^2} \right)^{\frac{1}{2} - \frac{p}{2}} \Gamma \frac{p+1}{2}, \quad (4.13)$$

где  $\Gamma(z)$  — это эйлерова гамма-функция<sup>2</sup>

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt.$$

Для нечетных моментов отношение равно 0. Для четных моментов

$$\frac{\mathbb{E} X^2}{\mathbb{E} |X|} = \sqrt{\frac{\pi}{2}} \sigma,$$

откуда

$$\frac{\sqrt{\mathbb{E} X^2}}{\mathbb{E} |X|} = \frac{\text{STD}}{\text{MAD}} = \sqrt{\frac{\pi}{2}};$$

1 Здесь использованы свойства гауссова распределения:

если  $p$  нечетно,  $\mathbb{E} X^p = 0$ ;

если  $p$  четно,  $\mathbb{E} X^p = \sigma^p (p-1)!! = \sigma^p 2^{p/2} \Gamma \frac{p+1}{2} / \sqrt{\pi}$ ;

$\mathbb{E} |X| = \sigma \sqrt{2/\pi}$ .

2 Поскольку для четного  $p$  эйлерова гамма-функция  $\Gamma \frac{p+1}{2} = 2^{-p/2} (p-1)!! \sqrt{\pi}$ , удобнее

$\mathbb{E} X^p / \mathbb{E} |X| = \sqrt{\pi/2} \sigma^{p-1} (p-1)!!$

$(\mathbb{E} X^p)^{1/p} / \mathbb{E} |X| = \sqrt{\pi/2} ((p-1)!!)^{1/p}$ .

отношение четвертого момента к среднему абсолютному отклонению

$$\frac{\mathbb{E} X^4}{\mathbb{E} |X|} = 3 \sqrt{\frac{\pi}{2}} \sigma^3$$

и т. д.

Для распределения по степенному закону с показателем хвоста  $\alpha = 3$ , скажем для  $t$ -распределения Стьюдента,

$$\frac{\sqrt{\mathbb{E} X^2}}{\mathbb{E} |X|} = \frac{\text{STD}}{\text{MAD}} = \frac{\pi}{2}.$$

Вернемся к другим показателям и определениям жирного хвоста при распределении по степенному закону с так называемыми «бесконечными» моментами, то есть с несуществующими моментами. Наша эвристика — рассматривать отношение момента к среднему абсолютному отклонению — работает только внутри выборки, не распространяясь за ее пределы.

**Случай Парето**<sup>1</sup>. Для стандартного распределения Парето с минимальным значением (и масштабом)  $L$ , плотностью вероятности  $f(x) = \alpha L^\alpha x^{-\alpha-1}$  и среднеквадратичным

отклонением  $\frac{\sqrt{\frac{\alpha}{\alpha-2}} L}{\alpha-1}$  имеем

$$\frac{\text{STD}}{\text{MAD}} = \frac{1}{2\sqrt{\alpha-2}(\alpha-1)} \frac{\alpha^{-1}}{\alpha^{\frac{1}{2}-\alpha}}, \quad (4.14)$$

центрируя MAD относительно математического ожидания.

**«Бесконечные» моменты.** Любой бесконечный момент любого порядка, например бесконечная дисперсия, в наблюдаемой выборке проявляется как вычисляемая величина, то есть как конечный момент, просто потому, что выборка конечна. Скажем, распределение Коши, имеющее неопределенное математическое ожидание, в конечной выборке даст измеримое математическое ожидание; другое дело, что разные выборки будут давать совершенно разные математические ожидания. На Рисунках 4.8 и 4.9 иллюстрируется «дрейф» моментов по мере поступления информации.

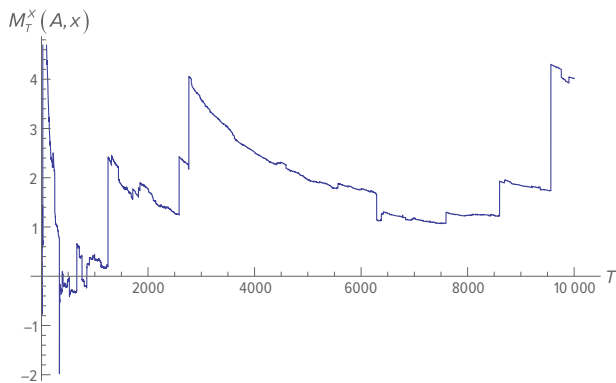
<sup>1</sup> Случайная величина Парето порождается средой, где неравномерности усиливаются до масштабов, сопоставимых с размерами всей среды; например, таково распределение богатства в условиях, когда действует принцип «деньги к деньгам», семейные богатства передавались наследникам из поколения в поколение и выросли до размеров, сравнимых с богатством всей страны.

Случайная величина  $X$  следует распределению Парето с коэффициентом масштаба  $L > 0$  и коэффициентом формы, он же показатель степени хвоста  $\alpha > 0$ , если  $X \in [L, \infty)$ , то есть имеет минимальное значение

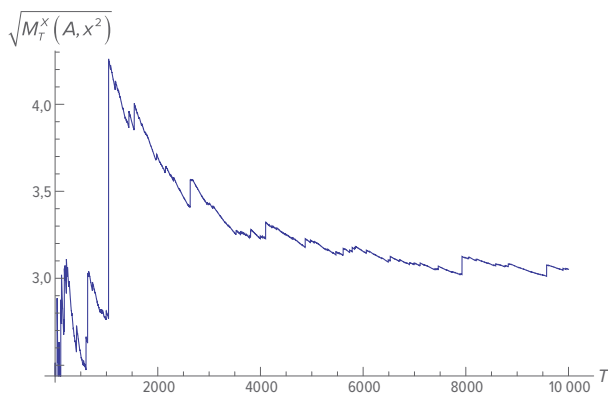
$L > 0$ , и функция распределения имеет вид степенного закона  $F_X(x) = 1 - \left(\frac{L}{x}\right)^\alpha$ ,  $x \geq L$ .

Плотность вероятности  $f_X(x) = \alpha L^\alpha x^{-\alpha-1}$  односторонняя с максимумом на краю  $x = L$ ; среднее  $\infty$  при  $\alpha \leq 1$ , иначе  $\frac{\alpha L}{\alpha-1}$ ; дисперсия  $\infty$  при  $\alpha \leq 2$ , иначе  $\frac{\alpha L^2}{(\alpha-1)^2(\alpha-2)}$ ; коэффициент асимметрии не существует при  $\alpha \leq 3$ ,

иначе  $\frac{2(1+\alpha)}{\alpha-3} \sqrt{\frac{\alpha-2}{\alpha}}$ ; коэффициент эксцесса не существует при  $\alpha \leq 4$ , иначе  $\frac{3(\alpha^3 + 9\alpha^2 - 24\alpha - 4)}{\alpha(\alpha-3)(\alpha-4)}$ .



**Рисунок 4.8:** Как изменяется математическое ожидание выборки для последовательности данных, когда для их распределения (распределение Коши) математическое ожидание не определено



**Рисунок 4.9:** Квадратный корень из второго момента выборки для последовательности данных, когда у их распределения бесконечная дисперсия. Наблюдаем кажущееся схождение, но только до очередного скачка

#### 4.4.5. Комментарий: почему среднееквадратическое отклонение нужно отправить в отставку, и немедленно!

Понятие среднееквадратического отклонения сбило с толку толпы ученых; пора отправить его в отставку из общего употребления и заменить на более эффективное среднее абсолютное отклонение. Среднееквадратическое отклонение (*Standard deviation*, STD) следует оставить математикам, физикам и специалистам по математической статистике, выводящим предельные теоремы. Нет научных резонов рекомендовать его для статистических исследований в нашу компьютерную эпоху, ибо от него больше вреда, чем пользы — особенно когда растущий класс специалистов в общественных науках применяет готовые статистические инструменты к научным проблемам.

Скажем, пусть кто-то попросил вас оценить «среднее отклонение за день» для температуры в вашем городе (или биржевой стоимости компании, или артериального давления вашего дядюшки) на протяжении пяти последних дней. Данные об отклонениях в эти пять дней такие:  $(-23, 7, -3, 20, -1)$ . Как вам оценить среднее отклонение?

Может быть, взять каждое наблюдение, возвести в квадрат, найти среднее по всем квадратам и извлечь из него квадратный корень? Или удалить знаки минус и посчитать среднее? Между этими двумя методиками серьезные различия. Первая оценит среднее отклонение как 15,7, а вторая — как 10,8. Первая в английской литературе

величается «стандартным» отклонением, а вторая — «средним» отклонением (*mean absolute deviation*, MAD). На самом деле вторая куда лучше соответствует «практическим задачам» и отражает реальность. Интуитивно мы часто ожидаем услышать именно о второй величине; нередко человеку сообщают среднеквадратическое отклонение, а он затем действует так, будто принял эту величину за среднее отклонение.

Причиной всему историческая случайность: в 1893 великий Карл Пирсон употребил термин «стандартное отклонение» (*standard deviation*) вместо употреблявшегося более громоздкого «среднеквадратическая ошибка» (*root mean square error*). И положил начало великому заблуждению: публика решила, что это значит среднее абсолютное отклонение. Предрассудок пустил корни и затвердел, как скала; всякий раз, когда какая-нибудь газета пыталась уточнить понятие рыночной «волатильности», она сбивалась в словесных описаниях на среднее абсолютное отклонение, даже если приводила формулу, по которой насчитала среднеквадратическое отклонение, более высокое.

И если бы только журналисты; я видел то же смешение в официальных документах Министерства торговли США и Федерального резерва — в заявлениях о рыночной волатильности, сделанных регуляторами рынка. Хуже того, мы с Голдстейном обнаружили многочисленных специалистов по анализу данных (зачастую с докторскими научными степенями), которые путали эти вещи в практических задачах.

Вот что натворил неудачный термин для неинтуитивного понятия. Психологический феномен склонность к упрощению (*attribute substitution*) толкает людей принимать STD за MAD, потому что о MAD легче думать — это эффект Линди<sup>1</sup>, хорошо известный мошенникам и иллюзионистам.

Вот наши аргументы за революцию.

1) MAD точнее оценивается по выборке и менее волатильно, чем STD, поскольку не взвешивает наблюдения, тогда среднеквадратическая ошибка приписывает большему отклонениям больший вес, переоценивая хвостовые события.

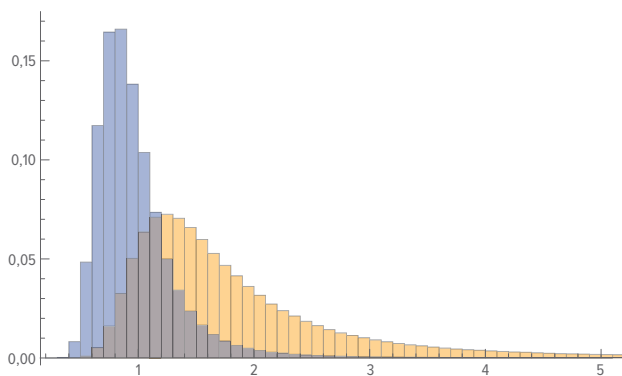
2) Мы часто используем STD в формулах, но в итоге приходим к необходимости преобразовать его в MAD (скажем, в финансах — для определения стоимости опциона). Поправка существенная. В Гауссовом мире STD примерно в 1,25 раза больше MAD — если точно, то в  $\sqrt{\frac{\pi}{2}}$  раз. Но нам надо выживать при стохастической волатильности, где STD нередко превосходит MAD в 1,6 раза.

3) Многие статистические явления и процессы характеризуются «бесконечной дисперсией» (то же знаменитое правило Парето 80/20), хотя обладают конечным и часто спокойно ведущим себя средним абсолютным отклонением. Во всех случаях, где существует STD, существует и MAD. Риска в эту сторону (бесконечное MAD при конечном среднеквадратическом отклонении) не существует.

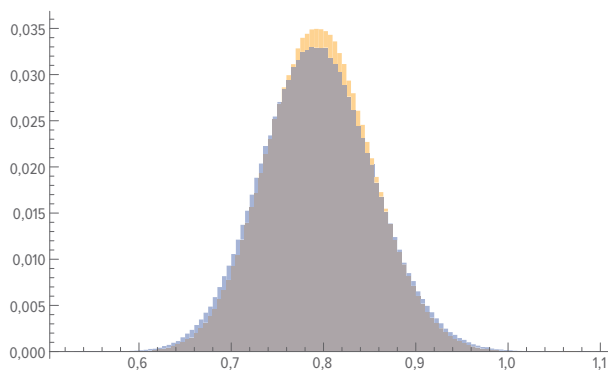
4) Многие экономисты побоялись рассмотреть модели с бесконечной дисперсией, подумав, что это значит бесконечное среднее абсолютное отклонение. Прискорбно, но факт. Когда великий Бенуа Мандельброт выдвинул свои модели с бесконечной дисперсией пятьдесят лет назад, экономисты испугались, потому что в их головах смешалось одно с другим.

---

1 См. определение *Lindy* в главке 5.0.2. — Прим. автора.



**Рисунок 4.10:** Как распределено среднее абсолютное отклонение (синим) и среднеквадратическое отклонение (желтым), если исходное распределение следует степенному закону с конечной дисперсией. Как обычно, у MAD более узкое распределение; в данном случае его даже трудно сравнить с распределением STD, имеющим бесконечную дисперсию (потому что квадрат случайной величины, распределенной по Парето с показателем  $\alpha$ , распределен по Парето с показателем  $\alpha/2$ ). Можно, впрочем, сравнить распределения MAD и STD по их среднему абсолютному отклонению: у распределения STD оно в 5 раз больше



**Рисунок 4.11:** Если исходное распределение гауссово, различие между формами пика распределений MAD и STD невелико. Здесь для наглядности оба распределения приведены к одинаковому среднему

Печально, что такая мелочь способна создать столько путаницы. Наши научные инструменты вышли далеко за нашу бытовую интуицию, и это стало проблемой для развития науки. В завершение приведу заявление сэра Роналда Фишера: «Статистик несет ответственность за понимание процедуры, которую применяет или рекомендует».

**Примечание.** Обычно нас учат, что если случайные величины  $X_1, \dots, X_n$  независимы, то

$$\mathbb{V}(X_1 + \dots + X_n) = \mathbb{V}X_1 + \dots + \mathbb{V}X_n$$

в силу линейности дисперсии. Если же нужна другая характеристика, ради суммирования ее приходится преобразовывать в линейную<sup>1</sup>. Как мы покажем, для гауссовой величины MAD  $X = \sqrt{\frac{2}{\pi}}\sigma$ , для t-распределения Стьюдента с 3 степенями свободы

нужен коэффициент  $\frac{2}{\pi}$  и так далее.

<sup>1</sup> Например, цена опциона по формуле Блэка — Шоулза вычисляется через дисперсию, но цену приводят в MAD; стрэддл «на деньгах» вычисляется как условное среднее абсолютное отклонение. Таким образом, мы переводим MAD в среднеквадратическое отклонение ради суммирования, а потом результат переводим обратно в MAD. — Прим. автора.

#### 4.5. ВИЗУАЛИЗИРУЕМ, КАК РОСТ $p$ ВЛИЯЕТ НА ИЗО-НОРМЫ

Рассмотрим область  $\mathcal{R}_{(d)}^{(p)}$  — множество всех векторов  $\mathbf{X} = (x_1, \dots, x_d)$ , для которых

$$\left( \sum_{i=1}^d |x_i|^p \right)^{\frac{1}{p}} \leq 1,$$

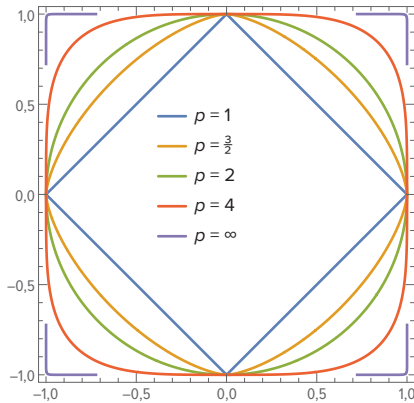
то есть границей служит единица<sup>1</sup>. Объем этого шара растет с ростом порядка нормы:

$$V_d^p = \int \cdots \int_{\mathbf{X} \in \mathcal{R}_{(d)}^{(p)}} 1 d\mathbf{X} = \frac{\left( 4\Gamma\left(1 + \frac{1}{p}\right) \right)^d}{\Gamma\left(\frac{d}{p} + 1\right)}.$$

На Рисунках 4.12 и 4.13 показаны два эффекта.

На первом показано, как с ростом порядка нормы растет занимаемая доля пространства<sup>2</sup>.

Второй показывает краешек так называемого проклятия размерности, весьма полезного принципа для задач в самых разных областях, и прежде всего — для оценки погрешностей модели. Сравним Рисунки 4.12 и 4.13: на первом (о размерности 2) самая малая область, при  $p = 1$ , занимает половину площади квадрата. Область при  $p = \infty$  занимает всю площадь квадрата. Значит, отношение величин единичных кругов с разными нормами достигает 1:2. На втором рисунке (о размерности 3) шар с нормой  $p = 1$  занимает  $\frac{4/3}{2^3} = \frac{1}{6}$  куба, тогда как при  $p = \infty$  опять занимает все место. Значит, отношение величин единичных шаров с разными нормами достигает 1:6. Так и отношение высших моментов к низшим растет с ростом размерности, и что происходит при больших размерностях, можно видеть на Рисунке 4.14.



**Рисунок 4.12:** Отношение круг/квадрат при растущем порядке нормы. Показаны изо-нормы вида  $(|x_1|^p + |x_2|^p)^{1/p} = 1$ . Видно, что площадь внутри изо-нормы (область, где норма  $\leq 1$ ) растет от  $V_2^1 = 2$  до

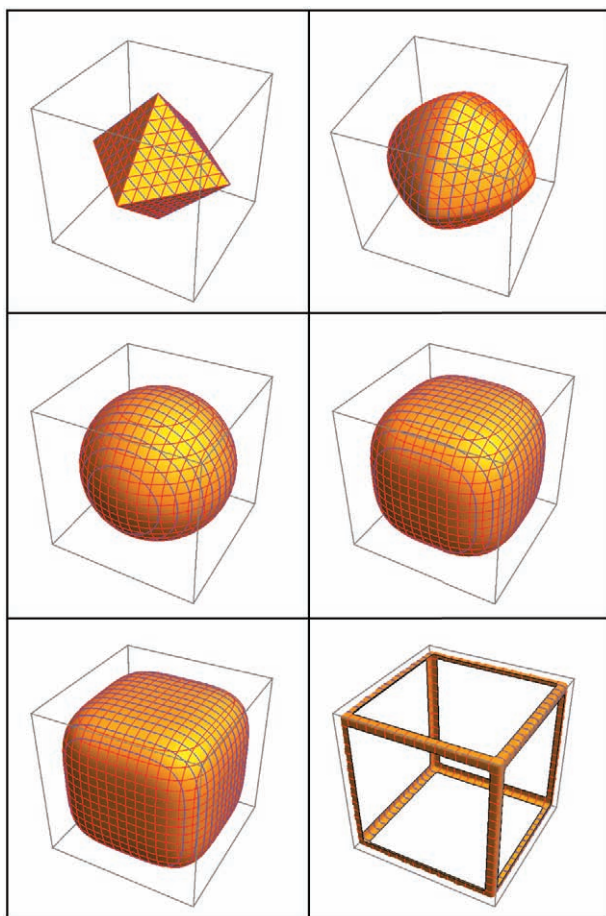
$$V_2^\infty = 4. \text{ Закон роста } V_2^p = \frac{\left( 4\Gamma\left(\frac{p+1}{p}\right) \right)^2}{\Gamma\left(\frac{p+2}{p}\right)}.$$

\* По этому закону  $V_2^2 = \pi$ . Другими словами, норма порядка 2 — это обычное расстояние, и круг этой нормы — обычный круг единичного радиуса.

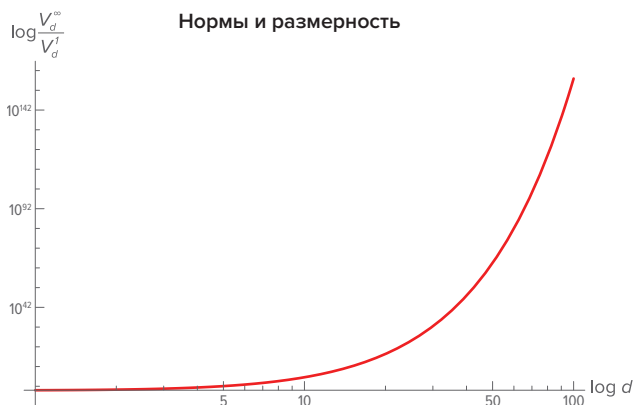
1 Эта область — единичный  $d$ -мерный «шар», если радиус понимать не обязательно как евклидово расстояние  $\ell_2$ , а как любую гёльдерову  $\ell_p$ -норму.

2 То есть доля в объеме  $d$ -мерного единичного куба.





**Рисунок 4.13:** Отношение шар/куб при растущем порядке нормы. Показаны изо-нормы вида  $(|x_1|^p + |x_2|^p + |x_3|^p)^{1/p} = 1$  для  $p = 1, \frac{3}{2}, 2, 3, 4$  и  $\infty$ . Объем области, где норма  $\leq 1$ , растет с  $\frac{4}{3}$  при  $p = 1$  до  $\frac{4\pi}{3}$  при  $p = 2$  (единичный шар) и далее до  $2^3$  при  $p = \infty$  (единичный куб); прирост получился в большее число раз, чем на Рисунке 4.12. Так работает проклятие размерности — с ростом  $n$  уменьшается объем при  $p = 1$  по отношению к объему при  $p = \infty$



**Рисунок 4.14:** Проклятие размерности, важнейший феномен в разных областях статистики, играет критическую роль для погрешности модели высокой размерности. С ростом размерности  $d$  отношение  $V_d^\infty$  к  $V_d^1$  демонстрирует взрывной рост. Если при  $d = 2$  оно составляет 2, то при  $d = 9$  оно уже шестизначное

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА: Остановимся здесь и представим книги по теории вероятностей.

Для общего представления о вероятности прочитайте книги Бореля [85], Колмогорова [145], Лозва [154], Феллера [92], [91].

О теории меры — Биллингсли [20].

**О субэкспоненциальности** — Питмана [196], Амбре и Голди (1982) [83], Амбре (1979, эта книга, по-видимому, близка к его докторской диссертации) [84], Чистякова (1964) [43], Голди (1978) [112] и Тойгелса [248].

**О распределениях экстремальных значений** — Амбре и др. [82], де Гаана и Феррейры [116].

**Об устойчивых распределениях** — Учайкина и Золотарева [257], Золотарева [271], Самородницкого и Самородницкого и Такку [209].

**О стохастических процессах** — Карацаса и Шрива [141], Оксендала [182], Варадхана [261].

## УРОВЕНЬ 2: СУБЭКСПОНЕНЦИАЛЬНЫЕ И СТЕПЕННЫЕ ЗАКОНЫ

В этой главе кратко представлены «истинно жирные хвосты» — классы субэкспоненциального и степенного закона (уже определенные в Главе 3) — и некоторые капризные особенности их поведения. Чисто субэкспоненциальные распределения (то есть без масштабируемости, вне класса степенного закона) составляют узкую категорию, куда попадают только пограничные экспоненциальные и гамма-распределения, такие как лапласово, и еще лог-нормальное.

### 5.0.1. Вернемся к ранжированию

В Таблице 5.1 уточнены ранги, показанные в Главе 3. Вспомним, что диапазон вероятностных распределений простирается от крайне тонких хвостов (Бернулли) до крайне жирных. Выбор категорий, в которые попадают те или иные распределения, в ряде случаев мотивирован разной сходимостью моментов:

1. С компактным носителем (невыврожденным).
2. Субгауссовы.
3. Субэкспоненциальные.
4. Степенного закона с показателем больше 2.
5. Степенного закона с показателем больше 1 и меньше или равным 2. Такие распределения отличаются, в частности, тем, что имеют конечное математическое ожидание — потому что показатель степени больше 1, — но не конечную дисперсию — потому что показатель степени меньше или равен 2.
6. Степенного закона с показателем меньше или равным 1.

Наш интерес в том, чтобы формально выделить категории распределений с преобладающим влиянием хвостовых наблюдений, то есть определить границу между Медиокристаном и Экстремистаном.

Центральную роль играет субэкспоненциальное распределение, расположенное между «тонкими» и «жирными» хвостами.

Оно определяется так.

Таблица 5.1: Ранжирование распределений

Класс	Описание
Истинно тонкие хвосты	Компактный носитель (напр., распределение Бернулли, биномиальное распределение)
Тонкие хвосты	Гауссово распределение, которое естественно получается при сложении тонкохвостых случайных величин, в силу центральной теоремы; также суммы случайных величин с компактным носителем до предельного перехода $n \rightarrow \infty$ , сами имеющие компактный носитель
Условно-тонкие хвосты	Гауссово приближение естественных феноменов
Хвосты на начальной ступени ожирения	Экссесс выше, чем у гауссианы, но при суммировании нескольких таких слагаемых сумма стремительно сходится к гауссиане
Субэкспоненциальное	(напр., логнормальное)
Суперкубический показатель $\alpha$	Условие Крамера нарушается при $t > 3$ , $\int e^{-tx} dF(x) = +\infty$
Бесконечная дисперсия	Устойчивые по Леви с $\alpha < 2$ , $\int e^{-tx} dF(x) = +\infty$
Неопределенный первый момент	Безнадега

В математике все точно: при тонких хвостах вероятность превышения, или функция выживания, должна быть экспоненциальна, без тонких хвостов — нет. Где проходит граница?

Естественной границей между Медиокристаном и Экстремистаном оказывается субэкспоненциальный класс, обладающий следующим свойством:

Пусть  $\mathbf{X} = X_1, \dots, X_n$  — последовательность независимых и одинаково распределенных случайных величин с носителем  $\mathbb{R}^+$  и интегральной функцией распределения  $F$ . Субэкспоненциальный класс распределений определяется таким критерием (см. [248], [196]):

$$\lim_{x \rightarrow \infty} \frac{1 - F^{*2}(x)}{1 - F(x)} = 2, \quad (5.1)$$

где  $F^{*2} = F' * F$  — интегральное распределение для суммы двух независимых случайных величин,  $X_1 + X_2$ . Смысл этого требования: вероятность того, что сумма  $X_1 + X_2$  превысит значение  $x$ , должна быть вдвое больше вероятности того, что это сделает одно отдельно взятое слагаемое. Значит, все случаи, когда сумма превысила  $x$ , объясняются тем, что это сделало большее из слагаемых, тогда как вклад другого слагаемого был несущественен.

Критерий можно обобщить: и в сумме  $n$  случайных величин доминирует одна наибольшая. Формально критерию субэкспоненциальности эквивалентны следующие два свойства [43], [84]. Для данного  $n \geq 2$  обозначим  $S_n = \sum_{i=1}^n X_i$  и  $M_n = \max_{1 \leq i \leq n} X_i$ .

$$\text{a) } \lim_{x \rightarrow \infty} \frac{\mathbb{P}(S_n > x)}{\mathbb{P}(X > x)} = n,$$

$$\text{b) } \lim_{x \rightarrow \infty} \frac{\mathbb{P}(S_n > x)}{\mathbb{P}(M_n > x)} = 1.$$

Оба они означают, что каждый раз, когда сумма  $S_n$  превысила  $x$ , она сделала это благодаря наибольшему слагаемому,  $M_n$ , то есть главную роль сыграло единственное хвостовое событие.

Интуитивно кажется, что с равным успехом от субэкспоненциальных распределений можно потребовать более медленного исчезновения хвостовых событий, чем в экспоненциальном распределении, у которого в той же области хвостовые события уже стали несущественными. Действительно, можно показать, что у субэкспоненциальных распределений нет экспоненциальных статистических моментов:

$$\int_0^{+\infty} e^{\varepsilon x} dF(x) = +\infty \quad (5.2)$$

для любых  $\varepsilon$  больше нуля. Однако обратное неверно, поскольку возможны распределения, не имеющие экспоненциальных моментов, и все же не удовлетворяющие критерию субэкспоненциальности.

Отметим, что если бы мы решили представить отклонения отрицательными значениями случайной величины  $X$ , то в силу симметрии получили бы тот же результат для экстремальных отрицательных значений, заменив  $x \rightarrow +\infty$  на  $x \rightarrow -\infty$ . Для случайных величин с двумя хвостами можно по отдельности рассмотреть положительную и отрицательную область определения.

### 5.0.2. Что такое пограничное распределение вероятностей?

Лучший способ нащупать нужное распределение вероятностей — это изобрести его. Чем мы и займемся в следующей главке, 5.0.3 — получим распределение, которое *по построению* расположено точно на границе между тонкими и жирными хвостами. Пока же обсудим требуемые свойства.

Пусть  $\bar{F}$  — функция выживания. Это отображение вида  $\bar{F}: \mathbb{R} \rightarrow [0, 1]$  такое, что

$$\lim_{x \rightarrow +\infty} \frac{\bar{F}(x)^n}{\bar{F}(nx)} = 1, \quad (5.3)$$

$$\lim_{x \rightarrow +\infty} \bar{F}(x) = 0$$

и

$$\lim_{x \rightarrow -\infty} \bar{F}(x) = 1.$$

Примечание: еще одно демаркационное свойство — отсутствие недооценки Лукреция, описанной в «Черном лебеде» (т. е. что при жирных хвостах будущие экстремальные события будут отличаться от прошлых экстремальных событий, и это различие растет с ростом жирности хвостов).

Пока посмотрим на уже выписанные демаркационные свойства. Пусть  $X$  — случайная величина, существующая в промежутке  $(0, \infty)$  или  $(-\infty, \infty)$ , а  $\mathbb{E}$  — оператор

математического ожидания в «практическом» (физическом) распределении. Согласно классическим результатам [82]:

$$\lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}(X | X > K) = \lambda. \quad (5.4)$$

- Если  $\lambda = 1$ , то случайную величину  $X$  относят к классу тонких хвостов  $D_1$ , и у нее есть характерный масштаб;
- Если  $\lambda > 1$ , то случайную величину  $X$  относят к классу  $D_2$ , классу с жирными хвостами и правильно меняющимися на бесконечности функциями, и у нее нет характерного масштаба;
- Если

$$\lim_{K \rightarrow \infty} (\mathbb{E}(X | X > K) - K) = \mu,$$

где  $\mu > 0$ , то случайную величину  $X$  относят к пограничному субэкспоненциальному классу.

Первый случай называется «эффект Линди»<sup>1</sup>, если случайная величина  $X$  — это время выживания. Данная тема изучена вне этого проекта, посвященного жирным хвостам. Смотрите о ней у Иддо Элиазара [77].

### 5.0.3. Давайте выдумаем распределение

Поворотной точкой для субэкспоненциального класса служит экспоненциальное распределение, однако его носитель  $[0, \infty)$ . Построим сами пограничное распределение с носителем  $(-\infty, \infty)$ <sup>2</sup>. Найдем функцию выживания  $\bar{F}: \mathbb{R} \rightarrow [0, 1]$  такую, что:

$$\forall x \geq 0, \quad \lim_{x \rightarrow +\infty} \frac{\bar{F}(x)^2}{\bar{F}(2x)} = 1, \quad \bar{F}'(x) \leq 0$$

и

$$\lim_{x \rightarrow +\infty} \bar{F}(x) = 0,$$

$$\lim_{x \rightarrow -\infty} \bar{F}(x) = 1.$$

<sup>1</sup> Термин сложился вокруг ресторана *Lindy's* на Бродвее, где обсуждали живучесть спектаклей и телепрограмм. Ранняя версия закона Линди предсказывала, что чем выше частота выхода телешоу, которое соглашается вести комик, тем скорее оно закроется; это утверждалось в статье «Lindy's Law» арт-критика Алберта Голдмана (*Albert Harry Goldman*, 1927–1994) в журнале *The New Republic* в 1964. В дальнейшем появились другие интуитивные приметы живучести, которые также назывались «законами Линди»; они, в частности, сулили долгое продолжение успеха у публики при наличии уже прожитой долгой истории успеха. Потом так же называли свои признаки живучести математики — сначала Бенуа Мандельброт в книге «Фрактальная геометрия природы» (*The Fractal Geometry of Nature*; 1982), затем Нассим Талеб в «Черном лебеде».

<sup>2</sup> Распределение Лапласа — экспоненциальное убывание по обе стороны от нуля — не обладает требуемым свойством, поскольку для него отношение квадрата вероятности к вероятности двойного значения составляет  $\frac{1}{2}$ . — Прим. автора.

Допустим, что найдется подходящая сигмоида<sup>1</sup>, использующая гиперболический тангенс

$$\bar{F}^k(x) = \frac{1}{2}(1 - \operatorname{th} kx), k > 0.$$

Используем ее как ядро, которое потом смикшируем с другими распределениями, чтобы получить нужный эксцесс.

Тогда плотность вероятности

$$f(x) = -\frac{\partial \bar{F}(x)}{\partial x} = \frac{1}{2} k \operatorname{sch}^2 k. \quad (5.5)$$

Характеристическая функция

$$\varphi(t) = \frac{\pi t \operatorname{csch} \frac{\pi t}{2k}}{2k}. \quad (5.6)$$

Поскольку все величины действительные, можно предположить, что математического ожидания 0, как и все нечетные моменты.

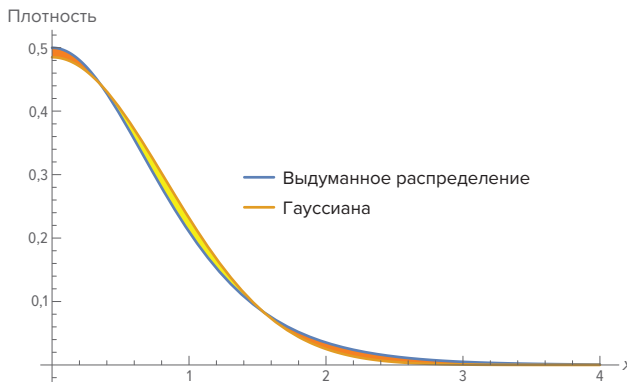
Второй момент будет

$$\lim_{t \rightarrow 0} (-i)^2 \frac{\partial^2}{\partial t^2} \frac{\pi t \operatorname{csch} \frac{\pi t}{2k}}{2k} = \frac{\pi^2}{12k^2},$$

а четвертый момент будет

$$\lim_{t \rightarrow 0} (-i)^4 \frac{\partial^4}{\partial t^4} \frac{\pi t \operatorname{csch} \frac{\pi t}{2k}}{2k} = \frac{7\pi^4}{240k^4},$$

так что эксцесс получится  $\frac{21}{5}$ ; хвосты будут лишь немногим жирнее, чем у гауссианы.



**Рисунок 5.1:** Сравним выдуманное распределение (на краю субэкспоненциальности) с гауссовым при одинаковой дисперсии ( $k = 1$ ). Мы добились субэкспоненциальных свойств, не уходя далеко от гауссианы

<sup>1</sup> Сигмоида (*sigmoid function*) — сглаженная ступенька, используется для сглаживания произвольной функции. Часто используются такие сигмоиды, как логистическая функция  $\frac{1}{1+e^{-x}}$  и гиперболический тангенс  $\operatorname{th} x \triangleq \frac{e^x - e^{-x}}{e^x + e^{-x}}$ .

### 5.1. УРОВЕНЬ 3: МАСШТАБИРУЕМОСТЬ И СТЕПЕННЫЕ ЗАКОНЫ

Перейдем к серьезным делам.

**Почему речь о степенных законах?** Нет числа теориям в оправдание степенным законам в тех или иных ситуациях, как будто речь идет об исключениях из теории вероятностей. Нет бы встать и сказать: все наоборот, степенные законы — в порядке вещей, а в оправданиях как особый случай нуждается гауссиана ([223]); но мы так заявили в книге *Антихрупкость* и покажем подробно в следующем томе *Технического цикла Incerto*). Это для тонких хвостов нужно искать механизм, выпуклый вверх-вниз отклик, который будет амортизировать шарахания хрупкости и антихрупкости.

#### 5.1.1. Масштабируемое и немасштабируемое — более глубокий взгляд на жирные хвосты

До сих пор, обсуждая жирные хвосты, мы оставались в ситуации конечных моментов. Для разных классов распределений, имеющих конечные моменты, отношение вероятностей  $\frac{\mathbb{P}(X > nK)}{\mathbb{P}(X > K)}$  по-разному зависит от  $n$  и от  $K$ . Для безмасштабных распределений, если  $K$  находится «где-то в хвосте», то есть для всех достаточно больших  $K$ , отношение  $\frac{\mathbb{P}(X > nK)}{\mathbb{P}(X > K)}$  перестает зависеть от  $K$  и зависит только от  $n$ . У таких распределений нет характерного масштаба, и вдали их хвост становится как в распределении Парето; другими словами, при больших  $x$

$$\mathbb{P}(X > x) = Cx^{-\alpha},$$

где  $\alpha$  — показатель хвоста и  $C$  — масштабирующая константа.

**Примечание:** Как видно из масштабирующих различий между распределениями Стьюдента и Парето, более строго распределение случайной величины  $X$  по степенному закону определяется условием

$$\mathbb{P}(X > x) = L(x)x^{-\alpha},$$

где «медленно меняющаяся» функция  $L$  удовлетворяет требованию

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1$$

для любого постоянного  $t > 0$ .

При больших  $x$  отношение  $\frac{\ln \mathbb{P}(X > x)}{\ln x}$  сходится к константе, а именно показателю хвоста  $-\alpha$ . Масштабируемое распределение должно в дважды логарифмическом масштабе показать наклон  $\alpha$  в хвостах, при  $x \rightarrow \infty$ . Сравним с гауссовым распределением (имеющим среднее квадратическое отклонение  $\sigma$  и математическое ожидание  $\mu$ ); в этот раз будем сравнивать не вероятность превышения, а плотность вероятности. Функция плотности  $f(x)$  гауссова распределения на логарифмической шкале

$$\ln f(x) = \frac{(x - \mu)^2}{2\sigma^2} - \ln \sigma \sqrt{2\pi} \approx -\frac{1}{2\sigma^2} x^2,$$

то есть уходит в  $-\infty$  быстрее, чем растёт  $\ln x$  при  $x \rightarrow \infty$ <sup>1</sup>.

<sup>1</sup> Следовательно, в дважды логарифмическом масштабе правый хвост кривой уйдет вниз быстрее, чем по прямой.

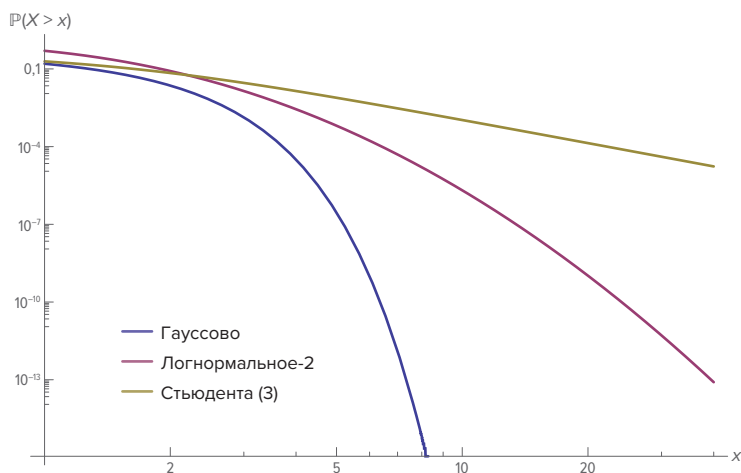


Таким образом, пока что мы получили интуитивное представление о различии между классами распределений. «Подлинно» жирные хвосты есть только у масштабируемых распределений, тогда как прочие при суммировании превращаются в гауссовы. И показатель хвоста — параметр асимптотический; мы можем так и не увидеть характерных хвостовых значений, и наша попытка оценить показатель хвоста забуксует на полдороге. Ясно видимые на Рисунке 5.2 различия относятся к идеальным распределениям из готового пакета; на практике процессы с большими отклонениями могут протекать сумбурно, переключаясь между разными показателями.

### Определение 5.1 (класс $\mathfrak{P}$ )

Класс  $\mathfrak{P}$  степенных законов (правильно меняющихся на бесконечности функций) определяется для НСВ  $X$  так:

$$\mathfrak{P} = \{X: \mathbb{P}(X > x) \sim L(x)x^{-\alpha}\} \quad (5.7)$$



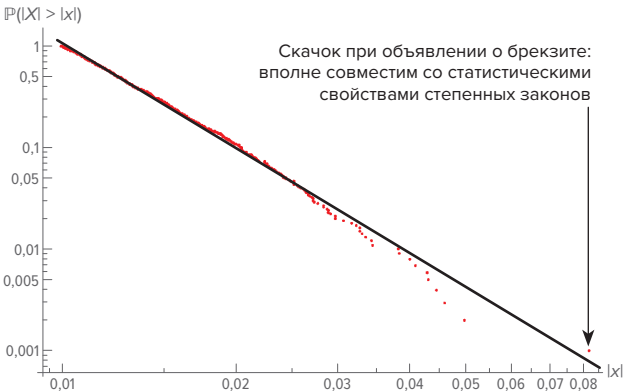
**Рисунок 5.2:** Три типа распределений. Дойдя до хвоста, видим, как распределение Стьюдента остается масштабируемым, а стандартное логнормальное некоторое время удерживается на промежуточной позиции и затем его наклон в дважды логарифмическом масштабе становится бесконечно крутым. Однако бойтесь логнормальных, они иногда приносят сюрпризы (Глава 8)

**Таблица 5.2:** Масштабируемость: сравнение правильно меняющихся функций, или степенных законов, с другими распределениями

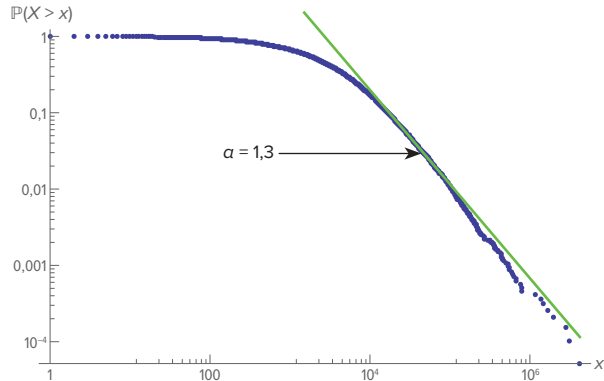
$k$	$\mathbb{P}(X > k)^{-1}$	$\frac{\mathbb{P}(X > k)}{\mathbb{P}(X > 2k)}$	$\mathbb{P}(X > k)^{-1}$	$\frac{\mathbb{P}(X > k)}{\mathbb{P}(X > 2k)}$	$\mathbb{P}(X > k)^{-1}$	$\frac{\mathbb{P}(X > k)}{\mathbb{P}(X > 2k)}$
	Гаусс	Гаусс	Стьюдент (3)	Стьюдент (3)	Парето (2)	Парето (2)
2	44	720	14,4	4,9	8	4
4	31600	$5,1 \times 10^{10}$	71,4	6,8	64	4
6	$1,01 \times 10^9$	$5,5 \times 10^{23}$	216	7,4	216	4
8	$1,61 \times 10^{15}$	$9 \times 10^{41}$	491	7,6	512	4

Продолжение таблицы 5.2

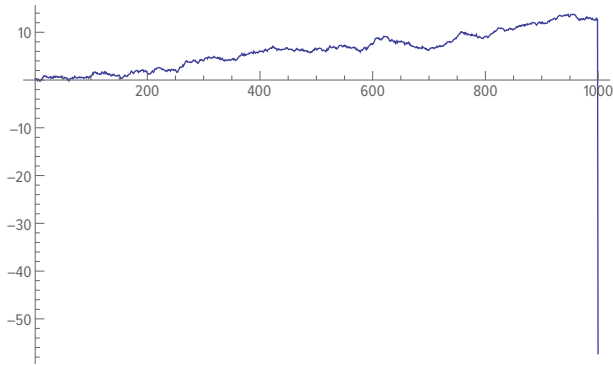
10	$1,31 \times 10^{23}$	$9 \times 10^{65}$	940	7,7	1000	4
12	$5,63 \times 10^{32}$	безнадега	1610	7,8	1730	4
14	$1,28 \times 10^{44}$	безнадега	2530	7,8	2740	4
16	$1,57 \times 10^{57}$	безнадега	3770	7,9	4100	4
18	$1,03 \times 10^{72}$	безнадега	5350	7,9	5830	4
20	$3,63 \times 10^{88}$	безнадега	7320	7,9	8000	4



**Рисунок 5.3:** В дважды логарифмическом масштабе показан график фунта стерлингов, британской валюты. Видно, что «серый лебедь» брекзита (скачок валюты при появлении неожиданных результатов референдума) укладывается в статистические свойства больших отклонений по степенному закону



**Рисунок 5.4:** Продажи книг. Ближняя часть хвоста под-держивает надежное пред-сказание продаж по рейтингу, и наоборот — и робастность сохраняется, пока не попыта-етесь вычислить ожидание по генеральной совокупности или высшие неусеченные моменты



**Рисунок 5.5:** Проблема индюка. Бывает, что ничто в прошлых свойствах не указывало на воз-можность скорого обвала

### 5.1.2. Серые лебеди

**Зачем нам t-распределение Стьюдента, если мы моделируем степенные законы?** Исклчительно ради нашего удобства. Мы отнюдь не полагаем, что порождающий события процесс отвечает t-распределению Стьюдента. Просто средняя часть распределения неважна для тех свойств, которые важны в интересующих нас сейчас ситуациях принятия решений.

Чем ниже показатель степени, тем меньшую роль играет центр. Чем выше показатель степени, тем больше t-распределение Стьюдента напоминает гауссово, вплоть до ситуации, где будет оправдано использование гауссианы.

В асимметричном случае могут пригодиться более сложные методики с использованием законов Леви; впрочем, без этих сложностей можно обойтись, используя просто два распределения Парето с двумя разными показателями степени для левого и правого хвостов.

**Проблемы с оценкой.** Заметим, что есть много методик, как оценить показатель степени, исходя из данных; это так называемая калибровка. Однако мы увидим, что судить о показателе степени хвоста трудно, что калибровка чревата ошибками из-за неэффективного сбора данных в хвостах. Вообще говоря, по данным выборки хвост рисуется тоньше, чем должен быть.

Мы обсудим эту проблему глубже в следующих главах.

## 5.2. НЕКОТОРЫЕ СВОЙСТВА СТЕПЕННЫХ ЗАКОНОВ

Два центральных свойства.

### 5.2.1. Сумма случайных величин

#### Свойство 1: Показатель степени хвоста для суммы

Пусть  $X_1, X_2, \dots, X_n$  — случайные величины, без требований независимости и одинакового распределения, и пусть каждый  $X_i$  следует распределению со своим асимптотическим показателем степени хвоста  $\alpha_i$  (мы предполагаем, что у случайных величин вне класса степенного закона будет асимптотическая альфа  $= +\infty$ ). Допустим далее, что нас интересует правый хвост распределения (наши доводы останутся в силе и применительно к левому хвосту). Дополнительные подробности смотрите в [99].

Рассмотрим взвешенную сумму  $S_n = \sum_{i=1}^n w_i X_i$ , причем все веса  $w_i$  строго положительны. Рассмотрим  $\alpha_S$ , показатель степени хвоста для суммы.

При любых  $w_i > 0$

$$\alpha_S = \min(\alpha_i).$$

Это очевидно для двух слагаемых, потому что при  $\alpha_2 \leq \alpha_1$  и  $w_2 > 0$

$$\lim_{x \rightarrow \infty} \frac{\ln(w_1 x^{-\alpha_1} + w_2 x^{-\alpha_2})}{\ln x} = \alpha_2.$$

Следствие: если в сумму добавить одно-единственное слагаемое с неопределенным (или бесконечным) средним, дисперсией или высшими моментами, то у всей суммы сделается неопределенное (или бесконечное) среднее, дисперсия или высшие моменты.

### Принцип 5.1 (Степенные законы + Тонкие хвосты = Степенные законы)

*При смешении в одной сумме величин, распределенных по степенному закону и тонкохвостых, результат будет распределен по степенному закону, независимо от подробностей состава.*

### 5.2.2. Преобразования

Следующее свойство звучит безобидно, но способно доставить кучу неприятностей.

#### Свойство 2

Пусть  $X$  — случайная величина с показателем степени хвоста  $\alpha$ . Для  $X^p$  показатель степени хвоста составит  $\frac{\alpha}{p}$ .

Тем самым эксцесс случайной величины с конечной дисперсией, но показателем степени хвоста  $< 4$  будет бесконечен<sup>1</sup>. Мы еще увидим, как это создает проблемы для стохастических моделей волатильности, когда в практике действительно встречается бесконечная дисперсия дисперсии.

Глядя на это, можно догадаться, не вдаваясь в технические выкладки, каким образом выпуклое вниз преобразование случайной величины утолщает хвост.

*Доказательство.* В общем случае подход такой. Пусть  $f(\cdot)$  — плотность распределения вероятностей, и  $\varphi(\cdot)$  — преобразование (с некоторыми ограничениями). Распределение преобразованной случайной величины (в предположении, что преобразование не изменяет носитель случайной величины):<sup>2</sup>

$$f(\varphi(x)) = \frac{f(\varphi^{-1}(x))}{\varphi'(\varphi^{-1}(x))}. \quad (5.8)$$

Допустим, что  $x > l$  и  $l$  велико (т. е. это точка, с точностью до порядка величины, после которой медленно меняющаяся функция «перестает изменяться»). Плотность

<sup>1</sup> Например, если случайная величина  $X$  имеет показатель хвоста 3,5, то у нее существует дисперсия  $\mathbb{V}X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$ , ведь у  $X^2$  показатель хвоста  $\frac{3,5}{2}$  больше 1, и  $\mathbb{E}X^2$  конечное. Однако эксцесс

$\text{Kurt } X = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{(\mathbb{V}X)^2} = \frac{\mathbb{E}X^4 - 4\mathbb{E}X\mathbb{E}X^3 + 6(\mathbb{E}X)^2\mathbb{E}X^2 - 3(\mathbb{E}X)^4}{(\mathbb{V}X)^2}$  уже будет бесконечным из-за бесконечности

$\mathbb{E}X^4$ , ведь у  $X^4$  показатель хвоста  $\frac{3,5}{4}$  меньше 1.

<sup>2</sup> Далее использована нотация  $f(\varphi(x))$  вместо  $f_{\varphi(x)}(\varphi(x))$ .

вероятности при таких значениях  $x$  упрощается до  $f(x) \propto x^{-\alpha-1}$ . Рассмотрим преобразование  $y = \varphi(x) = x^p$ ; обратной функцией для  $y = x^p$  будет  $x = y^{\frac{1}{p}}$ . Подставляя ее в знаменатель формулы (5.8), получаем  $\frac{1}{p} x^{\frac{1-p}{p}}$ .

Интегрируя выше  $l$ , получаем функцию выживания  $\mathbb{P}(Y > y) \propto y^{-\frac{\alpha}{p}+1}$ . ■<sup>2</sup>

### 5.3. КОЛОКОЛООБРАЗНЫЕ И ДРУГИЕ СТЕПЕННЫЕ ЗАКОНЫ

#### Влияние медленно меняющейся функции, расследование

Чем жирнее хвосты, тем меньше влияет «туловище» на моменты (которые, начиная с некоторой жирности, становятся бесконечными). Но при степенных законах с более тонкими хвостами область вне степенного хвоста (медленно меняющаяся часть) все-таки играет роль; «медленное изменение» более-менее формально определено в разделах 5.1.1 и 18.2.2. В этом разделе будет показано, насколько различаются формы у распределений, которые кажутся<sup>3</sup> одинаковыми.

Сравним двойное распределение Парето, имеющее плотность вероятности

$$f_p(X) = \begin{cases} \alpha(1+x)^{-\alpha-1}, & x \geq 0, \\ \alpha(1-x)^{-\alpha-1}, & x < 0 \end{cases},$$

и  $t$ -распределение Стюдента, имеющее тот же параметр центральности 0, параметр масштаба  $s$  и плотность вероятности

$$f_s(X) = \frac{\alpha^{\alpha/2} \left( \alpha + \frac{x^2}{s^2} \right)^{\frac{1}{2}(-\alpha-1)}}{sB\left(\frac{\alpha}{2}, \frac{1}{2}\right)},$$

где  $B(., .)$  — бета-функция Эйлера

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1} (1-t)^{b-1} dt.$$

1 В самом деле, для рассматриваемого преобразования по степенному закону  $\varphi(x) = x^p$  производная

$\varphi'(x) = px^{p-1}$  и обратное преобразование  $\varphi^{-1}(y) = y^{\frac{1}{p}}$ . У рассматриваемой исходной случайной величины  $X$  в далекой части хвоста, при  $x \rightarrow \infty$ , плотность вероятности  $f_X(x) \propto x^{-\alpha-1}$ , значит, для случайной величины  $Y = \varphi(X)$  плотность вероятности в далекой части хвоста, при  $y \rightarrow \infty$

$$f_Y(y) = \frac{f_X(\varphi^{-1}(y))}{\varphi'(\varphi^{-1}(y))} \propto \frac{(\varphi^{-1}(y))^{-\alpha-1}}{p(\varphi^{-1}(y))^{p-1}} = \frac{1}{p} (\varphi^{-1}(y))^{-\alpha-p} = \frac{1}{p} \left( y^{\frac{1}{p}} \right)^{-\alpha-p} = \frac{1}{p} y^{-\frac{\alpha+p}{p}}.$$

Соответственно функция выживания в далекой части хвоста, при  $y \rightarrow \infty$

$$\mathbb{P}(Y > y) = \int_y^{\infty} f_Y(y) dy \propto \int_y^{\infty} \frac{1}{p} y^{-\frac{\alpha+p}{p}} dy = -\frac{1}{p} \cdot \frac{p}{\alpha} \cdot y^{-\frac{\alpha}{p}} \Big|_y^{\infty} = \frac{1}{p} \cdot \frac{p}{\alpha} \cdot y^{-\frac{\alpha}{p}} \propto y^{-\frac{\alpha}{p}}.$$

2 Этим знаком обозначают, чаще в английской литературе, окончание доказательства. Вместо «ЧТД», «что и требовалось доказать».

3 Если судить по младшим статистическим моментам.

У нас есть два удобных способа сравнить эти распределения.

- Вывернуть распределения по хвостам: потребовать  $\lim_{x \rightarrow \infty} \frac{f_p(x)}{f_s} = 1$  и подобрать «равнохвостое» t-распределение Стьюдента с параметром масштаба

$$s = \left( \alpha^{1-\frac{\alpha}{2}} B\left(\frac{\alpha}{2}, \frac{1}{2}\right) \right)^{\frac{1}{\alpha}}.$$

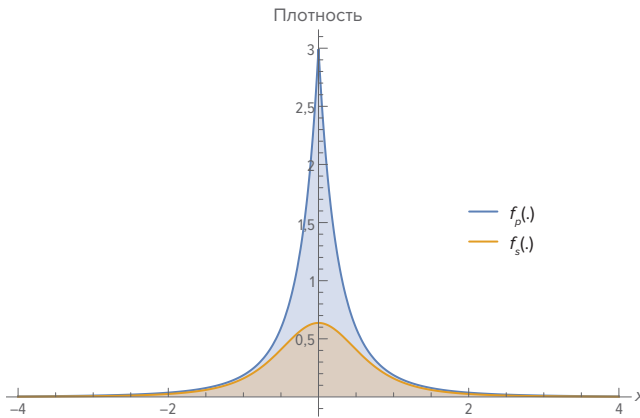
- Вывернуть распределения по среднеквадратичным отклонениям (если они конечны). Пока что у нас, при  $\alpha > 2$ ,

$$\mathbb{E} X_p^2 = \frac{2}{\alpha^2 - 3\alpha + 2} \text{ и } \mathbb{E} X_s^2 = \frac{\alpha}{\alpha - 2} \left( \alpha^{1-\frac{\alpha}{2}} B\left(\frac{\alpha}{2}, \frac{1}{2}\right) \right)^{\frac{2}{\alpha}}.$$

Растянем второй график так, что  $\sqrt{\mathbb{E} X_p^2} = \sqrt{k} \sqrt{\mathbb{E} X_s^2}$ :

$$k = \frac{2\alpha^{-2/\alpha} B\left(\frac{\alpha}{2}, \frac{1}{2}\right)^{-2/\alpha}}{\alpha - 1}.$$

Теперь мы совместили хвостами «колоколообразное» квазивыпуклое вверх распределение и «треугольное» дважды выпуклое вниз; см. Рисунок 5.6.



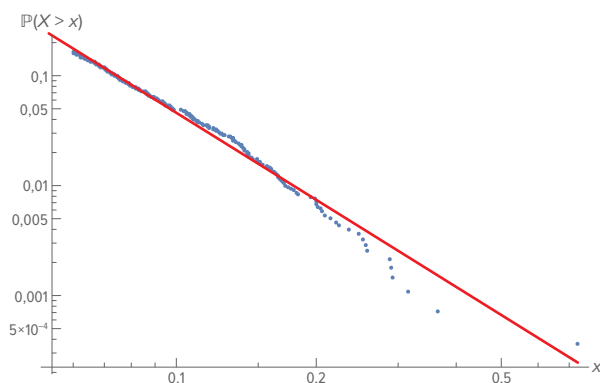
**Рисунок 5.6:** Сравним два симметричных распределения степенного закона: у одного медленно меняющаяся компонента короткоживущая, у другого действует в расширенной области. Все моменты в итоге одинаковы, хотя в центре, в области небольших отклонений, формы разные

#### 5.4. ИНТЕРПОЛЯЦИОННЫЕ ВОЗМОЖНОСТИ СТЕПЕННЫХ ЗАКОНОВ: ПРИМЕР

Рассмотрим заявки на пособие по безработице во время пандемии COVID-19: в марте 2020 безработица подскочила на много так называемых среднеквадратических отклонений. Считать ли этот скачок выбросом? Да, если смотреть на Рисунок 5.7 взглядом, обученным на примерах тонких хвостов. Но это неправильный взгляд. Как видно на Рисунке 5.8, показатель степени хвоста практически не изменился. Пусть масштаб распределения изменился, показатель степени оказался робастным, устойчивым к добавлению в выборку новых наблюдений.



**Рисунок 5.7:** Число заявок на пособие по безработице подскочило сверх ожиданий? Только для неопытных экономистов. Как показывает Рисунок 5.8, неожиданностей не случилось. Опытный взгляд (а-ля Бенуа Мандельброт) и раньше видел бы, что, хотя дисперсия небольшая, данное распределение отнюдь не гауссово



**Рисунок 5.8:** График Ципфа<sup>1</sup> с числом заявок на пособие по безработице. Не обязательно было дожидаться скачка в пандемию COVID-19 (крайняя справа точка), чтобы заметить степенной закон

## 5.5. СВЕРХЖИРНЫЕ ХВОСТЫ: РАСПРЕДЕЛЕНИЕ ЛОГ-ПАРЕТО

Матерь всех жирных хвостов — распределение лог-Парето — не представлено в обычных списках распределений, но мы сейчас сами выведем его из распределения Парето по аналогии с логнормальным распределением.

### Замечание 5: Заново открываем распределение лог-Парето

Если  $X \sim \mathcal{P}(L, \alpha)$  — случайная величина с распределением Парето, имеющим плотность вероятности

$$f^{(P)}(x) = \alpha L^\alpha x^{-\alpha-1}, \quad x \geq L$$

и функцию выживания

$$S^{(P)}(x) = L^\alpha x^{-\alpha},$$

<sup>1</sup> Графиком Ципфа называют график, линейный в дважды логарифмическом масштабе; такова плотность распределения, порожденного по закону Ципфа, или Парето.

то  $e^x \sim \mathcal{LP}(L, \alpha)$  — случайная величина с распределением лог-Парето, имеющим плотность вероятности

$$f^{(LP)}(x) = \frac{\alpha L^\alpha (\ln x)^{-\alpha-1}}{x}, x \geq e^L$$

и функцию выживания

$$S^{(LP)}(x) = L^\alpha (\ln x)^{-\alpha}.$$

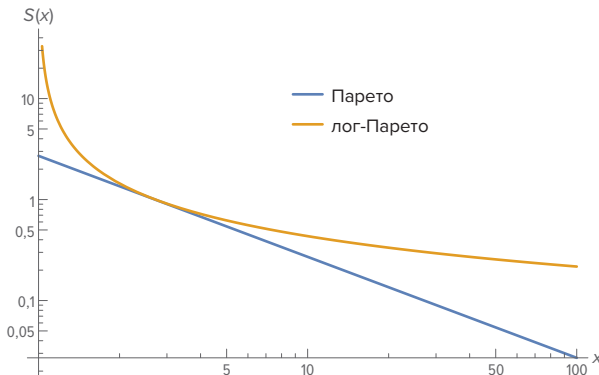
В то время как при обычном степенном законе на диаграмме в дважды логарифмическом масштабе график приближается к линейному склону, т. е.

$$\lim_{x \rightarrow \infty} \frac{\ln L^\alpha x^{-\alpha}}{\ln x} = -\alpha,$$

при лог-Парето наклон стремится к нулю:

$$\lim_{x \rightarrow \infty} \frac{\ln(L^\alpha (\ln x)^{-\alpha})}{\ln x} = 0,$$

и очевидно, что никакие моменты существовать не могут ни при каком параметре хвоста  $\alpha$ . Различие в асимптотическом поведении видно на Рисунке 5.9.



**Рисунок 5.9:** Сравним графики функций выживания для Парето и лог-Парето на дважды логарифмической диаграмме

## 5.6. ПСЕВДОСТОХАСТИЧЕСКАЯ ВОЛАТИЛЬНОСТЬ: РАССЛЕДОВАНИЕ

Мы упомянули ранее, в Главе 3, что заявления о событиях величиной «10 сигм» на самом деле означают, что мы не в гауссовом мире. Кроме того, мы обсудили проблему ненаблюдаемости распределения вероятностей: мы видим данные, а не порождающий процесс.

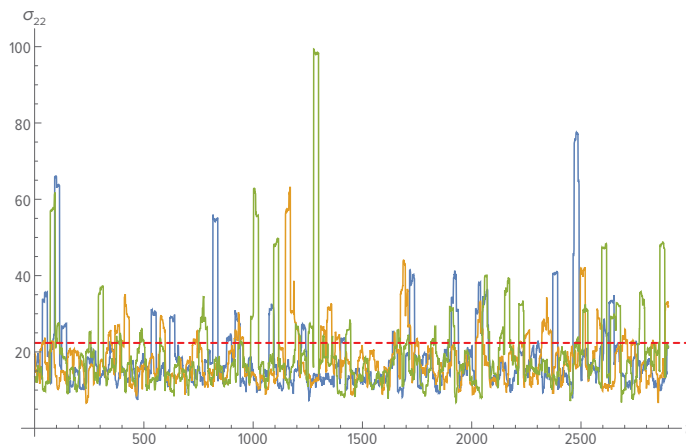
Поэтому так легко обмануться, приняв степенной закон за гетероскедастичный процесс. Задним числом всегда можно сказать: «Условная волатильность высока, и при новом среднеквадратическом отклонении данное событие оценивается уже не как 10 сигм, а всего лишь 3 сигмы».

Чтобы развенчать заявления этого рода, можно прибегнуть к рассуждению от обратной задачи: показать, как степенной закон постоянного масштаба может выда-



вать себя за гетероскедастичный процесс. В Приложении мы увидим, что опора на гетероскедастичность (т. е. изменяющуюся дисперсию)<sup>1</sup> в эконометрике имеет серьезные дефекты, поскольку дисперсия дисперсии лишена структуры.

На Рисунке 5.10 показана волатильность для доходности на некотором рынке, которая ужасно похожа на результат простого процесса со стохастической волатильностью; будем считать, что дисперсия ведет себя как случайная величина, и подберем параметры ее распределения<sup>2</sup>.



**Рисунок 5.10:** 22-дневная (за рабочий месяц) оценка исторической волатильности (среднеквадратического отклонения) доходности по ежедневным выборкам, сгенерированным t-распределением Стьюдента. Производит впечатление стохастической волатильности, хотя на самом деле масштаб распределения не изменялся

Пусть  $X$  — доходность с математическим ожиданием 0, масштабом  $\sigma$  и плотностью вероятности  $\varphi(\cdot)$ :

$$\varphi(X) = \frac{\left( \frac{\alpha}{\alpha + \frac{x^2}{\sigma^2}} \right)^{\frac{\alpha+1}{2}}}{\sqrt{\alpha} \sigma B\left(\frac{\alpha}{2}, \frac{1}{2}\right)}, \quad x \in (-\infty, \infty).$$

Преобразованная случайная величина  $Y = X^2$  (распределение которой даст нам распределение второго момента) будет иметь плотность

$$\psi(y) = \frac{\left( \frac{\alpha \sigma^2}{\alpha \sigma^2 + y} \right)^{\frac{\alpha+1}{2}}}{\sigma B\left(\frac{\alpha}{2}, \frac{1}{2}\right) \sqrt{\alpha y}}, \quad y \in (0, \infty),$$

1 От др.-греч. ἑτερος «другой» и σκέδασις «дисперсия».

2 Есть модель, стохастизирующая дисперсию, и есть модель, стохастизирующая среднеквадратическое отклонение; это две разные модели с разными ожиданиями. — Прим. автора.

и хвост этой функции можно преобразовать асимптотически в степенной закон с показателем степени  $\frac{\alpha}{2}$ . Характеристическую функцию  $\chi_Y(\omega) = \mathbb{E} e^{i\omega Y}$  можно записать как

$$\chi_Y(\omega) = \frac{1}{2B\left(\frac{\alpha}{2}, \frac{1}{2}\right)} \pi \operatorname{cosec} \pi \alpha \left( \frac{\sqrt{\pi} {}_1\tilde{F}_1\left(\frac{1}{2}; 1 - \frac{\alpha}{2}; -i\alpha\sigma^2\omega\right)}{\Gamma\frac{\alpha+1}{2}} - \right. \\ \left. - (-i\alpha\sigma^2\omega)^{a/2} {}_1\tilde{F}_1\left(\frac{\alpha+1}{2}; \frac{\alpha+2}{2}; -i\alpha\sigma^2\omega\right) \right). \quad (5.9)$$

Отсюда получаем среднее абсолютное отклонение второго момента:<sup>1</sup>

$\alpha$	MAD второго момента
$\frac{5}{2}$	$\frac{\sqrt[4]{\frac{5}{3}} 2^{3/4} \left( {}_2F_1\left(\frac{1}{4}, \frac{7}{4}; \frac{5}{4}; -\frac{5}{6}\right) + 3\left(\frac{6}{11}\right)^{3/4} \right) \sigma^2 \Gamma \frac{7}{4}}{\sqrt{\pi} \Gamma \frac{5}{4}}$
3	$\frac{6\sigma^2}{\pi}$
$\frac{7}{2}$	$\frac{5 \cdot 7^{3/4} \left( {}_7F_1\left(\frac{3}{4}, \frac{9}{4}; \frac{7}{4}; -\frac{7}{6}\right) - 3 {}_2F_1\left(\frac{7}{4}, \frac{9}{4}; \frac{11}{4}; -\frac{7}{6}\right) \right) \sigma^2 \Gamma \frac{5}{4}}{6 \cdot 6^{3/4} \sqrt{\pi} \Gamma \frac{7}{4}}$
4	$\frac{1}{7} (3\sqrt{21} - 7) \sigma^2$
$\frac{9}{2}$	$\frac{3 \sqrt[4]{\frac{3}{2}} \left( 6\left(\frac{2}{5}\right)^{3/4} - {}_6F_1\left(\frac{5}{4}, \frac{11}{4}; \frac{9}{4}; -\frac{3}{2}\right) \right) \sigma^2 \Gamma \frac{11}{4}}{5\sqrt{\pi} \Gamma \frac{7}{4}}$
5	$\frac{\sigma^2 \left( 7\sqrt{15} - 16 \arctg \sqrt{\frac{5}{3}} \right)}{6\pi}$

## ДАЛЕЕ

В следующей главе мы ринемся в высшие размерности. Некоторые следствия очевидны, другие не столь очевидны — например, существование корреляций даже там, где не существует ковариаций.

<sup>1</sup> Как обычно, мы не пользуемся таким показателем, как среднеквадратическое отклонение, поскольку он неустойчив и малоинформативен; вместо него берем среднее абсолютное отклонение. — Прим. автора.

## ЖИРНЫЕ ХВОСТЫ В ВЫСШИХ РАЗМЕРНОСТЯХ<sup>1</sup>

В этом обсуждении рассматривается максимально упрощенная работа с высокими размерностями. Мы познакомимся с (1) простым влиянием жирных хвостов на многомерную случайную величину, (2) эллиптичностью и распределениями, (3) случайными матрицами и связанными с ними распределениями собственных значений, (4) как можно смотреть на ковариацию и корреляции, когда моменты не существуют (например, в случае распределения Коши).

### 6.1. ТОЛСТЫЕ ХВОСТЫ ПРИ ВЫСОКОЙ РАЗМЕРНОСТИ, КОНЕЧНЫЕ МОМЕНТЫ

Построим интуитивную картину толстых хвостов от выпуклости вниз до масштаба подобно тому, как действовали в предыдущей главе, но для многомерных случайных величин.

Пусть  $\vec{X} = (X_1, X_2, \dots, X_m)$  — случайный вектор, или случайная величина размерности  $m \times 1$ , где все компоненты гауссовы. Рассмотрим совместное распределение вероятности с плотностью  $f(x_1, \dots, x_m)$ . Будем обозначать  $m$ -мерное нормальное распределение  $\mathcal{N}(\vec{\mu}, \Sigma)$ , где вместо математического ожидания вектор математического ожидания  $\vec{\mu}$ , вместо дисперсии ковариационная матрица  $\Sigma$ , плотность совместного распределения вероятностей

$$f(\vec{x}) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1} \sum^{-1} (\vec{x} - \vec{\mu})\right), \quad (6.1)$$

$\vec{x} = (x_1, x_2, \dots, x_m) \in \mathbb{R}^m$  и  $\Sigma$  — симметричная положительно-определенная<sup>1</sup> матрица размерности  $m \times m$ .

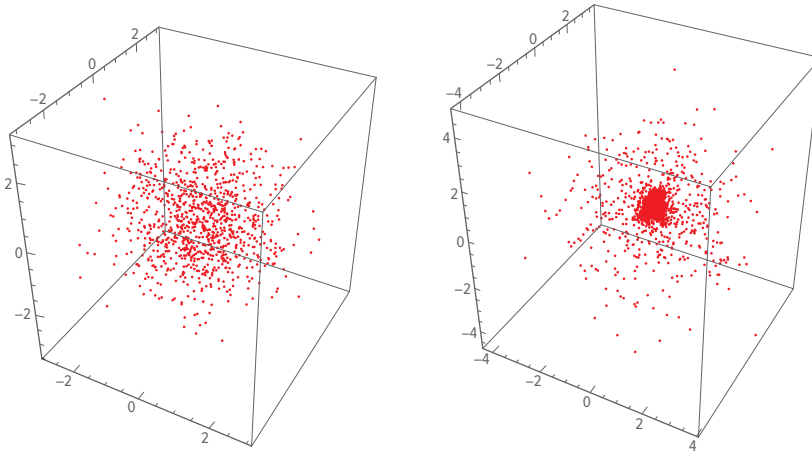
Для ужирнения хвостов мы можем применить упрощенную эвристику, сохраняющую дисперсию, как в главе 4.1:

<sup>1</sup> В английской литературе под положительно-определенной матрицей подразумевается матрица с действительными элементами, удовлетворяющая требованию  $z^T \Sigma z > 0$  для всякого столбца  $z$ . В русской литературе подразумевается более широкое толкование: эрмитова матрица, т. е. с комплексными элементами, но с такой комплексно-сопряженной матрицей, что  $\bar{\Sigma} = \Sigma^T$  и по-прежнему требуется  $z^T \Sigma z > 0$  для всякого столбца  $z$ .

$$f_a(\vec{x}) = \frac{1}{2} (2\pi)^{-m/2} |\Sigma_1|^{-1/2} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma_1^{-1}(\vec{x} - \vec{\mu})\right) + \\ + \frac{1}{2} (2\pi)^{-m/2} |\Sigma_2|^{-1/2} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma_2^{-1}(\vec{x} - \vec{\mu})\right), \quad (6.2)$$

где скалярный параметр  $a$  регулирует интенсивность вносимой стохастической волатильности:  $\Sigma_1 = \Sigma(1+a)$  и  $\Sigma_2 = \Sigma(1-a)$ <sup>1</sup>.

Отметим на Рисунке 6.1, что, как и в одномерном случае, наблюдения концентрируются в средней части распределения<sup>2</sup>.



**Рисунок 6.1:** Толстые хвосты в высших размерностях. 3-мерный случайный вектор с тонкими хвостами (слева) или толстыми хвостами (справа) при одной и той же дисперсии. Если в одномерном случае была колоколообразная кривая плотности вероятности, у которой по мере утолщения хвостов становился выше пик (над «туннелем»), здесь мы видим рост плотности точек вблизи центра

## 6.2. СОВМЕСТНАЯ ЖИРНОХВОСТОСТЬ И ЭЛЛИПТИЧНОСТЬ РАСПРЕДЕЛЕНИЙ

С переходом от одномерной случайной величины к случайному вектору появляется новый аспект в определении жирнохвостости.

**Что такое распределение с эллиптическими контурами?** Согласно стандартному определению [88], случайный вектор  $\mathbf{X}$  размерности  $p \times 1$  имеет эллиптическое распределение (или распределение с эллиптическими контурами) с параметрами положения  $\vec{\mu}$ , неотрицательной матрицей  $\Sigma$  и некоторой скалярной функцией  $\Psi$ , если характеристическая функция  $\varphi$  имеет вид

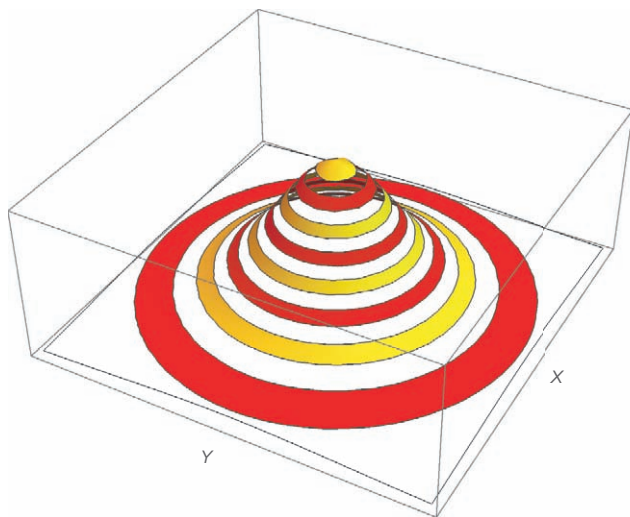
$$\varphi(t) = \exp(it' \vec{\mu}) \Psi(t' \Sigma t). \quad (6.3)$$

<sup>1</sup> Как и в одномерном случае, мы можем без малейшей потери общности предположить для простоты, что  $\vec{\mu} = (0, \dots, 0)$ . — Прим. автора.

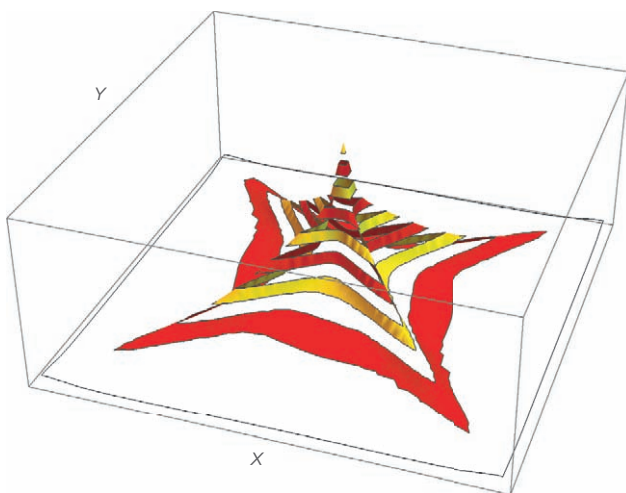
<sup>2</sup> Мы создали жирные хвосты, стохастизировав дисперсию, но сохранив постоянными корреляции; тем самым мы сохранили положительную определенность матрицы. — Прим. автора.

Есть эквивалентные определения, которые фокусируются на плотности; давайте пока ограничимся главным атрибутом — тем, что  $\Psi$  оказывается функцией одной-единственной ковариационной матрицы  $\Sigma$ .

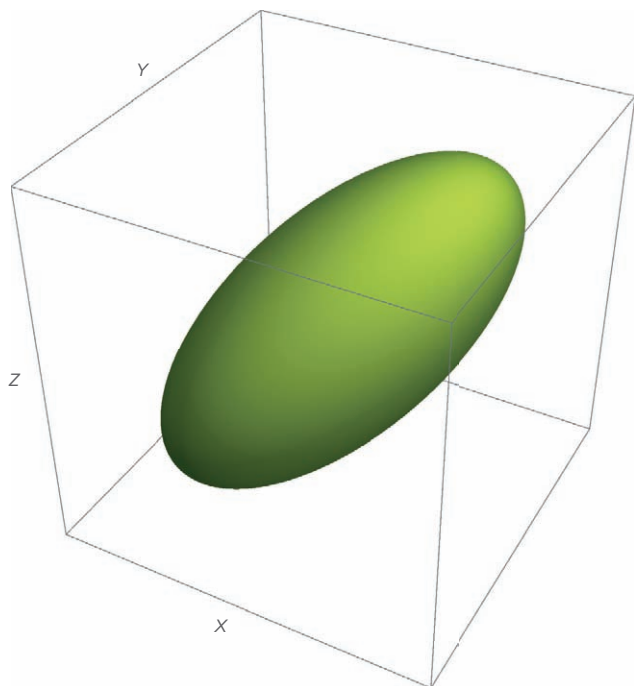
Интуитивное представление дает эллиптическая форма линий равной плотности на диаграмме распределения; посмотрите на 2-мерную диаграмму плотности (для совместного распределения двух случайных величин) и на 3-мерную диаграмму плотности (для совместного распределения трех случайных величин) на Рисунках 6.2 и 6.4. Неэллиптическое распределение нарушает эту форму, как показано на Рисунках 6.3 и 6.5.



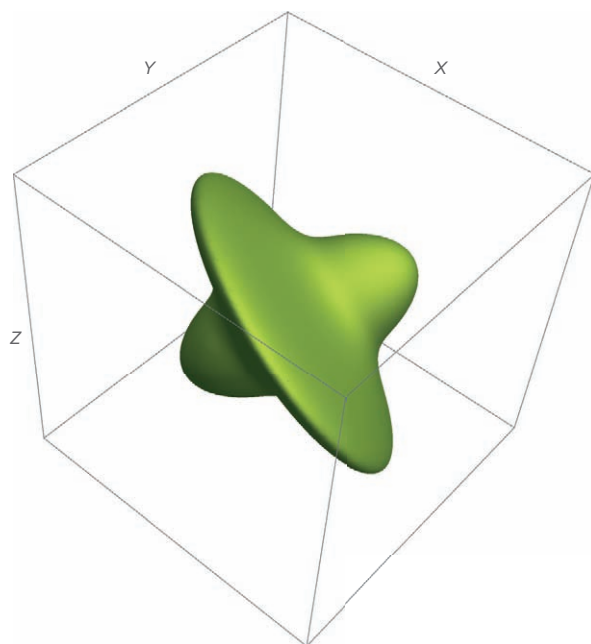
**Рисунок 6.2:** Эллиптические контуры совместной доходности, распределенной по степенному закону (t-распределение Стьюдента)



**Рисунок 6.3:** Неэллиптическая совместная доходность, полученная при стохастических корреляциях



**Рисунок 6.4:** Совместная доходность с эллиптическим контуром, полученным из многомерного распределения  $(X, Y, Z)$  как множество точек равной плотности



**Рисунок 6.5:** Неэллиптическая совместная НСВ, полученная при стохастических корреляциях из многомерного распределения  $(X, Y, Z)$  как множество точек равной плотности

Важнейшее свойство класса эллиптического распределения — замкнутость при линейных преобразованиях. Интуитивно его можно понять по аналогии с примером в Главе 3<sup>1</sup>, про суммарный рост или благосостояние двоих людей (это совместное распределение двух случайных величин). Совместные хвостовые события двойного размера чаще происходят при отклонениях двойного размера, чем при совпадении двух отклонений.

**Эллиптичность и коренные изъяны в финансовой теории.** Замкнутость при линейных преобразованиях ведет к заманчивым свойствам, помогающим строить портфели, и ко всем результатам портфельной теории (собственно, нельзя построить портфельную теорию без предположения об эллиптичности распределений).

В ситуации эллиптичности все портфели можно полностью характеризовать положением и масштабом в пространстве доходности, и любые два портфеля с одинаковыми положением и масштабом обладают одинаковыми распределениями доходности.

Ирония в том, что распределения, устойчивые по Леви, действительно эллиптичны — по определению.

Итак, эллиптичность (при условии конечной дисперсии) позволяет распространить результаты современной портфельной теории<sup>2</sup> (*modern portfolio theory*, МРТ) при так называемой «ненормальности», впервые описанной здесь: [183]; см. также [121]. Однако, как очевидно тем, кто работает со стохастической ковариацией, доходность отнюдь не эллиптична, ни по каким мыслимым меркам; см. Бушо и Шишпортик [42] и простые наглядные диаграммы на Рисунке Е.8, отслеживающие корреляцию между доходностью и S&P 500.

Простой учебный пример, использующий эвристику  $1 \pm a$ , мы представили в главе 4.1. Рассмотрим двумерное нормальное распределение с характеристической функцией

$$\Psi(t_1, t_2) = p e^{-\rho t_1 t_2 - \frac{t_1^2}{2} - \frac{t_2^2}{2}}.$$

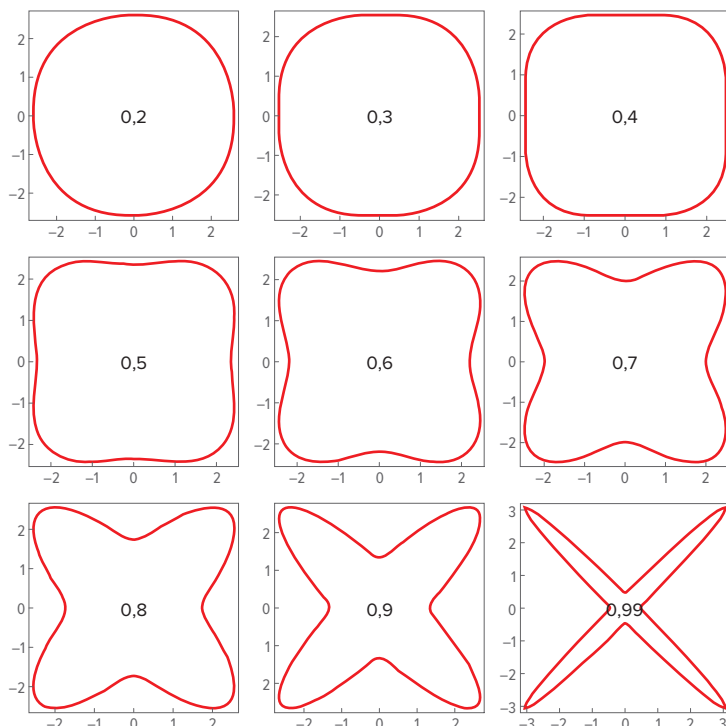
Теперь стохастизируем параметр  $\rho$ , задав с вероятностью  $p$  значение  $\rho_1$  и с вероятностью  $(1-p)$  значение  $\rho_2$ :

$$\Psi(t_1, t_2) = p e^{-\rho_1 t_1 t_2 - \frac{t_1^2}{2} - \frac{t_2^2}{2}} + (1-p) e^{-\rho_2 t_1 t_2 - \frac{t_1^2}{2} - \frac{t_2^2}{2}}. \quad (6.4)$$

На Рисунке 6.6 показаны результаты при  $p = \frac{1}{2}$  и различных  $\rho_1 = -\rho_2$ .

<sup>1</sup> См. Раздел 3.1.

<sup>2</sup> В русской литературе — портфельная теория Марковица.



**Рисунок 6.6:** Стохастическая корреляция для обычного двумерного нормального распределения. Показаны изолинии плотности для различных сочетаний параметров. Мы используем совсем простую технику формулы 6.4, переключаясь с вероятностью  $\frac{1}{2}$  между  $\rho_1 = \rho$  и  $\rho_2 = -\rho$  при различных  $\rho$

Рассуждая более формально, можно показать различие, возникающее при стохастической  $\Sigma$  между  $\Psi(t(\mathbb{E} \Sigma)t')$  и  $\mathbb{E} \Psi(t\Sigma t')$  в формуле 6.3.

### Диверсификация

Вспомним, что финансовая теория перестает работать при толстых хвостах (и никакие поправки не устранили проблему вне «чрезмерной подгонки», которую мы обсуждали в предыдущих главах). Отсутствие эллиптичности закрывает вопрос. Следствие: все методики на основе портфелей в духе Марковица, то есть основанные на идее диверсификации, не способны снижать риск, хотя способны порождать обманчиво сглаженную суточную волатильность. Добавление рычагов гарантирует взрыв конструкции в конечном итоге<sup>a</sup>.

<sup>a</sup> Это относится и к шарлатанскому «паритету рисков» (*risk parity*) и «сбору активов» (*asset gathering*), псевдотеоретической и псевдоакадемической дымовой завесе ради привлечения денег.



### 6.3. МНОГОМЕРНОЕ t-РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА

Многомерное t-распределение Стьюдента удобно для моделирования степенного закона, поскольку при  $\nu = 1$  схлопывается в распределение Коши. Другой вариант, многомерное устойчивое распределение, как мы увидим, не обладает плотностью.

Пусть  $\mathbf{X}$  — случайный вектор размерности  $p \times 1$  с многомерным t-распределением Стьюдента,  $\mathbf{X} \sim S_t(\mathbf{M}, \mathbf{\Sigma}, \nu)$ , где  $\mathbf{\Sigma}$  — матрица размерности  $p \times p$ ,  $\mathbf{M}$  — вектор длины  $p$  и  $\nu \geq 1$ ;<sup>1</sup> тогда плотность вероятности

$$f(\mathbf{X}) = \frac{\Gamma \frac{\nu+p}{2}}{\Gamma \frac{\nu}{2} \nu^{\frac{p}{2}} \pi^{\frac{p}{2}} |\mathbf{\Sigma}|^{\frac{1}{2}}} \left( \frac{(\mathbf{X}-\mathbf{M}) \cdot \mathbf{\Sigma}^{-1} \cdot (\mathbf{X}-\mathbf{M})}{\nu} + 1 \right)^{-\frac{1}{2}(\nu+p)}. \quad (6.5)$$

В простейшем случае, при  $p = 2$ ,  $\mathbf{M} = (0, 0)$  и  $\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ , получаем<sup>2</sup>

$$f(x_1, x_2) = \frac{\nu \sqrt{1-\rho^2} \left( \frac{-\nu \rho^2 + \nu - 2\rho x_1 x_2 + x_1^2 + x_2^2}{\nu - \nu \rho^2} \right)^{-\frac{\nu}{2}-1}}{2\pi(\nu - \nu \rho^2)}. \quad (6.6)$$

#### 6.3.1. Эллиптичность и независимость при толстых хвостах

Возьмем произведение двух плотностей Коши по  $x$  и по  $y$  (как мы уже делали на Рисунке 3.1):

$$f(x)f(y) = \frac{1}{\pi^2(x^2+1)(y^2+1)}, \quad (6.7)$$

которое, как мы видели в Главе 3 (на примере вероятности того, что два случайно выбранных человека имеют в сумме состояние 36 млн долларов), заведомо не эллиплично. Сравним с совместным распределением, имеющим коэффициент корреляции  $\rho$ :

$$f_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2} \left( y \left( \frac{y}{1-\rho^2} - \frac{\rho x}{1-\rho^2} \right) + x \left( \frac{x}{1-\rho^2} - \frac{\rho y}{1-\rho^2} \right) + 1 \right)^{3/2}}, \quad (6.8)$$

и, задав  $\rho = 0$  для ситуации без корреляции, получим

$$f_0(x, y) = \frac{1}{2\pi(x^2 + y^2 + 1)^{3/2}}, \quad (6.9)$$

<sup>1</sup> Со средним  $\mathbf{M}$ , если  $\nu > 1$ , с дисперсией  $\frac{\nu}{\nu-2} \mathbf{\Sigma}$ , если  $\nu > 2$ , и числом степеней свободы  $\nu$ .

<sup>2</sup> Пользуясь тем, что:  $\Gamma \frac{\nu+2}{2} / \Gamma \frac{\nu}{2} = \frac{\nu}{2}$ ;  $|\mathbf{\Sigma}| = \begin{vmatrix} 1 & \rho \\ \rho & 1 \end{vmatrix} = 1 - \rho^2$ ;  $\mathbf{\Sigma}^{-1} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$ ;  $(\mathbf{X}-\mathbf{M}) \cdot \mathbf{\Sigma}^{-1} \cdot (\mathbf{X}-\mathbf{M}) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{-2\rho x_1 x_2 + x_1^2 + x_2^2}{1-\rho^2}$ .

и это распределение эллиптическое. Наш пример иллюстрирует, каким образом отсутствие корреляции не означает независимости, поскольку

- независимость между случайными величинами  $X$  и  $Y$  определяется равенством

$$\frac{f(x, y)}{f(x)f(y)} = 1$$

- независимо от коэффициента корреляции. В классе эллиптических распределений двумерная гауссиана с коэффициентом 0 и независима, и некоррелирована. Это неприменимо к  $t$ -распределению Стьюдента или к распределению Коши.

Поясним, почему многомерное устойчивое распределение с нулевым коэффициентом корреляции не гарантирует независимость.

О случайном векторе  $\mathbf{X} = (X_1, X_2, \dots, X_k)$  говорят, что его многомерное распределение устойчиво, если устойчиво распределение каждой линейной комбинации компонент  $Y = a_1 X_1 + \dots + a_k X_k$ . Другими словами, для всякого постоянного вектора  $\mathbf{a} \in \mathbb{R}^k$  требуется устойчивое распределение одномерной случайной величины  $Y = \mathbf{a}^T \mathbf{X}$ . Если линейная комбинация всегда остается в том же классе, которому принадлежат комбинируемые компоненты, класс эллиптичен. По нашему построению, напротив, плотность  $f_0(x, y)$  не обязана совпадать с  $f(x)f(y)$ . Рассмотрим случай Коши, где плотность задана явно. Знаменатель произведения плотностей содержит дополнительный член,  $x^2 y^2$ <sup>1</sup>, который сдвигает изолинии плотности в том или ином направлении, как мы видели во вступительных примерах в Главе 3.

#### 6.4. ЖИРНЫЕ ХВОСТЫ И ВЗАИМНАЯ ИНФОРМАЦИЯ

Отметим, что при искусственном построении многомерных распределений взаимная информация отлична от 0, несмотря на независимость, поскольку отношение совместной плотности к произведению плотностей отлично от 1, несмотря на нулевой коэффициент корреляции  $\rho$ .

Какова взаимная информация при  $t$ -распределении Стьюдента (включая распределение Коши)?

$$\mathbb{I}(X, Y) = \mathbb{E} \ln \frac{f(x, y)}{f(x)f(y)},$$

где математическое ожидание берется от совместного распределения  $X$  и  $Y$ . Благодаря логарифмированию взаимная информация аддитивна (заметим, что если требуется

<sup>1</sup> В самом деле, для стандартных распределений Коши с плотностью  $f(x) = \frac{1}{\pi(1+x^2)}$  и  $f(y) = \frac{1}{\pi(1+y^2)}$  произведение  $f(x)f(y) = \frac{1}{\pi^2(x^2+1)(y^2+1)} = \frac{1}{\pi^2(x^2+y^2+1+x^2y^2)}$ .

считать информацию в битах, вместо натурального логарифма  $\ln$  нужно взять  $\log_2$  или оставить  $\ln$ , но результат потом разделить на  $\ln 2$ ).

Таким образом,  $\mathbb{I}(X, Y) = \mathbb{E} \ln f(x, y) - \mathbb{E} \ln f(x) - \mathbb{E} \ln f(y)$ , или  $\mathbb{H}X + \mathbb{H}Y - \mathbb{H}(X, Y)$ , где  $\mathbb{H}$  — энтропия, и  $\mathbb{H}(X, Y)$  — совместная энтропия.

Заметим, что для гауссовых случайных величин взаимная информация представима как  $-\frac{1}{2} \ln(1 - \rho^2)$  при любой параметризации. Поэтому для случайных величин  $X, Y \sim$  многомерное t-распределение Стьюдента с параметрами  $(\alpha, \rho)$  взаимная информация

$$\mathbb{I}_\alpha(X, Y) = -\frac{1}{2} \ln(1 - \rho^2) + \lambda_\alpha, \quad (6.10)$$

где

$$\begin{aligned} \lambda_\alpha = & -\frac{2}{\alpha} + \ln \alpha + 2\pi(\alpha+1) \operatorname{cosec} \pi\alpha + \ln B\left(\frac{\alpha}{2}, \frac{1}{2}\right) - \\ & -(\alpha+1)H_{-\frac{\alpha}{2}} + (\alpha+1)H_{-\frac{\alpha}{2}-\frac{1}{2}} - \ln 2\pi, \end{aligned} \quad (6.11)$$

где  $\operatorname{cosec}(\cdot)$  — косеканс,  $B(\cdot, \cdot)$  — бета-функция и  $H$  — гармоническое число, в общем случае  $H_n^{(r)} = \sum_{i=1}^n \frac{1}{i^r}$ , в нашем случае  $H_n = H_n^{(1)}$ . Заметим, что  $\lambda_\alpha \rightarrow 0$  при  $\alpha \rightarrow \infty$ .

В завершении этой главки подчеркнем, что показатели, связанные с энтропией, такие как взаимная информация, говорят куда больше, чем корреляция; взаимная информация способна обнаруживать нелинейности.

## 6.5. ЖИРНЫЕ ХВОСТЫ И СЛУЧАЙНЫЕ МАТРИЦЫ, КРАТКОЕ ОТСТУПЛЕНИЕ

У собственных значений матриц есть аналог гауссовой сходимости: распределение по полукруговому закону, как показано на Рисунке 6.9.

Пусть  $\mathbf{M}$  — симметричная матрица  $n \times n$ . Собственные значения — это такие  $\lambda_i$  при  $1 \leq i \leq n$ , что  $\mathbf{M} \cdot \mathbf{V}_i = \lambda_i \mathbf{V}_i$ , где  $\mathbf{V}_i$  — это  $i$ -тый собственный вектор.

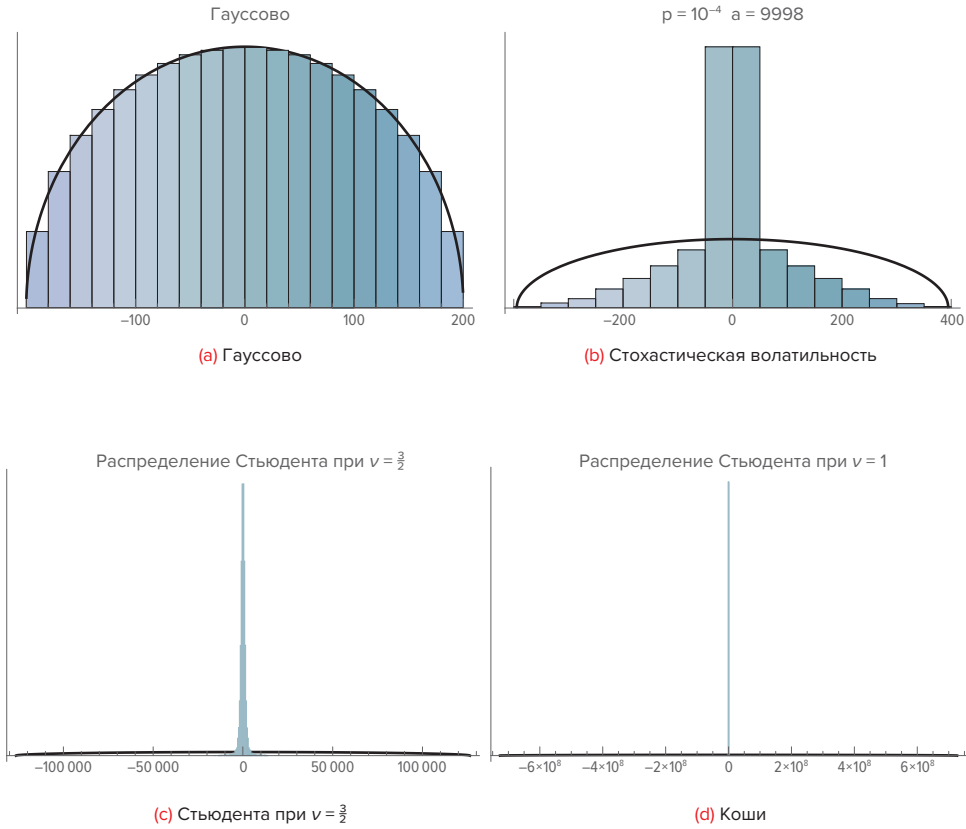
Распределение по полукруговому закону Вигнера с носителем  $[-R, R]$  и плотностью  $f$  в форме полуокружности радиуса  $R$  с центром в точке  $(0, 0)$  после полной нормализации имеет вид

$$f(\lambda) = \frac{2}{\pi R^2} \sqrt{R^2 - \lambda^2} \quad \text{при } -R \leq \lambda \leq R. \quad (6.12)$$

Это распределение возникает как предельное распределение собственных значений симметричной матрицы  $n \times n$  с конечными моментами, когда размер матрицы  $n$  стремится к бесконечности.

Ниже мы совершим экскурсию по «жирнохвостости» случайной матрицы и сходимости.

Таков эквивалент толстых хвостов для матриц. Ограничимся пока тем, что когда 4-й момент достигает гауссовых уровней (например, 3) для одномерной ситуации, то это эквивалентно тому, что собственные значения приближаются к полукруговому закону Вигнера.



**Рисунок 6.7:** Как различаются формы распределения собственных значений случайной матрицы, которое в Гауссовом случае следует полукруговому закону Вигнера. Случай Коши представим как  $t$ -распределение Стюдента при числе степеней свободы  $\nu = 1$

## 6.6. КОРРЕЛЯЦИЯ И НЕОПРЕДЕЛЕННАЯ ДИСПЕРСИЯ

Мы сейчас обсудим один парадокс: даже при бесконечных ковариациях корреляции остаются конечны. Правда, они информативны только при огромных выборках, та же проблема, что мы обсуждали в связи с методом главных компонент в Главе 3.

Вопрос: каким образом распределение с жирными хвостами в классе степенных законов  $\mathfrak{P}$  с бесконечным или неопределенным математическим ожиданием (и высшими моментами) может в многомерном случае, имея неопределенную (или бесконечную) ковариацию, иметь конечную корреляцию?

Рассмотрим распределение с носителем  $(-\infty, \infty)$ . У него нет моментов:  $\mathbb{E} X$  неопределенное,  $\mathbb{E} X^2 = \infty$ , ковариации нет,  $\mathbb{E} XY$  неопределенное. Но корреляция (не центральная) для  $n$  переменных ограничена  $-1$  и  $1$ .

$$r \triangleq \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \text{ при } n = 2, 3, \dots$$

В силу свойства субэкспоненциальности мы имеем  $\mathbb{P}(X_1 + \dots + X_n > x) \sim \mathbb{P}(\max(X_1, \dots, X_n) > x)$  при  $x \rightarrow \infty$ . Заметим, что класс степенного закона включен в субэкспоненциальный класс  $\mathfrak{S}$ .

Отсортируем переменные по абсолютному значению так, что  $|x_1| \leq |x_2| \leq \dots \leq |x_n|$ .

Пусть  $\kappa_1 = \sum_{i=1}^{n-1} x_i y_i$ ,  $\kappa_2 = \sum_{i=1}^{n-1} x_i^2$  и  $\kappa_3 = \sum_{i=1}^{n-1} y_i^2$ .

$$\lim_{x_n \rightarrow \infty} \frac{x_n y_n + \kappa_1}{\sqrt{x_n^2 + \kappa_2} \sqrt{y_n^2 + \kappa_3}} = \frac{y_n}{\sqrt{\kappa_3 + y_n^2}},$$

$$\lim_{y_n \rightarrow \infty} \frac{x_n y_n + \kappa_1}{\sqrt{x_n^2 + \kappa_2} \sqrt{y_n^2 + \kappa_3}} = \frac{x_n}{\sqrt{\kappa_2 + x_n^2}},$$

$$\lim_{\substack{x_n \rightarrow +\infty \\ y_n \rightarrow +\infty}} \frac{x_n y_n + \kappa_1}{\sqrt{x_n^2 + \kappa_2} \sqrt{y_n^2 + \kappa_3}} = 1,$$

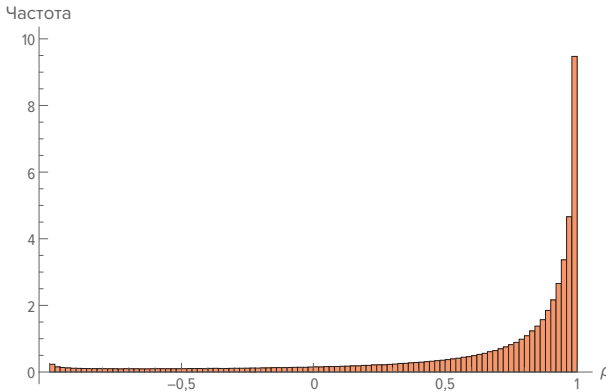
$$\lim_{\substack{x_n \rightarrow +\infty \\ y_n \rightarrow -\infty}} \frac{x_n y_n + \kappa_1}{\sqrt{x_n^2 + \kappa_2} \sqrt{y_n^2 + \kappa_3}} = -1$$

и

$$\lim_{\substack{x_n \rightarrow -\infty \\ y_n \rightarrow +\infty}} \frac{x_n y_n + \kappa_1}{\sqrt{x_n^2 + \kappa_2} \sqrt{y_n^2 + \kappa_3}} = -1$$

при всех  $n \geq 2$ .

Пример распределения корреляции показан на Рисунке 6.8. К сожалению, конечная корреляция не подразумевает низкую дисперсию: корреляция существует, но может оказаться бесполезна для статистических исследований из-за зашумленности и медленной сходимости.



**Рисунок 6.8:** Распределение корреляции по выборке, когда размер выборки  $10^3$ . Для двумерного  $t$ -распределения показатель степени  $\rho = \frac{2}{3}$ , корреляция  $\frac{3}{4}$ . Корреляция существует... однако бесполезна

## 6.7. ЖИРНОХВОСТЫЕ ОСТАТКИ В МОДЕЛЯХ ЛИНЕЙНОЙ РЕГРЕССИИ

Мы упоминали в Главе 3, что линейная регрессия неинформативна при жирных хвостах. Как ни прискорбно, она все равно практикуется. Например, величина дохода или благосостояния обычно распределена по степенному закону (что вызывает поток проблем, смотрите наше обсуждение индекса Джини в Главе 13). Баллы за тест IQ, напротив, распределены по Гауссу (по-видимому, тест специально настроен для полу-

чения такого распределения). Часто приходится видеть, как применяется регрессия с целью доказать зависимость первого от второго, в полном неведении о неприменимости данной методики.

Рассмотрим линейную регрессию, в которой независимая и зависимая случайные величины принадлежат разным классам распределения:

$$Y = aX + b + \varepsilon,$$

где  $X$  имеет стандартное гауссово распределение  $\mathcal{N}(0, 1)$  и ошибка  $\varepsilon$  распределена по степенному закону с  $\mathbb{E} \varepsilon = 0$  и  $\mathbb{E} \varepsilon^2 < +\infty$ . Никаких ограничений на параметры нет.

Очевидно<sup>1</sup>, что можно вычислить коэффициент детерминации  $R^2$  как 1 минус отношение ожидаемой суммы квадратов остатков регрессии к сумме квадратов отклонений, получив общий ответ на наш частный вопрос.  $X \sim \mathcal{N}(0, 1)$ , поэтому  $aX + b \sim \mathcal{N}(b, |a|)$  и

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^n (y_i - (ax_i + b + \varepsilon_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Можно доказать, что при больших  $n$

$$R^2 = \frac{a^2}{a^2 + \mathbb{E} \varepsilon^2} + o\left(\frac{1}{n^2}\right). \quad (6.13)$$

А при бесконечной дисперсии, очевидно,

$$\lim_{\mathbb{E} \varepsilon^2 \rightarrow +\infty} \mathbb{E} R^2 = 0.$$

Если случайная величина  $\varepsilon$  следует t-распределению Стьюдента с  $\alpha$  степенями свободы, очевидно, что  $\varepsilon^2$  следует распределению Фишера с числом степеней свободы  $(1, \alpha)$ , и хвост подчиняется степенному закону с показателем  $\frac{\alpha}{2}$ .

Заметим, что то же самое «ожидание» можно насчитать и просто для квадрата корреляции между  $X$  и  $Y$ . Например, положим  $\varepsilon \sim$  t-распределение Стьюдента с нулевым математическим ожиданием, масштабом  $\sigma$  и показателем хвоста  $\alpha > 2$  (как мы видели ранее, результаты не изменяются, коль скоро мы ограничиваемся случаями математического ожидания 0). Вычислим сначала корреляцию: числитель — это ковариация

1 В принятых в литературе обозначениях коэффициент детерминации — мера зависимости случайной величины  $Y$  от другой или других по подогнанному линейному закону — определяется как

$$R^2 = 1 - \sum_{i=1}^n (y_i - \hat{y}_i)^2 \bigg/ \sum_{i=1}^n (y_i - \bar{y})^2,$$

где  $y_i$  — значения зависимой случайной величины  $Y$ ,  $\hat{y}_i$  — значения, рассчитанные согласно подогнанному линейному закону и  $\bar{y}$  — среднее арифметическое по всем  $y_i$ . Также принято обозначать сумму квадратов как  $SS$  (от *sums of squares*); соответственно  $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$ .

2 В самом деле, плотность вероятности при распределении Фишера с числом степеней свободы  $(d_1, d_2)$

представима как  $f(x) = \frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}} \cdot \frac{1}{xB\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$ . Соответственно плотность вероятности при распределении Фи-

шера с числом степеней свободы  $(1, \alpha)$  представима как  $f(x) = \frac{\sqrt{x\alpha^\alpha}}{(x+\alpha)^{1+\alpha}} \cdot \frac{1}{xB\left(\frac{1}{2}, \frac{\alpha}{2}\right)} \sim x^{-1-\alpha/2}$  при  $x$  далеко в хвосте.

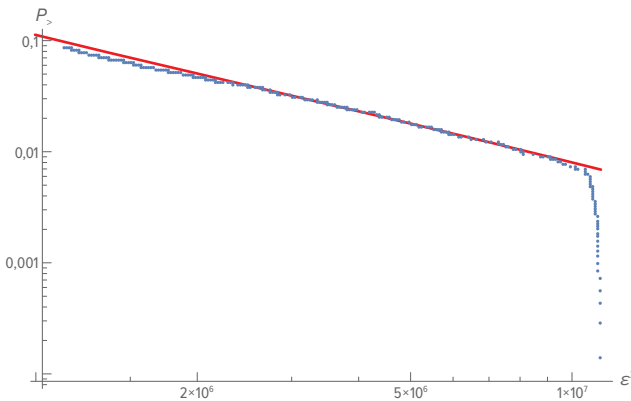
$\text{cov}(X, Y) = \mathbb{E}(aX + b + \varepsilon)X = a^1$ . Знаменатель, среднеквадратическое отклонение для  $Y^2$ , представим как  $\sqrt{\mathbb{E}((aX + \varepsilon) - a)^2} = \sqrt{\frac{2\alpha a^2 - 4a^2 + \alpha\sigma^2}{\alpha - 2}}$ . Т. о.,

$$\mathbb{E} R^2 = \frac{a^2(\alpha - 2)}{2(\alpha - 2)a^2 + \alpha\sigma^2}. \quad (6.14)$$

Переходя к описанному выше пределу,

$$\lim_{\alpha \rightarrow 2+0} \mathbb{E} R^2 = 0^3.$$

Мы скрупулезно рассматриваем  $\mathbb{E} R^2$  вместо обычного в детерминистском случае показателя  $R^2$ , поскольку у нас случайная величина — стохастическая и крайне зависима от выборки, достигая устойчивости только при больших — подчас астрономически больших —  $n$ . В самом деле, вспомним, что для любой выборки математическое ожидание всегда конечно, даже когда  $\varepsilon$  задают распределения Коши! Этот феномен проиллюстрирован на Рисунках 6.10 и 6.11. На практике, когда для  $R^2$  берут оценку максимального правдоподобия через  $\mathbb{E}\varepsilon^2$  с учетом  $\alpha$  (метод «теневого математического ожидания», в частности, в Главах 13 и 14), мы замечаем в примере с IQ, использованном на диаграмме, что остатки среднего по выборке составляют около половины оценки максимального правдоподобия<sup>4</sup>, отчего  $R^2$  падает еще ниже (практически до 0).



**Рисунок 6.9:** Так выглядит в дважды логарифмическом масштабе функция выживания для квадратов остатков  $\varepsilon^2$  от линейной регрессии IQ — доход по стандартным данным WLS<sup>5</sup>. Видно, что величина дохода винсоризована<sup>6</sup>. Отрезание хвостов создает иллюзию высокого  $R^2$ . Да и без отрезания хвоста коэффициент детерминации существенно завышается из-за свойств малой выборки при попытке оценить дисперсию степенного закона

1 В самом деле, по определению ковариация  $\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))$ . В нашем случае у случайных величин  $X$  и  $Y = aX + b + \varepsilon$  средние значения нулевые, так что  $\text{cov}(X, Y) = \mathbb{E}(XY) = \mathbb{E}(X(aX + b + \varepsilon)) = a\mathbb{E}X^2 + (b + \varepsilon)\mathbb{E}X = a$ .

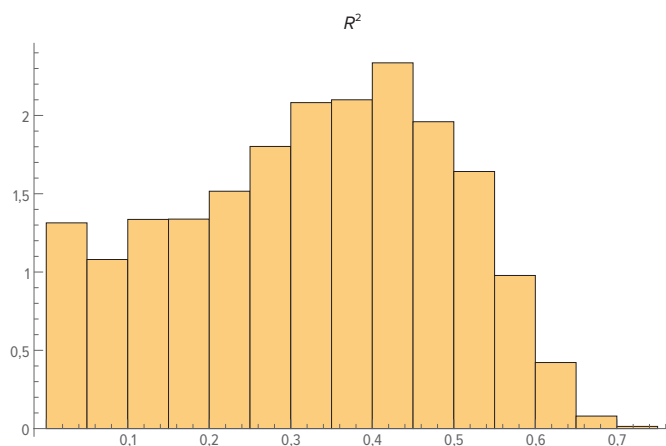
2 В самом деле, по определению корреляция  $R = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$ . В нашем случае  $X \sim \mathcal{N}(0, 1)$ , поэтому  $\sigma_X = 1$  и знаменатель просто  $\sigma_Y$ .

3 Разумеется, выражение (6.14) стремится к 0 и при  $\alpha \rightarrow 2$ . Односторонний предел  $\alpha \rightarrow 2+0$  лишний раз напоминает, что исследуются толстые хвосты, с показателем  $\alpha > 2$ .

4  $2,2 \cdot 10^9$  против  $1,24 \cdot 10^9$ . — Прим. автора.

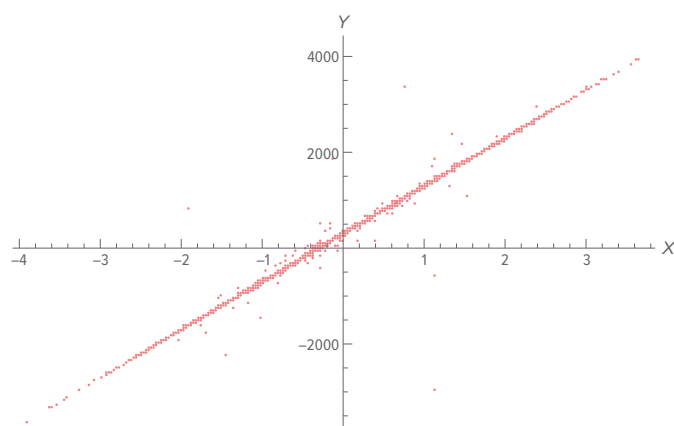
5 Сквозное исследование выпускников 1957 года школ в штате Висконсин (Wisconsin Longitudinal Studies, WLS), содержит данные по 2011 год.

6 Винсоризация — очистка статистических данных от сомнительных значений. Удаляются все значения, нарушающие некоторый порог (т. н. выбросы).



**Рисунок 6.10:** Пример с бесконечной дисперсией, где выборка поначалу показывает высокий  $R^2$ , хотя в итоге его значение 0. Вспомним, что  $R^2$  стохастичен. Ошибки в его оценке весьма похожи на проблему  $p$ -значений<sup>1</sup> в Главе 19 из-за усложнения метараспределения в промежутке  $[0, 1]$

Это обстоятельство лишает достоверности добрую половину статистических исследований взаимосвязи IQ — богатство и IQ — доход, аналогичных [268]; драматический пример видим на Рисунке 6.9. Если  $R$  в пределах  $[0, 1]$ , оценка приблизится к истинному значению крайне медленно; смотрите о проблеме  $p$ -значений в Главе 19.



**Рисунок 6.11:** Пример с распределениями Коши. Математическое ожидание  $R^2$  у данной регрессии нулевое, однако на малой выборке она показывает обманчиво высокий  $R^2 = 0,985$

<sup>1</sup> Кроме термина  $p$ -значение ( $p$ -value) в русской литературе также встречаются синонимы:  $p$ -уровень значимости и  $p$ -критерий. Имеется в виду готовность научного журнала признать, что в некотором исследовании гипотеза о некотором явлении подтверждена наблюдениями «статистически значимым» образом, если без этой гипотезы (говорят: при нулевой гипотезе) вероятность, что наблюдения все равно окажутся такими же или более резкими, — вероятность, называемая  $p$ -значением, — окажется ниже принятого в этом научном журнале порога, обычно 0,05 или 0,01.



**Свойство 3**

Если проверять регрессией зависимость толстохвостой случайной величины от тонкохвостой, коэффициент детерминации  $R^2$  получается завышенным и сходится только при огромных выборках, если вообще сходится.

Заметим, что иногда исследователи пытаются решить проблему, подвергнув толстохвостую случайную величину нелинейному преобразованию (например, прологарифмировав), в надежде добиться линейной зависимости. Если требуемое преобразование известно в точности, прием сработает — но если нет, то нет. Неудачное преобразование может привести к ошибкам, потому что корреляция вещь хрупкая и, в отличие от взаимной информации, неаддитивна и зачастую неинформативна. Вопрос исследован автором этой книги в [238].

**ДАЛЕЕ**

В Главе 8 мы изучим медленную сходимость случайных величин, распределенных по степенному закону, согласно закону больших чисел (ЗБЧ): замедление может оказаться в  $10^{13}$  раз по сравнению со сходимостью гауссовой случайной величины.

## А

# ОСОБЫЕ СЛУЧАИ ТОЛСТЫХ ХВОСТОВ

Для распределений с одной вершиной толстые хвосты — обычное явление: можно просмотреть десятки тысяч временных рядов социально-экономических случайных величин и ни разу не наткнуться на «платикуртическое»<sup>1</sup> распределение. Но для распределений с несколькими вершинами возможны сюрпризы.

### А.1. МУЛЬТИМОДАЛЬНОСТЬ И ЖИРНЫЕ ХВОСТЫ, или МОДЕЛЬ ВОЙНЫ И МИРА

Мы отмечали ранее в 4.1, что стохастизация (построение стохастической случайной величины из детерминистских случайных величин) даже при небольшом возмущении дисперсии делает хвосты распределения толще, как показывает рост эксцесса; при этом мы сохраняли прежнее среднее.

Но если стохастизировать и среднее (сохраняя его в среднем на прежнем уровне) и если достаточно широко расставить два стохастических значения среднего, добившись мультимодальности, то эксцесс, то есть четвертый статистический момент, может упасть. Сочетая разные стохастические дисперсии с разными стохастическими средними, мы получим разнообразие «режимов», каждый со своим набором возможностей.

При мультимодальности приходится либо признать, что понятие «толстых хвостов» теряет смысл, либо переосмыслить его так, что «середина», область вблизи математического ожидания, теряет свою важность [7, 156].

Начнем с того, что все мы на практике сталкиваемся с вещами, которые могут переключаться между разными режимами, или состояниями. Пусть во всех состояниях интересующая нас случайная величина имеет конечные статистические моменты, и пусть их структура такая: есть спокойный режим  $s_1$ , в котором математическое ожидание  $\mu_1$  и среднеквадратичное отклонение  $\sigma_1$ , и есть тревожный режим  $s_2$ , в котором математическое ожидание  $\mu_2$  и среднеквадратичное отклонение  $\sigma_2$ . Можно рассмотреть и много тревожных режимов, каждый со своей вероятностью  $p_i$ .

Пусть для простоты все распределения, между которыми мы переключаемся, гауссовы. Пусть с вероятностью  $\frac{1}{2}$  рассматриваемая случайная величина распределена

---

<sup>1</sup> С эксцессом меньше, чем у гауссова (от *kurtosis* — эксцесс).

в режиме  $X \sim \mathcal{N}(\mu_1, \sigma_1)$  и с вероятностью  $\frac{1}{2}$  рассматриваемая случайная величина распределена в режиме  $X \sim \mathcal{N}(\mu_2, \sigma_2)$ . Тогда

$$\text{эксцесс} = 3 - \frac{2((\mu_1 - \mu_2)^4 - 6(\sigma_1^2 - \sigma_2^2)^2)}{(\mu_1 - \mu_2)^4 + 2(\sigma_1^2 + \sigma_2^2)^2}. \quad (\text{A.1})$$

Как мы видим, эксцесс зависит от разности  $d = \mu_1 - \mu_2$ . В случае  $\sigma_1 = \sigma_2$ ,  $\mu_1 \neq \mu_2$  дробь окажется положительной и эксцесс окажется меньше, чем у обычной гауссианы. Чтобы сохранить эксцесс равным 3, потребуется сделать коэффициенты масштаба разными:

$$|d| = \sqrt[4]{6} \sqrt{\max(\sigma_1, \sigma_2)^2 - \min(\sigma_1, \sigma_2)^2},$$

тогда стохастизация математического ожидания будет компенсирована стохастизацией волатильности.

Допустим для простоты периодическую модель и что мы смотрим на хронологию конечного отрезка, видя все новые результаты. (Мы могли бы усложнить наш мысленный эксперимент, добавив матрицы переходов между режимами, но главный вывод не изменился бы.)

Характеристическая функция  $\varphi(t)$  смешанного распределения с  $N$  режимами

$$\varphi(t) = \sum_{i=1}^N p_i e^{-\frac{1}{2} t^2 \sigma_i^2 + i t \mu_i}.$$

При  $N = 2$  статистические моменты примут вид

$$M_1 = p_1 \mu_1 + (1 - p_1) \mu_2$$

$$M_2 = p_1(\mu_1^2 + \sigma_1^2) + (1 - p_1)(\mu_2^2 + \sigma_2^2)$$

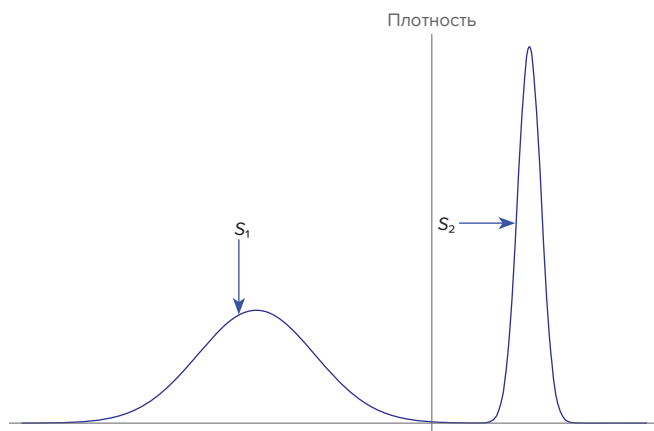
$$M_3 = p_1 \mu_1^3 + (1 - p_1) \mu_2(\mu_2^2 + 3\sigma_2^2) + 3p_1 \mu_1 \sigma_1^2$$

$$M_4 = p_1(6\mu_1^2 \sigma_1^2 + \mu_1^4 + 3\sigma_1^4) + (1 - p_1)(6\mu_2^2 \sigma_2^2 + \mu_2^4 + 3\sigma_2^4)$$

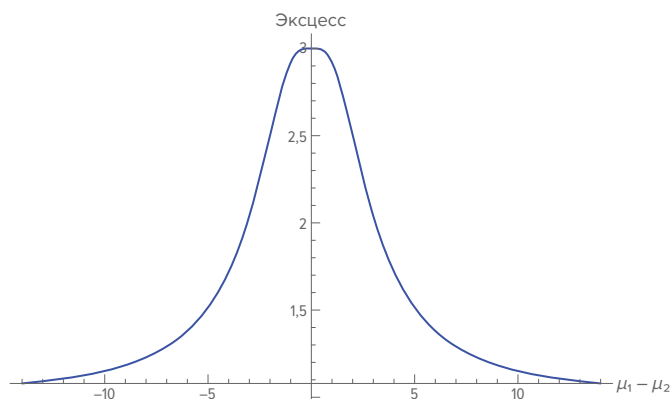
Рассмотрим те разновидности, где  $p_1 < 1 - p_1$ ,  $\mu_1 < \mu_2$ , а лучше  $\mu_1 < 0 < \mu_2$  и, главное,  $\sigma_1 > \sigma_2$ .

**Разновидность 1: Война и мир.** Спокойный период с положительным математическим ожиданием и очень низкой волатильностью, сменяющийся бурей с отрицательным математическим ожиданием и крайне низкой волатильностью.

**Разновидность 2: Условно-детерминистское состояние.** Рассмотрим облигацию стоимости  $B$ , по которой выплачивается процент  $r$  в конце одного-единственного периода. При реализации есть высокая вероятность получить обещанные  $B(1 + r)$  и есть некоторая вероятность дефолта. Получить ровно  $B$  маловероятно. Представим себе, что нет промежуточных ступеней между войной и миром: что это два отдельных, дискретных состояния. С облигациями не случается «немножко дефолт». Пропасть практически запрещает реализацию, равную или близкую к среднему. Собственно, при любом двумодальном распределении плотности вероятности  $p(x)$  значение плотности  $p(\mathbb{E}X)$  должно быть ниже, чем вблизи математических ожиданий отдельных



**Рисунок А.1:** Модель Война и мир. Эксцесс 1,7, намного ниже, чем у гауссианы



**Рисунок А.2:** Эксцесс опускается ниже гауссова в зависимости от расхождения пиков бимодального распределения (эксцесс 3 — у гауссова распределения)

режимов, так что  $\mathbb{P}(X = \mathbb{E}X) < \mathbb{P}(X = \mu_1)$  и  $< \mathbb{P}(X = \mu_2)$ <sup>1</sup>, но в нашем крайнем примере с облигациями  $\mathbb{P}(X = \mathbb{E}X)$  становится особенно мала. Хвостовым событием будет реализация, близкая к математическому ожиданию.

Та же идея приложима и к привязкам валюты, поскольку девальвации не бывают «умеренными», это волатильность типа всё-или-ничего, с «долиной» низкой плотности вероятности между двумя разными режимами.

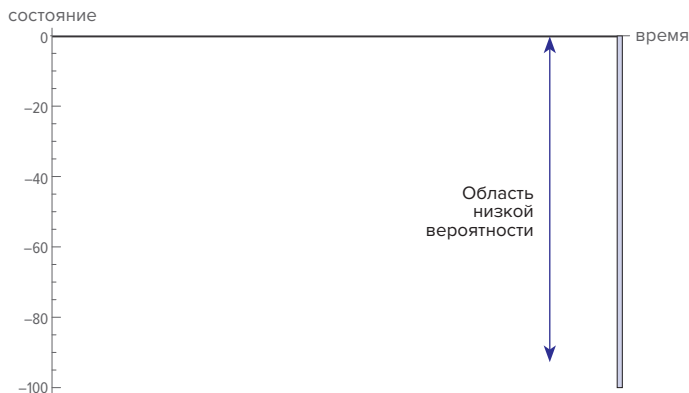
У доходности опциона бимодальность проявляется как рост премии для опционов «на деньгах» и понижение для опционов «вне денег», создавая прямую противоположность так называемой «улыбке волатильности»<sup>2</sup>.

Заметим, что чашка не бывает в промежуточном состоянии между разбитой и целой. Разбитую чашку можно считать поглощающим состоянием (в терминах цепи Маркова и переходных вероятностей), поскольку разбитые чашки не чинятся.

<sup>1</sup> Педантичнее было бы выписать  $\mathbb{P}(X \in \mathbb{E}(X) \pm dx) < \mathbb{P}(X \in \mu_1 \pm dx)$  и  $< \mathbb{P}(X \in \mu_2 \pm dx)$ ; впрочем, для дискретного примера с облигацией разницы никакой:  $\mathbb{P}(X \in \mu_1 \pm dx) = \mathbb{P}(X = \mu_1)$  и  $\mathbb{P}(X \in \mu_2 \pm dx) = \mathbb{P}(X = \mu_2)$ .

<sup>2</sup> Выпуклая вниз зависимость ожидаемой волатильности от страйк-цены.

И чашки редко оказываются «слегка разбитыми», как видно на Рисунке А.3.



**Рисунок А.3:** Если разбивается кофейная чашка, повреждение будет скорее большим, чем малым. Бьющаяся чашка существует по принципу всё или ничего. Тот же тип доходности преобладает на рынках при девальвациях (или ревальвациях), где сильное изменение вероятнее, чем слабое

### Краткий список прочих ситуаций с бимодальностью:

1. Привязка валюты.
2. Слияние.
3. Выбор профессии и итоговый заработок.
4. Конфликт: межличностный, общий, военный и вообще любая ситуация, где нет промежуточного состояния между гармоничными отношениями и боевыми действиями.
5. Каскад состояний.

## А.2. ПЕРЕХОДНЫЕ ВЕРОЯТНОСТИ: ЧТО МОЖЕТ РАЗБИТЬСЯ, РАЗОБЬЕТСЯ<sup>1</sup>

Пока что мы рассматривали модель одного-единственного периода, и это был реалистичный подход, поскольку новая информация может нарушить бимодальность, если двинуться дальше в будущее: мы ясно видим лишь один шаг. Однако давайте сделаем усилие и решим задачу, которая подарит нам одну идею касательно хрупкости. Предположим, что модель сохраняет структуру, и рассмотрим долговременное поведение при переходах между состояниями. Пусть  $P$  — матрица переходных вероятностей, где  $p_{i,j}$  относится к переходу из  $i$ -того состояния в  $j$ -тое за время  $\Delta t$ . Обозначая  $S(t)$  режим в момент времени  $t$ , это вероятность  $\mathbb{P}(S(t + \Delta t) = s_j | \mathbb{P}(S(t) = s_i))$ . Для двух состояний

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix}.$$

После  $n$  периодов, или шагов,

$$P^n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix},$$

где

$$a_n = \frac{(p_{1,1} - 1)(p_{1,1} + p_{2,2} - 1)^n + p_{2,2} - 1}{p_{1,1} + p_{2,2} - 2}$$

<sup>1</sup> Парафраз первого закона Мерфи «Все, что может испортиться, — портится».

$$b_n = \frac{(1-p_{1,1})(p_{1,1}+p_{2,2}-1)^n-1}{p_{1,1}+p_{2,2}-2}$$

$$c_n = \frac{(1-p_{2,2})(p_{1,1}+p_{2,2}-1)^n-1}{p_{1,1}+p_{2,2}-2}$$

$$d_n = \frac{(p_{2,2}-1)(p_{1,1}+p_{2,2}-1)^n+p_{1,1}-1}{p_{1,1}+p_{2,2}-2}$$

(мы избавились от обеих вероятностей  $p_{i, \neq i}$  заменив их на  $1-p_{i,i}$ ).

Рассмотрим крайний случай, где 1-е состояние — поглощающее,  $p_{1,1} = 1$ . Тогда

$$P^n = \begin{pmatrix} 1 & 0 \\ 1-p_{2,2}^n & p_{2,2}^n \end{pmatrix}$$

и «эргодические» вероятности

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

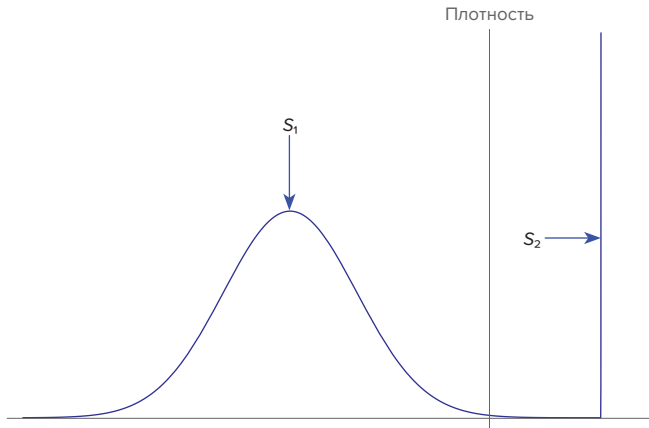
Смысл тот, что режим 1-го, поглощающего состояния  $s_1$  в конечном счете преобладает с вероятностью 1: что может разбиться, в конце концов разобьется.

Для «эргодической» матрицы

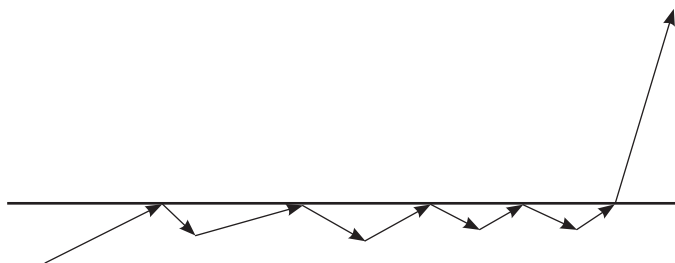
$$\lim_{n \rightarrow \infty} P^n = \pi \mathbb{1}^T,$$

где  $\mathbb{1}^T = (1 \ 1)$  — транспонированный единичный вектор и  $\pi$  — матрица собственных векторов. Собственные значения принимают вид  $\lambda = \begin{pmatrix} 1 \\ p_{1,1}+p_{2,2}-1 \end{pmatrix}$ , соответственно

$$\text{собственные векторы } \pi = \begin{pmatrix} 1 & 1 \\ \frac{1-p_{1,1}}{1-p_{2,2}} & 1 \end{pmatrix}.$$



**Рисунок А.4:** Модель доходности облигации или привязки валюты. В режиме 2, в точке привязки, ноль волатильности, детерминистская доходность; в режиме 1 хаос. В данном примере эксцесс  $K = 2,5$ . В отличие от этого примера пример с чашкой был вырожденным в обоих режимах



**Рисунок А.5:** Давление на привязку может давать распределение плотности вероятностей в виде функции Дирака, близкой к нулю в режиме «никакой девальвации» (что равносильно низкой волатильности). Финансовые кретины обычно принимают режим  $s_2$  за свидетельство низкой волатильности

**Часть II**

**ЗАКОН СРЕДНИХ ЧИСЕЛ**





## ПРЕДЕЛЬНЫЕ РАСПРЕДЕЛЕНИЯ, КОНСОЛИДАЦИЯ\*\*

В этой пояснительной главе мы продолжим сводить к общему языку литературу по предельным распределениям с учетом нашей цели; в некоторых случаях будем предлагать свои сокращенные версии. Познакомившись с законом больших чисел, мы покажем интуитивное понимание центральной предельной теоремы и проиллюстрируем ее разновидности доасимптотическим поведением разных распределений. Затем обсудим закон больших чисел применительно к высшим моментам. Более строгий и глубокий подход представим в следующей главе.

И закон больших чисел, и центральная предельная теорема дают частные ответы на общий вопрос: «Каково в пределе поведение суммы или среднего случайных величин, когда число слагаемых приближается к бесконечности?» Наш закон средних чисел, или закон предасимптотики, отвечает на вопрос: «А что, если число слагаемых не достигает бесконечности?»

### 7.1. НАПОМИНАНИЕ: СЛАБЫЙ И СИЛЬНЫЙ ЗБЧ

Обычное изложение такое. Пусть  $X_1, X_2, \dots$  — бесконечная последовательность независимых одинаково распределенных случайных величин, интегрируемых по Лебегу и имеющих математическое ожидание  $\mathbb{E} X_i = \mu$ . Ниже мы увидим, что можно несколько ослабить требование н. о. р. Среднее по выборке первых  $n$  величин

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$

сходится к тому же математическому ожиданию,  $\bar{X}_n \rightarrow \mu$ , при  $n \rightarrow \infty$ .

Предполагать конечную дисперсию не требуется, хотя, безусловно, конечность дисперсии и прочих высших моментов ускорит сходимость.

Сходимость происходит в двух режимах: сходимость по вероятности  $\xrightarrow{p}$  (из которой следует сходимость по распределению, но не наоборот) и более сильная  $\xrightarrow{\text{a.s.}}$ , сходимость почти наверное (аналог поточечной сходимости), она же сходимость почти всюду или почти всегда. Различие отвечает слабому и сильному ЗБЧ.

**Слабый ЗБЧ.** Слабый закон больших чисел (он же закон Хинчина<sup>1</sup>, он же закон Бернулли<sup>2</sup>) в кратком изложении сводится к тому, что вероятность, с которой среднее значение последовательности отклонится от среднего дальше чем на данный порог, уменьшается с удлинением последовательности.

То есть для любого положительного числа  $\varepsilon$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0.$$

Тем самым оценка среднего по все более длинным последовательностям *состоятельная*: сходится по вероятности к оцениваемой величине,

$$\bar{X}_n \xrightarrow{P} \mu \text{ при } n \rightarrow \infty.$$

Заметим, что стандартные доказательства опираются на неравенство Чебышёва: если у  $X$  конечная ненулевая дисперсия  $\sigma^2$ , то для любого действительного числа  $k > 0$

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

**Сильный ЗБЧ.** Сильный закон больших чисел гласит, что, когда число слагаемых  $n$  уйдет в бесконечность, вероятность того, что среднее сойдется к математическому ожиданию, равна 1:

$$\bar{X}_n \xrightarrow{\text{a.s.}} \mu \text{ при } n \rightarrow \infty.$$

То есть

$$\mathbb{P}(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1.$$

**Ослабление н. о. р.** Ослабим теперь условие одинаковости распределенности случайных величин. Как доказал Колмогоров<sup>3</sup>, для неодинаково распределенных слагаемых  $X_i$  достаточно потребовать существование конечного второго момента.

Что касается требования независимости, то некоторую слабую зависимость допустить можно. Традиционно формулируются такие условия, как, во-первых, опять-таки конечная дисперсия

$$1) \mathbb{V}(X_i) \leq c$$

и, во-вторых, некоторые ограничения на структуру ковариационной матрицы

$$2) \lim_{|i-j| \rightarrow +\infty} \text{Cov}(X_i, X_j) = 0.$$

1 Здесь имеется в виду, что в 1924 А. Я. Хинчин открыл закон повторного логарифма: если в последовательности н. о. р.  $\mu = 0$  и дисперсия  $\sigma^2 = 1$ , то при  $n \rightarrow \infty$

$$\sup \frac{S_n}{\sqrt{n \ln \ln n}} \xrightarrow{\text{a.s.}} \sqrt{2} \text{ и } \inf \frac{S_n}{\sqrt{n \ln \ln n}} \xrightarrow{\text{a.s.}} -\sqrt{2}.$$

2 Здесь имеется в виду, что в «Искусстве предположений» Якоба Бернулли (Ars conjectandi, 1713; книга издана посмертно под ред. Николая Бернулли) содержится первое доказательство ЗБЧ, для случая бинарных случайных величин. В той книге это было названо «золотой теоремой»; название закон больших чисел (*la loi des grands nombres*) ввел Пуассон в 1837 году.

3 В 1929 году Колмогоров обобщил закон повторного логарифма Хинчина, ослабив требование н. о. р.

Впрочем, оказалось, что (1) можно ослабить до требования

$$\sum_{i=1}^N \mathbb{V}(X_i) = o(n^2),$$

а (2) можно ослабить до требования

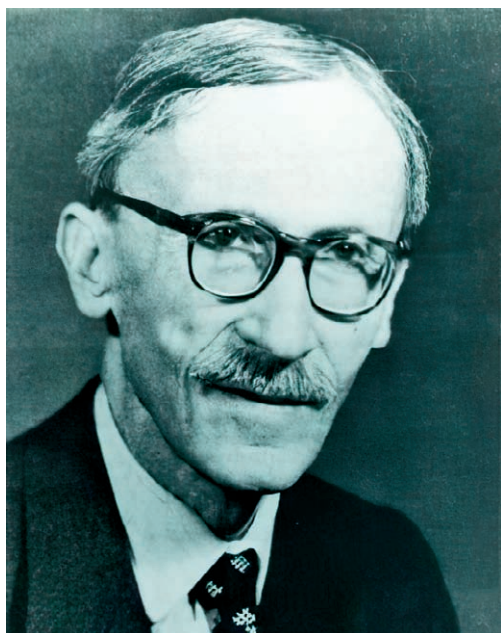
$$|\text{Cov}(X_i, X_j)| \leq \varphi(|i - j|), \text{ где } \frac{1}{n} \sum_{i=1}^N \varphi(i) \rightarrow 0.$$

См. Бернштейн<sup>1</sup> [19] (на французском) и Козлов<sup>2</sup> [148] (на русском)<sup>3</sup>.

**В чем наш интерес.** Нас, понятно, волнует «скорость» сходимости ЗБЧ, и мы будем ее рассматривать в этой и следующей главе. Заметим, что при сильном предположении н. о. р. не обязательно требовать конечную дисперсию, и мы переключимся с дисперсии на другой показатель рассеяния — среднее абсолютное отклонение.

## 7.2. ЦЕНТРАЛЬНЫЙ ПРЕДЕЛ В ДЕЙСТВИИ

Начнем с упрощенной версии обобщенной центральной предельной теоремы (ЦПТ) в формулировке Поля Леви (традиционный подход к ЦПТ и необходимую технику представим позже):



**Рисунок 7.1:** Поль Леви, 1886–1971, сформулировал обобщенную центральную предельную теорему

<sup>1</sup> Здесь ссылка на статью 1918 года.

<sup>2</sup> Здесь ссылка на статью 2004 года «О весовых средних значениях слабозависимых случайных величин», где у академика Козлова два соавтора: Татьяна Валерьевна Мадсен-Козлова из Ольборгского университета, Дания, и Алексей Александрович Сорокин, механико-математический факультет МГУ.

<sup>3</sup> Спасибо моему таинственному русскоговорящему помощнику с ником romanov на форуме *Mathematics Stack Exchange*. — Прим. автора.

### 7.2.1. Устойчивое распределение

Как и выше, пусть  $X_1, X_2, \dots, X_n$  — независимые и одинаково распределенные случайные величины. Рассмотрим их сумму  $S_n$ . Имеет место предел

$$\frac{S_n - a_n}{b_n} \xrightarrow{D} X_S, \quad (7.1)$$

где  $X_S$  следует устойчивому распределению  $S$ ,  $a_n$  и  $b_n$  — нормирующие постоянные и, как мы помним,  $\xrightarrow{D}$  обозначает сходимость по распределению (в данном случае сходимость к распределению случайной величины  $X$  при  $n \rightarrow \infty$ ). Свойства  $S$  мы должным образом определим и изучим в следующей главе. Пока ограничимся тем, что случайная величина  $X_S$  следует устойчивому (или  $\alpha$ -устойчивому) распределению, что записывается как  $X_S \sim S(\alpha_S, \beta, \mu, \sigma)$ , если ее характеристическая функция  $\chi(t) = \mathbb{E}e^{itX_S}$  имеет вид:

$$\chi(t) = e^{i\mu t X_S - |t\sigma|^{\alpha_S} \left(1 - \beta \operatorname{tg} \frac{\pi\alpha_S}{2} \operatorname{sgn} t\right)} \quad \text{при } \alpha_S \neq 1. \quad (7.2)$$

Ограничения:  $-1 \leq \beta \leq 1$  и  $0 \leq \alpha_S \leq 2$ <sup>1</sup>.

Обозначение устойчивого распределения подразумевает, что данное распределение (или класс распределений) устойчивы при суммировании: вы складываете случайные величины, следующие различным распределениям из класса  $\mathfrak{S}$ <sup>2</sup>, описанного в следующей главе (где все они сводимы к одному и тому же распределению с разными параметрами характеристической функции), и сумма получается с тем же распределением. На интуитивном уровне на  $\chi(t)^n$  можно смотреть как на результат подстановки в  $\chi(t)$  новых параметров:  $\mu \rightarrow n\mu$  и  $\sigma \rightarrow n^{\frac{1}{\alpha}}\sigma$ . Широко известные распределения в этом классе (который некоторые называют «бассейном») — гауссово, Коши и Леви, с параметром  $\alpha = 2, 1$  и  $\frac{1}{2}$  соответственно. У остальных распределений нет аналитического представления для плотности вероятности<sup>3</sup>.

### 7.2.2. Закон больших чисел для устойчивого распределения

Вернемся к закону больших чисел.

Согласно стандартным результатам, мы можем наблюдать работу закона больших чисел при устойчивом распределении, как проиллюстрировано на Рисунке 7.2:

1 Мы постараемся, где есть риск двусмысленности, обозначать как  $\alpha_S \in (0, 2]$  показатель предельного или идеального устойчивого распределения и как  $\alpha_p \in (0, \infty)$  соответствующий эквивалент распределения Парето (доасимптотического); в самоочевидном контексте обойдемся просто  $\alpha$ . — Прим. автора.

2 Это заглавная готическая «эс» (*mathematical fraktur capital s*).

3 На самом деле неаналитичность можно обойти, введя специальные функции; например, автор случайно открыл, что для устойчивого распределения  $S(\alpha, \beta, \mu, \sigma)$  с параметрами  $\alpha = \frac{3}{2}, \beta = 1, \mu = 0$  и  $\sigma = 1$  плотность вероятности выразима через функцию Эйри как

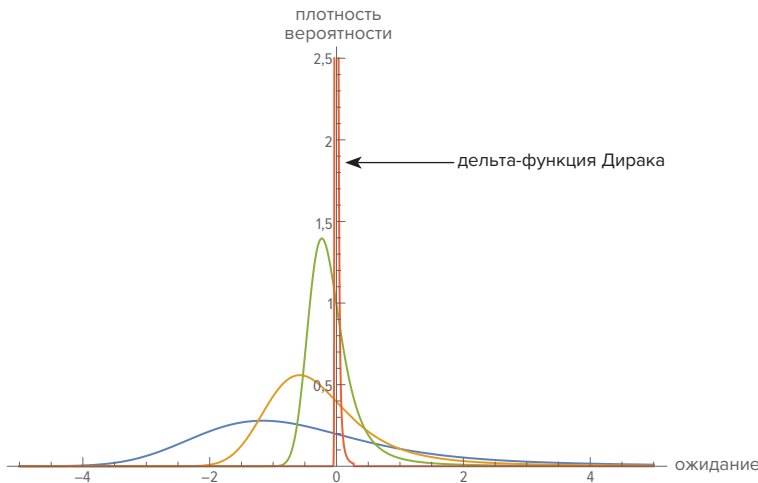
$$\text{PDF}(x) = - \frac{\sqrt[3]{2} e^{\frac{x^3}{27}} \left( \sqrt[3]{3} x \operatorname{Ai} \left( \frac{x^2}{3 \cdot 2^{2/3} \sqrt[3]{3}} \right) + 3 \sqrt[3]{2} \operatorname{Ai}' \left( \frac{x^2}{3 \cdot 2^{2/3} \sqrt[3]{3}} \right) \right)}{3 \cdot 3^{2/3}};$$

это представление используется далее в примере предельного распределения для сумм Парето. — Прим. автора.

$$\lim_{n \rightarrow +\infty} \chi \left( \frac{t}{n} \right)^n = e^{i\mu t}, \quad 1 < \alpha_s \leq 2, \quad (7.3)$$

то есть вырождается в характеристическую функцию дельта-функции Дирака; в самом деле, преобразование Фурье  $\mathcal{F}$  (здесь оно параметризовано, чтобы быть обратным к характеристической функции) дает:

$$\mathcal{F}_t(e^{i\mu t})(x) = \delta(x - \mu). \quad (7.4)$$



**Рисунок 7.2:** Закон больших чисел проявляется как сжатие распределения вокруг математического ожидания  $\mu$ , в пределе, вырождение в дираковский шест в точке математического ожидания

Мы еще увидим, как это работает «в реальном времени», при всех  $1 < n < +\infty$ , в нескольких вариантах, — в следующих разделах.

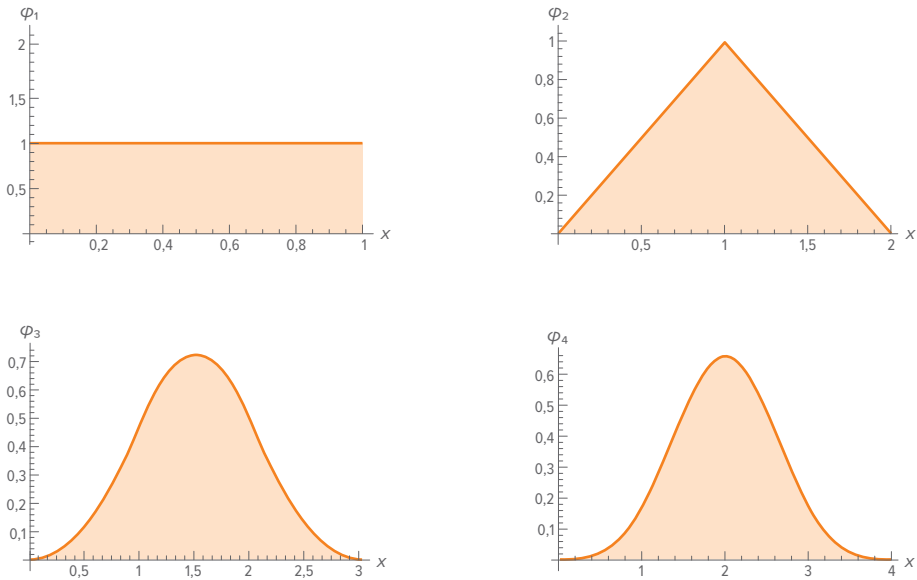
### 7.3. СКОРОСТЬ СХОДИМОСТИ ЦПТ: НАГЛЯДНЫЕ ОПЫТЫ

Заметим, что при конечной дисперсии случайных величин  $X$  устойчивое распределение случайной величины  $X_s$  будет гауссовым. Однако случайная величина  $X_s$  построена как предельный переход при  $n \rightarrow \infty$ , и возможны разные осложнения на пути к цели. Рассмотрим 4 случая, иллюстрирующие смысл ЦПТ и скорость сходимости.

#### 7.3.1. Быстрая сходимость: равномерное распределение

Рассмотрим равномерное распределение — простейшее из всех. Если случайная величина  $X_1$  пробегает отрезок  $[0, 1]$ , плотность вероятности будет просто  $\varphi_1(x) = 1$  при  $0 \leq x \leq 1$ , давая интеграл 1. Теперь прибавим к ней другую случайную величину  $X_2$ , независимую и с таким же распределением. У суммы  $X_1 + X_2$  распределение будет другим! Посмотрим на функцию  $\varphi_2(\cdot)$ , плотность вероятности для суммы, на Рисунке 7.3. График стал треугольным. Добавим еще одну переменную, и плотность вероятности  $\varphi_3$  для распределения суммы  $X_1 + X_2 + X_3$  станет колоколом всего-навсего при  $n = 3$  слагаемых<sup>1</sup>.

<sup>1</sup> Распределение суммы  $n$  равномерно распределенных н. о. р. называется распределением Ирвина — Холла (Irwin–Hall distribution).



**Рисунок 7.3:** Самая быстрая ЦПТ: равномерное распределение сходится к гауссову за несколько шагов. Рассмотрим число слагаемых 1, 2, 3, 4. Уже при 3 слагаемых видим отчетливый колокол

Сумма  $X_1 + X_2 + \dots$  независимых случайных величин  $X_1, X_2, \dots$ , каждая из которых равномерно распределена в отрезке  $[L, H]$ , имеет распределение

$$\varphi_n(x) = \sum_{k=1}^n (-1)^k \binom{n}{k} \left( \frac{x-L}{H-L} - k \right)^{n-1} \operatorname{sgn} \left( \frac{x-L}{H-L} - k \right) \text{ при } nL \leq x \leq nH^1.$$

### 7.3.2. Полузамедленная сходимость: экспоненциальные распределения

Рассмотрим сумму случайных величин с экспоненциальным распределением.

Исходная функция плотности

$$\varphi_1(x) = \lambda e^{-\lambda x}, x \geq 0,$$

а для  $n$  слагаемых<sup>2</sup>

$$\varphi_n(x) = \left( \frac{1}{\lambda} \right)^{-n} \frac{x^{n-1} e^{-\lambda x}}{\Gamma(n)}.$$

Заменяя  $x$  на  $n/\lambda$  (мы потом примем  $\lambda = 1$ , как в примерах на Рисунке 7.4), получаем в пределе

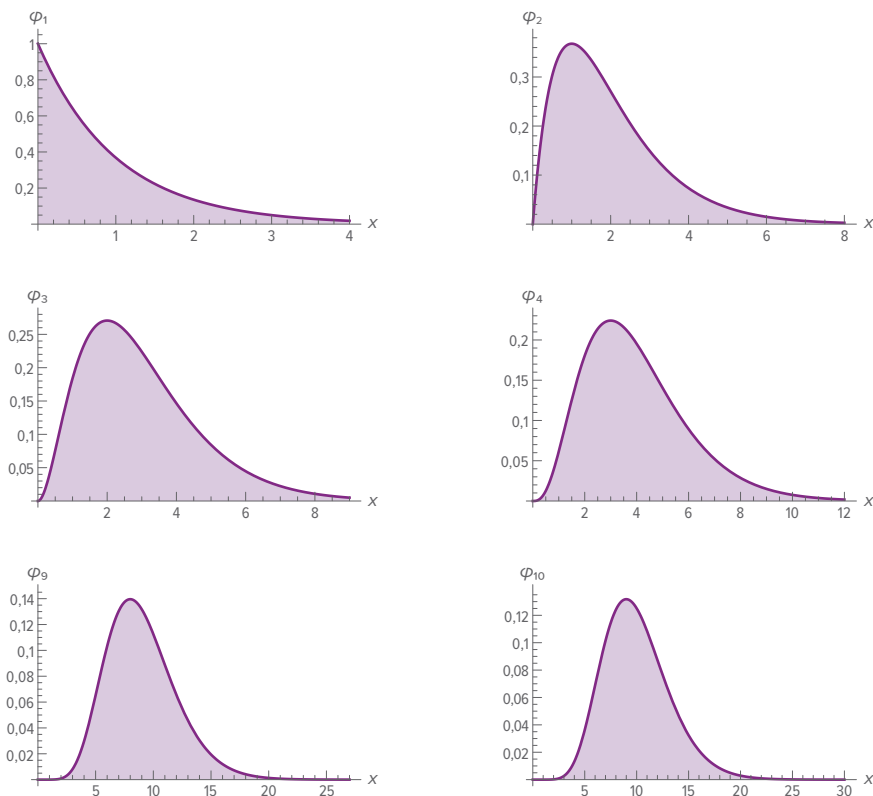
$$\left( \frac{1}{\lambda} \right)^{-n} \frac{x^{n-1} e^{-\lambda x}}{\Gamma(n)} \xrightarrow{n \rightarrow \infty} \frac{\lambda e^{-\frac{\lambda^2 \left( x - \frac{n}{\lambda} \right)^2}{2n}}}{\sqrt{2\pi\sqrt{n}}},$$

<sup>1</sup>  $\binom{n}{k} = \frac{n!}{k!(n-k)!}$  биномиальный коэффициент; в русской литературе также  $C_n^k$  — Прим. перев.

<sup>2</sup> Плотность вероятности для суммы в данном случае можно вывести просто сверткой; для более сложных случаев можно действовать через характеристические функции, как мы увидим на примере распределений Парето.

то есть плотность нормального распределения с математическим ожиданием  $\frac{n}{\lambda}$  и дисперсией  $\frac{n}{\lambda^2}$ .

Соответствующее продвижение к гауссиане, как видно на Рисунке 7.4, происходит медленнее, чем мы видели в предыдущем разделе; остатки изначальной асимметрии еще заметны даже при  $n = 10$ .



**Рисунок 7.4:** Экспоненциальное распределение  $\varphi$  с числом слагаемых, указанным в нижнем индексе. Сходимость замедлилась по сравнению с равномерным распределением, но еще хорошая

### 7.3.3. МЕДЛЕННЫЙ ПАРЕТО

Рассмотрим простейшее распределение Парето на промежутке  $[1, \infty)$ :

$$\varphi_1(x) = 2x^{-3}.$$

Обращая характеристическую функцию, выводим

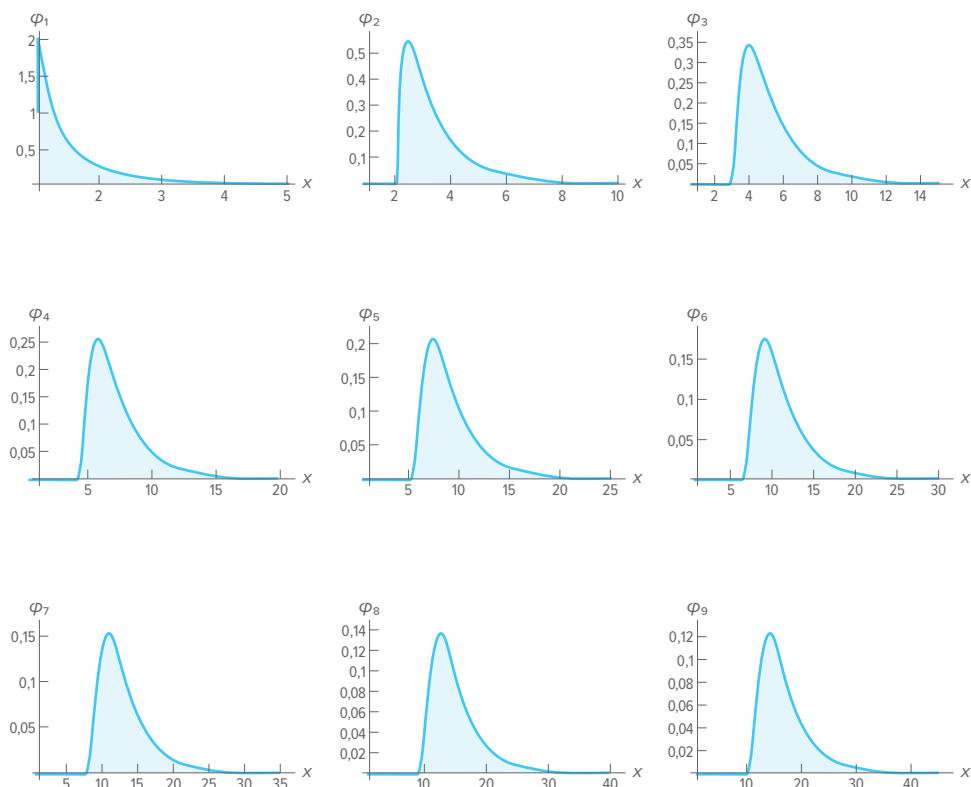
$$\varphi_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) (2E_3(-it))^n dt, \quad x \geq n,$$

где  $E_3(\cdot)$  — обобщенная модифицированная интегральная показательная функция,

которая определяется как  $E_n z = \int_1^{\infty} \frac{e^{t(-z)}}{t^n} dt$ .



Интегрировать придется численно (до сих пор никому не удалось аналитически представить сумму распределений Парето). Время вычислений растет по экспоненте, например 45 секунд тратится на  $n = 2$  и 24 часа на  $n = 50$ ; чтобы получить графики на Рисунке 7.5, нам пришлось прибегнуть к методу Монте-Карло.



**Рисунок 7.5:** Распределение Парето. Коэффициент асимметрии\* упорно не падает до нуля, хотя в данном случае распределение суммы сходится к гауссовому... в конце концов

\* Коэффициент асимметрии (англ. *skewness*) определяют как нормализованный третий центральный момент  $\tilde{\mu}_3 = \mu_3 / \sigma^3$ , где третий центральный момент  $\mu_3 = E(X - \mu)^3$ . — Прим. перев.

Вспомним равенство 7.1: для сходимости требуются нормирующие константы  $a_n$  и  $b_n$ . Из работы Учайкина и Золотарева имеем (сужая ситуацию для  $1 < \alpha_p \leq 2$ ):

$$\mathbb{P}(X > x) = cx^{-\alpha_p}$$

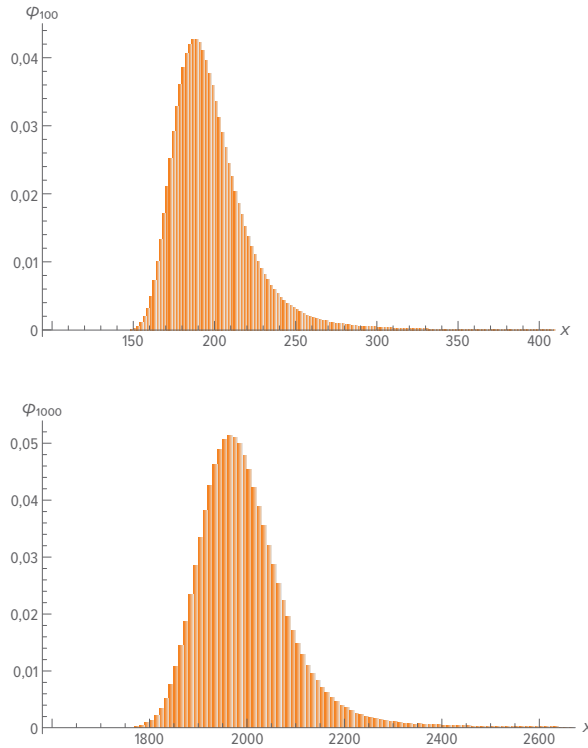
при  $x \rightarrow \infty$  (считайте пока  $c$  константой; в следующей главе мы представим более формально «медленно меняющуюся функцию»), и

$$\mathbb{P}(X < x) = d|x|^{-\alpha_p}$$

при  $x \rightarrow -\infty$ . Нормирующие константы принимают вид  $a_n = n \mathbb{E}X$  при всех  $\alpha_p > 1$  (за описанием других случаев отсылаю к [257], поскольку на практике они маловероятны), и

$$b_n = \begin{cases} \pi n^{\frac{1}{\alpha_p}} \left( 2 \sin \frac{\pi \alpha_p}{2} \Gamma \alpha_p \right)^{-\frac{1}{\alpha_p}} (c+d)^{\frac{1}{\alpha_p}} & \text{при } 1 < \alpha_p < 2, \\ \sqrt{c+d} \sqrt{n \ln n} & \text{при } \alpha_p = 2 \end{cases}, \quad (7.5)$$

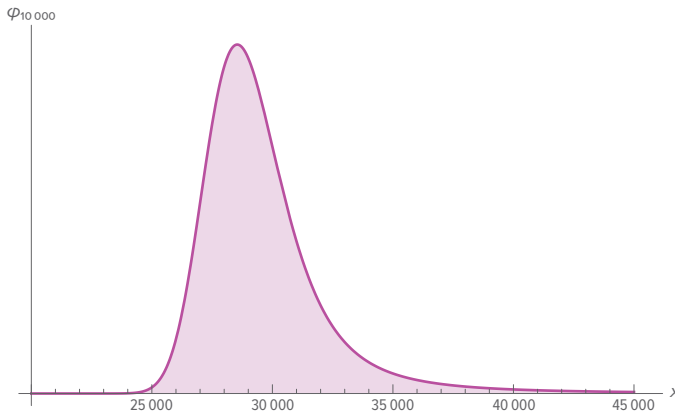
а параметр симметрии примет вид  $\beta = \frac{c-d}{c+d}$ . Разумеется, в случае, где у распределения Парето параметр  $\alpha_p$  превышает 2, сумма  $n$  слагаемых сходится к гауссиане.



**Рисунок 7.6:** Распределения Парето  $\phi_{100}$  и  $\phi_{1000}$  так и не приблизились к гауссиане, хотя при  $\alpha = 2$  это произойдет — если у вас хватит терпения и вы будете жить долго-долго

### 7.3.4. Полукубический Парето и его область сходимости

Интерес представляет случай  $\alpha = \frac{3}{2}$ . В отличие от ситуаций на Рисунке 7.3.1, распределение мало-помалу становится симметричным. Хотя, как мы обсудим в следующей главе, будет ошибкой присовокупить его к ситуациям устойчивого распределения. В определенном смысле слова оно еще более жирнохвостое.



**Рисунок 7.7:** Полукубическое распределение Парето так и не станет симметричным на практике. Здесь число слагаемых  $n = 10^4$

## 7.4. КУМУЛЯНТЫ И СХОДИМОСТЬ

Гауссово распределение, сходимостью к которому мы исследуем, имеет коэффициент асимметрии 0 и эксцесс (не приведенный) 3<sup>1</sup>. Поэтому мы можем изучить сходимость этих двух моментов, чтобы эвристически нащупать бассейн быстрого срабатывания ЦПТ.

**Определение 7.1** ( $p$ -тые избыточные кумулянты)

Пусть  $\chi(\omega)$  — характеристическая функция данного распределения,  $n$  — число слагаемых (независимых случайных величин),  $p$  — порядок момента. Обозначим отношение кумулянтов<sup>2</sup>, отвечающих  $p$ -тому и 2-му моменту:

$$K_n^p \triangleq \frac{(-i)^p \frac{\partial^p \ln \chi(\omega)^n}{\partial \omega^p} \Big|_{\omega=0}}{\frac{\partial^2 \ln \chi(\omega)^n}{\partial \omega^2} \Big|_{\omega=0}}.$$

$K_n^p$  при  $p > 2$  показывает превышение  $p$ -того момента распределения  $n$  независимых слагаемых относительно момента гауссианы; таким образом,  $K_n^4 = 0$  покажет, что при  $n$  слагаемых достигнута гауссовость.

1 Очевидно, оговорка адресована читателям, привыкшим, как в русской литературе, пользоваться не эксцессом, а коэффициентом эксцесса, то есть эксцессом минус 3.

2 В русской литературе кумулянты (*cumulants*) называются также полуинвариантами и семинвариантами. Кумулянты распределения случайной величины  $X$  аналогичны другим статистическим характеристикам:

Среднее, $\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx$	=	Первый момент	=	Первый кумулянт
Дисперсия, $\mathbb{V}X = \mathbb{E}(X - \mathbb{E}X)^2$	=	Второй центральный момент	=	Второй кумулянт
Коэффициент асимметрии, $\frac{\mathbb{E}(X - \mathbb{E}X)^3}{(\mathbb{V}X)^{3/2}}$	≠	Третий центральный момент, $\mathbb{E}(X - \mathbb{E}X)^3$	=	Третий кумулянт
Эксцесс, $\frac{\mathbb{E}(X - \mathbb{E}X)^4}{(\mathbb{V}X)^2}$	≠	Четвертый центральный момент, $\mathbb{E}(X - \mathbb{E}X)^4$	≠	Четвертый кумулянт, $\mathbb{E}(X - \mathbb{E}X)^4 - 3(\mathbb{V}X)^2$

Все кумулянты *кумулятивны*, т. е. для суммы  $n$  случайных н. о. р. величин  $p$ -тый кумулянт  $\kappa_n^p = n\kappa_1^p$ . При гауссовом распределении кумулянты порядка  $p \geq 3$  все нулевые, и нормализованная сумма н. о. р. случайных величин сходится к гауссиане, об этом можно судить по нулевым высшим кумулянтам.

**Замечание 6.** Заметим, что

$$\lim_{n \rightarrow \infty} K_n^p = 0$$

для всех распределений вероятностей вне класса степенного закона.

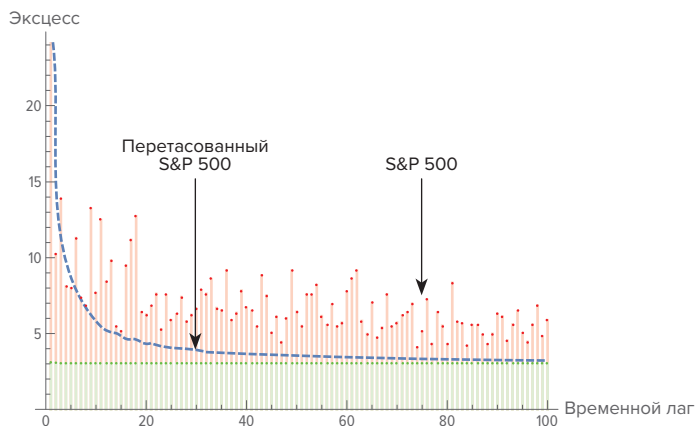
Заметим также, что  $\lim_{p \rightarrow \infty} K_n^p$  конечен для тонкохвостого класса. Другими словами, мы имеем четкое разделение классов на бассейн сходимости и бассейн несходимости моментов.

Для распределений вне бассейна степенного закона  $\forall p \in \mathbb{N}_{>2}$  отношение кумулянтов  $K_n^p$  падает как  $n^{p-2}$ .

Схематически доказательство опирается на  $\alpha$ -устойчивое распределение как предельный бассейн и на недифференцируемость порядка  $p$ , если он превышает показатель хвоста согласно формуле (8.4).

В Таблице 7.1 показано поведение кумулянтов  $K_n^p$  для суммы  $n$  величин.

Мы могли ожидать падения со скоростью  $\frac{1}{n^2}$  для стохастической волатильности (представленной гамма-распределением дисперсии). Однако, как показывает Рисунок 7.8, ничего похожего на столь быстрое падение не происходит. Ясно, что мы вне бассейна. Как видно из [228], при сложении экономических случайных величин сходимости эксцесса нет.



**Рисунок 7.8:** Визуальная диагностика сходимости для эксцесса S&P 500 за последние 17 000 наблюдений. Мы посчитали эксцесс при разных лагах для необработанного индекса S&P 500 и для перетасованных данных. Для необработанных данных сходимости 4-й нормы нет, а для перетасованных — есть. Можно предположить, что жирнохвостость присуща временной структуре данных, что волатильность образует комки по времени

1 Нормированных по  $K_n^2$ .

2 Имеется в виду приведенная сумма  $\sqrt{n}(\bar{X}_n - \mu) = \frac{(X_1 - \mu) + \dots + (X_n - \mu)}{\sqrt{n}}$ .

**Таблица 7.1:** Быстрота сходимости суммы  $n$  независимых случайных величин с тонкохвостым распределением — оценка по нормированным кумулянтам

Распред.	Пуассона <sup>1</sup> ( $\lambda$ )	Экспон. <sup>2</sup> ( $\lambda$ )	Гамма <sup>3</sup> ( $a, b$ )	Симметричная волатильность двух состояний <sup>4</sup> ( $\sigma_1, \sigma_2$ )	Гамма дисперсии <sup>5</sup> ( $a, b$ )
$K_n^2$	1	1	1	1	1
$K_n^3$	$\frac{1}{n\lambda}$	$\frac{2\lambda}{n}$	$\frac{2}{abn}$	0	0
$K_n^4$	$\frac{1}{n\lambda^2}$	$\frac{3!\lambda^2}{n}$	$\frac{3!}{a^2b^2n}$	$\frac{3(1-p)p(\sigma_1^2 - \sigma_2^2)^2}{n(p\sigma_1^2 + (1-p)\sigma_2^2)^3}$	$\frac{3b}{n}$

- 1 По Пуассону, с параметром  $\lambda$  распределено количество  $X$  событий в единичном промежутке времени, когда события происходят случайно, независимо друг от друга и со средней частотой  $\lambda$ . Распределение Пуассона — дискретное: наблюдения  $X$  — только натуральные числа.

Для дискретного распределения	Для непрерывного распределения
Вероятность $f(x) = \mathbb{P}(X=x), x \in \mathbb{N}$	Плотность вероятности $f(x) = \frac{\mathbb{P}(x < X < x+dx)}{dx}, x \in \mathbb{R}$
Среднее $\mathbb{E}X = \sum_{x \in \mathbb{N}} xf(x)$	Среднее $\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx$
Дисперсия $\mathbb{V}X = \sum_{x \in \mathbb{N}} (x - \mathbb{E}X)^2 f(x)$	Дисперсия $\mathbb{V}X = \int_{-\infty}^{\infty} (x - \mathbb{E}X)^2 f(x)dx$

Для распределения Пуассона с параметром  $\lambda$  вероятность  $f(x) = \lambda^x e^{-\lambda}/x!$ . Среднее  $\lambda$ , дисперсия опять  $\lambda$ , коэффициент асимметрии  $\lambda^{-1/2}$ , эксцесс  $3 + \lambda^{-1}$ . Все кумулянты  $\kappa^1 = \kappa^2 = \kappa^3 = \kappa^4 = \dots = \lambda$ . Для суммы  $n$  н. о. р. величин  $X_1 + \dots + X_n$  все кумулянты  $\kappa_n^1 = \kappa_n^2 = \kappa_n^3 = \kappa_n^4 = \dots = n\lambda$ . Таким образом,  $\lim_{p \rightarrow \infty} \kappa_n^p = n\lambda$  конечен.

- 2 Экспоненциально с параметром  $\lambda$  распределено время  $X$  ожидания следующего события, когда события происходят случайно, независимо друг от друга и со средней частотой  $\lambda$ . Плотность вероятности

$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{если } x \geq 0 \\ 0, & \text{иначе} \end{cases}$ . Среднее  $\lambda^{-1}$ , дисперсия  $\lambda^{-2}$ , коэффициент асимметрии 2, эксцесс 9. Кумулянты

$\kappa^1 = \lambda^{-1}, \kappa^2 = \lambda^{-2}, \kappa^3 = 2\lambda^{-3}, \kappa^4 = 6\lambda^{-4}, \dots, \kappa^p = \lambda^{-p}(p-1)!$ . Для суммы  $n$  н. о. р. величин  $X_1 + \dots + X_n$  кумулянты  $\kappa_n^1 = n\lambda^{-1}, \kappa_n^2 = n\lambda^{-2}, \kappa_n^3 = 2n\lambda^{-3}, \kappa_n^4 = 6n\lambda^{-4}, \dots, \kappa_n^p = n\lambda^{-p}(p-1)!, \lim_{p \rightarrow \infty} \kappa_n^p = \infty$ .

- 3 Гамма-распределение с параметром формы  $a > 0$  и параметром масштаба  $b > 0$  в частном случае  $a \in \mathbb{N}$  присуще случайному времени  $X$  ожидания  $a$  событий, когда события происходят случайно, независимо друг от друга и со средней частотой  $b^{-1}$ . В этом случае гамма-распределение называется также распределением Эрланга; плотность вероятности  $f(x) = x^{a-1} e^{-bx} b^a / (a-1)!$ . В общем случае  $a \in \mathbb{R}_{>0}$ ; плотность вероятности  $f(x) = \frac{x^{a-1} e^{-bx} b^a}{\Gamma(a)}$ . И для  $a \in \mathbb{N}$ , и в общем случае  $a \in \mathbb{R}_{>0}$  среднее  $\frac{a}{b}$ , дисперсия  $\frac{a}{b^2}$ , коэффициент асимметрии  $\frac{2}{\sqrt{a}}$ , эксцесс  $3 + \frac{6}{a}$ . Кумулянты  $\kappa^1 = \frac{a}{b}, \kappa^2 = \frac{a}{b^2}, \kappa^3 = \frac{2a}{b^3}, \kappa^4 = \frac{6a}{b^4}, \dots, \kappa^p = \frac{a(p-1)!}{b^p}$ . Для суммы  $n$  н. о. р.

$X_1 + \dots + X_n$  кумулянты  $\kappa_n^1 = \frac{an}{b}, \kappa_n^2 = \frac{an}{b^2}, \kappa_n^3 = \frac{2an}{b^3}, \kappa_n^4 = \frac{6an}{b^4}, \dots, \kappa_n^p = \frac{an(p-1)!}{b^p}, \lim_{p \rightarrow \infty} \kappa_n^p = \infty$ .

- 4 Имеется в виду  $f_X(x) = \frac{p}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{x^2}{2\sigma_1^2}\right) + \frac{1-p}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{x^2}{2\sigma_2^2}\right)$  в случае  $p = \frac{1}{2}$ .

- 5 Англ. *variance-gamma* — распределение вероятностей, представимое как нормальная смесь дисперсии-среднего, где в качестве взвешивающей плотности взята плотность гамма-распределения.

## 7.5. ПОВТОРИМ ТЕХНИКУ: ТРАДИЦИОННЫЕ ВЕРСИИ ЦПТ

Вспомним несколько теорий под общим названием ЦПТ.

**Стандартная<sup>1</sup> версия ЦПТ (Линдберга — Леви).** Пусть, как раньше, рассматривается последовательность  $X_i$  н. о. р. случайных величин с  $\mathbb{E}X_i = \mu$  и  $\mathbb{V}X_i = \sigma^2 < +\infty$ . Рассмотрим выборочное среднее  $\bar{X}_n$  первых  $n$  членов последовательности. Когда  $n$  приближается к бесконечности, приведенная сумма выборки  $\sqrt{n}(\bar{X}_n - \mu)$  сходится по распределению к гауссиане [20] [21]:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma^2).$$

Сходимость по распределению означает, что функция распределения для величины  $\sqrt{n}(\bar{X}_n - \mu)$  поточечно сходится к функции распределения  $\mathcal{N}(0, \sigma^2)$ ; другими словами, что для всякого действительного  $z$

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sqrt{n}(\bar{X}_n - \mu) \leq z) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq \frac{z}{\sigma}\right) = \Phi\left(\frac{z}{\sigma}\right), \quad \sigma > 0,$$

где  $\Phi(\cdot)$  — функция стандартного нормального распределения. Заметим, что эта сходимость равномерна по  $z$ , то есть

$$\lim_{n \rightarrow \infty} \sup_{z \in \mathbb{R}} \left| \mathbb{P}(\sqrt{n}(\bar{X}_n - \mu) \leq z) - \Phi\left(\frac{z}{\sigma}\right) \right| = 0,$$

где  $\sup$  — точная верхняя граница множества.

**ЦПТ Ляпунова.** У Ляпунова слагаемые предполагаются независимыми, но не обязательно одинаково распределенными. Кроме того, в посылке теоремы требуется, чтобы у случайных величин  $|X_i|$  существовали моменты некоторого порядка вида  $2 + \delta$  и чтобы скорость роста этих моментов ограничивалась условием Ляпунова, которое состоит в следующем.

Определим

$$s_n^2 = \sum_{i=1}^n \sigma_i^2.$$

Если для некоторого  $\delta > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}|X_i - \mu_i|^{2+\delta} = 0,$$

то сумма  $\frac{X_i - \mu_i}{s_n}$  при  $n$ , стремящемся к бесконечности, сходится по распределению к стандартной нормальной случайной величине:

$$\frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{D} \mathcal{N}(0, 1).$$

Если последовательность случайных величин удовлетворяет условию Ляпунова, она также удовлетворяет условию Линдберга, которое мы теперь рассмотрим; обратное же неверно.

<sup>1</sup> В русской литературе — классическая центральная предельная теорема.

**Условие Линдеберга.** По Линдебергу, удается достичь ЦПТ при более слабых допущениях. В тех же обозначениях, что выше: если верно, что

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E} \left( |X_i - \mu_i|^2 \cdot \mathbb{1}_{\{|X_i - \mu_i| > \varepsilon s_n\}} \right) = 0$$

для всех  $\varepsilon > 0$ , где  $\mathbb{1}$  — индикаторная функция<sup>1</sup>, то случайная величина

$$Z_n = \frac{1}{s_n} \sum_{i=1}^n (X_i - \mu_i)$$

сходится по распределению к гауссиане при  $n \rightarrow \infty$ .

Условие Линдеберга достаточное и, вообще говоря, избыточное — кроме случая, когда для рассматриваемой последовательности случайных величин

$$\max_{1 \leq k \leq n} \frac{\sigma_k^2}{\sigma_n^2} \rightarrow 0 \text{ при } n \rightarrow \infty.$$

В этом случае условие Линдеберга необходимое и достаточное, то есть оно выполнено тогда и только тогда, когда выполнено заключение центральной предельной теоремы.

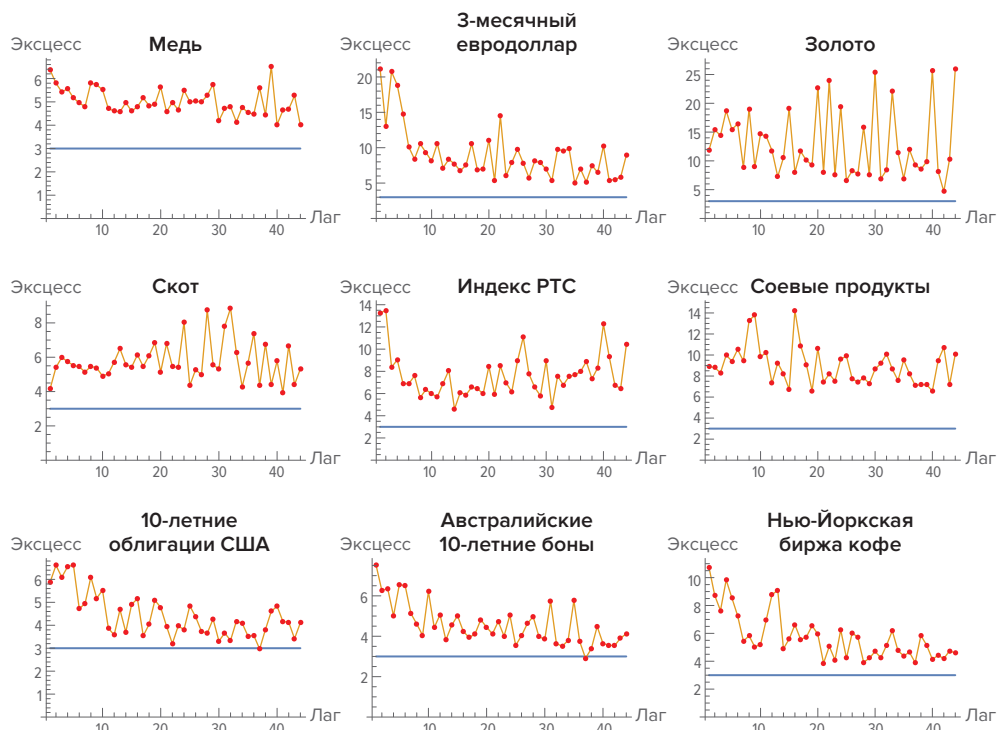
## 7.6. ЗАКОН БОЛЬШИХ ЧИСЕЛ ДЛЯ ВЫСШИХ МОМЕНТОВ

### 7.6.1. Высшие моменты

На критерий жирнохвостости можно смотреть как на применение закона больших чисел к высшим моментам и их сходимости. Можно посмотреть, как ведет себя кумулятивное среднее  $p$ -того момента — аналогично обычной наглядной проверке ЗБЧ, которую мы видели в Главе 3, только применительно не к самой случайной величине  $X$ , а к ее степени  $X^p$  (или к степени центрированной  $X$ ). Чтобы узнать, срабатывает ли закон больших чисел, посмотрим, приводит ли добавление наблюдений к сокращению изменчивости среднего (или дисперсии, если она существует). Когда момент не существует, мы увидим случайные скачки — то есть увидим, что даже большие разные выборки выдают разное среднее. Когда момент существует, добавление наблюдений приведет рано или поздно к тому, что скачки прекратятся.

Еще один наглядный метод — посчитать вклад максимального наблюдения в общую сумму и посмотреть, как он ведет себя с ростом  $n$ . Такой график называется MS [115], от *maximum to sum*; пример показан на Рисунке 7.10.

<sup>1</sup> Так в теории вероятностей называют то, что в теории множеств называется характеристической функцией. Это функция принадлежности элемента  $x$  из области определения заданному подмножеству  $A$  области определения:  $\mathbb{1}_A(x) \triangleq \begin{cases} 1, & \text{если } x \in A \\ 0, & \text{иначе} \end{cases}$ . Иногда вместо множества  $A$  указывают критерий принадлежности множеству  $A$ ; тогда вместо  $\mathbb{1}_{\{|X_i - \mu_i| > \varepsilon s_n\}}$  пишут  $\mathbb{1}_{|X_i - \mu_i| > \varepsilon s_n}$ .



**Рисунок 7.9:** Поведение 4-го момента при агрегации данных; представлены курсы различных ценных бумаг с окном сглаживания от 1 до 45 дней. Согласно сходимости распределения к гауссову эксцесс должен приблизиться к значению 3; фактически он этого не делает (использованы данные, собранные для [228]). Невозможно объявить, что такие данные становятся гауссовыми после снижения частоты выборки с ежедневной до ежемесячной

**Таблица 7.2:** Эксцесс  $K(t)$  при ежедневном, 10-дневном и 66-дневном окне  $t$  сглаживания для различных случайных величин

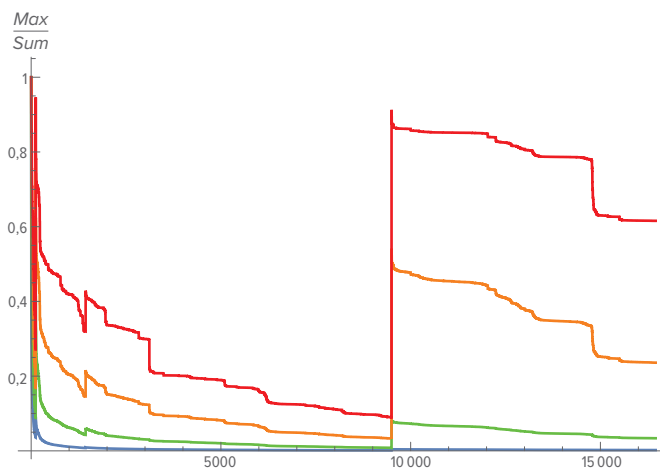
	K(1)	K(10)	K(66)	Максимум за квартал	Годы
10-летние казначейские облигации США	5,9	5,5	4,9	0,10	27
1-месячный евродоллар	41,5	28,0	6,0	0,31	19
30-летние бонды Казначейства США	5,6	4,7	3,9	0,02	32
3-месячный евродоллар	21,1	8,1	7,0	0,25	28
5-летние казначейские облигации США	5,1	3,2	2,5	0,06	21
Bovespa	24,9	5,0	2,3	0,27	16
СТ	7,8	4,8	3,7	0,25	48
Австралийские 10-летние казначейские бонды	7,5	6,2	3,5	0,08	25
Австралийские 3-летние казначейские бонды	7,5	5,4	4,2	0,06	21
Австралийский доллар / доллар США	6,3	3,8	2,9	0,12	22



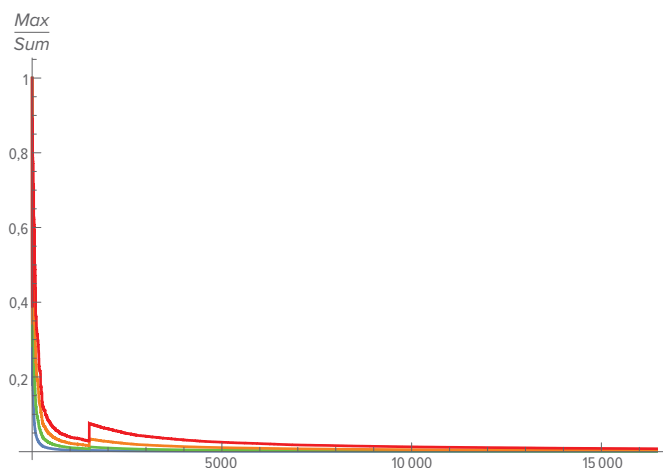
Продолжение таблицы 7.2

Британский фунт / доллар США	6,9	7,4	5,3	0,05	38
Евро / ранее немецкая марка	5,5	3,8	2,8	0,06	38
Евробонд	4,9	3,2	3,3	0,06	18
Золото	11,9	14,5	16,6	0,04	35
Иена / доллар США	9,7	6,1	2,5	0,27	38
Индекс CAC 40 (Франция)	6,5	4,7	3,6	0,05	20
Индекс FTSE	15,2	27,4	6,5	0,54	25
Индекс Nasdaq	11,4	9,3	5,0	0,13	21
Индекс S&P 500	38,2	7,7	5,1	0,79	56
Индекс Дакс	8,0	6,5	3,7	0,20	18
Индекс Никкей	52,6	4,0	2,9	0,72	23
Индекс РТС (Россия)	13,3	6,0	7,3	0,13	17
Канадский доллар	7,4	4,1	3,9	0,06	38
Короткий стерлинг	851,8	93,0	3,0	0,75	17
Кукуруза	9,4	8,0	5,0	0,18	49
Малая капитализация	6,1	5,7	6,8	0,06	17
Медь	6,4	5,5	4,5	0,05	48
Нью-Йоркская биржа какао	4,9	4,0	5,2	0,04	47
Нью-Йоркская биржа кофе	10,7	5,2	5,3	0,13	37
Природный газ	6,0	3,9	3,8	0,06	19
Пшеница	5,6	6,0	6,9	0,02	49
Сахар-сырец	9,4	6,4	3,8	0,30	48
Свиньи	4,5	4,6	4,8	0,05	43
Серебро	160,3	22,6	10,2	0,94	46
Скот	4,2	4,9	5,6	0,04	44
Соевое масло	5,5	7,0	4,9	0,11	47
Соевые продукты	8,9	9,8	8,5	0,09	48
Соя	7,1	8,8	6,7	0,17	47
Сырая нефть	29,0	4,7	5,1	0,79	26
Топочный мазут	20,0	4,1	4,4	0,74	31
Фондовый индекс Джакарты	40,5	6,2	4,2	0,19	16
Швейцарский франк	5,1	3,8	2,6	0,05	38
Японские гособлигации	17,2	16,9	4,3	0,48	24

Еще один наглядный метод — посчитать вклад максимального наблюдения в общую сумму и посмотреть, как он ведет себя с ростом  $n$ . Такой график называется MS [115], от *maximum to sum*; пример показан на Рисунке 7.10.



**Рисунок 7.10:** На графике MS видно, как вели себя кумулятивные моменты  $p = 1, 2, 3, 4$  для биржевого индекса S&P 500 за последние 60 лет, по конец 2018. График MS (Maxitum to sum, максимум к сумме) будет представлен в разделе 10.2.6



**Рисунок 7.11:** Проверка гауссовости для данных на Рисунке 7.10

## 7.7. СРЕДНЕЕ ОТКЛОНЕНИЕ ДЛЯ УСТОЙЧИВЫХ РАСПРЕДЕЛЕНИЙ

Подготовим один результат для следующей главы. Он касается нормы  $L^1$  для случаев конечного математического ожидания, но бесконечной дисперсии<sup>1</sup>. Очевидно, что по норме  $L^2$  оценить сжатие распределения вблизи математического ожидания невозможно.

Погрешность суммы по норме  $L^1$  определяется так. Пусть  $\vartheta(x)$  — функция Хевисайда (равная нулю для отрицательных аргументов и единице для положительных), и функция знака числа связана с ней как  $\operatorname{sgn} x = 2\vartheta(x) - 1$ . Характеристическая функция функции знака числа имеет вид:

$$\chi^{\operatorname{sgn} x}(t) = \frac{2i}{t}. \quad (7.6)$$

<sup>1</sup> Договоримся, что термин *бесконечный* относится к случаю однохвостого, ограниченного с одной стороны распределения; например, для произвольной случайной величины  $X$  величина  $X^2$  однохвостая, так что дисперсия  $\mathbb{E}X^2$  может быть бесконечной; термин *неопределенный* относится к случаю, когда случайная величина двуххвостая, например пресловутое распределение Коши. — Прим. автора.

Пусть  $\chi^d(\cdot)$  — характеристическая функция произвольного невырожденного распределения. Сверткой  $\chi^{\text{sgn } x} * (\chi^d)^n$  получим характеристическую функцию для позитивной вариации<sup>1</sup>  $n$  независимых слагаемых

$$\chi^m = \int_{-\infty}^{\infty} \chi^{\text{sgn } x}(t) (\chi^d(u-t))^n dt.$$

В нашем случае абсолютное отклонение вдвое больше положительных значений  $\chi$ :

$$\chi(|S_n|) = 2i \int_{-\infty}^{\infty} \frac{(\chi(t-u))^n}{t} du.$$

Это преобразование Гильберта для  $\chi$ , при котором под  $\int_{-\infty}^{\infty}$  понимается главное значение интеграла по Коши (Pinelis, 2015)[193]. В нашем случае, предполагая, что все независимые слагаемые — копии одного и того же распределения, можем заменить произведение  $(\chi(t))^n$  на  $\chi_s(t)$  — аналогичную характеристическую функцию, у которой масштаб изменен на  $\sigma_s = n^{1/a}$ , а параметр  $\beta$  оставлен прежним:

$$\mathbb{E}|X| = 2i \frac{\partial}{\partial u} \text{p.v.} \int_{-\infty}^{\infty} \frac{\chi_s(t-u)}{t} dt \Big|_{t=0}. \quad (7.7)$$

Перепишем затем преобразование Гильберта  $H$ , как описано в [193],

$$(Hf)(t) = \frac{2}{\pi i} \int_0^{\infty} (\chi_s(u+t) - \chi_s(u-t)) dt$$

в виде

$$(Hf)(t) = -i \frac{\partial}{\partial u} \left( 1 + \chi_s(u) + \frac{1}{\pi i} \int_0^{\infty} (\chi_s(u+t) - \chi_s(u-t) - \chi_s(t) + \chi_s(-t)) \frac{dt}{t} \right). \quad (7.8)$$

Рассмотрим устойчивое распределение, определенное в Разделе 7.2.1.

Внеся дифференцирование под интеграл и заменив переменную  $z = \ln t$ , находим

$$\begin{aligned} \mathbb{E}|X|_{(\tilde{\alpha}_s, \beta, \sigma_s, 0)} = & \int_{-\infty}^{\infty} 2i \alpha_s \exp(-(\sigma_s e^z)^{\alpha_s} - z) (\sigma_s e^z)^{\alpha_s} \left( \beta \operatorname{tg} \frac{\pi \alpha_s}{2} \sin \left( \beta \operatorname{tg} \frac{\pi \alpha_s}{2} (\sigma_s e^z)^{\alpha_s} \right) + \right. \\ & \left. + \cos \left( \beta \operatorname{tg} \frac{\pi \alpha_s}{2} (\sigma_s e^z)^{\alpha_s} \right) \right) dz, \end{aligned}$$

и этот интеграл легко берется. Получаем:

$$\mathbb{E}|X|_{(\tilde{\alpha}_s, \beta, \sigma_s, 0)} = \frac{\sigma_s}{2\pi} \Gamma \frac{\alpha_s - 1}{\alpha_s} \left( \left( 1 + i\beta \operatorname{tg} \frac{\pi \alpha_s}{2} \right)^{1/\alpha_s} + \left( 1 - i\beta \operatorname{tg} \frac{\pi \alpha_s}{2} \right)^{1/\alpha_s} \right). \quad (7.9)$$

## ДАЛЕЕ

В следующей главе представлена центральная концепция: как работать с законом средних чисел? Как переводить с языка распределений на язык других распределений?

<sup>1</sup> В русской литературе позитивную вариацию называют также зарядом.

## СКОЛЬКО НУЖНО ДАННЫХ? РАБОЧИЙ ПОКАЗАТЕЛЬ ЖИРНОХВОСТОСТИ<sup>†</sup>

В этой главе<sup>1</sup>, исследовательской, мы обсудим законы средних чисел. Мы представим рабочий показатель для одномерных односторонних распределений вероятности с конечным первым моментом, пробегающий диапазон  $[0, 1]$ , где 0 отвечает максимальной тонкохвостости (гауссиане), а 1 — максимальной жирнохвостости. Этот показатель продиктован вопросом: «Сколько нужно данных, чтобы выносить осмысленные суждения о наборе данных?»

Показатель каппа полезен, чтобы:

- оценивать размер выборки  $n$ , необходимой для статистической значимости за пределами гауссовой ситуации,
- измерять скорость сходимости к гауссиане (или устойчивому бассейну),
- сравнивать на практике данные из распределений разных классов жирнохвостости,
- оценивать при построении портфеля, сколько нужно ценных бумаг, чтобы диверсификация обеспечила заданный уровень устойчивости,
- понять, откуда взялись противоречия между атрибутами логнормального моделирования при разной параметризации дисперсии.

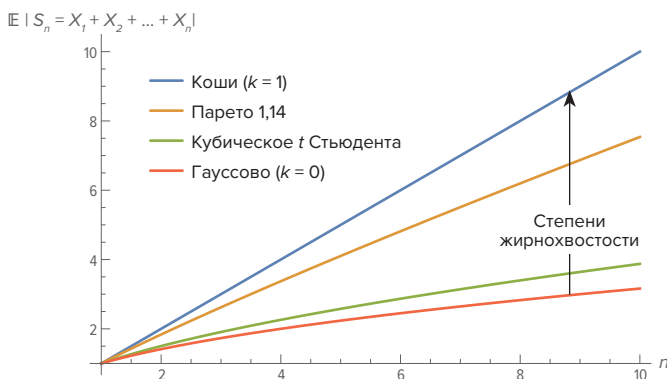
Асимптотическому поведению посвящена богатая литература, но существует дыра на месте изучения конечных, рабочих значений  $n$ .

<sup>1</sup> Автор больше всего благодарен точным комментариям Майкла Лулакиса, который к тому же предоставил дотошный вывод пределов  $k$  для  $t$ -распределения Стюдента и для логнормального распределения, а также терпению и мудрости Спираса Макридакиса. Глава вначале появилась как доклад на конференции *Extremes and Risks in Higher Dimensions* 12–16 сентября 2016 в Центре Лоренца в Лейдене и как статья в юбилейном сборнике Джима Гезерала в Курантовском институте математических наук в октябре 2017. Автор признателен Жан-Филипу Бушо, Джону Эйнемало, Паскуале Чирилло и другим. Лоренц де Гаан предложил заменить обозначение показателя с гаммы на каппу, чтобы избежать путаницы. Отдельная благодарность Колману Хамфри, Майклу Лолеру, Даниэлю Дюфрену и другим за обсуждение и идеи касательно вывода формул.

Предпосылки: известные показатели жирнохвостости: (1) показатель хвоста в классе степенного закона и (2) эксцесс для распределений с конечными моментами, — применимы не ко всем распределениям и не позволяют сравнивать данные из разных классов и систем параметров, а именно сравнивать распределение степенного закона за пределами бассейна устойчивости по Леви с распределением другого класса или сравнивать между собой распределения по степенному закону с разным числом слагаемых. Как сравнить между собой сумму ста случайных величин, если в одном случае они имеют  $t$ -распределение Стьюдента и три степени свободы, а в другом случае принадлежат классу устойчивости по Леви или логнормальному классу? Как сравнить между собой сумму ста случайных величин, с  $t$ -распределением Стьюдента и тремя степенями свободы, и одиночную случайную величину с  $t$ -распределением Стьюдента и двумя степенями свободы?

Мы предлагаем рабочий, эвристичный показатель, позволяющий сравнивать суммы  $n$  независимых величин любых распределений с конечным первым моментом. Метод основан на скорости сходимости конечной суммы, из  $n$  слагаемых, к закону больших чисел.

Мы выводим явные выражения или результаты моделирования и границы для таких распределений, как логнормальное, экспоненциальное, Парето и  $t$ -распределение Стьюдента при различных калибровках — в дополнение к общему виду класса распределений Пирсона.



**Рисунок 8.1:** Интуитивное представление о том, что измеряет  $k$ : как растет среднее отклонение суммы одинаковых независимых случайных величин  $S_n = X_1 + X_2 + \dots + X_n$  с ростом выборки и как можно доасимптотически сравнить распределения разных классов

## 8.1. ВВЕДЕНИЕ И ОПРЕДЕЛЕНИЯ

Как сравнить распределение Парето с хвостом  $\alpha = 2,1$  и тем самым с конечной дисперсией — и гауссово распределение? Асимптотически оба распределения принадлежат к классу правильно меняющихся функций распределения с конечным вторым моментом и при суммировании большого числа слагаемых дают гауссиану, но есть различия в доасимптотическом поведении. Стандартных способов сравнить эти распределения нет; показатели, основанные на высших моментах, такие как эксцесс, бесполезны. Столь же непросто сравнить распределение Парето, имеющее бесконечную дисперсию, и соответствующее предельное  $\alpha$ -устойчивое распределение (когда оба распределения имеют один и тот же индекс хвоста или показатель хвоста). И как сравнить жирнохвостость  $t$ -распределения Стьюдента с 3 степенями свободы — и устойчивое по Леви с показателем хвоста 1,95? У обоих распределений конечное математическое

ожидание; конечная дисперсия только у первого, но при малом числе слагаемых оно ведет себя даже более жирнохвостым образом, если судить по некоторым рабочим критериям.

**Критерий жирнохвостости.** Есть разные способы, как определить жирные хвосты и ранжировать распределения согласно тому или иному определению.

В узком классе распределений, у которых все моменты конечные, критерием служит эксцесс — по эксцессу легко сравнивать отличие того или иного распределения от гауссова, служащего нормой.

Для класса степенного закона критерием может служить показатель хвоста. Кроме того, можно использовать экстремальные значения и найти вероятность превысить максимальное значение с поправкой на масштаб (этот подход практикуется в теории экстремальных значений).

На практике жирнохвостость должна оценивать концентрацию случайной величины в важнейших наблюдениях, отвечая на вопрос:

— Какой вклад смогло внести в статистические параметры одно-единственное наблюдение?

Или в другой формулировке, с поправкой на масштаб (среднюю дисперсию):

— Какая доля национального богатства сосредоточилась в руках самого богатого жителя?

Мы здесь для своих целей используем следующий критерий, который тоже можно сопоставить с концентрацией:

— Сколько дополнительных данных (при таком-то распределении вероятностей) помогут повысить устойчивость наблюдаемого среднего?

Эта задача имеет смысл не только в статистике; ее можно понять и так:

— Насколько диверсификация ценных бумаг в структуре портфеля (при неизменной общей стоимости портфеля) повысит его устойчивость?

Наш показатель отличается от асимптотических показателей (в частности, от тех, что используются в теории экстремальных значений) тем, что является по своей сути доасимптотическим.

Это важное преимущество; в реальности и в реалистичных моделях асимптота не достигается.

**Что дает этот показатель?** Предлагаемый показатель,  $\kappa$ , дает следующее:

- Позволяет сравнивать суммы  $n$  величин, когда у них разные распределения при заданном числе слагаемых или одинаковое распределение при разных  $n$ , и оценивать доасимптотические свойства заданных распределений.
- Дает меру расстояния до предельного распределения, а именно до бассейна устойчивости по Леви (которому, в частности, принадлежит гауссово распределение).
- Для статистических выводов позволяет оценивать скорость срабатывания закона больших чисел как скорость изменения абсолютной погрешности математического ожидания, оцениваемого по выборочному среднему, при увеличении размера выборки  $n$ .
- Позволяет оценивать сравнительную жирнохвостость двух разных одномерных распределений, имеющих конечный первый момент.
- Позволяет заранее узнавать, сколько раз потребуется повторить обсчет для моделирования по методу Монте-Карло.

**Положение дел со статистическими выводами.** Задача оценивать скорость обычно игнорировалась (см. выше комментарии в Главе 3 о 9400 страницах *Энциклопедии математической статистики* [147]). Редко встретишь обсуждение вопроса, сколько времени уйдет на достижение асимптоты или как относиться к сумме  $n$  слагаемых, когда их много, но недостаточно для так называемой «нормальной аппроксимации».

Повторим наш девиз — «статистический параметр никогда не бывает стандартным». Цель нового показателя — оценить стандартность стандарта и отклонение от стандарта с точки зрения статистической значимости.

## 8.2. ПОКАЗАТЕЛЬ КАППА

### Определение 8.1 (показатель $\kappa$ )

Пусть  $X_1, \dots, X_n, \dots$  — н. о. р. случайные величины с конечным математическим ожиданием, то есть  $\mathbb{E} X < +\infty$ . Пусть  $S_n = X_1 + X_2 + \dots + X_n$  — частичная сумма. Пусть  $\mathbb{M}(n) = \mathbb{E}|S_n - \mathbb{E}S_n|$  — ожидаемое среднее абсолютное отклонение от среднего для суммы  $n$  слагаемых<sup>1</sup>. Определим скорость сходимости при росте числа слагаемых от  $n_0$  до  $n$ :

$$\kappa_{n_0, n} = \min \left\{ \kappa \mid \frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \left( \frac{n}{n_0} \right)^{\frac{1}{2-\kappa}} \right\}, \quad n_0 = 1, 2, \dots^2.$$

Поскольку  $n > n_0 \geq 1$ ,

$$\kappa_{n_0, n} = 2 - \frac{\ln n - \ln n_0}{\ln \frac{\mathbb{M}(n)}{\mathbb{M}(n_0)}}. \quad (8.1)$$

В дальнейшем для случая добавления одного-единственного слагаемого,  $n = n_0 + 1$ , будем коротко писать наш показатель как  $\kappa_{n_0}$ .

На такие «локальные» каппы можно разложить  $\kappa_{n_0, n}$ , или  $\kappa(n_0, n)$ , при сколь угодно большом числе добавленных слагаемых; процедура аналогична «пошаговому» расчету процента начислений:

<sup>1</sup> То есть  $\mathbb{M}(n)$  — сокращение вместо  $\mathbb{M}S_n$ , согласно определению  $\mathbb{M}$ , оператора среднего абсолютного отклонения.

<sup>2</sup> Случай  $\kappa = 0$  означает, что  $\frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \left( \frac{n}{n_0} \right)^{\frac{1}{2-0}} = \sqrt{\frac{n}{n_0}}$ , то есть с ростом  $n$ , числа случайных слагаемых, когда ожидаемая сумма растет пропорционально  $n$ , ожидаемые отклонения суммы растут намного медленнее, пропорционально корню из  $n$ . Это случай быстрой сходимости последовательности сумм. Противоположная крайность: случай  $\kappa = 1$ , — означает, что  $\frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \left( \frac{n}{n_0} \right)^{\frac{1}{2-1}} = \frac{n}{n_0}$ , то есть с ростом  $n$  ожидаемые отклонения суммы растут пропорционально  $n$  и не уменьшаются относительно ожидаемых значений суммы.

Такое поведение свойственно распределению Коши, с двухвостой одногорбой плотностью вероятности  $f(x) = \frac{1}{\pi \gamma \left( 1 + \left( \frac{x - x_0}{\gamma} \right)^2 \right)}$ , где  $x_0$  — медиана и мода, а  $\gamma$  — параметр масштаба. При распределении Коши среднее не определено, а дисперсия бесконечна, даже если параметр масштаба мал.

$$\kappa(n_0, n) = 2 - \frac{\ln n - \ln n_0}{\sum_{i=n_0}^{n-1} \frac{\ln(i+1) - \ln i}{2 - \kappa(i, i+1)}} \quad (8.2)^1$$

**Об использовании среднего отклонения.** Заметим, что для оценки разброса вокруг среднего значения мы используем среднее абсолютное отклонение, благодаря чему остаемся в пределах нормы  $L^1$  даже при отсутствии конечной дисперсии. На самом деле даже при наличии конечной дисперсии распределения степенного закона проявляют неустойчивый, невнятный второй момент. Среднее отклонение показывает себя здесь как более надежная мера. (Можно доказать, что среднее абсолютное отклонение вообще более эффективно, кроме узкого случая, когда эксцесс равен 3, то есть когда распределение гауссово. Смотрите более подробное обсуждение в [237]; о других преимуществах среднего абсолютного отклонения смотрите в [187].)

### 8.3. УСТОЙЧИВЫЙ БАССЕЙН СХОДИМОСТИ КАК ТОЧКА ОТСЧЕТА

#### Определение 8.2 (класс $\mathfrak{P}$ )

Класс  $\mathfrak{P}$  степенных законов, или правильно меняющихся функций, определяется как множество всех таких случайных величин  $X$ , для которых:

$$\mathfrak{P} = \{X \mid \mathbb{P}(X > x) \sim L(x) x^{-\alpha}\}, \quad (8.3)$$

где:

$\sim$  означает, что отношение выражения до тильды к выражению после тильды стремится к 1 при  $x \rightarrow \infty$ ;

функция  $L: [x_{\min}, +\infty) \rightarrow (0, +\infty)$  — медленно меняющаяся согласно критерию

$$\lim_{x \rightarrow +\infty} \frac{L(kx)}{L(x)} = 1 \text{ при любом } k > 0;$$

константа  $\alpha > 0$ .

Определим далее аттрактор суммы одинаково распределенных величин, которые в нашем случае будут с идентичными параметрами.

#### Определение 8.3 (устойчивый класс $\mathfrak{S}$ )

Случайная величина  $X$  следует устойчивому распределению (или  $\alpha$ -устойчивому), и про нее пишут  $X \sim S(\tilde{\alpha}, \beta, \mu, \sigma)$ , если ее характеристическая функция  $\chi(t) = \mathbb{E}e^{itX}$  имеет вид

<sup>1</sup> Если использовать сокращенную запись «локальной» каппы, то  $\kappa_{n_0, n} = 2 - \frac{\ln n - \ln n_0}{\sum_{i=n_0}^{n-1} \frac{\ln(i+1) - \ln i}{2 - \kappa_i}}$ .

Проверим эту формулу для легкого случая  $n = n_0 + 2$ . Согласно проверяемой формуле,

$$\kappa_{n_0, n_0+2} = 2 - \frac{\ln(n_0+2) - \ln n_0}{\frac{\ln(n_0+1) - \ln n_0}{2 - \kappa_{n_0}} + \frac{\ln(n_0+2) - \ln(n_0+1)}{2 - \kappa_{n_0+1}}}$$

Подставим вместо  $\kappa_{n_0}$  и  $\kappa_{n_0+1}$  выражения из определения (8.1):

$$\kappa_{n_0, n_0+2} = 2 - \frac{\ln(n_0+2) - \ln n_0}{\frac{\ln(n_0+1) - \ln n_0}{\ln \frac{\mathbb{M}(n_0+1)}{\mathbb{M}(n_0)}} + \frac{\ln(n_0+2) - \ln(n_0+1)}{\ln \frac{\mathbb{M}(n_0+2)}{\mathbb{M}(n_0+1)}}} = 2 - \frac{\ln(n_0+2) - \ln n_0}{\ln \frac{\mathbb{M}(n_0+2)}{\mathbb{M}(n_0)}},$$

что и требовалось.



$$\chi(t) = \begin{cases} \exp\left(i\mu t - |t\sigma|^{\tilde{\alpha}} \left(1 - i\beta \operatorname{tg} \frac{\pi\tilde{\alpha}}{2} \operatorname{sgn} t\right)\right), & \text{если } \tilde{\alpha} \neq 1 \\ \exp\left(it\left(\frac{2\beta\sigma \ln \sigma}{\pi} + \mu\right) - |t\sigma| \left(1 + \frac{2i\beta \operatorname{sgn} t \ln |t\sigma|}{\pi}\right)\right), & \text{если } \tilde{\alpha} = 1. \end{cases} \quad (8.4)$$

Теперь определим соответствующий устойчивый показатель  $\tilde{\alpha}$ :

$$\tilde{\alpha} \triangleq \begin{cases} \mathbb{1}_{\alpha < 2} \alpha + \mathbb{1}_{\alpha \geq 2} \cdot 2, & \text{если } X \text{ принадлежит } \mathfrak{F} \\ 2, & \text{иначе} \end{cases}. \quad (8.5)^1$$

Обсудим подробнее класс  $\mathfrak{S}$ .

### 8.3.1. Эквиваленты устойчивых распределений

Для всех  $n_0$  и  $n \geq 1$  в устойчивом классе  $\mathfrak{S}$  с  $\tilde{\alpha} \geq 1$ :

$$\kappa_{n_0, n} = 2 - \tilde{\alpha},$$

просто в силу того свойства, что

$$\mathbb{M}(n) = n^{1/\alpha} \mathbb{M}(1). \quad (8.6)$$

Это просто доказывает, что  $\kappa_{n_0, n} = 0$  для гауссова распределения.

Проблема предасимптотики для  $n$  слагаемых сводится к следующему:

- Каково распределение при  $n_0 = 1$  (начинаем ли мы со стандартного, хорошо изученного распределения)?
- Каково распределение при некотором числе  $n_0$  слагаемых?
- Сходится ли  $\kappa_n \rightarrow (2 - \tilde{\alpha})$ , и если да, то с какой скоростью?

Таблица 8.1: Каппа для двух слагаемых,  $\kappa_1$

Распределение	$\kappa_1$
t-распределение Стюдента ( $\alpha$ )	$2 - \frac{2 \ln 2}{2^{2-\alpha} \Gamma\left(\alpha - \frac{1}{2}\right) \Gamma\left(\frac{\alpha}{2}\right)^2 + \ln \pi}$
Экспоненциальное/Гамма	$2 - \frac{\ln 2}{2 \ln 2 - 1} \approx 0,21$
Парето ( $\alpha$ )	$2 - \ln 2 \left/ \ln \left( (\alpha-1)^{2-\alpha} \alpha^{\alpha-1} \int_0^{2/(\alpha-1)} -2\alpha^2(y+2)^{-2\alpha-1} \times \right. \right. \\ \left. \left. \times \left( \frac{2}{\alpha-1} - y \right) \left( B_{\frac{1}{y+2}}(-\alpha, 1-\alpha) - B_{\frac{y+1}{y+2}}(-\alpha, 1-\alpha) \right) dy \right) \right.$

<sup>1</sup> Другими словами,  $\tilde{\alpha} \triangleq \begin{cases} \alpha, & \text{если } \alpha < 2 \text{ и } X \in \mathfrak{F} \\ 2, & \text{иначе} \end{cases}$ .

Продолжение таблицы 8.1

Нормальное $(\mu, \sigma)$ , переключающееся на дисперсию $\sigma^2 a$ с вероятностью $p$	$2 - \ln 2 \Bigg/ \ln \left( \sqrt{2} \left( \sqrt{\frac{ap}{p-1} + \sigma^2} + p \left( -\sqrt{2} \sqrt{\frac{ap}{p-1} + \sigma^2} + \right. \right. \right. \\ + p \left( \sqrt{\frac{ap}{p-1} + \sigma^2} - \sqrt{2a \left( \frac{1}{p-1} + 2 \right) + 4\sigma^2} \right) + \sqrt{a + \sigma^2} \Bigg) + \\ \left. \left. + \sqrt{2a \left( \frac{1}{p-1} + 2 \right) + 4\sigma^2} \right) \Bigg/ \left( p \sqrt{a + \sigma^2} - (p-1) \sqrt{\frac{ap}{p-1} + \sigma^2} \right) \right)$
Логнормальное $(\mu, \sigma)$	$\approx 2 - \ln 2 \Bigg/ \ln \left( 2 \operatorname{erf} \frac{\sqrt{\ln \left( \frac{1}{2} (\exp \sigma^2 + 1) \right)}}{2\sqrt{2}} \Bigg/ \operatorname{erf} \frac{\sigma}{2\sqrt{2}} \right)$

Таблица 8.2: Сводка основных результатов

Распределение	$\kappa_n$
Экспоненциальное/Гамма	Выражается в явном виде
Логнормальное $(\mu, \sigma)$	Нет явного выражения для $\kappa_n$ , но есть для нижней и верхней границы (при низком или высоком $\sigma$ или $n$ ). При промежуточных $\sigma$ приближается распределением Пирсона IV типа
Парето $(\alpha)$ (Константа)	$\kappa_2$ выражается в явном виде (нижняя граница при всех $\alpha$ )
t-распределение Стьюдента $(\alpha)$	$\kappa_1$ выражается в явном виде при $\alpha = 3$

Таблица 8.3: Сравниваем Парето с t-распределением Стьюдента (при одинаковом показателе хвоста  $\alpha$ )

$\alpha$	Парето	Парето	Парето	Стьюдент	Стьюдент	Стьюдент
	$\kappa_1$	$\kappa_{1,30}$	$\kappa_{1,100}$	$\kappa_1$	$\kappa_{1,30}$	$\kappa_{1,100}$
1,25	0,829	0,787	0,771	0,792	0,765	0,756
1,5	0,724	0,65	0,631	0,647	0,609	0,587
1,75	0,65	0,556	0,53	0,543	0,483	0,451
2	0,594	0,484	0,449	0,465	0,387	0,352
2,25	0,551	0,431	0,388	0,406	0,316	0,282
2,5	0,517	0,386	0,341	0,359	0,256	0,227
2,75	0,488	0,356	0,307	0,321	0,224	0,189
3	0,465	0,3246	0,281	0,29	0,191	0,159
3,25	0,445	0,305	0,258	0,265	0,167	0,138
3,5	0,428	0,284	0,235	0,243	0,149	0,121
3,75	0,413	0,263	0,222	0,225	0,13	0,10
4	0,4	0,2532	0,211	0,209	0,126	0,093

### 8.3.2. Практическая значимость при достаточной выборке

**Доверительные интервалы:** Простое эвристическое правило гласит, что чем выше  $\kappa$ , тем драматичнее недостаточность доверительного интервала, рассчитанного обычным способом. Значения  $\kappa$  выше 0,15 говорят о существенной ненадежности «нормального приближения». Под сомнение поставлены многочисленные статьи об исследованиях жирнотелых областей.

Вычисления вроде тех, что сделаны в Таблице 8.2, позволяют сравнить различные распределения, имеющие различные системы параметров (различные распределения Парето с симметричным  $t$ -распределением Стюдента и, конечно, с гауссианой, у которой каппа постоянна и равна 0).

Как мы упомянули во введении, руководствуясь требуемым размером выборки для статистически значимых выводов, мы пытаемся спланировать  $n$ , число слагаемых. Однако нередко закон больших чисел используется не по делу; нам нужен более надежный показатель для размера выборки.

Во многих статьях при обсуждении финансовых вопросов, например в [99], в качестве бинарного признака жирнотелости используют конечность дисперсии: степенные хвосты с показателем больше 2 относят к «гауссову бассейну» и считают оправданием дисперсии и подобных ей показателей в финансовых расчетах. Намного естественнее будет отвести финансовым приложениям область конечного математического ожидания [229]. Наш показатель может быть полезен следующим образом.

Пусть  $X_{g,1}, X_{g,2}, \dots, X_{g,n_g}$  — последовательность гауссовых величин с математическим ожиданием  $\mu$  и масштабом  $\sigma$ . Пусть  $X_{v,1}, X_{v,2}, \dots, X_{v,n_v}$  — последовательность каких-то других случайных величин, отмасштабированных до того же значения  $\mathbb{M}(1)$ , а именно  $\mathbb{M}^v(1) = \mathbb{M}^g(1) = \sqrt{\frac{2}{\pi}}\sigma$ . Мы ищем значения  $n_v$ , соответствующие данному  $n_g$ .

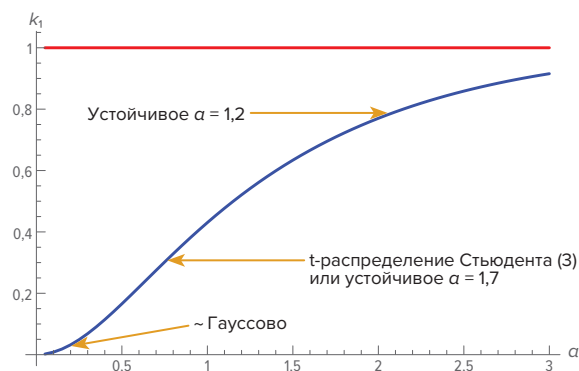
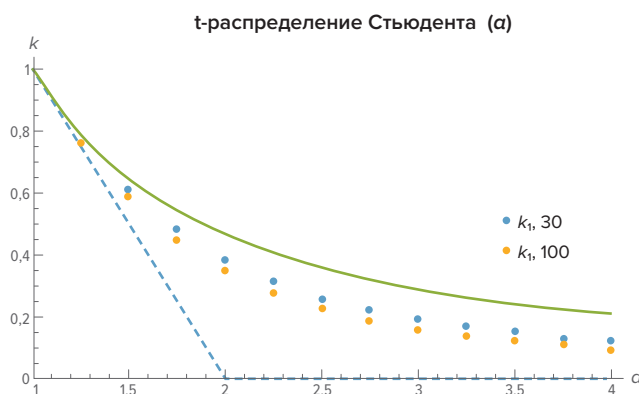
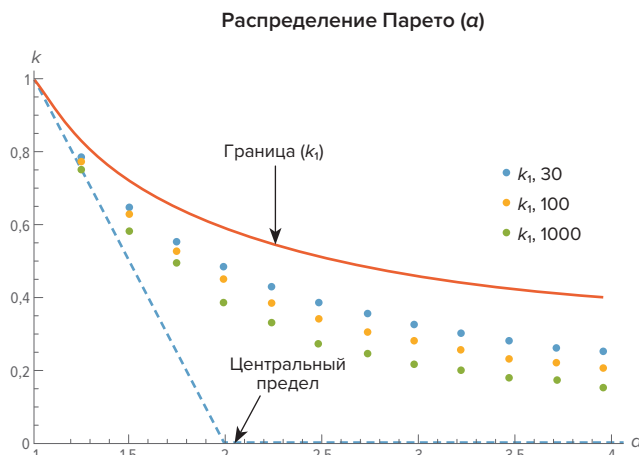
$\kappa_n$  характеризует одновременно и скорость сходимости согласно закону больших чисел, и скорость, при  $\kappa_n \rightarrow 0$ , сходимости слагаемых к гауссовым согласно центральной предельной теореме, как показано на Рисунке 8.2.

Необходимое число слагаемых

$$n_{\min} = \inf \left\{ n_v \left| \mathbb{E} \left| \sum_{i=1}^{n_v} \frac{X_{v,i} - \mu_v}{n_v} \right| \leq \mathbb{E} \left| \sum_{i=1}^{n_g} \frac{X_{g,i} - \mu_g}{n_g} \right|, n_v > 0 \right\} \quad (8.7)$$

можно вычислить, приняв  $\kappa_n = 0$  для гауссовой случайной величины и вернувшись от  $\kappa_n$  для заданного распределения согласно такому простому приближению:

$$n_v = n_g^{-1/(\kappa_1, n_g - 1)} \approx n_g^{-1/(\kappa_1 - 1)}, \quad n_g > 1. \quad (8.8)$$



**Рисунок 8.2:** Проверяем эффективность обобщенной центральной предельной теоремы: распределение Парето и t-распределение Стьюдента, в классе  $\mathcal{P}$ , с показателем хвоста  $\alpha$ ; капта должна сходиться к  $2 - (\mathbb{1}_{\alpha < 2} \alpha + \mathbb{1}_{\alpha \geq 2} \cdot 2)^*$ , или устойчивому классу  $\mathcal{S}$ . Мы наблюдаем, как медленно происходит эта сходимость, даже после того, как достигнута 1000 слагаемых. Этим обесценивается предположение Мандельброта, что распределение Парето с бесконечной дисперсией можно отнести к устойчивым распределениям

\* Здесь, очевидно, символ индикаторной функции  $\mathbb{1}$  легко заменить на традиционную запись:

$$2 - (\mathbb{1}_{\alpha < 2} \alpha + \mathbb{1}_{\alpha \geq 2} \cdot 2) = \begin{cases} 2 - \alpha, & \text{если } \alpha < 2 \\ 0, & \text{иначе} \end{cases},$$

и этот график показан пунктиром как центральный предел. — Прим. перев.

**Рисунок 8.3:** Логнормальное распределение ведет себя как гауссово при низких значениях  $\sigma$ , но быстро превращается в эквивалент степенного закона. Это демонстрирует несущестственность с практической точки зрения дебатов о распределении богатства: логнормальное оно (согласно правилу Жибра) или Парето (согласно закону Ципфа)

Приближение работает благодаря медленной сходимости. Например, для t-распределения Стьюдента с 3 степенями свободы ( $\alpha = 3$ ) требуется 120 наблюдений, чтобы получить такое же малое отклонение от среднего (при данной доверительной вероятности), какие для гауссова распределения достигаются за 30 наблюдений, то

есть в 4 раза быстрее. Для однохвостого распределения Парето с тем же показателем хвоста  $\alpha = 3$  потребуется 543 наблюдения, чтобы получить сопоставимость с выборкой 30 наблюдений гауссовой случайной величины; это в 4,5 раза больше, чем для распределения Стюдента. Таким образом:

- 1) конечная дисперсия — плохой критерий жирнохвостости (в статистике),
- 2) также плохо пытаться судить по показателю хвоста,
- 3) симметричное распределение Стюдента ведет себя не так, как распределение Парето; причина в колоколообразности распределения Стюдента (порожденного медленно меняющейся функцией) — это гасит отклонения в центре распределения.

Мы также можем извлечь некоторые контринтуитивные результаты. Согласно формуле 8.8, для проверки расхожего примера «Парето 80/20», соответствующего показателю хвоста около  $\alpha \approx 1,14$ , потребуется выборка в миллиард с лишним раз больше, чем для гауссова распределения.

## 8.4. ТЕХНИЧЕСКИЕ СЛЕДСТВИЯ

### 8.4.1. Некоторые странности асимметричных распределений

Если симметричное устойчивое распределение перекосить, показатель  $\kappa$  не изменится — другими словами,  $\kappa$  инвариантна к параметру  $\beta$  в формуле 8.4, устойчивой при суммировании. Однако простое распределение Парето, однохвостое, превосходит по жирности хвоста (в нашем смысле) свой симметричный эквивалент.

Это важное замечание, поскольку устойчивое распределение — это математический предельный объект, который невозможно наблюдать в реальной жизни, тогда как распределение Парето встречается на практике. Глубина различия между ними плохо осознается в литературе. Устойчивое распределение как заменитель Парето считается перспективным. В статье Учайкина и Золотарева [257] читаем:

Мандельброт привлек внимание к тому факту, что по ряду причин экстремальные устойчивые распределения (соответствующих  $\beta = 1$ ) удобнее для описания эмпирических принципов, чем распределения Ципфа — Парето. Можно видеть по многочисленным публикациям — и теоретическим, и прикладным, — как ширится признание идей Мандельброта среди экспертов. Таким образом, появляется надежда со временем подтвердить эмпирически нащупанные принципы в рамках математических моделей и прояснить механизмы, порождающие эти принципы.

Однако это разные звери, даже при большом числе слагаемых.

### 8.4.2. Скорость сходимости t-распределения Стюдента к гауссову бассейну

Мы доказываем в приложении — благодаря явному выражению для  $\kappa$  при сложении распределений Стюдента с  $\alpha = 3$ , или «кубических», как их называют в финансовой математике, — что скорость, с которой  $\kappa$  сходится к 0 при добавлении слагаемых, равна  $\frac{1}{\ln n}$ . Эта формула (наряду с квазизамкнутым выражением для плотности

суммы  $n$  кубических Стюдентов) дополняет результат Бушо и Поттерса [28] (см. также [214]), который заключается в следующем. Они решили разделить распределение на «гауссову зону», где плотность приближается гауссианой, и «зоны степенного закона» в хвостах, где оставлено первоначальное распределение: убывание по степенному закону. Точки перехода между зонами сдвигаются вправо и влево от центрального горба со скоростью  $\sqrt{n \ln n}$  стандартных отклонений, мучительно медленно. В центр распределения попадает больше слагаемых, чем в стороне от центра, и скорость сходимости согласно центральной предельной теореме различна для плотности вероятности в центре и в хвостах.

Дальнейшие исследования касаются сходимости распределения Парето к устойчивому по Леви, которая до настоящего времени исследовалась только численно.

### 8.4.3. Логнормальный хвост — ни тонкий, ни жирный

При наивном взгляде на Рисунок 8.2 при низких значениях параметра  $\sigma$  логнормальное распределение ведет себя подобно гауссову, а при высоких  $\sigma$  — как своего рода однохвостое Коши (точнее, как устойчивое распределение с  $\alpha = 1, \beta = 1$ )<sup>1</sup>, и  $\kappa$  все ближе подходит к 1. Напрашивается объяснение, почему обычно бесплодны дебаты о том, к какому распределению отнести ту или иную случайную величину — Парето или логнормальному; примеры смотрите в спорах о богатстве [162], [53], [54]. А именно: такие дебаты обычно не имеют отношения к реальности. Как заметил П. Чирилло [44], множество примеров, объясняемых как Парето, объяснимы и как логнормальные с высокой дисперсией; на практике различия в статистике не так велики, как считается.

### 8.4.4. Возможна ли отрицательная каппа?

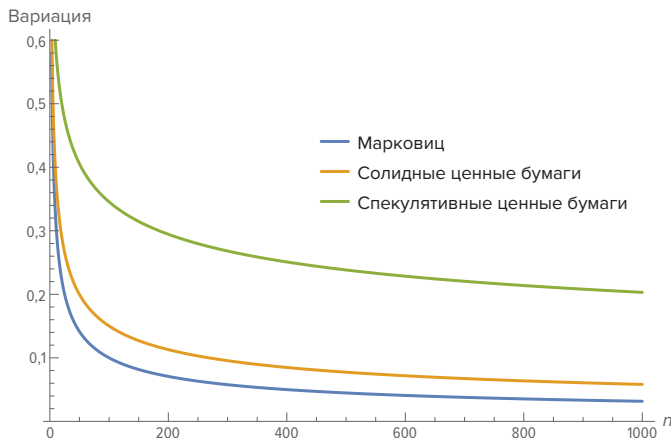
Подобно тому, как эксцесс для смеси гауссиан с разным математическим ожиданием (в отличие от стохастической волатильности) может опуститься ниже 3 (и дать отрицательный коэффициент эксцесса, определяемый как избыточный эксцесс по сравнению с гауссовым), показатель каппа может стать отрицательным — при отрицательном коэффициенте эксцесса. Для этого требуется бимодальность (то есть переключение между разными математическими ожиданиями при фиксированной дисперсии, при немалом расстоянии между модами, если считать его в стандартных отклонениях). По-видимому, в одnogорбых распределениях так не бывает. Подробности и вывод формул представлены в приложении.

## 8.5. ЗАКЛЮЧЕНИЕ И СЛЕДСТВИЯ

В то время как предельные теоремы (закон больших чисел и центральный предел) касаются поведения при  $n \rightarrow +\infty$ , нас интересуют конечные и точно заданные  $n$ , большие и малые.

<sup>1</sup> Говоря об устойчивых распределениях, обозначают  $\alpha \in (0, 2]$  параметр устойчивости,  $\beta \in [-1, 1]$  коэффициент асимметрии,  $c \in (0, \infty)$  параметр масштаба и  $\mu \in (-\infty, \infty)$  параметр сдвига. Совершенно симметричное (и, следовательно, двуххвостое) устойчивое распределение с параметрами  $\alpha = 1$  и  $\beta = 0$  представимо как распределение Коши с медианой и модой  $x_0 = \mu$  и параметром масштаба  $\gamma = c$ . В более общем случае уподобить распределению Коши устойчивое распределение можно в том смысле, что у него нет среднего при  $\alpha \leq 1$  и дисперсия бесконечна при  $\alpha < 2$ .

Мы можем извлечь ряд рабочих следствий:



**Рисунок 8.4:** Коротко о том, как работает эвристика  $1/n$ : требуется гораздо больше ценных бумаг в структуре портфеля, чтобы добиться того сокращения риска, которое получается согласно Марковицу. Мы для простоты приняли, что ценные бумаги не зависят друг от друга; в реальности возможен коллективный обвал, что еще больше усиливает эффект

### 8.5.1. Портфельная псевдостабильзация

Наш метод, естественно, применим и к диверсификации при формировании портфеля ценных бумаг: добавить новую ценную бумагу в портфель ради устойчивости — все равно что добавить наблюдение ради статистической значимости. Вопрос «Сколько вам нужно данных?» переводится на финансовый язык как «Сколько вам нужно разных ценных бумаг?». Понятно, что используемый в наше время метод формирования портфеля Марковица [166] (которым, судя по всему, сам Марковиц для своего собственного портфеля не пользовался [178]) применим только для  $k$  вблизи 0; инвесторы же пользуются эвристикой, выпуклой вниз, — иначе они бы недооценивали хвостовые риски и «лопались» (как лопнул портфельноориентированный хедж-фонд *Long Term Management* в 1998 [236] [250]).

Мы заметили выше, что для распределения Парето в районе «80/20» требуется в  $10^9$  раз больше наблюдений, чем для гауссова. Представьте себе, что вы руководствуетесь критериями современной портфельной теории и узнаете, что на самом деле для заданного снижения риска необходимо дополнительно расширить портфель... на 8 порядков величины! Чего стоит, в ситуации распределения Парето 80/20, стратегия простого расширения портфеля?

Парето — не Парето, но практически никакие ценные бумаги не имеют столь тонкого хвоста, как гауссиана, если судить по простейшему критерию эксцесса [228]; это значит, что формирование портфеля по Марковицу никогда не дает оптимального решения. Когда оказалось, что агенты мудро применяют нечто вроде (если пренебречь шумом) эвристики  $\frac{1}{n}$ , то психологи отнесли такое поведение к так называемой систематической ошибке восприятия, или смещенной оценке. Однако в дальнейшем этот вывод был разоблачен как «ложное смещение»: неоптимальной как раз была модель, с которой сравнивал поведение испытуемых психолог, тогда как испытуемые, привыкшие принимать ответственные решения, судили здраво. О склонности инвесторов к «чрезмерной» диверсификации, по сравнению с предположительно оптимальным поведением, объявили Бенарци и Талер [18]; в [16] это поведение описано так: «Если

есть  $n$  опций, раздели свой актив поровну по всем опциям. Мы назвали подобную эвристику Правилom  $1/n$ ». В действительности максимальная диверсификация по меньшей мере не хуже, чем стандарт формирования портфеля (см. критику в работах Уиндклиффа и Бойля [265] и [62]). Грубо говоря, равномерно распределенный портфель превосходит по эффективности S&P 500 по самым разным методам оценки. К сожалению, даже последние две работы недооценивают масштаб отличий и влияния жирных хвостов; мы же теперь можем оценить его более-менее точно. На Рисунке 8.5 влияние жирных хвостов на ценные бумаги сравнивается с тем, что предлагает Марковиц.

Ложный диагноз смещенного восприятия — один из многих примеров, как лица, определяющие политику, подталкивают народ к ложной рациональности [236] и провоцируют многократно превышать разумный портфельный риск.

Несколько замечаний о рисках финансового портфеля. У индекса S&P 500 показатель  $\kappa$  небольшой, около 0,2. Однако речь о корзине  $n = 500$  ценных бумаг, недифференцированных по весу и без учета корреляции между разными ценными бумагами, что по сути завышает вес устойчивых акций. Для отдельных акций значения  $\kappa$  выше: от 0,3 до 0,7; это означает настоятельную необходимость «чрезмерной» диверсификации.

Аналогичным образом данный показатель служит до некоторой степени руководством при обработке данных для прогнозов: с его помощью можно оценить, например, сколько на самом деле нужно лет наблюдения, чтобы судить об изменении климата; см. [160].

### 8.5.2. Другие аспекты статистических выводов

Пока мы рассматривали только одномерные случайные величины. При высших размерностях потенциальная область исследования — эквивалентный подход к распределению многомерных жирнохвостых случайных величин, для которых захват выборки не подчиняется распределениям случайных матриц, таким как распределение Марченко — Пастура или распределение Уишхарта. В нашей ситуации добавление случайных величин с трудом подавляет шум в случайных матрицах.

### 8.5.3. Последний комментарий

Хотя мы продолжаем приговаривать, что «статистика не бывает стандартной», в наших руках — эвристические методы, которыми можно оценить, где и в какой степени мы удаляемся от стандарта.

## 8.6. ПРИЛОЖЕНИЕ, ВЫВОД ФОРМУЛ, ДОКАЗАТЕЛЬСТВА

Приведем вывод некоторых формул.

### 8.6.1. Кубическое $t$ -распределение Стьюдента (гауссов бассейн)

$t$ -распределению Стьюдента с тремя степенями свободы уделяется особое внимание в литературе, поскольку оно доминирует в финансовой математике [99]. Оно имеет конечную дисперсию, и поэтому его часто приближают гауссовым, хотя это неправильно. Асимптотически оно в конце концов превращается в гауссово, но сам по себе этот факт вовсе не обещает скорой сходимости. Мандельброт и Талейб [165] отмечают, что кубическое распределение скорее можно уподобить степенному закону, если гово-



речь о распределении экстремальных значений, и здесь мы разовьем это замечание благодаря явному выражению для распределения плотности вероятностей суммы.

Пусть  $X$  — случайная величина с распределением плотности  $p(x)$ :

$$p(x) = \frac{6\sqrt{3}}{\pi(x^2 + 3)^2}, x \in (-\infty, +\infty). \quad (8.9)^1$$

### Предложение 8.1

Пусть  $Y$  — сумма  $X_1, \dots, X_n$ ,  $n$  тождественных экземпляров случайной величины  $X$ . Пусть  $\mathbb{M}(n)$  — среднее абсолютное отклонение от среднего для  $n$  слагаемых. Скорость

сходимости  $\kappa_{1,n} = \min \left\{ \kappa \left| \frac{\mathbb{M}(n)}{\mathbb{M}(1)} = n^{\frac{1}{2-\kappa}} \right. \right\}$  имеет вид

$$\kappa_{1,n} = 2 - \frac{\ln n}{\ln(e^n n^{-n} \Gamma(n+1, n) - 1)}, \quad (8.10)$$

где  $\Gamma(., .)$  — так называемая верхняя неполная гамма-функция  $\Gamma(a, z) = \int_z^\infty dt t^{a-1} e^{-t}$ .

Соответственно среднее абсолютное отклонение

$$\mathbb{M}(n) = \begin{cases} \frac{2\sqrt{3}}{\pi} & \text{при } n = 1, \\ \frac{2\sqrt{3}}{\pi} (e^n n^{-n} \Gamma(n+1, n) - 1) & \text{при } n > 1 \end{cases}. \quad (8.11)$$

Выведем эти формулы отдельно для плотности вероятности и отдельно для среднего абсолютного отклонения.

Характеристическая функция для  $n$  слагаемых

$$\varphi(\omega) = (1 + \sqrt{3} |\omega|)^n e^{-n\sqrt{3}|\omega|}.$$

Плотность вероятности случайной величины  $Y$

$$p(y) = \frac{1}{\pi} \int_0^\infty (1 + \sqrt{3}\omega)^n e^{-n\sqrt{3}\omega} \cos \omega y d\omega.$$

Все интегралы для  $\mathbb{M}(n)$  берутся, если приложить усилия, и в результате получаем формулу 8.11. Поскольку в литературе этой формулы не найти, будет полезным при-

<sup>1</sup> То есть случайная величина  $X$  следует t-распределению Стьюдента с 3 степенями свободы.

В самом деле, t-распределение Стьюдента с  $\nu > 0$  степенями свободы присуще случайной величине  $\frac{\bar{X} - \mu}{S/\sqrt{\mu+1}}$ , где  $\bar{X}$  — среднее  $(\nu+1)$  гауссовых н. о. р. случайных величин  $X_1, X_2, \dots, X_{\nu+1}$ , имеющих среднее  $\mu$  и порождающих дисперсию выборки  $S^2 = \frac{1}{\nu} \sum_{i=1}^{\nu+1} (X_i - \bar{X})^2$ . Плотность вероятности t-распреде-

ления Стьюдента  $f(x) = \frac{\Gamma \frac{\nu+1}{2}}{\sqrt{\nu\pi} \Gamma \frac{\nu}{2}} \left( 1 + \frac{x^2}{\nu} \right)^{-\frac{\nu+1}{2}}$ . В случае трех степеней свободы плотность вероятности

$$\frac{2}{\pi\sqrt{3} \left( 1 + \frac{x^2}{3} \right)^2} = \frac{2 \cdot 9}{\pi\sqrt{3}(3+x^2)^2} = \frac{6\sqrt{3}}{\pi(x^2+3)^2}.$$

вести здесь также побочное следствие: распределение плотности вероятности случайной величины  $Y$  можно записать как

$$p(y) = \frac{e^{n - \frac{iy}{\sqrt{3}}} \left( e^{\frac{2iy}{\sqrt{3}}} E_{-n} \left( n + \frac{iy}{\sqrt{3}} \right) + E_{-n} \left( n - \frac{iy}{\sqrt{3}} \right) \right)}{2\sqrt{3}\pi}, \quad (8.12)$$

где  $E_{(\cdot)}(\cdot)$  — обобщенная модифицированная интегральная показательная функция

$$E_n z = \int_1^\infty \frac{e^{-tz}}{t^n} dt.$$

Отметим при этом следующие тождества (из обновленного Абрамовица и Стегана<sup>1</sup> [69]):

$$n^{-n-1} \Gamma(n+1, n) = E_{-n} n = e^{-n} \frac{(n-1)!}{n^n} \sum_{m=0}^n \frac{n^m}{m!}.$$

Что касается асимптотики, имеем следующий результат (предложенный Михалисом Лулакисом). Перепишем 8.11 в виде<sup>2</sup>

$$\mathbb{M}(n) = \frac{2\sqrt{3}n!}{\pi n^n} \sum_{m=0}^{n-1} \frac{n^m}{m!}.$$

Заметим далее, что

$$e^{-n} \sum_{m=0}^{n-1} \frac{n^m}{m!} = \frac{1}{2} + O\left(\frac{1}{\sqrt{n}}\right).$$

(Это можно заметить по поведению суммы случайных величин, распределенных по Пуассону и сходящихся к гауссовому распределению согласно центральной предельной теореме. Если  $X_n$  — случайная величина с распределением Пуассона с параметром  $n$ <sup>3</sup>, то  $\mathbb{P}(X_n < n) = e^{-n} \sum_{m=0}^{n-1} \frac{n^m}{m!}$ . Поскольку сумма  $n$  независимых случайных величин, распределенных по Пуассону с параметром 1, распределена по Пуассону с параметром  $n$ , то центральная предельная теорема утверждает, что распределение вероятностей случайной величины  $Z_n = (X_n - n) / \sqrt{n}$  стремится к стандартному нормальному распределению. Следовательно,  $\mathbb{P}(X_n < n) = \mathbb{P}(Z_n < 0) \rightarrow 1/2$  при  $n \rightarrow \infty$ <sup>4</sup>. Другой подход, в котором доказывается, что  $1 + \frac{n}{1!} + \frac{n^2}{2!} + \dots + \frac{n^{n-1}}{(n-1)!} \sim \frac{e^n}{2}$ <sup>5</sup>, см. в [179].)

1 Так называют математический справочник Национального института стандартов и технологий США, первое издание 1964 года под редакцией Абрамовица и Стегана.

2 При  $n > 1$  формула 8.11 принимает вид  $\mathbb{M}(n) = \frac{2\sqrt{3}}{\pi} (e^n n^{-n} \Gamma(n+1, n) - 1)$ . С другой стороны, формула из справочника  $n^{-n-1} \Gamma(n+1, n) = e^{-n} \frac{(n-1)!}{n^n} \sum_{m=0}^n \frac{n^m}{m!}$  равносильна  $e^n n^{-n} \Gamma(n+1, n) - 1 = \frac{n!}{n^n} \sum_{m=0}^n \frac{n^m}{m!} - 1$ . В последней формуле вычитание 1 эквивалентно остановке суммирования на предпоследнем слагаемом,  $m = n-1$ , поскольку последнее слагаемое в сумме, при  $m = n$  — это  $\frac{n^m}{m!} = \frac{n^n}{n!}$ , что при умножении на множитель перед суммой дает  $\frac{n!}{n^n} \cdot \frac{n^n}{n!} = 1$ .

3 То есть имеет функцию вероятности  $\mathbb{P}(X_n = m) = e^{-n} \frac{n^m}{m!}$ .

4 Предложил Роберт Израэль (Robert Israel) на форуме *Math Stack Exchange*. — Прим. автора.

5 Не путать с бесконечным рядом  $1 + \frac{n}{1!} + \frac{n^2}{2!} + \dots = e^n$ .

Пользуясь тем, что  $\lim_{n \rightarrow \infty} \frac{n! e_n}{n^n \sqrt{n}} = \sqrt{2\pi}$ <sup>1</sup>, получаем следующую точную асимптотику:

$$\lim_{n \rightarrow \infty} (\ln n) \kappa_{1,n} = \frac{\pi^2}{4},$$

откуда  $\kappa$  уходит в ноль (то есть среднее становится гауссовым) со скоростью  $\frac{1}{\ln n}$ , мучительно медленно. Другими словами, даже при  $10^6$  слагаемых поведение суммы не становится гауссовым вопреки интуитивной догадке Б. Мандельброта [165].

### 8.6.2. Логнормальные суммы

Проследив поведение кумулянтов для  $n$  слагаемых, заключаем, что при малом параметре масштаба  $\sigma$  сумма ведет себя как гауссова случайная величина, а при большом  $\sigma$  — как логнормальная, причем в обоих случаях мы имеем явное выражение для  $\kappa_n$ .

У логнормального распределения (с параметрами  $\mu$  и  $\sigma$ ) нет явного выражения для характеристической функции. Но можно выразить кумулянты  $K_p$  всех порядков  $p$ , применив рекурсию, и для нашего случая суммы  $n$  тождественных копий случайной величины  $X_i$  кумулянт суммы  $K_p^n = K_p\left(\sum_n X_i\right) = nK_p(X_1)$ .

Кумулянты

$$K_1^n = n e^{\mu + \frac{\sigma^2}{2}}$$

$$K_2^n = n(e^{\sigma^2} - 1) e^{2\mu + \sigma^2}$$

$$K_3^n = n(e^{\sigma^2} - 1)(e^{\sigma^2} + 2) e^{3\mu + \frac{3\sigma^2}{2}}$$

$$K_4^n = \dots$$

позволяют вычислить

$$\text{коэффициент асимметрии} = \frac{\sqrt{e^{\sigma^2} - 1}(e^{\sigma^2} + 2)e^{\frac{2\mu + \sigma^2}{2} - \mu - \frac{\sigma^2}{2}}}{\sqrt{n}},$$

$$\text{эксцесс} = 3 + \frac{e^{2\sigma^2} \left( e^{2\sigma^2} (e^{2\sigma^2} + 2) + 3 \right) - 6}{n}.$$

Благодаря кумулянтам и моментам сразу доказывается, что

$$\lim_{n \rightarrow +\infty} \kappa_{1,n} = 0,$$

$$\lim_{\sigma \rightarrow 0} \kappa_{1,n} = 0,$$

и параметр  $\kappa$  удается ограничить явным выражением. Для этого обозначим  $\kappa_{1,n}^*$  те его значения, при которых суммирование логнормальных случайных величин сохраняет логнормальность плотности и первые два момента. Тогда

$$0 \leq \kappa_{1,n}^* \leq 1,$$

<sup>1</sup> Этот факт более известен как формула Стирлинга для факториала больших чисел:

$$n! \sim \sqrt{2\pi} \left(\frac{n}{e}\right)^n, \text{ то есть } \lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi} \left(\frac{n}{e}\right)^n} = 1.$$

$$\kappa_{1,n}^* = 2 - \frac{\ln n}{\frac{n \operatorname{erf} \frac{\sqrt{\ln \frac{n+e^{\sigma^2}-1}{n}}}{2\sqrt{2}}}{\ln \frac{\operatorname{erf} \frac{\sigma}{2\sqrt{2}}}{2\sqrt{2}}}}.$$

**Эвристическая попытка.** Помимо прочих эвристических подходов, мы можем за два шага увидеть, что (1) при высоких значениях  $\sigma$  показатель  $\kappa_{1,n} \rightarrow \kappa_{1,n}^*$ , поскольку закон больших чисел замедляется, и что (2)  $\kappa_{1,n}^* \rightarrow 1$  при  $\sigma \rightarrow \infty$ .

**Доказательство Лулакиса.** Верхнюю границу, к которой  $\kappa_{1,n}$  приближается при высокой дисперсии, формально вывел Михалис Лулакис<sup>1</sup>, и мы кратко перескажем здесь суть.

Начнем с тождества

$$\mathbb{E}|X-m| = 2 \int_m^\infty (x-m)f(x)dx = 2 \int_m^\infty \bar{F}_X(x)dx,$$

где  $f(\cdot)$  — плотность вероятности,  $m$  — среднее, а  $\bar{F}_X(\cdot)$  — функция выживания. Далее заметим, что

$$\mathbb{M}(n) = 2 \int_{nm}^\infty \bar{F}_X(x)dx.$$

Допустим, что  $m = \frac{\sigma^2}{2}$ , или  $X = \exp\left(\sigma Z - \frac{\sigma^2}{2}\right)$ , где  $Z$  — стандартная нормированная случайная величина. Пусть  $S_n$  — сумма  $X_1 + \dots + X_n$ ; получаем  $\mathbb{M}(n) = 2 \int_n^\infty \mathbb{P}(S_n > t)dt$ . Пользуясь субэкспоненциальностью ([196]), ограничим

$$\mathbb{P}(S_n > t) \geq \mathbb{P}\left(\max_{0 \leq i \leq n} X_i > t\right) \geq n\mathbb{P}(X_1 > t) - \binom{n}{2}\mathbb{P}(X_1 > t)^2.$$

В последнем выражении слагаемое  $n\mathbb{P}(X_1 > t) \rightarrow 1$  при  $\sigma \rightarrow \infty$ , и второе слагаемое стремится к нулю (использовано неравенство Гёльдера<sup>3</sup>).

Пропуская несколько шагов, видим, что  $\liminf_{\sigma \rightarrow \infty} \frac{\mathbb{M}(n)}{\mathbb{M}(1)} \geq n$ , притом что, с другой стороны, есть граница  $\frac{\mathbb{M}(n)}{\mathbb{M}(1)} \leq n$ . Значит, при  $\sigma \rightarrow \infty$  предел  $\frac{\mathbb{M}(n)}{\mathbb{M}(1)} = n$ , откуда  $\kappa_{1,n} \rightarrow 1$  при  $\sigma \rightarrow \infty$ .

1 В редакции бумажной версии; Лулакис дал строгое доказательство вместо эвристического вывода. — Прим. автора.

2 Первое неравенство в этой цепочке неравенств следует из того, что сумма положительных слагаемых  $S_n = X_1 + \dots + X_n$  обязательно превысит  $t$  в тех случаях, когда наибольшее из них превысит  $t$ . Второе неравенство следует из того, что вероятность превышения  $t$  хотя бы одним из  $n$  н. о. р. слагаемых представима как  $1 - (1-p)^n = np - \binom{n}{2}p^2 + \dots$ , где  $p = \mathbb{P}(X_1 > t)$  — вероятность того, что первое слагаемое превысит  $t$ ,

а  $\binom{n}{2} = \frac{n(n-1)}{2}$  — биномиальный коэффициент.

3 Неравенство Гёльдера в теории вероятностей формулируют в терминах математического ожидания абсолютной величины произведения случайных величин  $X$  и  $Y$ , следующих распределениям в  $L^p$  (пространство функций с конечной интегрируемой  $p$ -той степенью):

$$\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{1/p} \cdot (\mathbb{E}|Y|^q)^{1/q}, \text{ если } p+q=1 \text{ и существуют конечные } \mathbb{E}|X|^p \text{ и } \mathbb{E}|Y|^q,$$

и, следовательно:

$\mathbb{E}|X|^r \leq (\mathbb{E}|X|^s)^{r/s}$  при  $0 < r < s$ . Впрочем, как видно из предыдущего примечания, можно обойтись и более простыми соображениями.

**Вычисления при помощи семейства Пирсона.** Для удобства вычислений при небольшом параметре  $\sigma$  (ниже  $\approx 0,3$ ) можно использовать семейство Пирсона — хотя логнормальное распределение не принадлежит классу Пирсона (этому классу принадлежит нормальное распределение; впрочем, мы достаточно близки к нему для приближенных вычислений). Интуитивно понятно, что при низкой  $\sigma$  может хватить первых четырех моментов, потому что больших отклонений нет; другое дело — при высокой  $\sigma$ , когда приходится все-таки работать с логнормальным распределением.

Применение класса Пирсона распространено в таких областях, как информатика и теория коммуникаций, где ему посвящена богатая литература: о суммировании логнормальных случайных величин смотрите статью Не и Ченя [180], а о IV типе распределения Пирсона — [41], [65].

Семейство распределений Пирсона определяется через следующее дифференциальное уравнение для масштабированной плотности вероятности  $f$ :

$$f'(x) = -\frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2} f(x). \quad (8.13)$$

Заметим, что параметры  $a_0$ ,  $b_2$  и др. влияют на тип распределения Пирсона — в нашем случае нужен тип IV, — и на выражения для среднего отклонения через  $n$ ,  $\sigma$  и  $\mu$ .

Пусть  $m$  — математическое ожидание. Как доказали Диаконис и другие в [67], пользуясь старым трюком де Муавра, и Судзуки в [221], можно получить явные выражения для среднего абсолютного отклонения. Вновь используя тождество

$$\mathbb{E}|X - m| = 2 \int_m^{\infty} (x - m) f(x) dx$$

и интегрируя по частям, получаем

$$\mathbb{E}|X - m| = \frac{2(b_0 + b_1 m + b_2 m^2)}{a_1 - 2b_2} f(m). \quad (8.14)$$

Для подбора параметров используем кумулянты логнормальной  $n$ -суммы. Задав  $a_1 = 1$  и  $m = \frac{b_1 - a_0}{1 - 2b_2}$ , получаем параметры

$$\begin{aligned} a_1 &= e^{\mu + \frac{\sigma^2}{2}} \left( -12n^2 + (3 - 10n)e^{4\sigma^2} + 6(n-1)e^{\sigma^2} + 12(n-1)e^{2\sigma^2} - \right. \\ &\quad \left. - (8n+1)e^{3\sigma^2} + 3e^{5\sigma^2} + e^{6\sigma^2} + 12 \right) / \left( 2 \left( 6(n-1) + e^{2\sigma^2} \left( e^{\sigma^2} (5e^{\sigma^2} + 4) - 3 \right) \right) \right), \\ b_2 &= e^{2\sigma^2} (e^{\sigma^2} - 1) (2e^{\sigma^2} + 3) / \left( 2 \left( 6(n-1) + e^{2\sigma^2} \left( e^{\sigma^2} (5e^{\sigma^2} + 4) - 3 \right) \right) \right), \\ b_1 &= (e^{\sigma^2} - 1) e^{\mu + \frac{\sigma^2}{2}} \left( e^{\sigma^2} \left( e^{\sigma^2} \left( -4n + e^{\sigma^2} (e^{\sigma^2} + 4) + 7 \right) - 6n + 6 \right) + 6(n-1) + \right. \\ &\quad \left. + 12(n-1) \right) / \left( 2 \left( 6(n-1) + e^{2\sigma^2} \left( e^{\sigma^2} (5e^{\sigma^2} + 4) - 3 \right) \right) \right), \quad b_0 = n(e^{\sigma^2} - 1) e^{2(\mu + \sigma^2)} \left( e^{\sigma^2} \left( -2(n-1)e^{\sigma^2} - \right. \right. \\ &\quad \left. \left. - 3n + e^{3\sigma^2} + 3 \right) + 6(n-1) \right) / \left( 2 \left( 6(n-1) + e^{2\sigma^2} \left( e^{\sigma^2} (5e^{\sigma^2} + 4) - 3 \right) \right) \right). \end{aligned}$$

**Полиномиальные разложения.** Другие методы, такие как разложения Грама — Шарлье, использованные, например, Шлахером [210] и Больё [14], оказались не так полезны для получения  $\kappa_n$ . При высоких значениях  $\sigma$  приближения становятся неустойчивыми к добавлению многочленов Эрмита высших порядков. Смотрите обзоры у Дюфрена [70] и [71].

### 8.6.3. Экспоненциальное распределение

Экспоненциальное распределение имеет жирный хвост «начального уровня», на самой границе области жирных хвостов:

$$f(x) = \lambda e^{-\lambda x}, x \geq 0.$$

Сверткой суммы  $X_1, X_2, \dots, X_n$  получаем — рекуррентно, опираясь на  $f(y) = \int_0^y f(x)f(y-x)dx = \lambda^2 y e^{-\lambda y}$ :

$$f_n(z) = \frac{\lambda^n z^{n-1} e^{-\lambda z}}{(n-1)!}, \quad (8.15)$$

что представляет собой гамма-распределение; получаем среднее отклонение для  $n$  слагаемых:

$$\mathbb{M}(n) = \frac{2e^{-n} n^n}{\lambda \Gamma(n)}, \quad (8.16)$$

откуда

$$\kappa_{1,n} = 2 - \frac{\ln n}{n \ln n - n - \ln \Gamma(n) + 1}. \quad (8.17)$$

Можно видеть, что асимптотическое поведение столь же медленное (аналогично случаю Стюдента), хотя экспоненциальное распределение балансирует на грани суб-экспоненциальности:

$$\lim_{n \rightarrow \infty} \kappa_{1,n} \ln n = 4 - 2 \ln 2\pi.$$

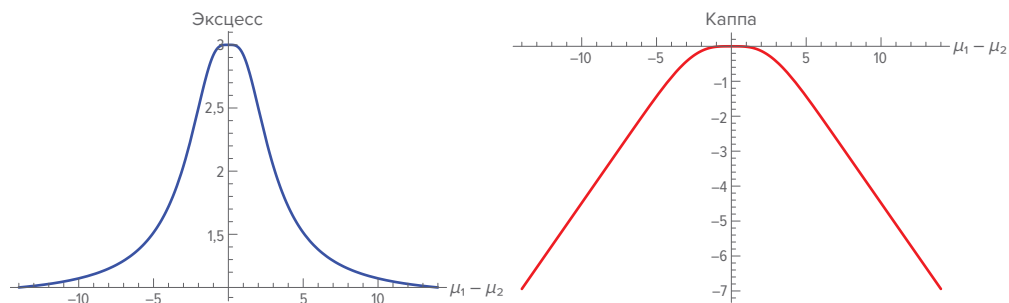
### 8.6.4. Отрицательная капша, отрицательный эксцесс

Рассмотрим простой случай переключения между гауссовыми распределениями с разными математическими ожиданиями и дисперсиями: с вероятностью  $\frac{1}{2}$  случайная величина  $X \sim \mathcal{N}(\mu_1, \sigma_1)$  и с вероятностью  $\frac{1}{2}$  случайная величина  $X \sim \mathcal{N}(\mu_2, \sigma_2)$ .

Разные  $\mu_1$  и  $\mu_2$  могут дать распределение с хвостами тоньше, чем у гауссианы, и двумя горбами; эффект становится драматическим, если расстояние между вершинами достигает нескольких стандартных отклонений. Пусть  $d = \mu_1 - \mu_2$  и  $\sigma = \sigma_1 = \sigma_2$  (чтобы минимизировать эксцесс); тогда показатель

$$\kappa_1 = \frac{\ln 4}{\ln \pi - 2 \ln \frac{\sqrt{\pi} d e^{\frac{d^2}{4\sigma^2}} \operatorname{erf} \frac{d}{2\sigma} + 2\sigma e^{\frac{d^2}{4\sigma^2}} + 2\sigma}{d e^{\frac{d^2}{4\sigma^2}} \operatorname{erf} \frac{d}{2\sqrt{2}\sigma} + 2\sigma \sqrt{\frac{2}{\pi}} e^{\frac{d^2}{8\sigma^2}}} + 2, \quad (8.18)$$

и, как можно видеть, отрицателен при широком зазоре  $\mu_1 - \mu_2$ .



**Рисунок 8.5:** Отрицательный эксцесс\* из Раздела А.3 и соответствующая каппа

\* Строго говоря, эксцесс не отрицательный, а меньше 3. Отрицателен коэффициент эксцесса. — Прим. перев.

## ДАЛЕЕ

Далее рассмотрим несколько простых диагностик для степенных законов применительно к S&P 500. Мы покажем различия между наивными методами и использованием оценочных функций метода максимального правдоподобия, который можно экстраполировать на хвосты.

## ЭКСТРЕМАЛЬНЫЕ ЗНАЧЕНИЯ И СКРЫТЫЕ ХВОСТЫ<sup>\*\*†</sup>

Scilicet et fluvius qui visus maximus ei,  
Qui non ante aliquem maiorem vidit, et ingens  
Arbor homoque videtur et omnia de genere omni  
Maxima quae vidit quisque, haec ingentia fingit<sup>1</sup>.

*О природе вещей*, книга VI, строки 674–677

У распределения данных с жирным хвостом есть скрытая часть, невидимая по прошлым выборкам. Прошлые экстремальные значения (максимум или минимум) — плохой предиктор будущих экстремальных значений; понятно, что случаются рекорды, превосходящие прошлые крайние отметки, и в Главе 3 уже описывалась *недооценка Лукреция*: глупо считать самой большой на свете рекой или горой самую большую из виденных тобой рек или гор.

В этой главе, после краткого введения в теорию экстремальных значений, мы сосредоточимся на ее применении к жирным хвостам. Когда данные распределены по степенному закону, максимум, взятый по  $n$  наблюдениям, следует распределению, которое нетрудно построить с нуля. Мы покажем, что на практике для случайных величин, распределенных по степенному закону, максимальным аттрактором (*maximum domain of attraction*, MDA) служит, асимптотически, распределение Фреше<sup>2</sup>.

В более общем случае теория экстремальных значений дает строгий подход к обработке экстремальных значений и экстраполяции за простой максимум. Мы приводим некоторые результаты по «скрытому среднему»; с его игнорированием связаны разнообразные ошибки в литературе об управлении рисками.

1 Тем, кто не видел еще величайшего; также громадны / Будут для них человек, или дерево, или другие / Вещи, пока не пришлось повстречать им еще величайших (*пер. И. Рачинского*).

2 Распределение Фреше случайной величины  $X$  на носителе  $(m, \infty)$  при коэффициенте формы  $\rho > 0$ , коэффициенте масштаба  $s > 0$  и произвольном положении минимума  $m$  определяется как

$$\mathbb{P}(X \leq x) = \exp\left(-\left(\frac{x-m}{s}\right)^{-\rho}\right) \text{ при } x > m, \text{ иначе } 0.$$

$$\text{Плотность вероятности } \frac{\rho}{s} \left(\frac{x-m}{s}\right)^{-1-\rho} \exp\left(-\left(\frac{x-m}{s}\right)^{-\rho}\right).$$

Среднее  $m + s\Gamma(1-1/\rho)$  при  $\rho > 1$ , иначе бесконечно.

Дисперсия  $s^2(\Gamma(1-2/\rho) - (\Gamma(1-1/\rho))^2)$  при  $\rho > 2$ , иначе бесконечна.

Коэффициент асимметрии конечен при  $\rho > 3$ , эксцесс при  $\rho > 4$ .





**Рисунок 9.1:** Древнеримский поэт-философ Лукреций (вверху). Высокая вода (внизу): считать ли уровень наводнения в Париже в 1910 максимальным значением? Очевидно, каждый рекорд побивает предыдущий рекорд, который держался какое-то время. Грамотное управление рисками должно оценивать, какой запас разумно заложить на случай нового рекорда в будущем. Но мы из века в век снова и снова повторяем недооценку Лукреция



### 9.1. ПРЕДВАРИТЕЛЬНОЕ ВСТУПЛЕНИЕ К ТЭЗ

Пусть  $X_1, \dots, X_n$  — независимые случайные величины с распределением вероятностей  $F(\cdot)$  по Парето.

Мы можем получить точное распределение максимума (или минимума). Распределение вероятностей максимума по  $n$  переменным будет

$$\mathbb{P}(X_{\max} \leq x) = \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) = \mathbb{P}(X_1 \leq x) \cdot \dots \cdot \mathbb{P}(X_n \leq x) = F(x)^n, \quad (9.1)$$

это вероятность того, что все величины окажутся меньше или равны  $x$ . Соответствующая плотность вероятности — это первая производная,  $\psi(x) = \frac{\partial F(x)^n}{\partial x}$ .

Распределение экстремального значения имеет дело с максимумом среди случайных величин при  $x \rightarrow x^*$ , где «правый край распределения»  $x^* = \sup \{x \mid F(x) < 1\}$  принадлежит максимальному аттрактору (*maximum domain of attraction*, MDA) [116]. Другими словами,

$$\max (X_1, \dots, X_n) \xrightarrow{P} x^*,$$

где  $\xrightarrow{P}$  означает сходимость по вероятности. В новой формулировке центральный вопрос стал таким: найти распределение  $x^*$ . Мы начали с того, что известно точное распределение  $F(\cdot)$  случайной величины  $X_1$ , и в таком случае инженерам достаточно того выражения для искомого распределения, которое содержится в формуле 9.1. В самом деле, возможно набрать проверочную статистику — вооружившись терпением, вычислительными мощностями и желанием исследовать вопрос. Это единственный надежный способ исследовать предасимптотику, то есть «происходящее при малых  $n$ , когда  $x$  еще не достигает  $x^*$ ».

Но для исследования статистики полезно понять общую доасимптотическую структуру.

Теорема Фишера — Типпетта — Гнеденко (Амбре и др. [82], де Гаан и Феррейра [116]) утверждает следующее. Если существуют последовательности «нормирующих» констант  $a_n > 0$  и  $b_n \in \mathbb{R}$ , такие, что

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x) \text{ при } x \rightarrow \infty^1, \quad (9.2)$$

то функция

$$G(x) \propto \exp(-(1+\xi x)^{-1/\xi}),$$

где  $\xi$  — индекс экстремальных значений и управляет поведением хвоста распределения. Функция  $G$  называется (обобщенным) распределением экстремальных значений (*generalized extreme value distribution*, GED). Подсемейства, определенные по крите-

1 Очевидно, здесь  $M_n = \max_{1 \leq i \leq n} X_i$ , как в Разделе 2.2.11.

рию  $\zeta = 0$ ,  $\zeta > 0$  и  $\zeta < 0$ , представляют собой соответственно семейства распределений Гумбеля<sup>1</sup>, Фреше и Вейбулла.<sup>2</sup>

**Распределение Гумбеля (тип 1).** Здесь принимается  $\zeta = 0$ ; точнее, выполняется предельный переход от  $\exp(-1 + \zeta x)^{-1/\zeta}$  при  $\zeta \rightarrow 0$ :

$$G(x) = \exp\left(-\exp\left(-\frac{x-b_n}{a_n}\right)\right) \text{ при } x \in \mathbb{R},$$

и у распределения  $M_n$  экспоненциальный хвост.

**Распределение Фреше (тип 2).** Здесь  $\xi = \frac{1}{\alpha}$ , и

$$G(x) = \begin{cases} 0 & x \leq b_n \\ \exp\left(-\left(\frac{x-b_n}{a_n}\right)^{-\alpha}\right) & x > b_n \end{cases},$$

и у распределения  $M_n$  правый хвост степенного закона, как мы видели ранее. Обратите внимание, что  $\alpha > 0$ .

**Распределение Вейбулла (тип 3).** Здесь  $\xi = -\frac{1}{\alpha}$  и

$$G(x) = \begin{cases} \exp\left(-\left(-\left(\frac{x-b_n}{a_n}\right)\right)^\alpha\right) & x < b_n, \\ 1 & x \geq b_n \end{cases},$$

и у распределения  $M_n$  область определения ограничена справа (т. е. максимумом). Обратите внимание, что и здесь  $\alpha > 0$ .

### 9.1.1. Любой хвост степенного закона ведет к Фреше

Сейчас мы начнем действовать скорее как инженеры, чем математики, и рассмотрим два существующих распределения, Парето и Фреше, чтобы понять, может ли одно сходиться к другому, — другими словами, вывести Фреше другим способом, из асимптотических свойств степенных законов.

Приведенные ниже рассуждения можно обобщить на любую случайную величину с хвостом Парето, рассмотренным выше «точки Караматы» — той точки, где медленно меняющаяся функция удовлетворительно приблизилась к константе.

1 Стандартное распределение Гумбеля положительной случайной величины  $X$  определяется как

$$\mathbb{P}(X \leq x) = \exp(-e^{-x});$$

плотность вероятности  $\exp(-(x + e^{-x}))$ ; среднее равно постоянной Эйлера — Маскерони  $\gamma \approx 0,58$ ; дисперсия  $\frac{\pi^2}{6}$ ; коэффициент асимметрии  $\approx 1,14$ ; избыточный эксцесс  $\frac{12}{5}$ .

2 Стандартное распределение Вейбулла положительной случайной величины  $X$  определяется как

$$\mathbb{P}(X \leq x) = 1 - \exp(-x^k),$$

где коэффициент формы  $k > 0$ ; плотность вероятности  $kx^{k-1}\exp(-x^k)$ ; среднее  $\Gamma(1 + 1/k)$ ; дисперсия  $\Gamma(1 + 2/k)$ ; коэффициент асимметрии и эксцесс конечны.

Функция распределения вероятностей Парето с минимальным значением (и коэффициентом масштаба)  $L$  и показателем хвоста  $\alpha$

$$F(x) = 1 - \left(\frac{L}{x}\right)^\alpha,$$

поэтому функция плотности вероятности для максимума по  $n$  наблюдениям

$$\psi(x) = \frac{\alpha n \left(\frac{L}{x}\right)^\alpha \left(1 - \left(\frac{L}{x}\right)^\alpha\right)^{n-1}}{x}. \quad (9.3)$$

Функция плотности вероятности Фреше

$$\varphi(x) = \alpha \beta^\alpha x^{-\alpha-1} \exp(\beta^\alpha (-x^{-\alpha})). \quad (9.4)$$

Найдем такой большой  $x$ , при котором эти две функции склеиваются; другими словами, такое число  $x^*$ , что  $\psi(x^*) \rightarrow \varphi(x^*)$ . Поскольку

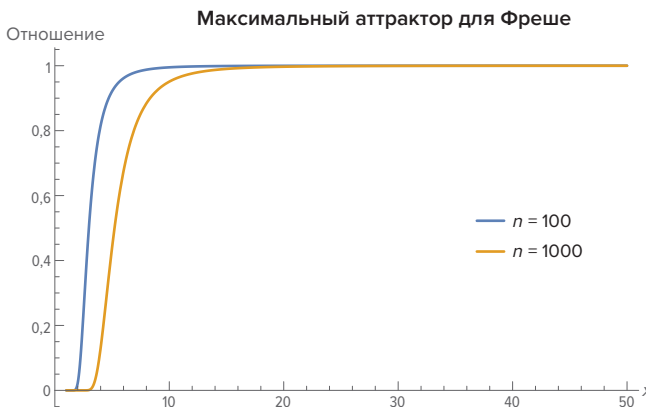
$$\lim_{x \rightarrow \infty} \frac{\psi(x)}{\varphi(x)} = n \left(\frac{1}{\beta}\right)^\alpha L^\alpha, \quad (9.5)$$

при большом  $x$  можно считать, что  $\beta = L n^{1/\alpha}$ . Равенство 9.5 показывает, как  $\alpha$  хвоста сохраняется при преобразованиях распределения:

#### Свойство 4

Показатель хвоста для максимума среди  $n$  о. р. случайных величин тот же, что у самих этих случайных величин.

«Где именно» можно на практике приблизить одно другим, показано на рисунке 9.2.



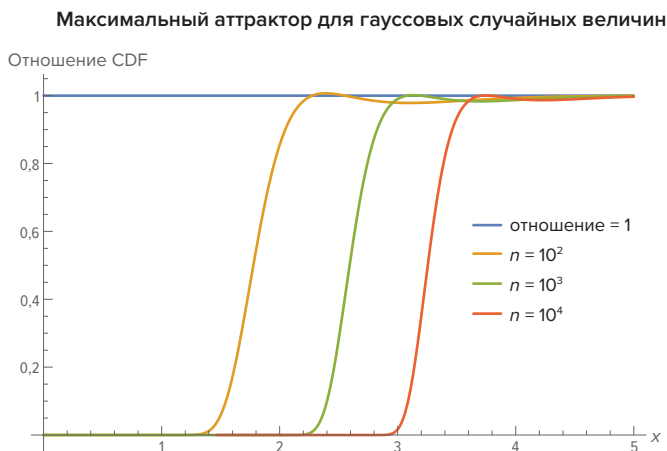
**Рисунок 9.2:** Отношение данной функции распределения вероятностей к распределению Фреше. Наглядно прослеживается приемлемый уровень приближения, и видно, как  $x$  вступает в максимальный аттрактор (Maximum Domain of Attraction, MDA). Здесь  $\alpha = 2$ ,  $L = 1$ . Заметим, что отношение функций плотности вероятностей даст ту же картину, в отличие от гауссова распределения, как мы увидим ниже

### Свойство 5

Для крайних значений по степенному закону получается точная асимптотическая подгонка.

#### 9.1.2. Гауссов случай

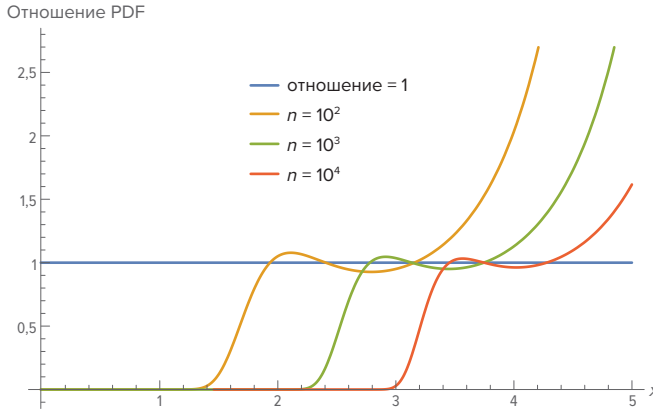
Случай Фреше совсем простой, потому что степенные законы легко анализируются, и удается получить параметры предельных распределений. Для гауссова и прочих распределений требуются более изощренные выкладки, и обычно приходится подбирать нормирующие константы  $a_n$  и  $b_n$  приближенными методами на основе квантильных функций<sup>1</sup>. Основополагающая статья Фишера и Типпета [94] предупреждает, что «от нормального распределения приближение к предельному распределению происходит крайне медленно» (это место цитируется в статье Газуля и др. [101]).



**Рисунок 9.3:** Поведение гауссовых случайных величин; к ним, в отличие от степенных законов, трудно подобрать хорошее приближение аналитической формулой. Вдоль оси у отложено отношение точной кумулятивной функции распределения максимального значения среди  $n$  случайных величин к параметризованному аналитическому приближению согласно ТЭЗ

<sup>1</sup> Квантильной функцией случайной величины  $X$  с кумулятивной функцией распределения вероятностей  $F(x) = \mathbb{P}(X < x)$  называется обратная функция,  $Q(p) = \inf\{x \in \mathbb{R} \mid p \leq F(x)\} = F^{-1}(p)$ , где  $p \in [0, 1]$ . То, что с вероятностью  $p$  случайная величина  $X$  достигнет или превысит значение  $x$ ,  $p = F(x)$ , равносильно тому, что  $1 - p$  верхних значений случайной величины  $X$  достигают или превосходят величину  $x$ ,  $Q(p) = x$ . В приближенных вычислениях промежутков  $[0, 1]$  разбивают на  $q$  равных интервалов и вместо непрерывной квантильной функции  $Q(p)$  используют набор  $q$ -квантилей, значений вида  $Q(k/q)$ , где  $k = 1, 2, \dots, q - 1$ . 100-квантили называются процентиями, и квантиль вида  $Q(k/100)$ , где  $k = 1, 2, \dots, 99$ , называется  $k$ -тым процентилем и обозначается  $P_k$ . Диапазон от 25-го процентия до 75-го,  $[P_{25}, P_{75}]$ , называется диапазоном «средних фифти»; половина всех наблюдений случайной величины  $X$  попадает в этот диапазон, а оставшаяся половина наблюдений разбита на равное число наблюдений, когда случайная величина  $X$  оказалась меньше  $P_{25}$ , и число наблюдений, когда случайная величина  $X$  оказалась больше  $P_{75}$ . 4-квантили называются квартилями и обозначаются  $Q_k$ ; четверть всех наблюдений случайной величины  $X$  оказывается меньше первого квартиля  $Q_1$ , еще четверть наблюдений — больше третьего квартиля  $Q_3$ . Оставшаяся половина наблюдений попадает в диапазон  $[Q_1, Q_3]$ , называемый межквартильным диапазоном; он совпадает с диапазоном «средних фифти»,  $[Q_1, Q_3] = [P_{25}, P_{75}]$ . Первый и единственный 2-квантиль совпадает с медианой случайной величины  $X$ : одна половина наблюдаемых значений оказывается меньше, чем  $Q(1/2)$ , а другая — больше. С медианой также совпадают второй квартиль и 50-й процентиль.

Максимальный аттрактор для гауссовых случайных величин



**Рисунок 9.4:** Та же история, что на Рисунке 9.3, представлена функциями плотности вероятности. Невозможно добиться хорошего приближения в хвостах

Ниже мы будем искать нормирующие константы для гауссианы, основываясь на статье [120] и более поздних исследованиях.

Пусть  $M_n = a_n x + b_n$  в формуле 9.2. Допустим далее, что максимальные значения  $M_n$  следуют распределению экстремальных значений согласно ТЭЗ, с кумулятивной функцией  $\exp(-e^x)$ , и что минимальные значения следуют зеркальному распределению Гумбеля, которое получается преобразованием из распределения  $-M_n$ , где  $\frac{M_n - b_n}{a_n}$  следуют распределению Гумбеля с кумулятивной функцией  $1 - \exp(-e^x)$ <sup>1</sup>. Параметризованная кумулятивная функция для  $M_n$  имеет вид  $\exp\left(-e^{-\frac{x-b_n}{a_n}}\right)$ .

Удобное упрощение следует из следующего приближения:<sup>2</sup>  $a_n = \frac{b_n}{b_n^2 + 1}$  и  $b_n = -\sqrt{2} \operatorname{erfc}^{-1} 2\left(1 - \frac{1}{n}\right)$ , где  $\operatorname{erfc}^{-1}$  — обратная дополнительная функция ошибок.

**Свойство 6**

Для исследования хвостового риска и других свойств сугубо предпочтительно работать с точным распределением гауссовых случайных величин; для  $n$  величин распределение максимума выводится из кумулятивной функции стандартной гауссианы  $F^{(g)}$ :

$$\frac{\partial F^{(g)}(K)}{\partial K} = \frac{e^{-\frac{K^2}{2}} 2^{\frac{1}{2}-n} n \operatorname{erfc}\left(-\frac{K}{\sqrt{2}}\right)^{n-1}}{\sqrt{\pi}}, \quad (9.6)$$

где  $\operatorname{erfc}$  — дополнительная функция ошибок.

1 Мы следуем соглашению, где распределение Гумбеля принимается только для минимальных значений, а для максимальных значений используется распределение ТЭЗ с собственными параметрами. — Прим. автора.

2 Амбре и др. в [82] предлагает  $a_n = \frac{1}{\sqrt{2 \ln n}}$ ,  $b_n = \sqrt{2 \ln n} - \frac{\ln \ln n + \ln 4\pi}{2\sqrt{2 \ln n}}$ , и второе слагаемое для  $b_n$  требуется только для больших  $n$ . Это приближение порядка  $\sqrt{\ln n}$ . — Прим. автора.

### 9.1.3. Теорема Пикэндса — Балкемы — де Гаана

Функция распределения условного эксцесса эквивалентна по плотности условному математическому ожиданию отклонения эксцесса Линди [116, 190] — мы воспользуемся ею в Главе 16.

Рассмотрим неизвестную функцию распределения  $F$  случайной величины  $X$ ; мы хотим оценить условную функцию распределения  $F_u$  для величин  $X$ , превысивших некоторый порог  $u$ , которая определяется как

$$F_u(y) = \mathbb{P}(X - u \leq y \mid X > u) = \frac{F(u + y) - F(u)}{1 - F(u)} \quad (9.7)$$

при  $0 \leq y \leq x^* - u$ , где  $x^*$  — это конечная или бесконечная правая граница исходного распределения  $F$ . Тогда существует измеримая функция  $G(u)$  такая, что

$$\lim_{u \rightarrow x^*} \sup_{0 \leq x < x^* - u} |F_u(x) - G_{\xi, \sigma(u)}(x)| = 0 \quad (9.8)$$

и наоборот; здесь  $G_{\xi, \sigma(u)}(x)$  — это обобщенное распределение Парето (*generalized Pareto distribution, GPD*):

$$G_{\xi, \sigma(u)}(x) = \begin{cases} 1 - (1 + \xi x / \sigma)^{-1/\xi}, & \text{если } \xi \neq 0 \\ 1 - \exp(-x / \sigma), & \text{если } \xi = 0 \end{cases} \quad (9.9)$$

Если  $\xi > 0$ , то  $G_{\xi, \sigma(u)}(\cdot)$  — это распределение Парето. Если  $\xi = 0$ , то  $G_{0, \sigma(u)}(\cdot)$ , как мы видели выше, — это экспоненциальное распределение. Если  $\xi = -1$ , то  $G_{-1, \sigma(u)}(\cdot)$  — это равномерное распределение.

Теорема дает возможность делать статистические выводы из данных, изолируя выбросы. Дополнительные подробности смотрите в обсуждении войн и тенденций в уровне насилия в Главе 16.

### 9.2. Невидимый хвост при степенном законе

Пусть  $K_n$  — максимум в выборке из  $n$  независимых одинаково распределенных величин в классе степенного закона;  $K_n = \max \{X_1, X_2, \dots, X_n\}$ . Пусть  $\varphi(\cdot)$  — плотность исходного распределения. Мы можем разложить каждый момент на два слагаемых, одно из которых «скрытое», потому что учитывает значения выше  $K_n$ , как показано на Рисунке 9.5:

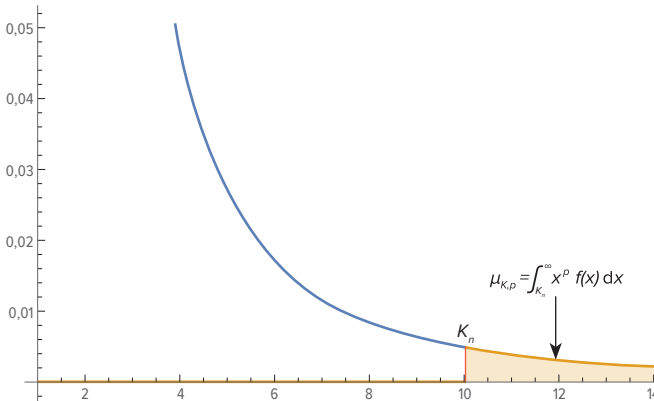


Рисунок 9.5: Момент порядка  $p$  выше  $K_n$ .

$$\mathbb{E} X^p = \underbrace{\int_L^{K_n} x^p \varphi(x) dx}_{\mu_{0,p}} + \underbrace{\int_{K_n}^L x^p \varphi(x) dx}_{\mu_{n,p}},$$

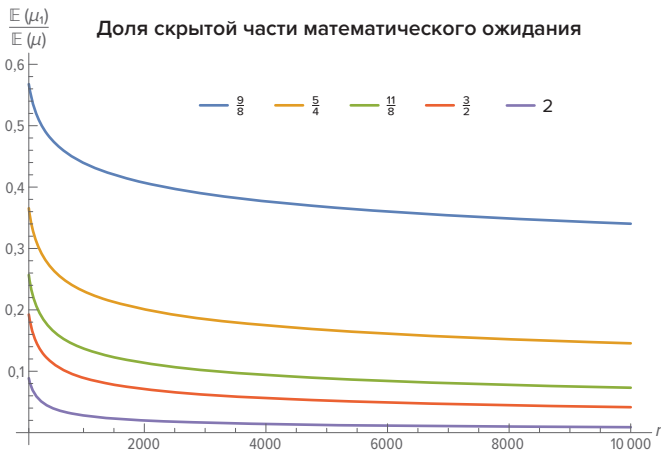
где  $\mu_0$  отражает видимую часть распределения и  $\mu_n$  — скрытую.

Мы также можем выполнить нормализацию и использовать как эмпирическое распределение функцию  $\varphi_e$ . Поскольку

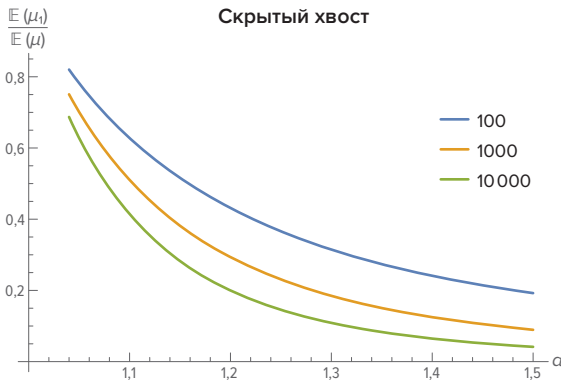
$$\underbrace{\left( \int_L^{K_n} \varphi_e(x) dx - \int_{K_n}^{\infty} \varphi(x) dx \right)}_{\text{С поправкой}} + \int_{K_n}^{\infty} \varphi(x) dx = 1, \quad (9.10)$$

можно использовать производную Радона — Никодима

$$\mathbb{E} X^p = \int_L^{K_n} x^p \frac{\partial \mu(x)}{\partial \mu_e(x)} \varphi_e(x) dx + \int_{K_n}^{\infty} x^p \varphi(x) dx. \quad (9.11)$$



**Рисунок 9.6:** Доля скрытой части в математическом ожидании для нескольких значений показателя хвоста  $\alpha$



**Рисунок 9.7:** Доля скрытой части в математическом ожидании для нескольких размеров выборки  $n$

### Предложение 9.1

Пусть  $K^*$  — точка, в которой функция выживания случайной величины  $X$  достаточно приблизилась к степенному хвосту с постоянным множителем, то есть  $\mathbb{P}(X > x) \approx L^{-\alpha} x^{-\alpha}$ .

При допущениях, гарантирующих  $K > K^*$ , распределение скрытого момента  $\mu_{K,p}$  для  $n$  наблюдений имеет плотность  $g_{(\cdot, \cdot)}(\cdot)$ :



$$g_{(n,p,\alpha)}(z) = nL^{\frac{\alpha p}{p-\alpha}} \left( z - \frac{pz}{\alpha} \right)^{\frac{p}{\alpha-p}} \exp \left( n \left( -L^{\frac{\alpha p}{p-\alpha}} \right) \left( z - \frac{pz}{\alpha} \right)^{\frac{-\alpha}{p-\alpha}} \right) \quad (9.12)$$

при  $z \geq 0$ ,  $p > \alpha$  и  $L > 0$ .

Математическое ожидание  $p$ -того момента в хвосте выше  $K$  при  $K > L > 0$  можно выразить как

$$\mathbb{E} \mu_{K,p} = \frac{\alpha(L^p - L^\alpha K^{p-\alpha})}{\alpha - p}. \quad (9.13)$$

Заметим, что распределение функции выживания выборки (то есть момента порядка  $p = 0$ ) — это экспоненциальное распределение с плотностью вероятности:

$$g_{(n,0,\alpha)}(z) = n e^{-nz}, \quad (9.14)$$

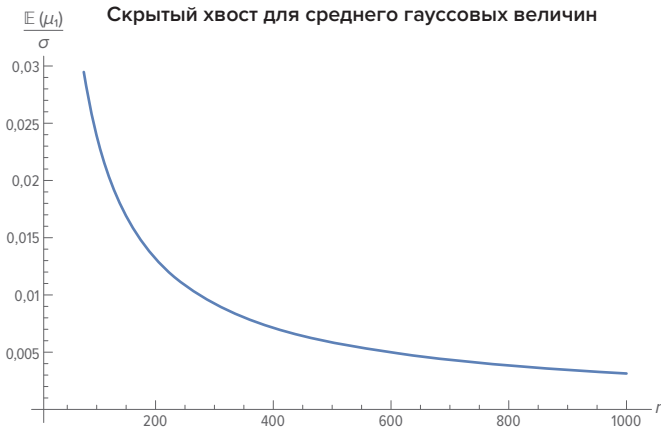
зависящей, как мы видим, только от  $n$ . Вероятность выброса в эмпирическом распределении не зависит от жирности хвостов.

Чтобы получить среднее, достаточно взять интеграл со стохастическим нижним пределом  $K > K_{min}$ :

$$\int_{K_{min}}^{\infty} \left( \underbrace{\int_{K_n}^{\infty} x^p \varphi(x) dx}_{\mu_{K,p}} \right) f_K(K) dK.$$

Чтобы получить среднее по всему распределению  $g_{(n,p,\alpha)}(z)$ , разложим среднее распределения Парето с коэффициентом масштаба  $L$ , приняв  $K_{min} = L$ .

Стандартным преобразованием: заменой переменной  $K \sim \mathcal{F}(\alpha, L n^{1/\alpha})^1$  на распределение Фреше с плотностью  $f_K(K) = \alpha n K^{-\alpha-1} L^\alpha \exp n \left( -\frac{L}{K} \right)^\alpha$ , — получаем нужный результат.



**Рисунок 9.8:** Отношение скрытой части среднего к среднему квадратическому отклонению в зависимости от размера выборки  $n$

<sup>1</sup> Очевидно, имеется в виду распределение Фреше с коэффициентом формы  $\alpha$ , коэффициентом масштаба  $L n^{1/\alpha}$  и минимальным значением 0.

### 9.2.1. Сравнение с нормальным распределением

Пометим параметры гауссовой случайной величины верхним индексом (g). Обозначая

плотность вероятности как  $\varphi^{(g)}$ , запишем  $\mu_K^{(g)} = \int_K^\infty \varphi^{(g)}(x) dx = \frac{2^{\frac{p-1}{2}} \Gamma\left(\frac{p+1}{2}, \frac{K^2}{2}\right)}{\sqrt{\pi}}$ . Как мы

видели выше, чем проходить через Гумбеля (точнее, ТЭЗ, или зеркального Гумбеля), удобнее извлечь точное распределение максимального значения из кумулятивного распределения стандартной гауссианы  $F^{(g)}$ :

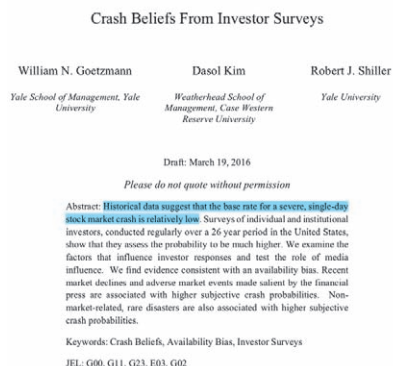
$$\frac{\partial F^{(g)}(K)}{\partial K} = \frac{e^{-\frac{K^2}{2}} 2^{\frac{1}{2}-n} n \operatorname{erfc}\left(-\frac{K}{\sqrt{2}}\right)^{n-1}}{\sqrt{\pi}},$$

где  $\operatorname{erfc}$  — дополнительная функция ошибок.

При  $p = 0$  ожидание «невидимого хвоста»  $\approx \frac{1}{n}$ .

$$\int_0^\infty \frac{e^{-\frac{K^2}{2}} 2^{-n-\frac{1}{2}} n \Gamma\left(\frac{1}{2}, \frac{K^2}{2}\right) \left(\operatorname{erf} \frac{K^2}{2} + 1\right)^{n-1}}{\pi} dK = \frac{1-2^{-n}}{n+1}.$$

### 9.3. ПРИЛОЖЕНИЕ: ЭМПИРИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ НЕ ЭМПИРИЧНО



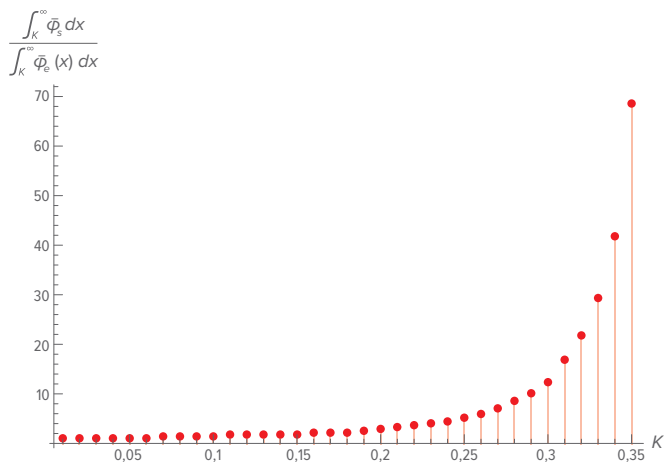
**Рисунок 9.9:** Еще раз о смещенной оценке частоты, на этот раз в другую сторону. «Базовая частота» является эмпирической оценкой с учетом худших прошлых наблюдений и содержит ошибку, выявленную в [227] как та же смещенная оценка, о которой писал древнеримский автор Лукреций в поэме «О природе вещей»: думать, что самая высокая гора, которая встретится в будущем, равна самой высокой горе, виденной ранее. Цитируется<sup>1</sup> без разрешения авторов, которые об этом извещены

1 Перевод: «Предрассудки о крахе по опросам инвесторов (Уильям Н. Гецманн, Йельская школа менеджмента, Йельский университет; Дасол Ким, Школа менеджмента Уэзерхеда, Кейсовский университет Западного резервного района; Роберт Дж. Шиллер, Йельский университет). Черновая версия 19 марта 2016. Просьба не цитировать без разрешения. Резюме: судя по историческим данным, базовая частота обвалов на фондовой бирже низка. Опросы индивидуальных инвесторов и инвестиционных фондов, проводимые в США в течение 26 лет, показывают завышенную оценку такой вероятности. Мы исследовали факторы, влияющие на ответы инвесторов, и проверили влияние СМИ. Мы нашли, что достаточным объяснением может служить эвристика доступности [склонность завышать оценку вероятности события, если аналогичные события легко вспомнить]. Недавние спады на рынке и неблагоприятные события на фондовой бирже, которые муслировались финансовыми изданиями, коррелируют с более высокими субъективными оценками вероятности краха. Не связанные с рынком катастрофы также коррелируют с преувеличенной боязнью краха. Ключевые слова: предрассудки о крахе, эвристика доступности, опросы инвесторов».

Преобладающая путаница касательно непараметрического эмпирического распределения проистекает из следующего сильного свойства: с ростом  $n$  и стабилизацией накопленной эмпирической гистограммы частот ошибки на ней рано или поздно становятся гауссовыми случайными величинами, *даже если исходное распределение жирнохвостое* (речь о случайной величине с бесконечной областью определения). Поскольку кумулятивное распределение вероятности — или функция выживания, неважно, — равномерно непрерывны на промежутке  $[0, 1]$  и, согласно теореме<sup>1</sup> Донскера, последовательность  $\sqrt{n} (F_n(x) - F(x))$ , где  $F_n$  — наблюдаемое кумулятивное распределение либо функция выживания для  $n$  слагаемых и  $F$  — истинное кумулятивное распределение либо функция выживания, сходится по распределению к нормальному распределению со средним 0 и дисперсией  $F(x)(1 - F(x))$ . И можно найти даже более сильные виды сходимости, по теореме Гливленко — Кантелли.

Благодаря этому замечательному свойству можно подумать, что и влияние хвостов на распределение сходится столь же быстро, независимо от исходного распределения. Тем более что дисперсия  $F(x)(1 - F(x))$  (как у эмпирического кумулятивного распределения, так и у эмпирической функции выживания) падает в области крайних значений... Только это падение не сопровождается падением ее влияния на выигрыш. На самом деле — такова уж природа крайних значений — ошибка под влиянием хвостов растет, если мы умножаем величину отклонения на его вероятность.

Для биржевых индексов в США, хотя первый метод осуждают как нелепый, расчеты по второму методу ведут к недооценке влияния хвостов на выигрыш от 5 до 70 раз, как показано на Рисунке 9.10. Мы вернемся к этой теме в Главе 11, обсуждая различие между бинарной и непрерывной платежной функцией и путаницу между вероятностью и практическим выигрышем, когда исходное распределение с жирным хвостом.



**Рисунок 9.10:** Показан относительный вклад хвоста с оценкой методом CVar по сравнению с (сглаженной) оценкой эмпирического распределения. Глубокий хвост недооценивается сегодняшними методами до 70 раз, хотя они считаются «эмпирическими»

<sup>1</sup> Функциональная ЦПТ, расширение центральной предельной теоремы на функции. Известна также как принцип инвариантности.

# В СКОРОСТЬ РОСТА И РЕЗУЛЬТАТ ПРИНАДЛЕЖАТ РАЗНЫМ КЛАССАМ РАСПРЕДЕЛЕНИЙ

Автор и Паскуале Чирилло показали, что число смертей от пандемий следует степенным законам с показателем хвоста заведомо ниже 1. Это значит, что вся информация сосредоточена в хвосте. Поэтому, если нет серьезных причин игнорировать общую и недвусмысленную статистику (в стиле «здесь особый случай»), не следует основывать решения по управлению рисками на поведении ожидаемого среднего или на точечной оценке.

Возник следующий парадокс: число смертей  $X_t$  за период с  $t_0$  по  $t$  распределено по Парето с неопределенным средним. Тем не менее скорость экспоненциального роста для числа смертей распределена не так! Она окажется тонкохвостой, с экспоненциальным распределением или подобным.

Чирилло и Талеб (2020) [48] (СТ) показали методами теории экстремальных значений, что пандемиям свойственен хвост с  $\alpha < 1$ , если рассматривать  $X_T$ , число смертей на некоторый день  $T$  в будущем, с функцией выживания  $P(X > x) = L(x)x^{-\alpha}$ . Примем для простоты, что при минимальном значении  $L$  функция  $L(x) \sim L$ , и функция выживания

$$\mathbb{P}(X > x) = Lx^{-\alpha}. \quad (\text{B.1})$$

## В.1. ЗАГАДКА

Рассмотрим обычную модель

$$X_t = X_0 e^{r(t-t_0)}, \quad (\text{B.2})$$

где

$$r = \frac{1}{t-t_0} \int_{t_0}^t r_s ds \quad (\text{B.3})$$

и  $r_s$  — это мгновенная интенсивность<sup>1</sup>. Нормализуем распределение к  $L = 1$ . Можно доказать следующее (предполагая, как выше, что у  $X_t$  есть функция выживания согласно формуле 13.13):

<sup>1</sup> В отличие от мгновенной скорости в кинематике (англ. *instantaneous velocity*),  $v = \frac{dX}{dt}$ , мгновенная интенсивность событий в статистике (англ. *instantaneous rate*)  $r = \frac{d \ln X}{dt} = \frac{v}{X}$ . В случае постоянной скорости  $X(t) = X(0) + vt$ ; в случае постоянной интенсивности  $X(t) = X(0)e^{rt}$ .

**Теорема 1**

Если  $r$  определена на промежутке  $(-\infty, \infty)$ , то плотность вероятности  $\varphi$  отмасштабированной интенсивности  $\rho = r(t - t_0)$  можно параметризовать как

$$\varphi(\rho) = \begin{cases} \frac{1}{2b} \exp\left(-\frac{\rho}{b}\right), & \text{если } \rho \geq 0 \\ \frac{1}{2b} \exp\frac{\rho}{b}, & \text{иначе} \end{cases},$$

где  $b = \frac{1}{\alpha}$ .

Если  $r$  определена на промежутке  $(0, \infty)$ , то плотность вероятности

$$\varphi(\rho) = \begin{cases} \alpha e^{-\alpha\rho}, & \text{если } \rho \geq 0 \\ 0, & \text{иначе} \end{cases}.$$

Получились разные варианты экспоненциального распределения, или распределения Лапласа (т. н. двойного экспоненциального распределения)<sup>1</sup>.

**Замечание 7**

*Следствие: будет наивностью просто перенести свойства интенсивности роста  $r$  на сам рост  $X_T$ , поскольку даже малые ошибки в  $r$  приведут к взрывным ошибкам в оценке экспоненциального роста  $X_T$ .*

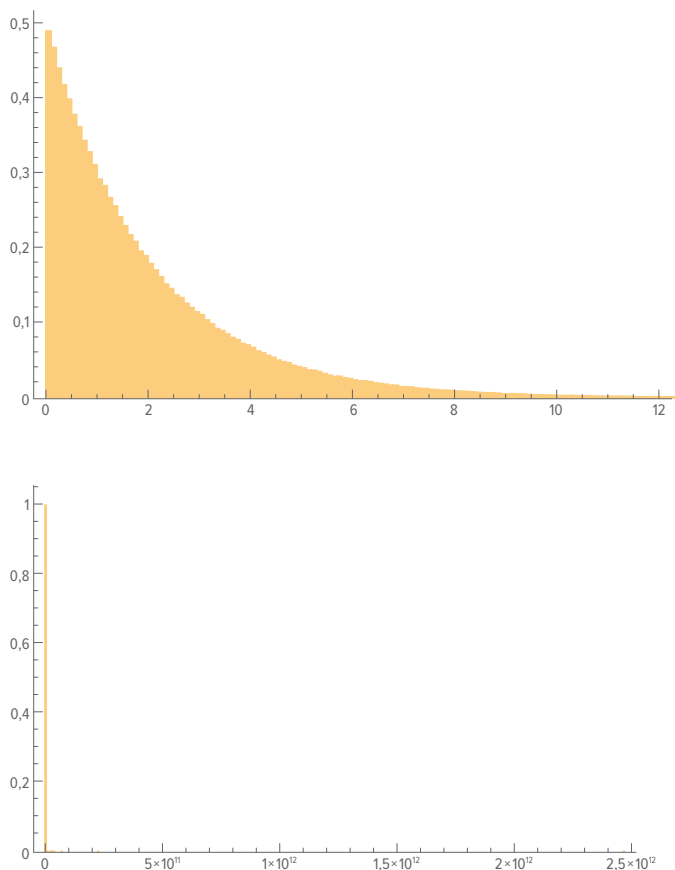
Верно и обратное: если  $r$  распределена экспоненциально,  $X_T$  должна быть распределена по Парето, как в формуле 13.13.

Схема вывода путем замены переменных такая. Пусть  $r$  следует распределению с плотностью  $\varphi$  с областью определения  $(a, b)$ ; при некоторых стандартных условиях  $u = g(r)$  следует новому распределению с плотностью

$$\psi(u) = \frac{\varphi(g^{(-1)}(u))}{g'(g^{(-1)}(u))}$$

и областью определения  $[g(a), g(b)]$ .

<sup>1</sup> По Лапласу распределена разность  $X = Y_2 - Y_1$  между двумя периодами  $Y_2$  и  $Y_1$  ожидания следующего события, если события случаются независимо друг от друга со средней частотой  $\lambda$ . Плотность вероятности лапласовой случайной величины  $X$  имеет вид  $f(x) = \frac{1}{2\lambda} e^{-\lambda|x|}$ . Таким образом, в области  $x > 0$  плотность вероятности  $f(x) = \frac{1}{2\lambda} e^{-x/\lambda}$  аналогична экспоненциальному распределению, а в области  $x < 0$  график этой функции представляет собой зеркальное отражение графика положительной области. Поэтому распределение Лапласа называют также двойным экспоненциальным распределением. В целом график двухвостый с максимумом в точке  $x = 0$ ; среднее 0; дисперсия  $2\lambda^2$ ; коэффициент асимметрии 0; коэффициент эксцесса 6.



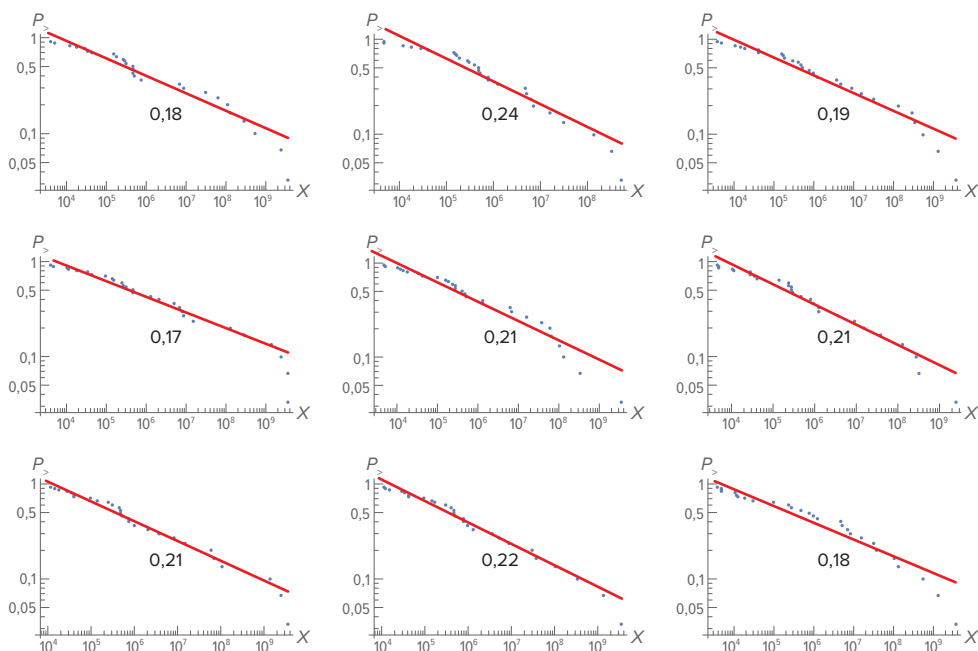
**Рисунок В.1:** Вверху — гистограмма  $10^6$  реализаций  $r$  из экспоненциального распределения с коэффициентом  $\lambda = \frac{1}{2}$ . Внизу — гистограмма  $X = e^r$ . Распределения получились, мягко говоря, разные.

Экссесс выборки соответственно 9 и  $10^6$  (и в теории второй экссесс вообще бесконечный); все значения второго распределения уступают единственному большому отклонению

## В.2. У ПАНДЕМИЙ ИСКЛЮЧИТЕЛЬНО ЖИРНЫЕ ХВОСТЫ

На Рисунке В.2 показано, что степенной закон с низким показателем  $\alpha$  получается независимо от выборки данных. В [48] мы пользовались теорией экстремальных значений, но на графике показан предварительный анализ, не вошедший в статью. Столь низкого показателя хвоста мы больше нигде не видели. Вывод: эпидемиологические исследования важны для науки, но для политических решений нужно опираться на ТЭЗ или на простые соображения осторожности — как рекомендация удалить раковую опухоль пораньше, пока это обойдется дешево<sup>1</sup>.

<sup>1</sup> Большой ошибкой будет строить политику, исходя из прогноза по одной точке — а, как мы показываем в Главе 11, заведомо ложным будет и прогноз функции выживания. Не следует полагаться на прогнозы, пусть и составленные «по всем правилам науки», за пределами бинарных пари. — *Прим. автора.*



**Рисунок В.2:** Возьмем 60 крупнейших пандемий и будем случайно выбирать половину из них. Нормализуем данные по текущей численности населения. Свойства и параметры Парето устойчивы к таким возмущениям. ТЭЗ дает чуть больший показатель хвоста, но все равно заведомо ниже 1. Столь низкий показатель хвоста авторы впервые видят в своих исследованиях

## С ПРИНЦИП БОЛЬШОГО ОТКЛОНЕНИЯ, ВКРАТЦЕ

Вернемся к границе Крамера и кратко представим литературу о ней. Суть противопоставления высоких выбросов богатым выбросам в Разделе 3.1 в том, что при некоторых условиях падение вероятности выбросов в хвосте происходит по экспоненте. Это важное условие в управлении рисками — как мы упоминали выше, принцип катастроф требует этого быстрого спада, иначе диверсификация будет бесполезна.

Принцип большого отклонения помогает понять, как ведет себя экспоненциальный хвост и почему наши бизнесы не терпят крах при тонких хвостах... и, что не менее важно, почему это может случиться при жирных хвостах или при нарушении условия Крамера [118].

Пусть  $M_N$  — среднее по последовательности реализаций  $N$  случайных величин, одинаково распределенных. При большом  $N$  рассмотрим хвостовую вероятность:

$$\mathbb{P}(M_N > x) \approx e^{-N I(x)},$$

где  $I(\cdot)$  — функция Крамера, или функция интенсивности<sup>1</sup> (Варадхан [260], Дембо и Зеитунни [59]). Если известно распределение  $X$ , то преобразованием<sup>2</sup> Лежандра получаем  $I(x) = \sup_{\vartheta > 0} \{\vartheta(x) - \lambda(\vartheta)\}$ , где  $\lambda(\vartheta) = \ln \mathbb{E} e^{\vartheta(X)}$  — производящая функция кумулянтов<sup>3</sup>.

1 В 1930-е годы Крамер определил эту функцию для данной случайной величины  $X$  именно этим условием: что среднее по большому числу  $N$  экземпляров случайной величины  $X$  имеет функцию выживания, которая в хвостовой области (при больших  $x$ ) имеет экспоненциальный спад  $\approx e^{-N I(x)}$ . В 1998 Дембо и Зеитунни обобщили функцию Крамера, определив ее не на пространстве состояний случайной величины  $X$ , а на произвольном вероятностном пространстве.

2 Преобразование Лежандра для заданной функции  $\lambda(\vartheta)$  и ее производной  $\lambda'(\vartheta)$  строит функцию  $I(x)$  с производной  $I'(x)$  так, что старая производная принимается за новую переменную  $x$ , а старая переменная — за новую производную  $I'(x)$ , то есть  $\lambda'(I'(x)) = x$  и  $I'(\lambda'(\vartheta)) = \vartheta$ . Новая функция находится как  $I(x) = \sup \{x\vartheta - \lambda(\vartheta)\}$ .

3 В русской литературе производящая функция кумулянтов чаще называется просто логарифмом характеристической функции.



Поведение функции  $\mathcal{H}(x)$  показывает, каков вклад одного-единственного события в общий ожидаемый выигрыш. (Оно связано с условием Крамера, требующим, чтобы существовали экспоненциальные моменты.)

Особый случай — для случайных величин Бернулли; для этого класса дискретных случайных величин есть строгая оценка — оценка Чернова.

## ПРОСТОЙ СЛУЧАЙ: ОЦЕНКА ЧЕРНОВА

Выигрыш по бинарному опциону подчиняется строгим ограничениям. Пусть  $(X_i)_{1 \leq i \leq n}$  — последовательность независимых испытаний Бернулли, принимающих значения в множестве  $\{0, 1\}$ , причем вероятности  $\mathbb{P}(X = 1) = p$  и  $\mathbb{P}(X = 0) = 1 - p$ . Рассмотрим сумму  $S_n = \sum_{1 \leq i \leq n} X_i$ , имеющую ожидаемое среднее  $\mathbb{E}S_n = np = \mu$ . Обозначим  $\delta$  расстояние от среднего; оценка Чернова дает следующее. Для любого  $\delta > 0$

$$\mathbb{P}(S_n \geq (1 + \delta)\mu) \leq \left( \frac{e^\delta}{(1 + \delta)^{1 + \delta}} \right)^\mu,$$

а для любого  $0 \leq \delta \leq 1$

$$\mathbb{P}(S_n \geq (1 + \delta)\mu) \leq \exp\left(-\frac{\mu\delta^2}{3}\right)^1.$$

Оценим вероятность того, что при  $n$  бросаний монеты результат на  $\delta = 50\%$  превзойдет истинное среднее.

Поскольку  $p = \frac{1}{2}$  и  $\mu = \frac{n}{2}$ , искомая вероятность  $\mathbb{P}(S_n \geq \frac{3}{2} \frac{n}{2}) \leq \exp\left(-\frac{\mu\delta^2}{3}\right) = e^{-n/24}$ ; при  $n = 1000$  такое случается в одной серии из  $1,24 \times 10^{18}$ .

*Доказательство.* Оценка Маркова<sup>2</sup> ограничивает вероятность  $\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}X}{c}$ . Заменим  $X$  положительной функцией  $g(X)$ :  $\mathbb{P}(g(X) \geq g(c)) \leq \frac{\mathbb{E}g(X)}{g(c)}$ . Далее мы используем функцию  $g(X) = e^{\omega X}$ .

Рассмотрим расстояние от среднего, равное  $(1 + \delta)$  при  $\delta > 0$ . Полагая  $\omega > 0$ , получим оценку

$$\mathbb{P}(S_n \geq (1 + \delta)\mu) = \mathbb{P}(e^{\omega S_n} \geq e^{\omega(1 + \delta)\mu}) \leq e^{-\omega(1 + \delta)\mu} \mathbb{E}e^{\omega S_n}. \quad (\text{C.1})$$

Последний сомножитель,  $\mathbb{E}e^{\omega S_n} = \mathbb{E}e^{\omega \sum X_i} = \mathbb{E}(e^{\omega X_i})^n$ , в силу независимости от момента остановки<sup>3</sup>, представим как  $(\mathbb{E}e^{\omega X})^n$ <sup>4</sup>.

1 Для этого случая есть еще более строгая оценка Чернова,  $\mathbb{P}(S \geq (1 + \delta)\mu) \leq \exp\left(-\frac{\mu\delta^2}{2 + \delta}\right)$ .

2 Она же неравенство Маркова (*Markov's inequality*), для случайной величины  $X$ , о которой известно лишь то, что  $X \geq 0$  и существует  $\mathbb{E}X$ . Эта оценка названа в честь Андрея Андреевича Маркова, хотя открыта его учителем, Пафнутием Львовичем Чебышёвым. Оценкой Чебышёва называется более строгая оценка, учитывающая второй момент.

3 Очевидно, имеется в виду, что свойства последовательности  $(X_i)_{1 \leq i \leq n}$  исследуются при заданном  $n$ . Иными были бы свойства серии подбрасываний монеты, если задать сумму  $S_n = \sum_{1 \leq i \leq n} X_i$  и продолжать бросать монету, пока не достигнут этот результат; тогда число подбрасываний монеты  $n$  будет случайной величиной, зависящей от случайных величин  $X_i$ .

4 Более аккуратная запись этого рассуждения длиннее. Представим  $e^{\omega \sum X_i} = e^{\omega X_1} \dots e^{\omega X_n}$ . Поскольку все случайные величины  $X_1, \dots, X_n$  независимы друг от друга, все случайные величины  $e^{\omega X_1}, \dots, e^{\omega X_n}$  независимы друг от друга. Поэтому  $\mathbb{E}(e^{\omega X_1} \dots e^{\omega X_n}) = \mathbb{E}e^{\omega X_1} \dots \mathbb{E}e^{\omega X_n} = (\mathbb{E}e^{\omega X})^n$ .

Основание степени  $\mathbb{E}e^{\omega X} = 1 - p + pe^{\omega}$ . Поскольку  $1 + x \leq e^x$ <sup>1</sup>,

$$\mathbb{E}e^{\omega S_n} \leq e^{\mu(e^{\omega} - 1)}.$$

Подставляя в С.1, получаем, что

$$\mathbb{P}(S_n \geq (1 + \sigma)\mu) \leq e^{-\omega(1 + \delta)\mu} e^{\mu(e^{\omega} - 1)}. \quad (\text{C.2})$$

Чтобы ужесточить оценку, минимизируем правую часть подбором  $\omega$ .

Из  $\omega^* \in \left\{ \omega \left| \frac{\partial e^{\mu(e^{\omega} - 1) - \omega(1 + \delta)\mu}}{\partial \omega} = 0 \right. \right\}$  получаем<sup>2</sup>  $\omega^* = \ln(1 + \delta)$ .

Откуда открывается оценка  $e^{\delta\mu} (1 + \delta)^{-(1 + \delta)\mu}$ <sup>3</sup>.

Известно обобщение оценок Чернова, сделанное Хефдингом [130], — оценка распространена на случайные величины, не обязательно распределенные по Бернулли.

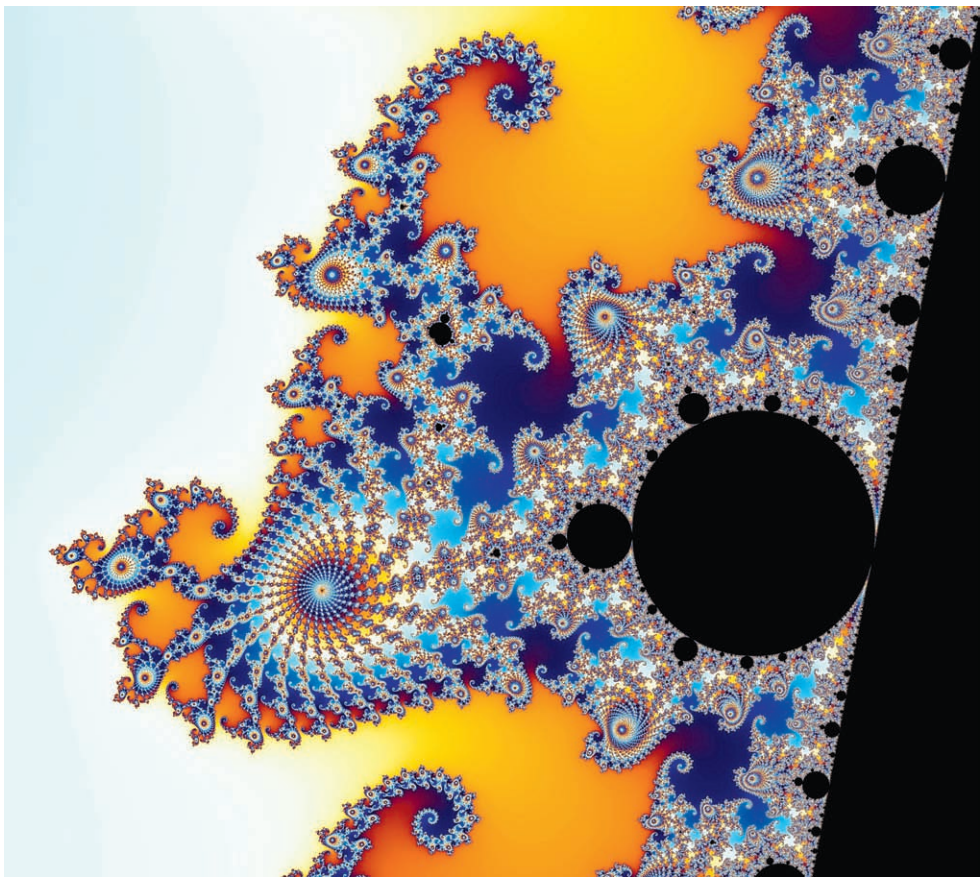
1 В данном случае в знаменитое неравенство вместо  $x$  подставляется  $-p + pe^{\omega}$ , так что получается  $1 - p + pe^{\omega} \leq e^{-p + pe^{\omega}}$ . Обе части положительны, так что их можно возвести в степень  $n$  и получить  $(\mathbb{E}e^{\omega X})^n = (-p + pe^{\omega})^n \leq (e^{-p + pe^{\omega}})^n = e^{-np + npe^{\omega}}$ . Поскольку мы обозначали ожидаемое среднее  $\mathbb{E}S_n = np = \mu$ , можно коротко записать  $(\mathbb{E}e^{\omega X})^n \leq e^{\mu(-1 + e^{\omega})}$ .

2 В самом деле, ищем точку минимума по переменной  $\omega$  для выражения вида  $e^{\mu P}$ , где  $P = e^{\omega} - 1 - \omega(1 + \delta)$ . Частная производная  $\frac{\partial}{\partial \omega} e^{\mu P}$  проходит через ноль, меняя знак с отрицательного на положительный, только если это делает  $\frac{\partial}{\partial \omega} P = e^{\omega} - (1 + \delta)$ , то есть при  $e^{\omega} = 1 + \delta$ .

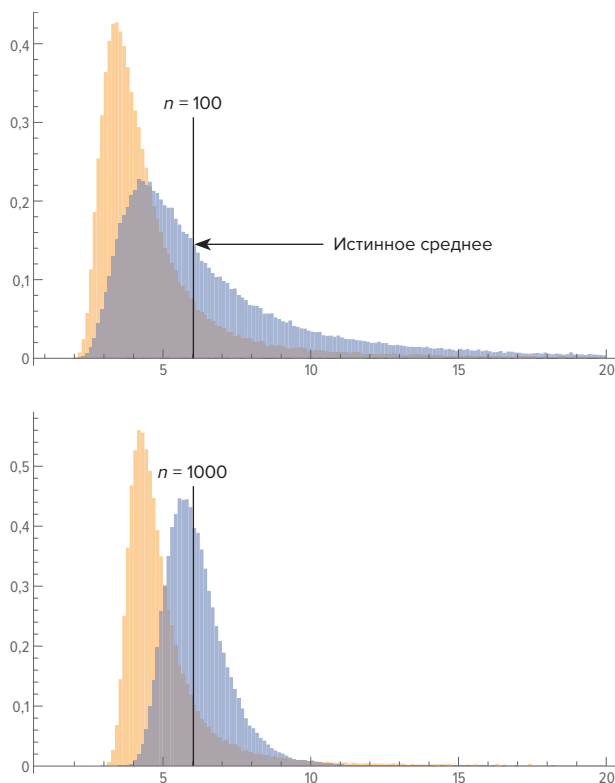
3 В самом деле, подставим  $\ln(1 + \delta)$  вместо  $\omega$  в оценку:  $e^{-\omega(1 + \delta)\mu} e^{\mu(e^{\omega} - 1)} = e^{-(\ln(1 + \delta))(1 + \delta)\mu} e^{\mu(e^{\ln(1 + \delta)} - 1)} = (1 + \delta)^{-(1 + \delta)\mu} e^{\mu(1 + \delta - 1)}$ .

## D

### КАЛИБРОВКА В СИТУАЦИИ ПАРЕТО



**Рисунок D.1:** Великий человек Бенуа Мандельброт связал фрактальную геометрию со статистическими распределениями через самоподобие во всех масштабах. Когда его попросили объяснить, что он исследует, он ответил: «rugosité», то есть «шероховатость»... и у него ушло пятьдесят лет на то, чтобы понять это. Картина Морской конек — фрагмент множества Мандельброта, построенный пользователем Wolfgang Beyer для Wikipedia Commons



**Рисунок D.2:** Методом Монте-Карло (при  $10^5$  испытаниях) проверено, насколько точно и надежно оценивается среднее: по выборочному среднему (методы 1 и 2) — гистограммы персикового цвета, методом максимального правдоподобия<sup>1</sup> (метод 3) — синие гистограммы, если дано распределение Парето с  $\alpha = 1,2$  и размер выборки увеличивают с  $n = 100$  до 1000.

Видно, что ММП в итоге приближается к точной оценке распределения, тогда как выборочное среднее по методам 1 и 2 продолжает занижать среднее асимметрично распределенных данных. Чтобы достичь той же точности, выборку пришлось бы увеличить до  $10^7$  или больше

<sup>1</sup> В методе максимального правдоподобия параметр распределения оценивается по наблюдаемой выборке так. Пусть монету бросали три раза, и выпали Орел, Орел, Решка; пусть бросания были независимыми, но честность монеты под вопросом: правда ли вероятность выпадения Орла  $p = \frac{1}{2}$ , неизвестно; требуется оценить «орловость» монеты, исходя из наблюдаемой выборки.

Рассмотрим т. н. функцию правдоподобия  $\mathcal{L}(\theta) = \mathbb{P}((X_1, X_2, X_3) = (\text{Орел}, \text{Орел}, \text{Решка}) \text{ при } p = \theta) = \theta\theta(1-\theta) = \theta^2 - \theta^3$ . Поскольку производная  $\frac{d}{d\theta} \mathcal{L}(\theta) = 2\theta - 3\theta^2$  меняет знак с положительного на отрицательный в точке  $\hat{\theta} = \frac{2}{3}$ , это точка максимума функции правдоподобия и, согласно ММП, это и есть оценка «орловости»  $p$  данной монеты. Заметим, что по столь короткой серии наблюдений получается функция правдоподобия со слабо выраженным максимумом: правдоподобие гипотезы, что орловость монеты была  $p = \frac{2}{3}$ , составляет  $\mathcal{L}\left(\frac{2}{3}\right) = \frac{4}{27} \approx 0,148$  и не намного превосходит, например, правдоподобие гипотезы честной монеты с «орловостью»  $p = \frac{1}{2}$ , ее правдоподобие  $\mathcal{L}\left(\frac{1}{2}\right) = \frac{1}{8} = 0,125$ .

Чтобы найти точку максимума  $\hat{\theta}$  функции правдоподобия  $\mathcal{L}(\theta)$ , достаточно найти точку максимума ее логарифма  $\ln \mathcal{L}(\theta)$ ; в самом деле, логарифм возрастает, так что точка максимума логарифмической функции правдоподобия  $\ln \mathcal{L}(\theta)$  та же самая, что у функции правдоподобия  $\mathcal{L}(\theta)$ . Выгода работы с  $\ln \mathcal{L}(\theta)$  вместо  $\mathcal{L}(\theta)$  возникает, когда  $\mathcal{L}(\theta)$  представлена произведением и, следовательно, логарифмическая функция правдоподобия представима суммой логарифмов. В нашем примере  $\ln \mathcal{L}(\theta) = \ln \theta\theta(1-\theta) = 2 \ln \theta + \ln(1-\theta)$ ; производная  $\frac{d}{d\theta} \ln \mathcal{L}(\theta) = \frac{2}{\theta} - \frac{1}{1-\theta} = \frac{2-3\theta}{\theta(1-\theta)}$  меняет знак с положительного на отрицательный в точке  $\hat{\theta} = \frac{2}{3}$ .

Начнем с повторения.

**Определение D.1** (Класс степенного закона  $\mathfrak{P}$ )

Случайная величина  $X$ , определенная на  $\mathbb{R}$ , принадлежит  $\mathfrak{P}$ , классу медленно меняющихся функций, он же класс хвостов Парето, он же класс хвостов степенного закона, если спад функции выживания при больших  $|X|$  асимптотически представим степенью с постоянным показателем  $\alpha$  или двумя разными показателями  $\alpha$  и  $\alpha'$  для левого и правого хвоста, то есть

$$\mathbb{P}(X > x) = L(x)x^{-\alpha} \quad (\text{D.1})$$

для правого хвоста и

$$\mathbb{P}(-X > x) = L(x)x^{-\alpha'} \quad (\text{D.2})$$

для левого хвоста, где  $\alpha, \alpha' > 0$  и функция  $L: (0, \infty) \rightarrow (0, \infty)$ , медленно меняющаяся, что определяется как

$$\lim_{x \rightarrow \infty} \frac{L(kx)}{L(x)} = 1$$

при всех  $k > 0$ .

Что с таким распределением удобно, так это то, что параметр  $\alpha$  следует обратному<sup>1</sup> гамма-распределению, которое быстро сходится к гауссову, не требуя больших  $n$  для оценки.

Проблема со степенными законами, даже имеющими так называемую конечную дисперсию, та, что конечность дисперсии на практике не спасает, как мы видели в Главе 8. Это иллюстрирует Рисунок D.2, где мы видим подгонку с переменным успехом.

## D.1. РАСПРЕДЕЛЕНИЕ ВЫБОРОЧНОЙ ОЦЕНКИ ПОКАЗАТЕЛЯ ХВОСТА

Рассмотрим стандартное распределение Парето для случайной величины  $X$  с плотностью вероятности:<sup>2</sup>

$$\varphi_X(x) = \alpha L^\alpha x^{-\alpha-1}, \quad x > L. \quad (\text{D.3})$$

Изменим масштаб так, что  $L = 1$ <sup>3</sup>.

1 Обратному гамма-распределению следует  $X^{-1}$ , если случайная величина  $X$  следует гамма-распределению. Если  $X$  — случайное время ожидания  $a$  событий, которые происходят независимо друг от друга со средним периодом  $b$ , то  $aX^{-1}$  — случайная частота, измеренная по первым  $a$  событиям. Как и для гамма-распределения, коэффициент формы  $a > 0$ , коэффициент масштаба  $b > 0$  и все значения  $X^{-1} > 0$ ; плотность вероятности  $f(x) = \frac{x^{-a-1} e^{-b/x} b^a}{\Gamma(a)}$  имеет максимум при  $x = \frac{b}{a+1}$ . Моменты существуют при достаточно больших

$a$ : среднее  $\frac{b}{a-1}$  при  $a > 1$ , дисперсия  $\frac{b^2}{(a-1)^2(a-2)}$  при  $a > 2$ , коэффициент асимметрии  $\frac{4\sqrt{a-2}}{a-3}$  при  $a > 3$ , эксцесс  $3 + \frac{6(5a-11)}{(a-3)(a-4)}$  при  $a > 4$ .

2 Подразумевается, что  $X > L$ ; поэтому при  $x < L$  плотность вероятности  $\varphi_X(x) = 0$ .

3 Соответственно,  $\varphi_X(x) = \begin{cases} \alpha x^{-\alpha-1}, & x > 1 \\ 0 & x < 1 \end{cases}$ .

При таком минимальном значении функция правдоподобия

$$\mathcal{L}(\alpha) = \prod_{i=1}^n \alpha x_i^{-\alpha-1};$$

максимизируя ее логарифм

$$\ln \mathcal{L}(\alpha) = n \ln \alpha - (\alpha + 1) \sum_{i=1}^n \ln x_i,$$

находим, что  $\hat{\alpha} = \frac{n}{\sum_{i=1}^n \ln x_i}$ <sup>1</sup>.

Рассмотрим средний логарифм  $l = \frac{1}{n} \sum_{i=1}^n \ln X_i$ . Его характеристическая функция<sup>2</sup>

$$(\psi(t))^n = \left( \int_1^{\infty} \varphi_X(x) \exp \frac{it \ln x}{n} dx \right)^n = \left( \frac{\alpha n}{\alpha n - it} \right)^n$$
<sup>3</sup>,

то есть представима как характеристическая функция гамма-распределения с коэффициентами  $\left(n, \frac{1}{an}\right)$ <sup>4</sup>. Стандартный результат показывает, что  $\hat{\alpha} \triangleq \frac{1}{l}$  будет следовать обратному гамма-распределению с плотностью вероятности

$$\varphi_{\hat{\alpha}}(a) = \frac{\left(\frac{\alpha n}{a}\right)^n \exp \frac{-\alpha n}{a}}{a \Gamma(n)}, \quad a > 0$$
<sup>5</sup>.

1 В самом деле, производная  $\frac{d}{d\alpha} \ln \mathcal{L}(\alpha) = \frac{d}{d\alpha} \left( n \ln \alpha - (\alpha + 1) \sum_{i=1}^n \ln x_i \right) = \frac{n}{\alpha} - \sum_{i=1}^n \ln x_i$  меняет знак с положительного на отрицательный в такой точке  $\hat{\alpha}$ , что  $\frac{n}{\hat{\alpha}} - \sum_{i=1}^n \ln x_i = 0$ .

2 Здесь  $\psi(t)$  — характеристическая функция случайной величины  $\frac{\ln X}{n}$ . Поскольку характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых и поскольку слагаемые  $\frac{\ln X_i}{n}$  распределены одинаково, характеристическая функция суммы представима как  $(\psi(t))^n$ .

3 В самом деле, поскольку  $X > 1$ , случайная величина  $Y = \frac{\ln X}{n} > 0$  и ее характеристическая функция  $\psi(t) = \mathbb{E} e^{itY} = \int_0^{\infty} \varphi_Y(y) e^{ity} dy$ , где плотность вероятности  $\varphi_Y(y) = \varphi_X(e^{ny}) \frac{d}{dy} (e^{ny})$ . Заменяя  $e^{ny} = x$  и  $e^{ity} = x^{it/n}$ , получаем  $\varphi(t) = \int_1^{\infty} \varphi_X(x) x^{it/n} dx = \int_1^{\infty} \alpha x^{-\alpha-1+it/n} dx = \alpha \left. \frac{x^{-\alpha+it/n}}{-\alpha+it/n} \right|_1^{\infty} = \frac{\alpha n}{\alpha n - it}$ .

4 В самом деле, для гамма-распределения с коэффициентами формы и масштаба  $k$  и  $\vartheta$  характеристическая функция  $\psi(k, \vartheta^{-1})(t) = \left(1 - \frac{it}{\vartheta}\right)^{-k}$ . Подставляя  $(k, \vartheta^{-1}) = \left(n, \frac{1}{an}\right)$ , получаем  $\psi\left(n, \frac{1}{an}\right)(t) = \left(1 - \frac{it}{an}\right)^{-n} = \left(\frac{\alpha n}{\alpha n - it}\right)^n$ .

5 В самом деле, если случайная величина  $X \sim \text{Gamma}(k, \vartheta^{-1})$ , то случайная величина  $Y = X^{-1} \sim \text{Inv-Gamma}(k, \vartheta)$  с плотностью вероятности  $\frac{\vartheta^k}{\Gamma k} y^{-k-1} \exp\left(-\frac{\vartheta}{y}\right)$ . Подставляя  $X = l$ ,  $Y = \hat{\alpha}$ ,  $y = a$ ,  $k = n$ ,  $\vartheta = an$ , получаем для  $\hat{\alpha}$  плотность вероятности  $\frac{(\alpha n)^n}{\Gamma n} a^{-n-1} \exp\left(-\frac{\alpha n}{a}\right) = \frac{(\alpha n / a)^n}{a \Gamma n} \exp\left(-\frac{\alpha n}{a}\right)$ .

**Устранение систематической ошибки.** Поскольку  $\mathbb{E} \hat{\alpha} = \frac{n}{n-1} \alpha^1$ , выберем другую случайную величину, несмещенную:

$$\hat{\alpha}' = \frac{n-1}{n} \hat{\alpha},$$

для которой, после масштабирования<sup>2</sup>, распределение получится такое:

$$\varphi_{\hat{\alpha}'}(a) = \frac{\left(\frac{\alpha(n-1)}{a}\right)^{n+1} \exp \frac{-\alpha(n-1)}{a}}{\alpha \Gamma(n+1)} \quad 3.$$

**Усечение при  $\alpha > 1$ .** Учитывая, что при  $\alpha \leq 1$  среднее не существует, ограничимся теми распределениями Парето, где  $\alpha = 1 + \varepsilon$ ,  $\varepsilon > 0$ . Тогда в выборке для оценки показателя участвуют только значения строго больше 1. Ограничим выборку, используя точку отсечения снизу  $\varepsilon > 0$ , то есть требуя  $\frac{n-1}{\sum \ln x_i} > 1 + \varepsilon$ . Ожидаемая оценка станет

$$\mathbb{E}(\hat{\alpha} \mid \hat{\alpha} > 1 + \varepsilon), \text{ и ее плотность вероятности будет } \varphi_{\hat{\alpha}'}(a) = \frac{\varphi_{\hat{\alpha}'}(a)}{\int_{1+\varepsilon}^{\infty} \varphi_{\hat{\alpha}'}(a) da} \quad 4.$$

Таким образом, распределение оценок показателя при условии, что он превышает 1, принимает вид:

$$\varphi_{\hat{\alpha}'}(a) = \frac{\left(\frac{\alpha n^2}{a(n-1)}\right)^n \exp \frac{\alpha n^2}{a - an}}{a \left( \Gamma n - \Gamma \left( n, \frac{\alpha n^2}{(n-1)(1+\varepsilon)} \right) \right)}, \quad a > 1 + \varepsilon. \quad (\text{D. 4})$$

В итоге, как видно на Рисунке D.2, полученное стандартными методами среднее для  $\alpha$  хвоста может хорошо приближать однохвостовое распределение Парето.

1 В самом деле, если случайная величина  $Y \sim \text{Inv-Gamma}(k, \vartheta)$ , то  $\mathbb{E} Y = \frac{\vartheta}{k-1}$ . Подставляя  $Y = \hat{\alpha}$ ,  $k = n$ ,  $\vartheta = \alpha n$ , получаем  $\mathbb{E} \hat{\alpha} = \frac{\alpha n}{n-1}$ . Таким образом, пока число наблюдений  $n < \infty$ , полученная по ММП оценка  $\hat{\alpha}$  будет в среднем завышена по сравнению с показателем  $\alpha$  распределения Парето в  $\frac{n}{n-1}$  раз.

2 Масштабировать в  $s$  раз случайную величину  $X$  значит рассмотреть случайную величину  $sX = Y$  и ее значения  $sx = y$ . Для случайной величины  $Y$  плотность вероятности по-прежнему имеет площадь 1 под всей кривой, но по сравнению с плотностью вероятности случайной величины  $X$  график растянут в  $s$  раз по горизонтали и сжат в  $s$  раз по вертикали:  $\varphi_Y(y) = \frac{1}{s} \varphi_X\left(\frac{y}{s}\right)$ .

3 С точностью до постоянного множителя. Масштабируя случайную величину  $\frac{n}{n-1} \hat{\alpha} = \hat{\alpha}'$  и ее значения  $\frac{n}{n-1} a = a'$ , а также пользуясь свойством гамма-функции  $n \Gamma n = \Gamma(n+1)$ , получаем выражение для плотности вероятности

$$\varphi_{\hat{\alpha}'}(a') = \frac{n}{n-1} \varphi_{\hat{\alpha}}\left(\frac{n}{n-1} a'\right) = \frac{n}{n-1} \cdot \frac{\left(\frac{\alpha(n-1)}{a'}\right)^n \exp \frac{-\alpha(n-1)}{a'}}{\frac{n}{n-1} a' \Gamma n} = \frac{n}{n-1} \cdot \frac{\left(\frac{\alpha(n-1)}{a'}\right)^{n+1} \exp \frac{-\alpha(n-1)}{a'}}{\alpha \Gamma(n+1)}.$$

4 Точнее,  $\varphi_{\hat{\alpha}'}(a) = \begin{cases} \frac{\varphi_{\hat{\alpha}'}(a)}{\int_{1+\varepsilon}^{\infty} \varphi_{\hat{\alpha}'}(a) da}, & a > 1 + \varepsilon \\ 0, & \text{иначе} \end{cases}$ .

## ПЕЧАЛЬНО, НО ФАКТ: ДИАГНОСТИКА S&P 500<sup>†</sup>

Это диагностический тур<sup>1</sup> по свойствам биржевого индекса S&P 500 на протяжении его истории. Мы устроим серию тестов и рассмотрим полученную статистическую картину. Очевидно, что доходность распределена по степенному закону (с некоторыми усложнениями, например с асимметрией между подъемом и спадом), откуда следует, что и здесь обычные методы анализа несостоятельны. Мы обращаем внимание на такие вещи, как:

- поведение эксцесса при накоплении наблюдений (по мере расширения временного окна);
- поведение условного математического ожидания  $\mathbb{E}(X \mid X > K)$  при различных значениях  $K$ ;
- график отношения максимума к сумме (*maximum to sum*, MS);
- падения (максимальные колебания в данном временном окне);
- экстремальные значения — правда ли они независимы.

И подтверждаем диагноз: весь класс анализов в  $L_2$ , включая современную портфельную теорию, факторный анализ, GARCH, условную дисперсию и стохастическую волатильность, методологически (и практически) несостоятельны.

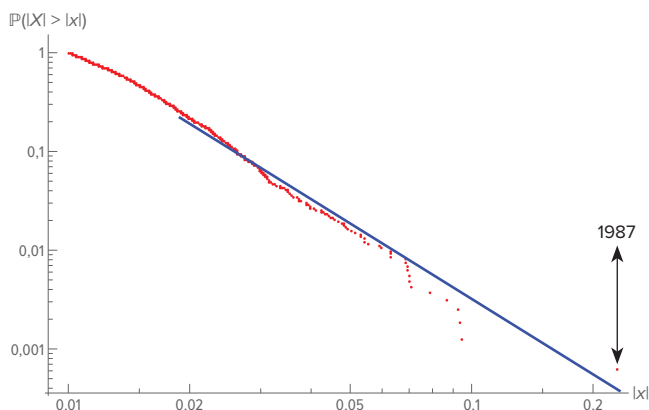
### 10.1. ПРИНАДЛЕЖНОСТЬ КЛАССУ ПАРЕТО И МОМЕНТЫ

**Проблема.** Как мы говорили в Прологе, переход от тонкохвостости к жирнохвостости не похож на простую смену фасона. Беспринципные финансисты и экономисты заявляют, что услышали предупреждение о жирнохвостости, после чего спокойно продолжают в прежнем духе, отмахиваясь от таких последствий, как медлительность закона больших чисел и провал попыток статистически значимой оценки среднего и высших моментов по доступной выборке (можно также упомянуть провал эргодичности). Горы неинформативных аналитических публикаций в области инвестиций все растут и растут.

<sup>1</sup> Эта глава больше опирается на рисунки, чем на текст, потому что главными доводами служит отсутствие сходимости на графиках.



Принадлежность классу Парето четко определяется по отсутствию некоторого высшего момента, что проявляется как отсутствие сходимости согласно ЗБЧ.



**Рисунок 10.1:** Наглядная проверка принадлежности классу Парето на обычном графике в дважды логарифмическом масштабе, где вдоль горизонтальной оси отложена абсолютная доходность, а вдоль вертикальной — функция выживания.

Если удалить одну-единственную точку, отвечающую краху 1987 года, можно было бы надеяться на логнормальный класс или еще какое-то смешанное жирнохвостое распределение вне класса степенного закона. Потому что функция выживания начинает вертикально падать, что характерно для асимптотически экспоненциального хвоста. Если бы не одно маленькое «но»...

### Замечание 8

Поскольку:

1) класс правильно меняющихся функций не содержит моментов выше  $\alpha$ , а конкретнее

- если  $p > \alpha$ , то  $\mathbb{E}X^p = \infty$ , когда  $p$  четно или случайная величина  $X$  определена на области с одним хвостом, и
- момент  $\mathbb{E}X^p$  не определен, когда  $p$  нечетно и случайная величина  $X$  определена на области с двумя хвостами,

и

2) распределения вне класса правильно меняющихся функций содержат все моменты:  $\forall p \in \mathbb{N}^+, \mathbb{E}X^p < \infty$ ,

$\exists p \in \mathbb{N}^+$  такое, что момент  $\mathbb{E}X^p$  не определен или бесконечен  $\Leftrightarrow X \in \mathfrak{P}$ .

Изучим теперь способы обнаруживать «бесконечные» моменты. С этой задачей связана путаница, поскольку моменты конечной выборки по определению конечны и измеримы, в вероятностном пространстве считающей меры<sup>1</sup>.

Мы будем обращать внимание на несходимость моментов. Пусть  $\|X\|_p$  — взвешенная  $p$ -норма:

$$\|X\|_p \triangleq \left( \frac{1}{n} \sum_{i=1}^n |X_i|^p \right)^{\frac{1}{p}}.$$

По свойству степенных законов

$$\mathbb{E}X^p \not\prec \infty \Leftrightarrow \|X\|_p \text{ расходится.}$$

**Вопрос.** Каким образом принадлежность классу хвостов степенного закона (с показателем  $\alpha \leq 4$ ) отменяет многие методики из  $L^2$ ?

В Разделе 5.10 показано распределение среднего отклонения второго момента для степенного закона с конечной дисперсией. Говоря просто, в ситуации, когда не существует четвертый момент и бесконечны высшие моменты, у дисперсии ее второй момент сам имеет бесконечную дисперсию, создавая прежние проблемы с попытками выборочной оценки: если в случае степенного закона при  $\alpha$  больше 1, но ненамного, среднее существует, но так и не будет получено из наблюдений, то в ситуации бесконечного третьего момента наблюдаемый второй момент так и не станет информативен, потому что так и не сойдется к своему истинному значению.

## 10.2. КРИТЕРИИ СХОДИМОСТИ

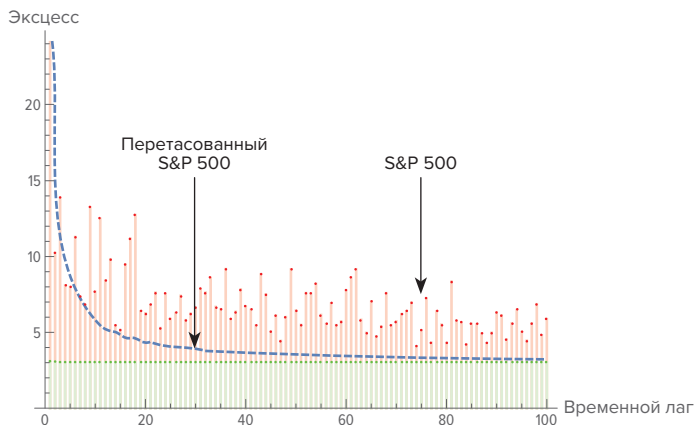
Законы сходимости помогают *исключить* некоторые классы распределений вероятности.

### 10.2.1. Критерий 1: эксцесс при агрегации

Если бы эксцесс существовал, то по мере удлинения временного окна он сходил бы к гауссову эксцессу. Поэтому мы проверяем расчетные доходности за все более долгие временные лаги, как показано на Рисунке 10.2.

**Результат.** Вердикт по Рисунку 10.2 тот, что эксцесс за месяц не снизился по сравнению с ежедневным, и вообще добавление данных не обваливает эксцесс. В дальнейшем можно ожидать спада  $\sim n^{-1}$ . Можем смело исключить целый ряд классов, включая стохастическую волатильность простого формата, например гамма-дисперсию. Далее мы рассмотрим технические детали вопроса и доказательную силу таких наблюдений.

<sup>1</sup> Считающая мера множества  $A$  определяется как  $\mu_A = \begin{cases} |A|, & A \text{ конечно} \\ +\infty, & \text{иначе} \end{cases}$ . В классической теории вероятностей рассматривалось конечное множество  $\mathcal{F}$  всех возможных элементарных событий, которые считались равновероятными, и распределение вероятности случайной величины  $X$  понималось как  $\mathbb{P}(X < x) = \frac{\mu\{a \in \mathcal{F} \mid X(a) < x\}}{\mu \mathcal{F}}$ . В современной теории вероятностей, по Колмогорову, ту же формулу применяют в любом измеримом.



**Рисунок 10.2:** Повторим изображение, которое уже давали в Главе 7: визуальную диагностику сходимости для эксцесса S&P 500 по последним 1700 наблюдениям. Мы вычисляем эксцесс при разном выборе временного лага сначала для исходного S&P 500, затем для перетасованных данных. В то время как для исходных данных 4-я норма не сходится, для перетасованного ряда она несомненно сходится. Отсюда понятно, что жирнохвостьность связана с временной структурой данных, а именно с кластерами волатильности. Смотрите Таблицу 7.1, где показано ожидаемое падение со скоростью  $\frac{1}{n}$  для тонкохвостых распределений

Вот типичное недоразумение. В заметке «Чему Талеб может научиться у Марковица» [255] Джек Л. Трейнор, один из основателей портфельной теории, защищает свою область тем доводом, что даже если данные кажутся жирнохвостыми «на короткой дистанции», на некоей «долгой дистанции» они еще станут гауссовыми. Увы, не станут. (Мы добавим проблему с эргодичностью, которая сгладит, если не устранил вовсе различие между долгой дистанцией и короткой дистанцией.)

Причина в том, что мы не можем надеяться на гауссовость, если эксцесс бесконечен, даже если низшие моменты существуют. Кроме того, при  $\alpha \approx 3$  центральная предельная теорема работает медленно, и какая-то надежность наступит при  $n$  порядка  $10^6$ , недоступном в существующей хронологии рынков [27].

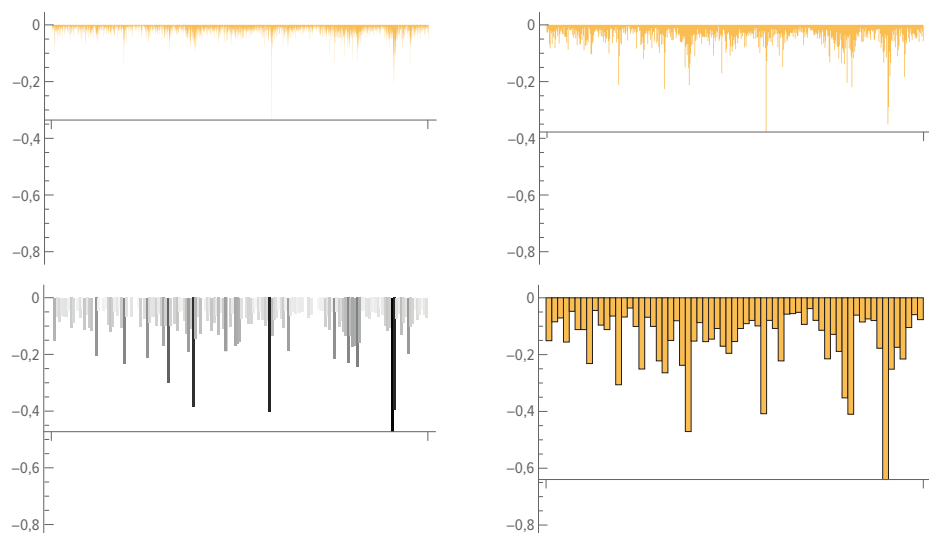
### 10.2.2. Максимальные падения

Пусть для некоторого временного ряда оценок актива  $S$ , сделанных в моменты  $(t_0, t_0 + \Delta t, \dots, t_0 + n\Delta t)$ , нас интересует поведение величины

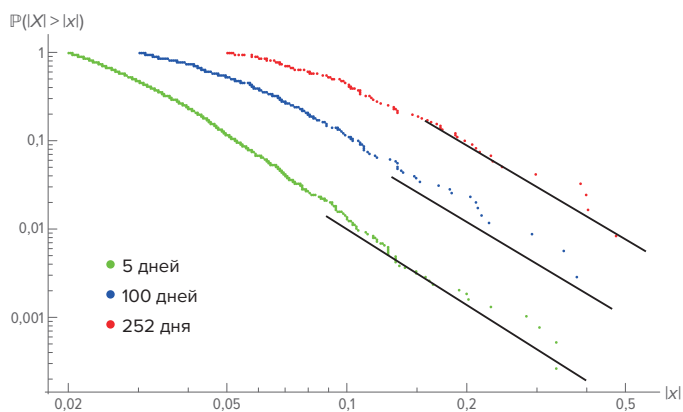
$$\delta(t_0, t, \Delta t) = \max_{i=0}^n \left( S_{t_0+i\Delta t} - \min_{j=i+1}^n S_{t_0+j\Delta t} \right)^1. \quad (10.1)$$

Мы можем рассмотреть относительное падение, взяв логарифм этой величины, как мы поступаем с доходностью. Окно  $n$  для наблюдения падений может быть 5, 100, 252 дня. Как видно на Рисунке 10.4, падения распределены по Парето.

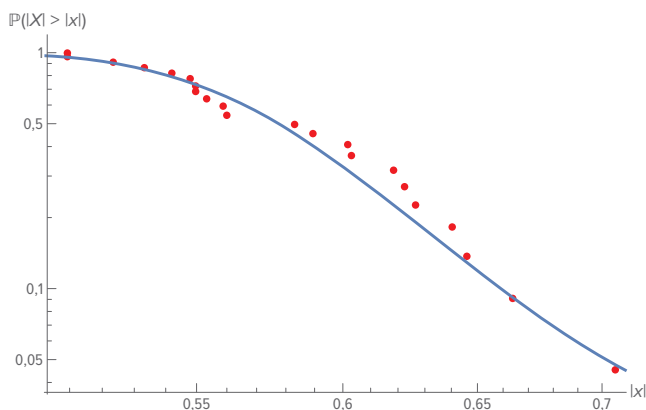
<sup>1</sup> По-видимому, здесь  $t = t_0 + n\Delta t$ .



**Рисунок 10.3:** Падения в окнах  $n = 5, 30, 100$  и  $252$  дня. Максимальные падения — это отклонения, которые отслеживаются формулой 10.1. Мы используем здесь логарифм минимального  $S$  в окне  $n$  дней после того, как было достигнуто заданное значение  $S$



**Рисунок 10.4:** Принадлежность падений классу Парето и масштаб



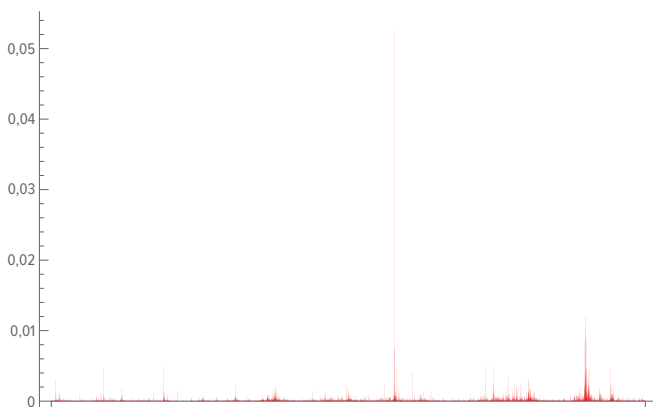
**Рисунок 10.5:** Подгонка устойчивого распределения к эмпирическому распределению падений

### 10.2.3. Эмпирическая каппа

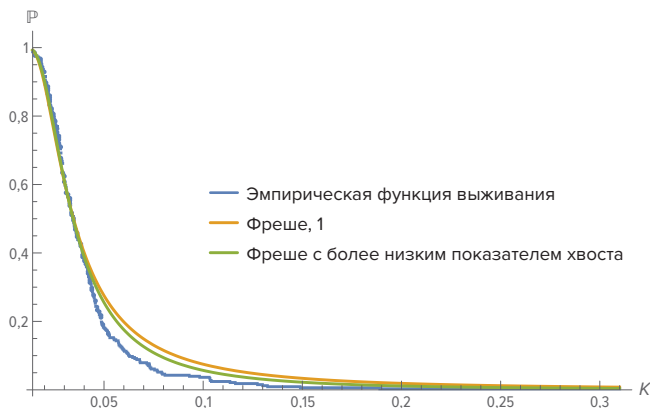
Бутстрэпом<sup>1</sup> по нашей формуле для каппы из главы 8:

$$\kappa_{n_0, n} = 2 - \frac{\ln n - \ln n_0}{\ln \frac{\mathbb{M}(n)}{\mathbb{M}(n_0)}}, \quad (10.2)$$

— получаем такие эмпирические оценки (обозначая для краткости  $\kappa_n = \kappa_{1, n}$ ):



**Рисунок 10.6:** Квадрат доходов по 16 500 наблюдениям S&P 500. Никакой GARCH (1, 1) не воспроизведет подобную зазубренность (великий человек Бенуа Мандельброт называл такие графики «рюгоситэ»<sup>2</sup>)



**Рисунок 10.7:** Каппа- $n$  по эмпирическим оценкам

Видно, что эмпирическая каппа распределена по степенному закону и что для отрицательных доходов  $\alpha < 3$ .

<sup>1</sup> Бутстрэп (англ. *bootstrap*) в статистике — численное исследование распределения по данной выборке путем многократной генерации подвыборок методом Монте-Карло.

<sup>2</sup> Фр. *rugosité*, шероховатость.

### 10.2.4. Проверка 2: условное математическое ожидание избытка

**Результат:** вердикт по этому тесту тот, что, как видно на Рисунке 10.8, условное математическое ожидание  $X$  (а также  $-X$ ) при условии избытка над  $K$  остается пропорциональным  $K$  даже при больших  $K$ .

#### Определение 10.1

Пусть  $K$  из  $\mathbb{R}^+$  — условное математическое ожидание относительного избытка:

$$\varphi_K^+ \triangleq \frac{\mathbb{E}(X | X > K)}{K},$$

$$\varphi_K^- \triangleq \frac{\mathbb{E}(X | X > K)}{K}.$$

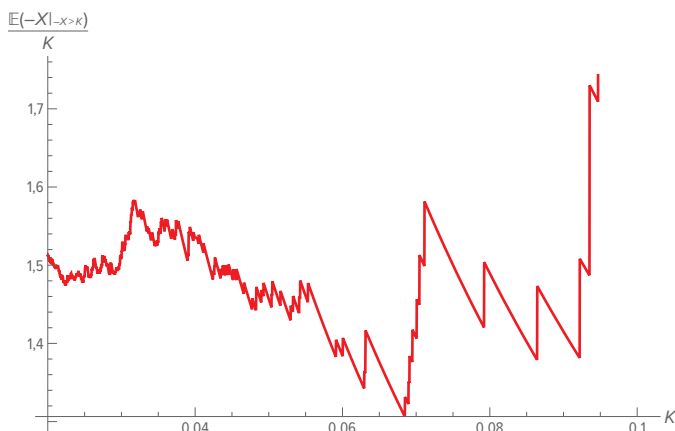
Тогда для распределений вне бассейна степенных законов

$$\lim_{K \rightarrow \infty} \varphi_K = 0,$$

а для распределений из Определения 1

$$\lim_{K \rightarrow \infty} \varphi_K / K = \frac{\alpha}{1 - \alpha}.$$

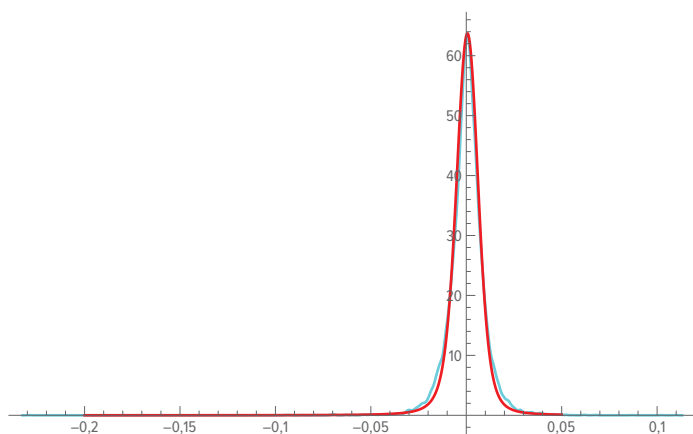
Обратите внимание на закон<sup>1</sup> Ван дер Вейка.



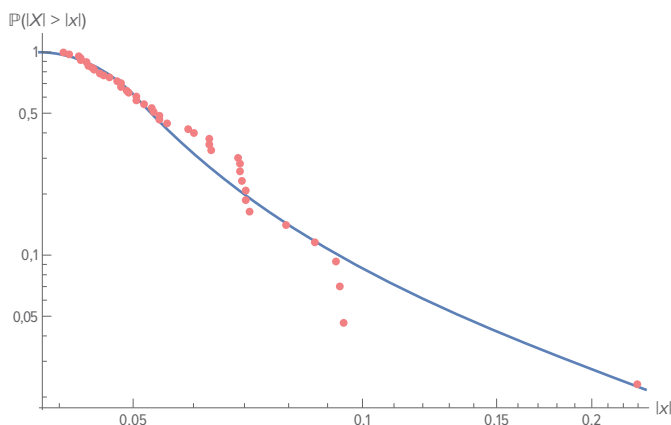
**Рисунок 10.8:**  
«Тест Линди», или  
Condexp<sup>2</sup> — проверяется  
масштабируемость,  
используя условное мате-  
матическое ожидание  
ниже  $K$ : при больших  $K$   
оно должно упасть

<sup>1</sup> Ключевое свойство распределения Парето, т. н. принцип 80/20: что 80 % богатств сосредоточены в руках 20 % семейств и что среди этих богатейших семейств 80 % их богатств сосредоточены в руках 20 % богатейших среди богатейших, и так далее, — можно обобщить для произвольного распределения Парето с показателем  $\alpha > 1$ : существует доля  $p \leq \frac{1}{2}$  богатейших семейств такая, что в их руках сосредоточено  $1 - p$  богатств, и среди этих  $p$  семейств богатейшие  $p^2$  семейств сосредоточено в своих руках  $(1 - p)^2$  богатств, и так далее, причем  $\frac{\alpha}{\alpha - 1} = \frac{\ln p}{\ln(1 - p)} = \frac{\ln p^2}{\ln(1 - p)^2} = \dots$ . Общий вариант иногда называют законом Ван дер Вейка.

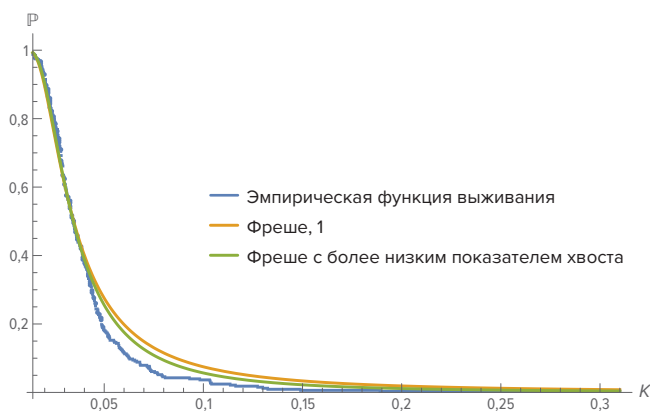
<sup>2</sup> От *conditional expectation*, условное математическое ожидание.



**Рисунок 10.9:** Под эмпирическое распределение убедительно подгоняется устойчивое по Леви при  $\alpha_l = 1,62$



**Рисунок 10.10:** Хвосты подгоняются даже под устойчивое распределение с бесконечным средним, при  $\alpha_l = 1$



**Рисунок 10.11:** Корректировка эмпирического распределения при помощи Фреше для S&P 500

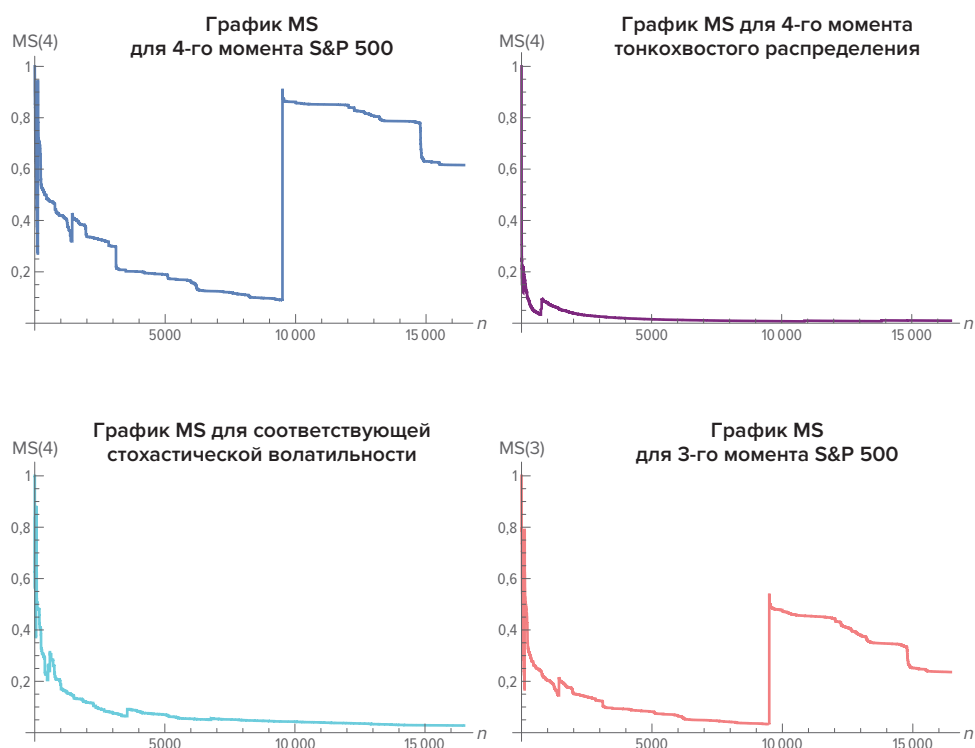
Рисунок 10.8 показывает, что условное ожидание не падает при больших значениях порога, что несовместимо с распределениями, отличными от Парето.

### 10.2.5. Проверка 3: неустойчивость 4-го момента

Главный довод в [228] тот, что за 50 лет наблюдений S&P 500 одно-единственное наблюдение создает  $> 80\%$  эксцесса. Аналогичный эффект виден и для других социально-экономических показателей фондовой биржи, таких как цены на золото, нефть, серебро и софты<sup>1</sup>. Такая зависимость эксцесса от выборки означает, что четвертый момент неустойчив, не существует.

### 10.2.6. Проверка 4: график MS

Дополнительный способ узнать, существует ли  $\mathbb{E}X^n$ , — исследовать сходимость по закону больших чисел (или ее отсутствие), изучив поведение высших моментов в данной выборке. Удобно пользоваться графиком отношения максимума к сумме (*maximum to sum*, MS), показанным на Рисунке 10.12. График MS опирается на последовательность в законе больших чисел [184] с того места, где встретилось максимальное



**Рисунок 10.12:** Сравним график  $MS(p)$ , или «закон больших чисел для  $p$ -тых моментов» при  $p = 4$  для S&P 500 и для гауссианы и стохастической волатильности при таком же эксцессе ( $\approx 30$ ) за период наблюдений. Сходимость, если она вообще есть, не наступает ни за какое разумное время. График MS при  $p = 3$  для S&P 500 сравним с  $p = 4$  для гауссианы. Можно смело утверждать, что 4-й момент бесконечен и 3-й момент неопределенный

<sup>1</sup> Они же «мягкие товары» (*soft commodities*). Это те товары на фьючерсных рынках, которые не добываются, а выращиваются: сахар, кукуруза, хлопок, кофе, какао, зерно, апельсины.



значение случайной величины. Если для последовательности неотрицательных н. о. р. случайных величин  $X_1, X_2, \dots, X_n$  моменты  $\mathbb{E} X^p < \infty$  при  $p = 1, 2, 3, \dots$ , то

$$R_n^p = \frac{M_n^p}{S_n^p} \xrightarrow{\text{a.s.}} 0 \text{ при } n \rightarrow \infty,$$

где  $S_n^p = \sum_{i=1}^n X_i^p$  — частичная сумма, а  $M_n^p = \max(X_1^p, \dots, X_n^p)$  — частичный максимум. (Заметим, что этот подход можно обобщить и на нечетные моменты случайной величины произвольного знака; достаточно заменить  $X$  на абсолютную величину данной случайной величины.)

Покажем для сравнения график MS для гауссовой случайной величины и для t-распределения Стьюдента с показателем хвоста 3. Видим, что S&P 500 демонстрирует типичные характеристики резкого степенного закона и за 16 000 наблюдений (за 50 лет) так и не падает до точки, где может начаться работа закона больших чисел.

### 10.2.7. Рекорды и экстремальные значения

Метод рекордов Гумбеля состоит в следующем (Амбре и др., [82]). Пусть  $X_1, X_2, \dots$  — дискретный временной ряд, в котором максимум за период  $t \geq 2$  определяется как  $M_t = \max(X_1, X_2, \dots, X_t)$ . Обозначим  $N_{1,t}$  число рекордов за первые  $t$  точек данных:

$$N_{1,t} = 1 + \sum_{k=2}^t 1_{X_k > M_{k-1}}. \quad (10.3)$$

Независимо от распределения случайных величин  $X_k$ , математическое ожидание  $\mathbb{E} N_t$  равно  $t$ -му гармоническому числу<sup>1</sup>  $H_t = \sum_{i=1}^t \frac{1}{i}$ , а дисперсия равна разности  $H_t - H_t^2$ . Заметим, что последовательность гармонических чисел отличается выпуклостью вверх и медленным, логарифмическим ростом, описываемым приближенной формулой  $\ln t + \gamma$ , где  $\gamma$  — постоянная<sup>2</sup> Эйлера — Маскерони. Приближение таково, что  $\frac{1}{2(t+1)} \leq H_t - \ln t - \gamma \leq \frac{1}{2t}$  (сайт *Wolfram Mathworld* [263]).

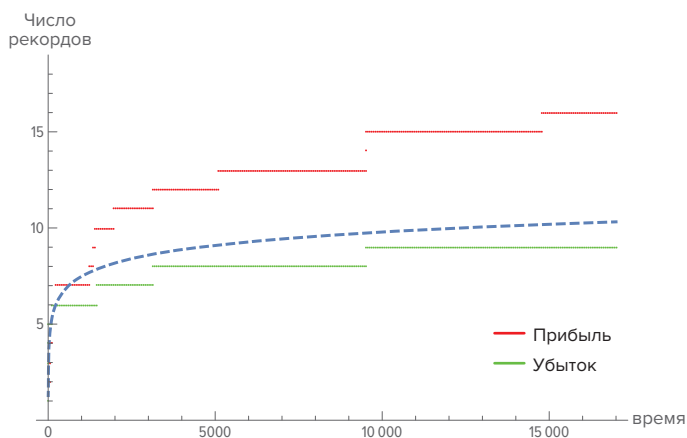
#### Замечание 9

Описанный выше критерий независимости по Гумбелю — достаточное условие того, что экстремальные отрицательные значения логарифмической доходности<sup>3</sup> S&P 500 сходятся к MDA распределения экстремальных значений.

<sup>1</sup> То есть частичной сумме гармонического ряда  $1 + \frac{1}{2} + \frac{1}{3} + \dots$  по  $t$ -й член.

<sup>2</sup> Постоянная Эйлера — Маскерони  $\gamma = -\int_0^\infty \frac{\ln x}{e^x} dx \approx 0,577$  введена в 1735 Эйлером и изучена в 1790 Маскерони.

<sup>3</sup> Логарифмическая доходность (*log-return*, *logarithmic return*), она же непрерывная кумулятивная доходность (*continuously compounded return*), она же сложный процент (*force of interest*) определяется для заданного периода времени как  $R_{\log} = \ln \frac{V_e}{V_b}$ , где  $V_e$  и  $V_b$  — стоимость финансового актива соответственно на конец и начало периода. Для произвольного периода  $t$  используется логарифмическая норма доходно-



**Рисунок 10.13:** Проверка по рекордам показывает независимость для экстремальных отрицательных доходов и зависимость для положительных. Число рекордов при данном числе независимых наблюдений растет со временем как гармоническое число  $H_t$  (пунктирная линия),  $\approx$  по логарифмическому закону, но здесь для положительных доходов рост происходит более чем на 2,5 стандартного отклонения быстрее, так что для экстремальных прибылей предположить независимость нельзя. Данный тест не позволяет судить о зависимости вне экстремальных значений

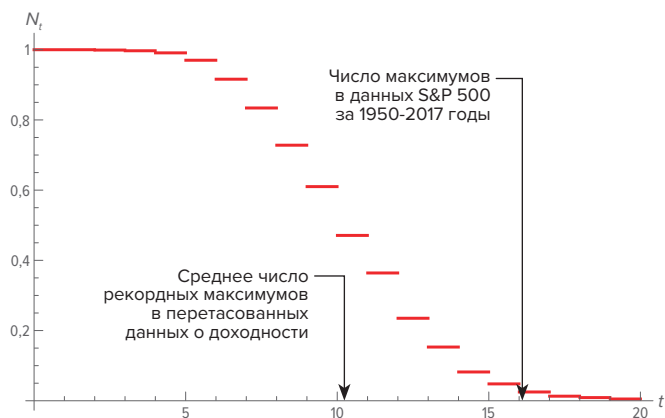
**Ряды целиком.** Мы перетасовали S&P 500 (применили бутстрэп с размером подвыборки, равным всем исходным  $\approx 17\,000$  точек, и запретом выбирать одну точку несколько раз; и так  $10^3$  раз) и проследили число рекордов по всем подвыборкам. Как показано на Рисунках 10.14 и 10.15, среднее составило 10,4 (что отвечает приближению гармоническим числом с соответствующим стандартным отклонением.) Функция выживания  $S(\cdot)$  при числе рекордов  $N_{1,7 \times 10^4} = 16$  принимает значение  $S(16) = \frac{1}{40}$ , что позволяет отбросить гипотезу о независимости положительных экстремальных значений.

С другой стороны, число отрицательных экстремальных значений составило 9, что отличается от математического ожидания 10,3 всего на  $\frac{1}{2}$  стандартного отклонения, и отвергнуть независимость не удастся.

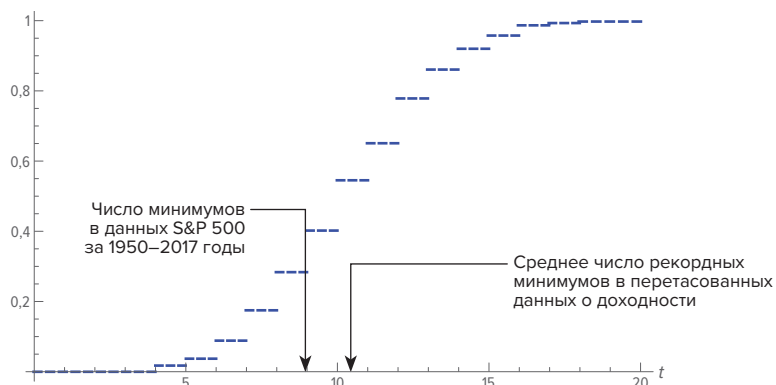
**Подрекорды.** Если обрабатывать данные не единым блоком за весь период  $T$ , а разбить его на подпериоды, получим (благодаря выпуклости вверх меры и неравенству Йенсена)  $T/\delta$  наблюдений числа подрекордов вида  $N_{t_1+\delta, t_1+\delta+\Delta}$ . Возьмем  $\Delta = 10^3$  и  $\delta = 10^2$ , и из исходных  $T \approx 17 \times 10^3$  дней получим 170 подпериодов. Результат показан на Рисунке 10.16; он не позволяет отвергнуть независимость ни для положительных, ни для отрицательных подрекордов.

---

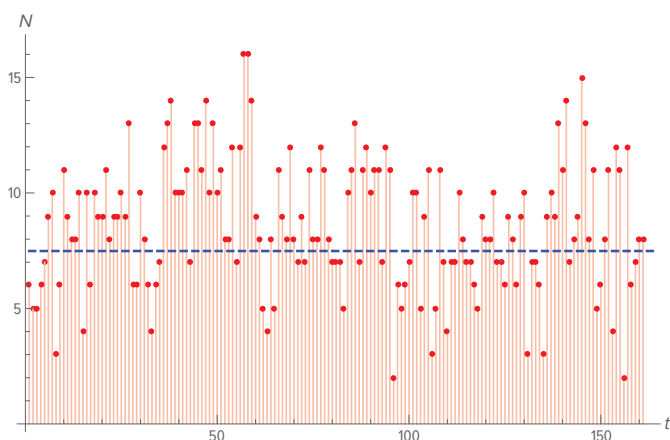
сти (logarithmic rate of return)  $r_{\log} = \frac{\ln V_e / V_b}{t}$ . Этот показатель постоянный, пока стоимость финансового актива растет пропорционально текущей стоимости, то есть по экспоненциальному закону  $V_e = V_b \exp r_{\log} t$ .



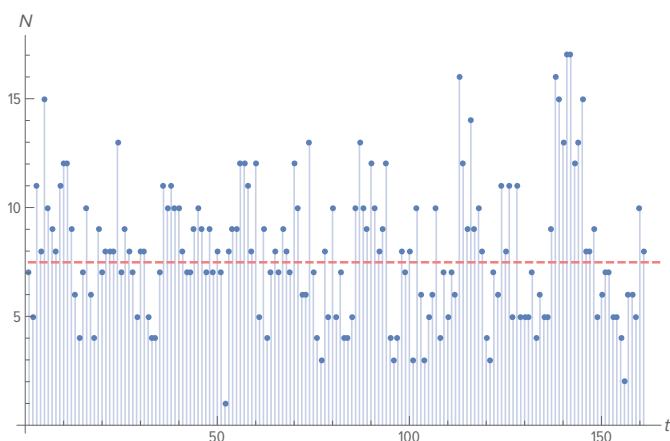
**Рисунок 10.14:** Функция выживания для числа рекордов — положительных максимумов, полученных из  $10^3$  перетасованных (и тем самым лишенных временной структуры) полных выборок данных S&P 500. Значение выше 16 (столько рекордных максимумов наблюдалось в данных S&P 500 за всю историю индекса) встретилось всего в  $\frac{1}{40}$  полных выборок



**Рисунок 10.15:** Кумулятивное распределение рекордных отрицательных экстремумов для перетасованного (103 раза) S&P 500, как на предыдущем рисунке. Масса выше 9 (наблюдавшееся число рекордных минимумов SP500 за этот период) составила  $\frac{2}{5}$



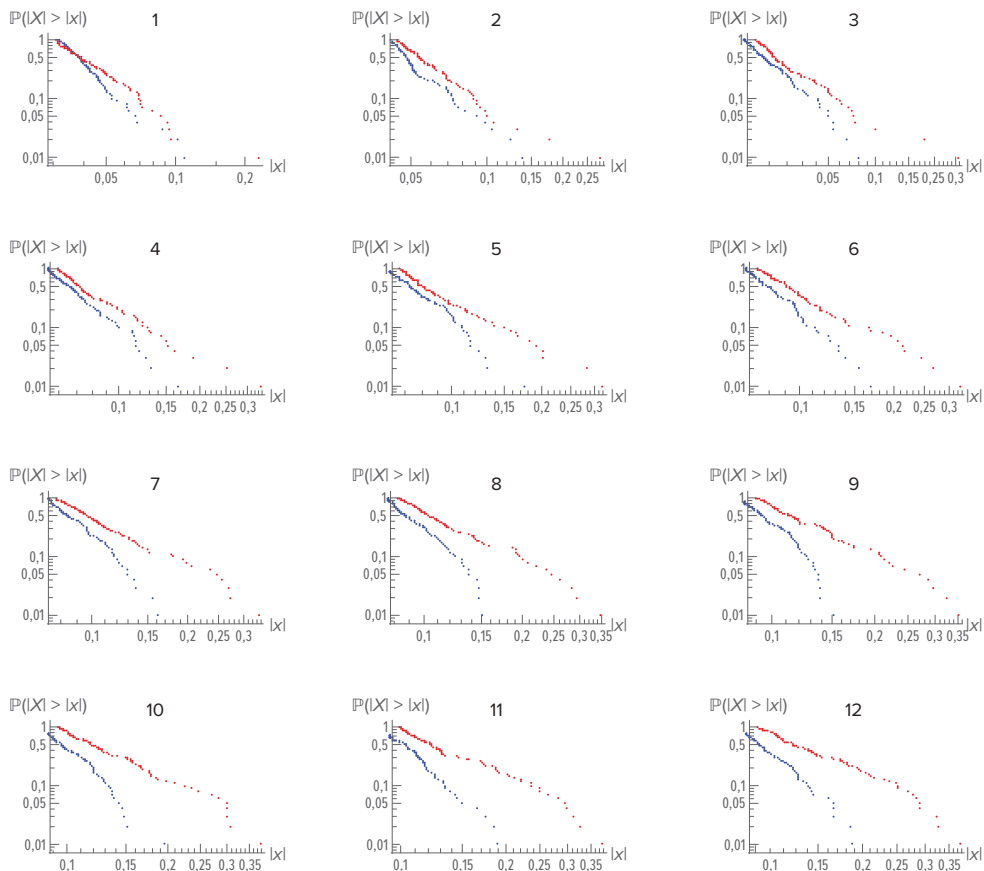
**Рисунок 10.16:** Используем короткий период  $t = 1000$  дней регистрации рекордных максимумов (график вверх) и минимумов (внизу) и выполним перекрывающиеся наблюдения. Математическое ожидание, то есть гармоническое число  $H_{1000}$ , показано пунктиром



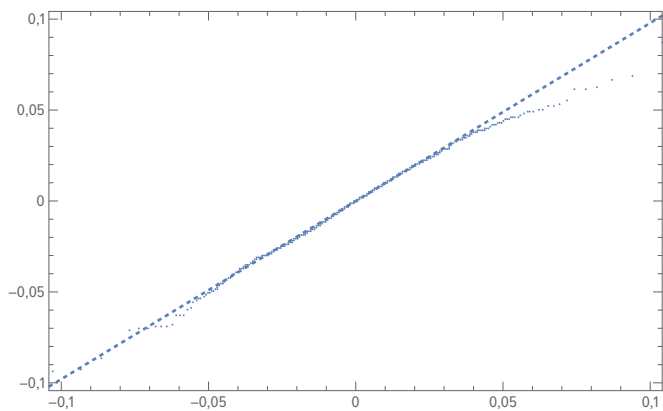
**Выводы из подрекордов.** По крайней мере для отрицательных рекордов на коротких подпериодах применимы методы EVT.

### 10.2.8. Асимметричность хвостов справа и слева

На Рисунке 10.17 замечаем асимметрию: левый хвост жирнее правого. Это будет головной болью для тех, кто пытается строить точные модели процесса, но вряд ли важно для тех, кто занимается рисками или торгует опционами.



**Рисунок 10.17:** Представим отдельно положительные (синим) и отрицательные (красным) логарифмические доходности и используем перекрывающиеся накопленные доходности от 1 до 12. Видно, что отрицательные доходности следуют степенному закону, тогда как принадлежность правого хвоста классу Парето остается сомнительной



**Рисунок 10.18:**  $t$ -распределение Стьюдента против эмпирического распределения S&P 500 в координатах  $Q - Q^*$ : левый хвост подгоняется, правый — нет

\* График в координатах квантиль — квантиль (*quantile — quantile*).

### 10.3. ЗАКЛЮЧЕНИЕ: ПЕЧАЛЬНО, НО ФАКТ

В этой главе нам удалось исследовать простую тему: доходность по индексу S&P 500 (охватывающему львиную долю капитализации фондового рынка США) распределена по степенному закону, а любые попытки подбирать для нее другие модели нарушают принцип линейки Витгенштейна. Принятые стандартные подходы — современная портфельная теория (*Modern Portfolio Theory*, MPT) и демагогия о якобы присущей людям склонности переоценивать вероятность хвостовых событий относительно «базовой частоты катастроф» (*base crash rate*) — сплошная лженаука, хотя вышло уже > 70 000 статей и сформировались целые армии исследователей, даже не считая  $10^6$  общеэкономических статей, результаты которых основаны на «дисперсии» и «корреляции». Придется свыкнуться с тем, что эти показатели — туфта. Как говорили древние, *dura lex sed lex*<sup>1</sup>, а на современном жаргоне:

Печально, но факт.

---

<sup>1</sup> «Закон суров, но это закон» — изречение древнеримского юриста Ульпиана, III век н. э. Автор перефразирует как *It is what it is*.

## Е ПРОБЛЕМА С ЭКОНОМЕТРИКОЙ

Неладно что-то в эконометрике: ее статьи почти никогда не отражают реальность. Две проверки надежности в Главе 10: одна для параметрических методов, другая для робастной статистики, — показывают, что вся область фундаментально прогнила, что ее методы ненадежны и не годятся ни для каких рискованных решений. От прикладников льются потоки заявлений о причинах провалов — противоречивые, поверхностные. Это резюме результатов в [228], без технических деталей.

Экономические переменные устроены так, что одно-единственное наблюдение из 10 000 — за один день из 40 лет — отвечает в значительной степени за величину «эксцесса», стандартной характеристики «жирных хвостов» для распределений с конечными моментами. Тем самым оно одновременно характеризует и отклонение изучаемого распределения от стандартного гауссова, и влияние событий в далеком хвосте на свойства распределения в целом. Для фондового рынка США один-единственный день, крах 1987 года, определил 80 % эксцесса за период с 1952 по 2008 год. Та же проблема обнаруживается с процентными ставками, обменными курсами, ценами на товары и другими величинами. Повторяя исследование на других периодах с другими величинами, видим вопиющую неустойчивость эксцесса. Проблема не столько в том, что данные содержат «жирные хвосты», а исследователи о них то и дело забывают; проблема в том, что узнать «величину жирнохвостости» стандартными методами невозможно. В принципе невозможно<sup>1</sup>.

<sup>1</sup> Макроэкономические величины, такие как число заявок на пособия по безработице в США в неделю, традиционно казались отслеживаемыми в стенах (уродливых и унылых) экономических департаментов. В итоге основанные на них модели разлетелись вдребезги. С началом эпидемии нового коронавируса Covid-19 число заявок на пособия по безработице претерпело «неожиданные» скачки в «тридцать среднеквадратических отклонений», а эксцесс (логарифм изменений) подскочил с 8 до > 550 после одного-единственного наблюдения в апреле 2020. Чуть ли не все выборочные высшие моменты определяются одной-единственной точкой данных, и чем выше момент, тем сильнее этот эффект — приходится признать, что высшие моменты не существуют, и нет информативного низшего момента, и величину следует признать распределенной по степенному закону.

Такая роль хвоста отменяет всю историю макроэкономического моделирования, как и политические курсы, основанные на заключениях экономистов, которые пользовались характеристиками из Медиокристана. И даже если экономисты в кругах взаимного цитирования еще не осознали, что гнали туфту, до остальных уже дошло. На момент написания этой книги специалисты начали понимать, что хвосты



Рисунок Е.1: © Stefan Gasic

Следствие то, что для ряда инструментов, основанных на возведении величин в квадрат (более строго, на использовании нормы  $L^2$ ), включая такие вещи из учебников, как среднеквадратическое отклонение, дисперсия, корреляция и регрессия, — для всех этих вещей применение в экономике научно несостоятельно (за редкими исключениями, где величина ограничена). Так называемые « $p$ -значения», приводимые в исследованиях как оценка значимости, бессмысленны для экономических и финансовых величин. Даже более сложные методы стохастического исчисления, применяемые в финансовой математике, в экономике не работают, за исключением нескольких узких областей.

### Е.1. ЭФФЕКТИВНОСТЬ СТАНДАРТНЫХ ПАРАМЕТРИЧЕСКИХ ОЦЕНОК РИСКА

Результаты большинства статей в экономике основаны на этих стандартных статистических методах, и, следовательно, трудно ожидать от них воспроизводимости. Ее, собственно, и нет. И это те инструменты, которыми обосновывают призывы к большим рискам. Альтернативные методы тоже не дают надежных характеристик для редких событий — разве что позволяют иногда выяснить, что события в далеком хвосте недооценены; своей оценки эти методы не предлагают.

Как описано в [228], рассмотрим логарифмическую доходность

$$X_{t, i\Delta t} \triangleq \ln \frac{P(t)}{P(t-i\Delta t)}$$

и максимальный биквадрат по выборке из  $n$  наблюдений с шагом  $\Delta t$

$$\text{Max}_{i=0}^n X_{t, i\Delta t}^4.$$

Пусть  $Q(n)$  — вклад максимального биквадратного отклонения:

$$Q(n) \triangleq \frac{\text{Max}_{i=0}^n X_{t, i\Delta t}^4}{\sum_{i=0}^n X_{t, i\Delta t}^4}.$$

---

весьма жирны, и политика должна основываться на экстремальных значениях, пользуясь теорией экстремальных значений (*extreme value theory*, EVT), и что пропасть между гауссовыми моделями и моделями степенного закона еще глубже, когда рассматриваются экстремальные значения. — Прим. автора.



Заметим, что для целей данного исследования практически все равно, какой взять эксцесс — центральный или нецентральный.

Для Гауссова случая величина  $Q$  будет распределена как квадрат хи-квадрата, и для суммы  $10^4$  наблюдений вклад максимального,  $Q(10^4)$ , должен был составить около  $0,008 \pm 0,0028$ . Мы же видим, что фактическое распределение 4-го момента, напротив, таково, что

$$\mathbb{P}(X > \text{Max}_{i=1}^n x_i^4) \approx \mathbb{P}\left(X > \sum_{i=1}^n x_i^4\right)^1.$$

Вспомним наивное представление о четвертом моменте как выразителе устойчивости второго момента и о втором моменте — как выразителе устойчивости исследуемой характеристики от выборки к выборке.

Заметим, что снимок за другой период времени покажет, что экстремальные значения создаются другими величинами, тогда как эти величины, показывавшие высокие максимальные значения эксцесса, упадут — простое следствие неустойчивости данной характеристики для разных рядов и разных периодов времени.

**Описание набора данных.** Все макроданные относятся к реально торгуемым товарам на август 2008, то есть к таким, для которых есть фактическая цена закрытия, и она соответствует проведенным транзакциям — и тем самым установлена рынком, а не бюрократическими оценками. Товары здесь охватывают процентные ставки, валюты и фондовые риски.

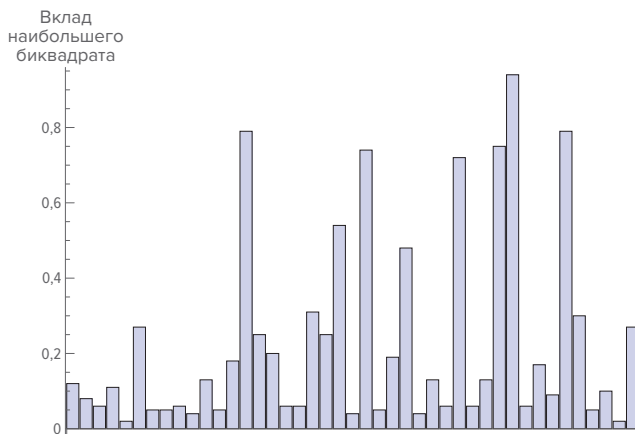


Рисунок Е.2: Наибольшие биквадраты среди ценных бумаг из Таблицы Е.1

1 То есть вклад максимального слагаемого не то что не близок к 0,8 %; он сравним со всей суммой.

**Таблица Е.1:** Максимальный вклад в четвертый момент, внесенный наблюдением за единственным днем

Ценные бумаги	Макс. $Q$	Годы
Серебро	0,94	46
Индекс S&P 500	0,79	56
Сырая нефть	0,79	26
Короткий стерлинг <sup>1</sup>	0,75	17
Топочный мазут	0,74	31
Индекс Никкей	0,72	23
Индекс FTSE <sup>2</sup>	0,54	25
Японские гособлигации <sup>3</sup>	0,48	24
1-месячный евродоллар <sup>4</sup>	0,31	19
Сахар-сырец	0,3	48
Иена	0,27	38
Индекс Bovespa <sup>5</sup>	0,27	16
3-месячный евродоллар <sup>6</sup>	0,25	28
СТ	0,25	48
Индекс Дакс <sup>7</sup>	0,2	18



**Рисунок Е.3:** Эксцесс по неперекрывающимся периодам для долларовых депозитов за пределами США

1 Стерлинговый курс покупки краткосрочных векселей.

2 Англ. *Financial Times Stock Exchange Index* — индекс, рассчитываемый газетой *Financial Times* совместно с Лондонской фондовой биржей.

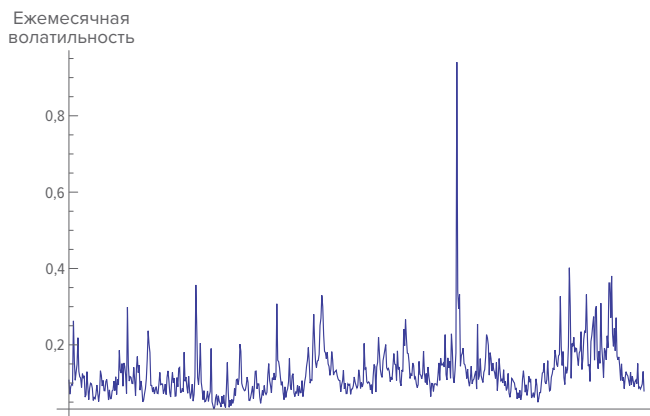
3 Облигации государственного займа Японии.

4 Депозиты в долларах, помещенные в банки за пределами США.

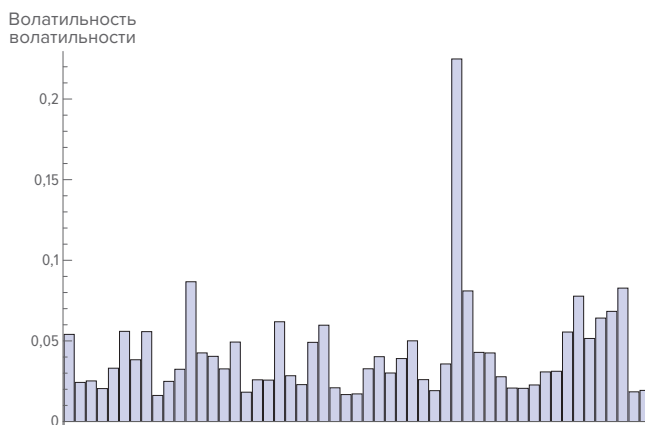
5 Англ. *Bolsa de Valores de Sao Paulo* — фондовая биржа Сан-Паулу.

6 Евродоллары на 3 месяца.

7 Нем. *Deutsche Aktien Index, DAX* — индекс акций на Франкфуртской фондовой бирже.



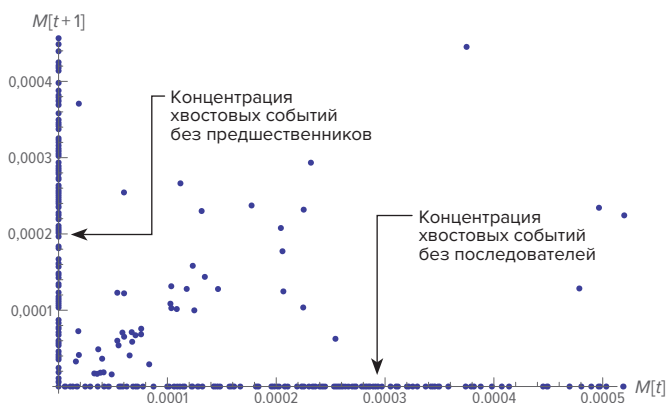
**Рисунок Е.4:** Волатильность S&P 500 за месяцы (в среднеквадратических отклонениях). Кроме ограничения нулем, никакой структуры в данных не видно. Это стандартное поведение



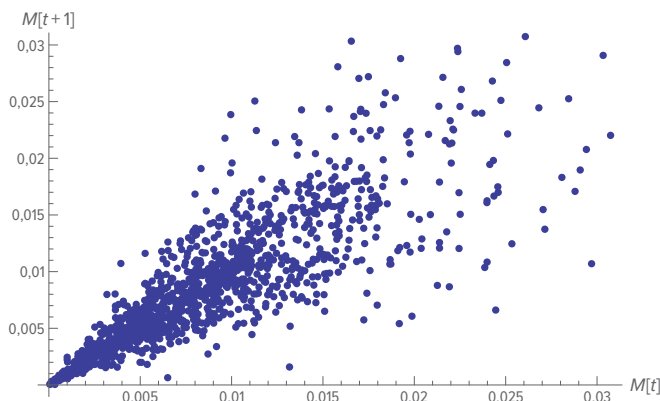
**Рисунок Е.5:** Ежемесячная волатильность волатильности по тем же данным из Таблицы Е.1. Как можно было ожидать, никакой устойчивости

## Е.2. ЭФФЕКТИВНОСТЬ СТАНДАРТНЫХ НЕПАРАМЕТРИЧЕСКИХ ОЦЕНОК РИСКА

Есть ли сходство между прошлыми и будущими хвостовыми событиями? Приведенные ниже проверки непараметрические, то есть основаны только на эмпирических распределениях вероятности.



**Рисунок Е.6:** Абсолютное отклонение  $M[t]$  стоимости акций сравнивается со следующим абсолютным отклонением  $M[t+1]$ , если превышен заданный порог 4 %; видно, что большие отклонения редко когда имеют предшественников или последователей — и так на протяжении 50 лет



**Рисунок Е.7:** «Правильное» распределение случайной величины — предиктор «правильного» распределения среднего отклонения. Сравним абсолютное отклонение  $M[t]$  со следующим,  $M[t+1]$ , для макроэкономических данных...

Пока что мы оставались в одномерном пространстве. Глядя на многомерные характеристики, такие как ковариационные матрицы, видим, что ситуация только ухудшается. Вернемся к проблеме обработки ошибок модели при оптимизации на основе среднего и дисперсии.

Когда  $x_t$  — векторы в  $\mathbb{R}^N$ , проблемы с чувствительностью к изменениям в ковариационных матрицах делают эмпирически наблюдаемые моменты и условные моменты крайне неустойчивыми. Для вектора хвостовые события гораздо труднее калибровать, и они растут с ростом размерности.

#### **Отклики на данный момент со стороны мейнстримных экономистов и эконометристов**

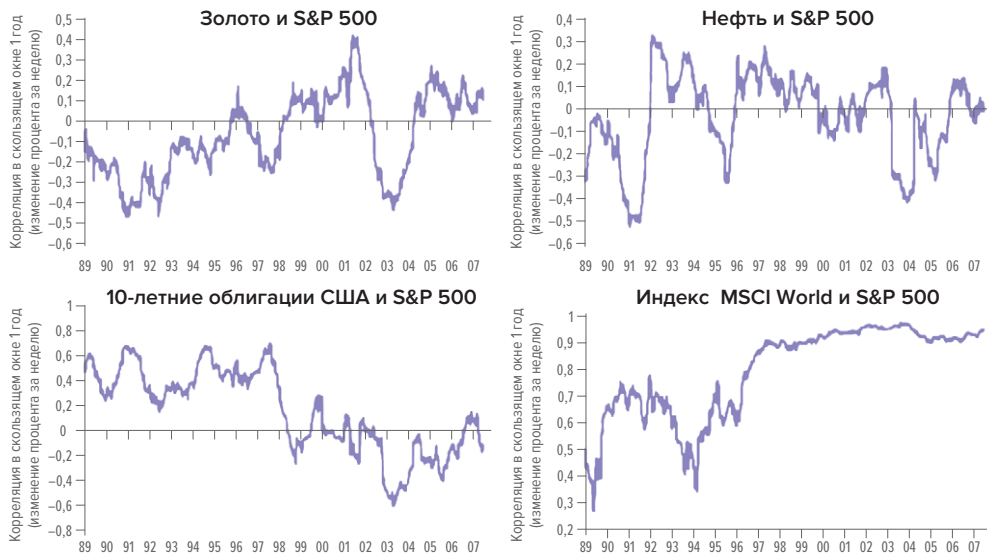
Никаких объяснений, почему они продолжают использовать STD, регрессию, GARCH, стоимость-под-риском и тому подобные методы.

**Проблема песо**<sup>1</sup>. Бенуа Мандельброт говаривал, что под любое распределение можно подогнать распределение Пуассона, если добавлять стохастические скачки частоты. Аналогичным образом по  $n$  точкам данных можно построить многочлен степени  $n-1$ . Если в вашей модели нужно подбирать параметры, она не отвечает распределению по степенному закону.

Многие исследователи ссылаются на «выбросы» и «проблему песо»<sup>2</sup>, как бы признавая жирные хвосты (или важную роль хвостов в распределении), но игнорируют их аналитически — разве что строят модели Пуассона, но их невозможно откалибровать, кроме того случая, когда скачки конвенциональной модели Пуассона тонкохвостые. Мы придерживаемся прямо противоположного подхода: не замечать выбросы под ковер, а поместить в центр построений. Другими словами, подобно тому, как

<sup>1</sup> Кажется, это заголовок от другого замечания. Обычно «проблемой песо» (*Peso problem*) называется разница в рыночной цене якобы похожих финансовых товаров, которая кажется загадочной, если не предположить опасений внезапного обвала. Впервые вопрос был поставлен в 1970-е годы Милтоном Фридманом в отношении процентных ставок на депозиты в банках США и Мексики: они оставались разными на протяжении 20 лет привязки песо к доллару, и была возможность зарабатывать, просто переводя вложения в песо... пока в 1976 он не рухнул на 46 % — как только мексиканское правительство отпустило курс в свободное плавание.

<sup>2</sup> Проблема песо — это обнаружение выброса в финансировании; так стали называть все выбросы и вообще случаи необъяснимого поведения какой-то величины в эконометрике. — *Прим. автора.*



**Рисунок Е.8:** Проблемы есть и с корреляциями; их причиняют как неустойчивость дисперсии отдельных компонент, так и эффект перемножения случайных величин. При такой стохастичности корреляций нет никакого смысла использовать методы на основе ковариации, включая портфельную теорию

FAA<sup>1</sup> и FDA<sup>2</sup>, изучая безопасность, фокусируются на предотвращении катастроф, мы заметем под ковер обычные наблюдения и оставим одни экстремальные значения как единственный надежный способ управления рисками. Это важно не только в вопросах безопасности: в любой области, если аналитика и политика торпедируются хвостовыми событиями, значит, непригодны.

### Проблему песо путают с проблемой черного лебедя:

«Черные лебеди (Талеб, 2007) стали мемом, означающим катастрофы столь редкие, что не поддаются анализу согласно стандарту статистических выводов. Мы находим данную перспективу малополезной, поскольку она предполагает состояние беспомощного невежества и обрекает нас на беззащитность перед ударами неведомого».

Эндрю Лоу (очевидно, не прочитавший книгу, которую цитирует)

**Своя шкура на кону.** Можно удивляться, почему методы эконометрики, оказавшиеся в корне ошибочными, до сих пор в ходу и почему взрослые люди, университетские ученые, участвуют в подобных художествах: учитывают обычные наблюдения и скрывают эффекты высших порядков. Невелика честь воспользоваться редкостью

1 Федеральное управление гражданской авиации США (Federal Aviation Administration).

2 Управление по санитарному надзору за качеством пищевых продуктов и медикаментов в США (Food and Drug Administration).

катастроф и ограничить окно данных так, чтобы предстать во всем белом вопреки принципиальным порокам.

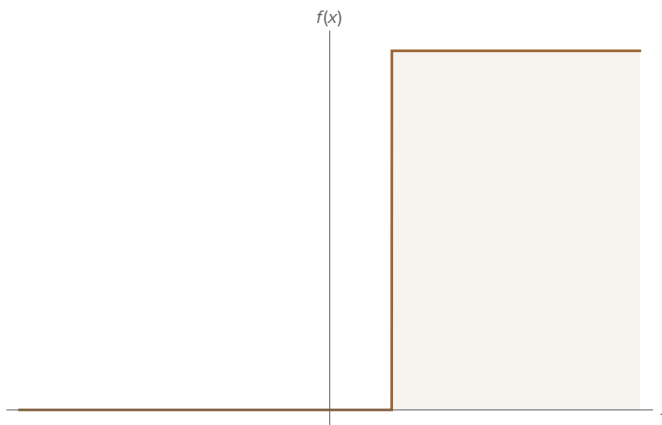
Причина, по которой эта нездоровая ситуация длится и длится, в том, что работчики управления рисками и академические экономисты не рискуют собственной шкурой — они не страдают из-за неверной оценки риска, страдают другие. Пока одним позволено наносить ущерб другим безнаказанно, первые могут продолжать «гнать зайца». (Подробнее смотрите в статье Талеба и Сандиса [246] и в книге Талеба [236].)

## Ф ОСОБЕННОСТИ МАШИННОГО ОБУЧЕНИЯ

Изучая торговлю опционами, мы узнаем, что любую одномерную функцию можно представить линейной комбинацией опционов «колл» и «пут», добавив к опциону временную стоимость. Опцион превращается в строительный блок. Точнее говоря, прибыль, построенную на опционе, можно выразить как  $S = \sum_{i=1}^n \omega_i C(K_i, t_i)$ , где  $C$  — цена (предполагаемая цена) опциона «колл»,  $\omega$  — вес,  $K$  — страйк-цена и  $t$  — время до окончания срока опциона. Для европейского опциона каждый колл принесет прибыль  $\max(S - K, 0)$  в момент окончания  $t^a$ .

С другой стороны, нейросети и нелинейная регрессия, предшественники машинного обучения, фокусировались на функции Хевисайда, сглаженной до S-образной сигмоиды. Под выборку можно подогнать много разных сигмоид.

<sup>a</sup> По-видимому, здесь трейдеры независимо открыли теорему Цыбенко, она же универсальная теорема аппроксимации; первоначально сформулирована для сигмоидных функций, как обсуждается ниже (Цыбенко [52]).

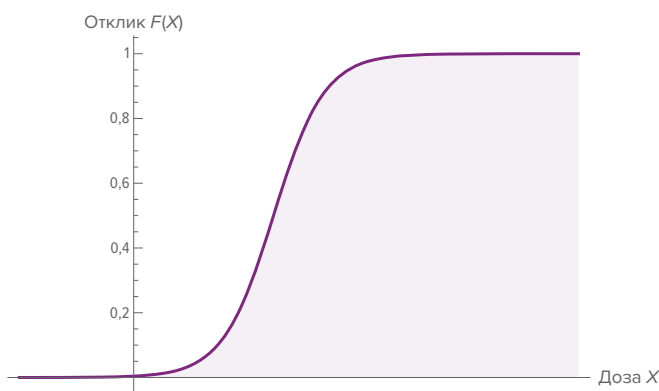


**Рисунок F.1:**  $\vartheta$ -функция Хевисайда. Обратите внимание, что это прибыль от «бинарного опциона», и ее можно представить как  $\lim_{\Delta K \rightarrow 0} \frac{C(K) - C(K + \Delta K)}{\Delta K}$

Таким образом, мы обсуждаем жирнохвостость и возможные строительные блоки для ее моделирования. Машинное обучение в статистике переключилось с агрегаций S-образных кривых на так называемые функции ReLU<sup>1</sup>, они же функции подъема<sup>2</sup>, которые работают в точности как опционы колл. Ученые обнаружили, что благодаря экстраполируемости с такими функциями удобнее обрабатывать хвостовые события вне выборки (внутри выборки неожиданных хвостовых событий нет по определению).

Что такое сигмоида? Рассмотрим платежную функцию на Рисунке F.7. Ее можно приблизить формулой  $S: (-\infty, \infty) \rightarrow (0, 1)$ ,  $S(x) = \frac{1}{2} \operatorname{th} \frac{kx}{\pi} + \frac{1}{2}$ <sup>3</sup> или, точнее, трехпараметрической формулой  $S_i: (-\infty, \infty) \rightarrow (0, a_i)$ ,  $S_i(x) = \frac{a_i}{\exp(c_i - b_i x) + 1}$ <sup>4</sup>. Ту же ступеньку

можно приблизить кумулятивной функцией нормального распределения  $\mathcal{N}(\mu, \sigma)$ , и тогда управлять сглаженностью будет  $\sigma$  (приближая ступеньку к хевисайдовой на Рисунке F.7 в пределе  $\sigma \rightarrow 0$ ). Сигмоида — это ступенька Хевисайда после сглаживания и, возможно, задания границ.



**Рисунок F.2:** Сигмоидная функция — сглаженная ступенька Хевисайда 9. Заметим, что даже сглаженная ступенька ограничена слева и справа областями насыщения

1 Англ. *Rectified Linear Unit, ReLU* — синоним линейно нарастающей функции, который в английской литературе по искусственным нейронным сетям применяется к функции активации нейрона. Дословно «линейный выпрямитель»; противопоставляется сглаженному подъему, например «выпрямителю с гауссовой ступенькой» (*Gaussian Error Linear Unit, GELU*)  $\frac{1}{2}x \left( 1 + \operatorname{erf} \frac{x}{\sqrt{2}} \right) = x\Phi(x)$ , где  $\operatorname{erf}$  — функция ошибок и  $\Phi$  — гауссова функция распределения. — Прим. перев.

2 Англ. *Ramp function* — линейно нарастающая функция, или функция подъема  $R(x) \triangleq \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} = \max(0, x) = x\mathbb{1}_{x>0}$ . Этой функции следует на протяжении одного периода пилообразная (*sawtooth*) функция  $\operatorname{arctg} \operatorname{tg} x$ .

3 В самом деле, сглаженная ступенька гиперболического тангенса  $\operatorname{th} y = \frac{e^y - e^{-y}}{e^y + e^{-y}} = 1 - \frac{2}{e^{2y} + 1}$  дает подъем от  $-1$  до  $1$ , и середина подъема приходится на точку  $y = 0$  с производной  $1$ . В приведенной формуле  $S(x)$  ступенька гиперболического тангенса сжимается по высоте в  $2$  раза и затем сдвигается вверх на  $\frac{1}{2}$ ; кроме того, она сжимается по горизонтали в  $\frac{k}{\pi}$  раз заменой  $y = \frac{kx}{\pi}$ .

4 Здесь, очевидно, сглаженная ступенька гиперболического тангенса преобразована так, что подъем происходит от  $0$  до  $a_i$  и середина подъема приходится на точку  $x = c_i/b_i$  с производной  $a_i b_i/4$ .



Можно строить составные S-образные функции из  $n$  слагаемых  $\chi''(x) = \sum_{i=1}^n \omega_i S_i(x)$ , как в Разделе F.3. Однако:

### Замечание 10

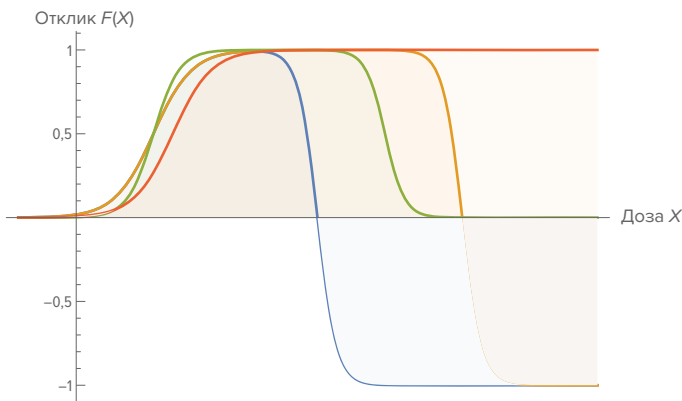
Чтобы значения функции  $\chi_n(x)$  пробегали область  $[0, \infty)$ , или  $(-\infty, 0]$ , или  $(-\infty, \infty)$ , необходимо  $n \rightarrow \infty$ .

Чтобы получить неограниченную функцию, потребуется бесконечно много слагаемых. И повсюду, где «эмпирическое распределение» ограничено максимумом, последнее наблюдение будет соответствовать плоской части сигмоиды. (Определение эмпирического распределения смотрите в Разделе 3.4.)

Рассмотрим платежные функции опциона. На Рисунке F.4 показана прибыль от обычного опциона на момент окончания срока; она по определению соответствует линейному выпрямителю (*Rectifier Linear Unit, ReLU*) в машинном обучении. Функция на Рисунке F.5 получена так. Рассмотрим функцию  $\rho: (-\infty, \infty) \rightarrow [k, \infty)$  с параметрами  $K$  и  $p$ :

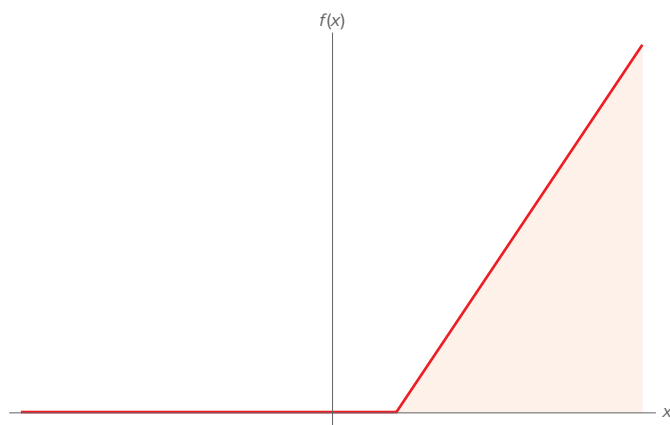
$$\rho_{K,p}(x) = k + \frac{\ln(e^{p(x-K)} + 1)}{p}, \text{ где } K \in \mathbb{R}^1. \quad (\text{F.1})$$

Можно просуммировать несколько таких функций,  $\sum_{i=1}^n \rho_{K_i, p_i}(x)$ , чтобы подогнать под нелинейную функцию, и это, собственно, повторяет то, что мы делали с опционами колл — параметры  $p_i$  позволяют сгладить зависимость величины от времени.

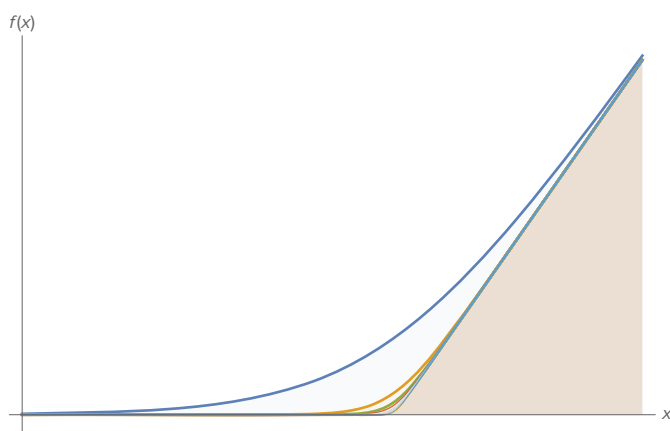


**Рисунок F.3:** Сумма сигмоид всегда ограничена; если нужно смоделировать платежную функцию с открытой областью, не выходящую на насыщение, потребуется бесконечно много слагаемых

<sup>1</sup> Здесь функция подъема сдвинута по вертикали на  $k$ , начало подъема сдвинуто в  $x = K$  и функция сглажена при малых значениях  $p$ . При больших  $p$  функция  $\rho_{K,p}(x) \approx \begin{cases} x, & x > K \\ 0, & x \leq K \end{cases}$ .



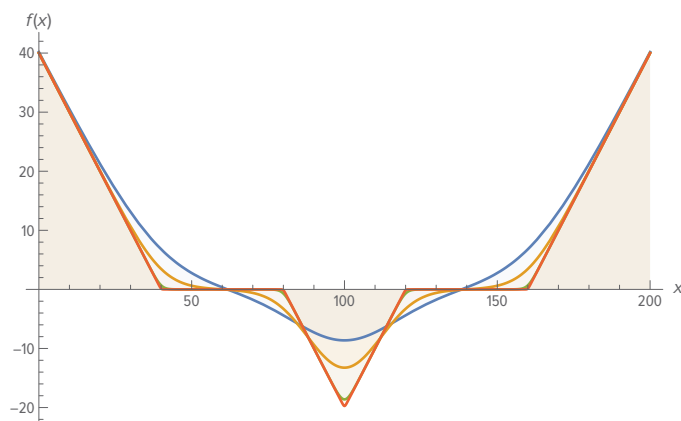
**Рисунок F.4:** Платежная функция опциона на момент истечения срока, открытая справа



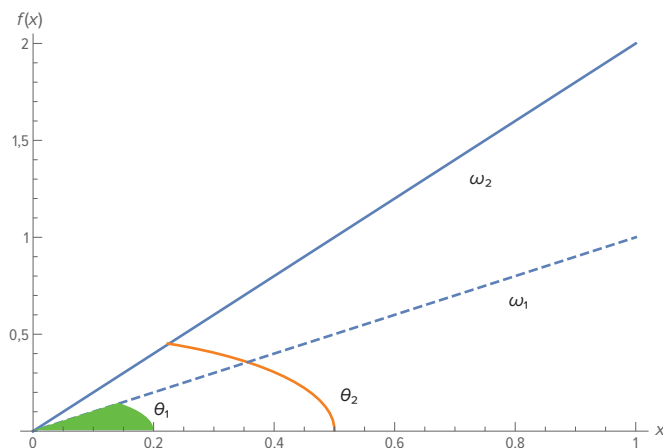
**Рисунок F.5:** Функция  $\rho$  из формулы 11.18, при  $k = 0$ . Сглаживание платежной функции регулируется параметром  $\rho$

### F.0.1. Калибровка по углам

Из Рисунка F.6 видно, что весовым коэффициентам  $\omega_i$  в формуле  $S = \sum_{i=1}^n \omega_i C(K_i, t_i)$  соответствует тангенс создаваемого угла, если вес положителен (как на Рисунке F.7), или величина, противоположная тангенсу дополнительного угла.



**Рисунок F.6:** Такую бабочку удастся породить суммой не сигмOID, а платежных функций опционов, или ReLU, — с открытыми хвостами с обеих сторон и с перевернутами производной и второй производной. Это яркий пример поведения, которое трудно описать словами, но в котором может разобратся опционный дилер и программа машинного обучения



**Рисунок F.7:** Как работает связь  $\omega = \tan \theta$ . Подбирая углы на каждом участке, можно представить нелинейную функцию суммой платежных функций опционов

## Резюме

Все одномерные нелинейные функции можно представлять взвешенными суммами платежных функций опционов с различными страйк-ценами, и для программ машинного обучения это более точный способ представить хвосты, чем через суммы сигмOID (которые и сами приходится представлять сетью длинных и коротких опционов с близкими страйками). Веса можно получить недвусмысленным образом из углов между графиками функций и осями декартовых координат.

**Часть III**

**ПРЕДСКАЗАНИЯ, ПРОГНОЗЫ  
И НЕОПРЕДЕЛЕННОСТЬ**



## КАЛИБРОВКА ВЕРОЯТНОСТИ ПРИ ЖИРНЫХ ХВОСТАХ<sup>†</sup>

Как влияют на эффективность вложений ваши способности к бинарным вероятностным прогнозам? Мы очертим различие между (одномерными) бинарными предсказаниями, они же «мнения» (о том, что данное «событие» скорее произойдет или скорее не произойдет), и непрерывными платежными функциями (с учетом объема прибылей или убытков из-за некоторого события), возникающими на практике. Мы покажем, что эти два понятия смешивают и что путаница самым прискорбным образом сказалась на литературе о принятии решений.

Вклад в путаницу вносят следующие эффекты:

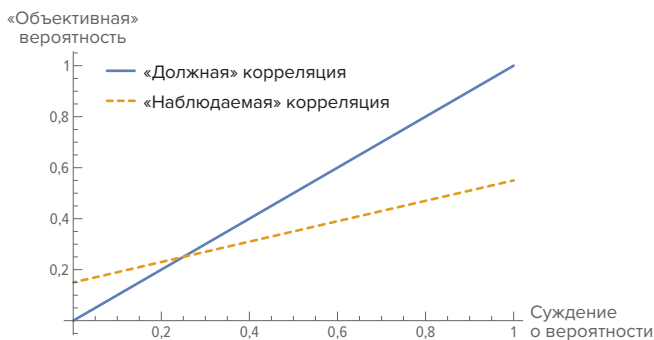
А) Псевдонаучная трактовка психологических исследований — в частности, мнимое открытие, будто люди склонны переоценивать хвостовые вероятности и редкие события и чрезмерно беречься от биржевого краха, экологических бедствий и т. п. На поверку обнаруженное «смещение» говорит лишь о неверной оценке рисков самими психологами. Сюда же относятся демагогические попытки пристегнуть доводы Хайека к защите рынка предсказаний.

В) «Хороший предиктор» в бинарном пространстве не обязательно будет хорошо работать в небинарном пространстве, и наоборот, особенно в нелинейной ситуации. В ряде классов распределений значение бинарного предиктора срабатывает как обратный предиктор. Если распределение принадлежит неизвестному классу или реалистичному классу со сложным поведением, шансы на применимость бинарного предиктора тают.

С) Программы машинного обучения (*Machine Learning*, ML) нужно использовать с умом. Некоторые нелинейные платежные функции трудны для словесного описания и текстовых прогнозов, но в подходящем формате успешно захватываются программами ML и переводятся в формулировку опционных договоров.

Д) Важны тонкости в методах оценки. Оценки на соревнованиях Макридакиса  $M_4$  и  $M_5$ <sup>1</sup>, по-видимому, лучше отражают реальные величины, чем оценка по Брайеру<sup>2</sup>.

В приложении показаны математические свойства и точное распределение для различных платежных функций, а также точное распределение для оценки Брайера, полезное при проверке значимости и достаточности выборки.



**Рисунок 11.1:** «Типичные паттерны», объявленные и описанные в [13], — пример исследования по психологии принятия решений, приписывающего человеку склонность переоценивать маловероятные события. Главные открытия опубликованы в 1977 и 1978 годах в [152] и [153]. Заметим, что в левой части графика, там, где усматривают переоценку, располагаются такие события, как наводнения, торнадо, ботулизм, — то есть такие, для которых характерны распределения с толстым хвостом серьезных последствий. Резонно предположить, что (1) опрашиваемые ставят себя на место участников событий и включают в оценку вероятности свою озабоченность последствиями, (2) опрашивающие, напротив, далеки от этих событий и недооценивают их вероятность

## 11.1. НЕПРЕРЫВНЫЕ И ДИСКРЕТНЫЕ ПЛАТЕЖНЫЕ ФУНКЦИИ: ОПРЕДЕЛЕНИЯ И КОММЕНТАРИИ

### Пример 11.1 («Мнениями и (бинарными) прогнозами сыт не будешь»)

В первом томе цикла *Incerto* (Одуроченные случайностью, 2001 [226]) в повествовании от лица трейдера, когда менеджер спрашивает: «Ваш прогноз рынка — подъем или спад?» — рассказчик уверенно отвечает: «Подъем». Потом начальник смотрит на сделанные фирмой вложения и негодует: рассказчик шортил<sup>3</sup>, то есть предполагал прибыль в случае спада на рынке.

1 Соревнования Макридакиса (*Makridakis Competitions, M Competitions*) — с 1982 открытый конкурс между методами прогнозирования временных рядов. На  $M_4$  (2020) победил Slawek Smyl, на  $M_5$  (2021–2022) победил YeonJun In. На  $M_4$  использовалась «гибридная» оценка успешности вероятностных прогнозов — средняя масштабированная интервальная (*Mean Scaled Interval Score, MSIS*).

2 Оценка по Брайеру, она же «показатель Брайера» (*Brier Score, BS*), — показатель успешности вероятностных прогнозов  $n$  бинарных событий, эквивалентная среднеквадратическому отклонению:  $BS \triangleq \frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2$ , где  $f_i$  и  $o_i$  — соответственно предсказанная вероятность исхода и наблюдавшийся исход  $i$ -того бинарного события.

3 Шорти́ть (*be short of the market*) — продавать ценные бумаги, которых в данный момент еще не имеешь; их придется вскоре купить, чтобы выполнить условия сделки, но если они еще раньше подешевеют, получишь прибыль.

*Рассказчику приходится долго разяснять боссу кажущееся противоречие между бинарным мнением о том, что рынок скорее поднимется, чем опустится, и количественной оценкой — после учета небольшой вероятности большого падения оказалось, что короткие позиции дают положительное математическое ожидание, и рациональное поведение — вкладываться в шорты. Как говорят трейдеры: «Тебя кормят не прогнозы, а  $P/L$ <sup>1</sup>» (и еще: «Прогнозы не монетизируешь»).*

Когда отдача от вложений расходится с предсказаниями, это происходит потому, что предсказания — словесное упрощение, в котором многомерный объект сжимается в одномерный. Чтобы выразить ошибку менеджера в терминах специалиста по теории принятия решений, нам придется устранить путаницу между такими основными понятиями, как, с одной стороны, бинарное событие (с которым связан нулевой статистический момент) и, с другой стороны, математическое ожидание результата с учетом вероятности события (с которым связан первый момент, а в случае нелинейной платежной функции — и высшие моменты). И мы убедимся, что платежные функции первого и второго рода в одних ситуациях похожи, а в других — нет.

## Комментарий 11.1

*Коротко говоря, если для калибровки вероятности бинарного события достаточно оценить нулевой момент, то для практических случайных величин необходимы все моменты (за исключением игровых ставок и таких искусственных ситуаций, как психологический эксперимент; в этих ситуациях платежная функция специально упрощена). Коренное свойство толстых хвостов — большие (и даже «бесконечные») высшие моменты, роль которых может оказаться сколь угодно велика.*

### 11.1.1. Отходим от словесных формулировок

Анекдот про трейдера математически тривиальный (хотя речь о распространенной ошибке); есть более серьезные трудности с пониманием принятия решений и управления рисками, особенно когда платежная функция сложная или нелинейная (и связана с высшими моментами). Таким образом, важны правильные математические формулировки в договорах и политике вложений, не ограниченные упрощенными словесными формулировками. Нам нужно разобраться с рядом серьезных вопросов о распределении вероятностей.

## Определение 11.1 (Событие)

*Непрерывная случайная величина — это функция*

$$X: \Omega \rightarrow \mathbb{R},$$

*определенная на множестве возможных исходов  $\Omega$  вероятностного пространства  $(\Omega, \mathcal{F}, \mathbb{P})$ ; пишут  $X(\omega)$ , где исход  $\omega \in \Omega$ .*

*Вероятность же приписывается событию, под которым понимается измеримое подмножество (мера  $\mathcal{F}$  не обязательно считающая) множества возможных исходов  $\Omega$ ; измеримость события означает, что его меру можно выразить через значение одной или нескольких величин.*

<sup>1</sup> *Profit/loss* — отношение чистой прибыли к убыткам.



**Определение 11.2** (Бинарный прогноз / платежная функция)

*Бинарный прогноз (мнение, платежная функция) — это случайная величина, принимающая два значения, то есть*

$$X: \Omega \rightarrow \{X_1, X_2\},$$

где реализации  $X_1, X_2 \in \mathbb{R}$ .

Другими словами, речь о ситуации, где двухэлементное множество чисел (такое как  $\{0, 1\}$ ,  $\{-1, 1\}$  и т. п.) исчерпывает различаемые исходы события. Иногда говорят, что событие может либо произойти, либо не произойти; иногда рассматривают платежную функцию, но она принимает только два конечных значения (задан некоторый фиксированный платеж, если событие произойдет, и другой фиксированный платеж, если событие не произойдет). Далее всюду, где не оговорим другое, будем предполагать множество платежей  $\{0, 1\}$ .

Примеры, где бинарная платежная функция встречается на практике:

- Ставка в казино, лотерея, бросок монеты, игровая среда, бинарный опцион (например, с выплатой фиксированной суммы, если биржевой курс упадет ниже заданного пункта, и без выплаты, если не упадет; считается разновидностью азартной игры)<sup>1</sup>.
- Голосование, где происходит выбор из двух вариантов (напр., референдумы, выборы президента Соединенных Штатов). Однако экономические последствия бинарных выборов бинарной величиной уже не будут<sup>2</sup>.
- Врачебный прогноз, выздоровеет ли пациент; как вариант — проживет ли заданное время. Однако продолжительность жизни бинарной величиной не будет; также не будет бинарной величиной срок выживания с данным заболеванием или условная ожидаемая продолжительность жизни. Эпидемиологические показатели исключаются все.
- Приобретет ли пользователь с известным сетевым профилем хотя бы одну единицу заданного товара за заданный период времени — однако количество купленного товара или число купленных единиц товара бинарной величиной не будет.

**Комментарий 11.2.** (Бинарное мнение эквивалентно бинарной платежной функции)

*Бинарное «мнение» должно взаимно-однозначно отображаться на платеж (путем масштабирования или нормализации, чтобы получить вероятность), — этой идеей*

<sup>1</sup> Розничная продажа бинарного опциона обычно используется как азартная игра и запрещена во многих юрисдикциях; например, ее запретила Европейская организация по ценным бумагам и рынкам (*European Securities and Markets Authority*, ESMA, [www.esma.europa.eu](http://www.esma.europa.eu)), а Соединенные Штаты приравняли к интернет-гэмблингу, что вызвало жалобу со стороны специалистов по принятию решений — см. Эрроу и др. [3]. Мы же считаем данный запрет оправданным, поскольку ставки практически не помогают экономике — в отличие от финансовых рынков, куда доступ открыт широкой публике и где естественные риски удается надлежащим образом купировать. — *Прим. автора.*

<sup>2</sup> Обратим внимание на отсутствие стихийно сложившихся финансовых рынков для бинарных опционов на основе непрерывных величин. Долгое время бинарных опционов не было вовсе, а когда они стали появляться, в период 1993–1998, то, как автору известно из личного опыта, мотивом было хитроумное уклонение от налогов. — *Прим. автора.*

мы обязаны де Финетти [57], который объявил, что «мнение» и «предсказание» (одного из двух исходов) отображаются на эквивалент математического ожидания бинарной случайной величины и ставки с платежами  $\{0, 1\}$ . Мнение отображается на цену, при которой игроки соглашаются покупать и продавать ставку. Если система ваших мнений — «голландская книга», то есть противоречит аксиомам теории вероятностей, иначе называемым правилами арбитража, то вы делаете ставки, которые гарантируют вам убытки.

**Определение 11.3** (Непрерывная платежная функция на открытом носителе, из реального мира)

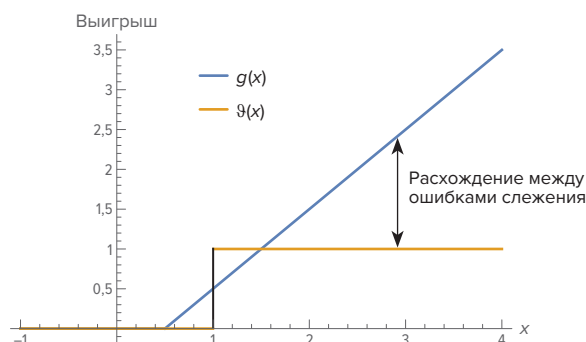
$$X: \Omega \rightarrow [a, \infty) \vee (-\infty, b] \vee (-\infty, \infty).$$

Непрерывная платежная функция принимает значения не на конечном множестве, а на интервале. Соответствующая случайная величина может быть неограниченной или ограниченной только с одной стороны (однохвостая случайная величина).

**Примечание.** Будем рассматривать только два сорта случайных величин: бинарные и неограниченные (то есть на некомпактном носителе) непрерывные. Почему только их? Дискретные платежные функции с многочисленными значениями поглощаются классом непрерывных случайных величин на основе обычных доводов аппроксимации. Если значений мало, например как у тернарной платежной функции на множестве  $\{-1, 0, 3\}$ , поведение аналогично бинарному случаю (и платежная функция представима как сумма бинарных случайных величин). Наконец, когда случайная величина формально ограничена снизу и сверху, но верхняя граница далекая — примером служит число жертв в катастрофе, — формально компактный носитель не мешает на практике анализировать поведение этой случайной величины как неограниченной [46].

Примеры практических ситуаций, где платежная функция непрерывна:

- Число жертв войны, ущерб при землетрясении, медицинские счета и т. д.
- Магнитуда биржевого краха, серьезность рецессии, уровень инфляции
- Выигрыш от той или иной стратегии
- Продажи и прибыль от нового продукта
- Любые вещи, покрываемые страховкой



**Рисунок 11.2:** Сравняем платежную функцию для бинарного пари — это ступенька Хевисайда  $g(\cdot)$  — и непрерывную неограниченную отдачу от вложений  $g(x)$ . Очевидно, нет никакой математической возможности построить сопоставимые производные таких платежных функций для хеджирования

Большинство естественных и социально-экономических случайных величин — непрерывные, причем их носитель не компактен или, по крайней мере, не требуется работать с конкретной верхней границей.



**Рисунок 11.3:** Путаница между вероятностью и математическим ожиданием результата глупо укоренилась среди специалистов по психологии и финансам. © Stefan Gasic

### Пример 11.2

Предсказательная аналитика в бинарном пространстве  $\{0, 1\}$  может успешно предсказать, например, что некий Яннис Пападопулос, клиент онлайн-магазина, имеющий данную историю активности, скорее приобретет данный товар, например обручальное кольцо, или что скорее не приобретет. Однако предсказание «благоприятного исхода» может быть столь же обманчиво для оценки потенциального нового продукта, как предсказание подъема рынка в анекдоте про трейдера. Продажи обычно распределены с толстым хвостом, и даже низкая вероятность успеха может быть достаточна для решения в пользу нового продукта. Сравните это с договорами венчурных инвестиций или торговли опционами: опцион «вне денег» может быть привлекательной покупкой, даже когда шансы получить от него прибыль меньше 1 из 1000.

Еще важнее то, что ошибка слежения за предсказанием вероятности не отображается в ошибку слежения за эффективностью. В отличие от  $\lambda^{(M_4)}$ .

Это различие хорошо знакомо опционным трейдерам, поскольку есть такие деривативы, как «бинарные» договоры, по которым платежная функция принимает значения в множестве  $\{0, 1\}$  (например, если основной актив  $S$  превысит страйк-цену  $K$ ), а есть «ванильные» опционы, с платежной функцией  $\max(S - K, 0)$ . Доход по ним пробегает диапазон  $[0, \infty)$ ... а то и  $(-\infty, 0)$ , потому что продавец опциона, в связи с неограниченностью вложений, подвергается риску банкротства. Существенные математические и экономические различия между этими двумя видами опционов — тема книги «Динамическое хеджирование: как управлять ванильными и экзотическими опционами» [225]. Поскольку бинарные деривативы суть пари с фиксированной ставкой, а у обычных опционов полноценная платежная функция, первые и вторые несводимы друг к другу (и не могут компенсировать друг друга в хедж-фонде), особенно при жирных хвостах и неопределенных параметрах — то есть эффективность одного инструмента не переводится в эффективность другого. Вот о чем знают специалисты по финансовой математике и вот какое знание почему-то не проникает в литературу по теории принятия решений.

**Комментарий 11.3** (Теория деривативов)

Этот подход опирается на теорию и практику деривативов (или опционов), где есть разные типы договоров: (1) с бинарной платежной функцией (владелец получает фиксированную сумму в случае некоторого события) и (2) «ванильные» (обычные опционы с непрерывной платежной функцией). Практически невозможно хеджировать договоры одного типа договорами другого типа [225]. Более того, если страйк-цена  $K$  опциона «колл» находится в хвосте распределения, то стоимость этого опциона и стоимость пари с той же страйк-ценой противоположным образом реагируют на увеличение эксцесса распределения при сохранении первых трех моментов, а в лог-нормальной ситуации — на увеличение неопределенности путем масштабирования распределения.

**Комментарий 11.4** (Перечни условий)

Заметим, что благодаря так называемым перечням условий, необходимых по юридическим и математическим соображениям, практика финансовых деривативов разного рода позволяет исследователю построить достоверное, юридически подтвержденное отображение платежных функций со всеми математическими, статистическими и экономическими подробностями.

Существовали разногласия между рынками прогнозов и реальными финансовыми рынками. Мы можем доказать, что рынки прогнозов интересны только азартным игрокам, а для хеджирования вложений в экономику бесполезны.

Вот математическое обоснование различий и невозможности хеджирования. Пусть  $X$  — случайная величина в  $\mathbb{R}$ , и рассматривается пари или прогноз с платежной функцией  $\vartheta_K: \mathbb{R} \rightarrow \{0, 1\}$ ,

$$\vartheta_K(x) = \begin{cases} 1, & x \geq K \\ 0, & \text{иначе} \end{cases} \quad (11.1)$$

и платежная функция  $g: \mathbb{R} \rightarrow \mathbb{R}$  вложений в обычные опционы. Производная  $\frac{\partial}{\partial x} \vartheta_K(x) = \delta(x - K)$ , дельта-функция Дирака с пиком в точке  $K$ , а производная  $\frac{\partial}{\partial x} g(x)$  по меньшей мере однократно дифференцируема при  $x \geq K$  (производная либо постоянна, в случае линейной платежной функции, либо кусочно-постоянна, если, как у многих опционов, отдача от вложений в опционы кусочно-линейная выше  $K$ ). Значит, задача построить сопоставимые деривативы для уменьшения дисперсии не имеет решения<sup>1</sup>. Этот довод проиллюстрирован на Рисунке 11.2.

### 11.1.2. При жирных хвостах нет стандартной величины «коллапса», «катастрофы» или «успеха»

Тот факт, что разговор о «событии» сопровождается некоторой неопределенностью относительно подразумеваемой магнитуды события, имеет математические последствия. Между тем даже в статьях 2019 года все еще встречаем порочную практику

<sup>1</sup> Чтобы представить бинарными функциями непрерывную платежную функцию на некомпактном носителе, потребуется бесконечный ряд пари, и это хоронит надежду преобразовать рынок прогнозов в финансовый рынок. У распределений на компактном носителе все моменты всегда конечные, но реальная жизнь устроена иначе. — Прим. автора.

бинаризовать событие на носителе  $[0, \infty)$ ; в недавней статье о калибровке мнений используется формулировка: «...если человек утверждает, что США стоят на пороге экономического коллапса или что надвигается климатическая катастрофа...», как будто «коллапс» экономики или «катастрофу» с климатом можно представить как событие из  $\{0, 1\}$ ; на самом деле под такими словами могут подразумеваться вещи разных масштабов, и, чтобы понимать, о чем речь, необходимо задавать характерную величину. Когда хвосты жирные, у распределения нет собственного характерного масштаба. Соответственно не существует «типичного» коллапса или катастрофы; словесные прогнозы и мнения в такой области не калибруются.

Различие между областью тонких хвостов и областью жирных хвостов можно представить так.

#### Определение 11.4 (Характерный масштаб)

Пусть  $X$  — случайная величина с носителем  $(0, \infty)$  или  $(-\infty, \infty)$  и реалистичным распределением; пусть  $\mathbb{E}$  — оператор математического ожидания. Согласно классическим результатам [82], предельное поведение распределения в хвосте:

$$\lim_{K \rightarrow \infty} \frac{1}{K} \mathbb{E}(X |_{X > K}) = \lambda \quad (11.2)$$

позволяет выделить следующие варианты:

- Если  $\lambda = 1$ , случайная величина  $X$  относится к классу тонких хвостов  $D_1$  и имеет характерный масштаб.
- Если  $\lambda > 1$ , случайная величина  $X$  относится к классу правильно изменяющихся распределений с жирным хвостом  $D_2$  и не имеет характерного масштаба.
- Если

$$\lim_{K \rightarrow \infty} (\mathbb{E}(X |_{X > K}) - K) = \mu,$$

где  $\mu > 0$ , случайная величина  $X$  относится к пограничному экспоненциальному классу.

Можно пояснить проблему так. Бинарный договор не способен хеджировать вас на случай «коллапса», поскольку неизвестно заранее, каких масштабов ожидать «коллапс» и какой должна быть номинальная стоимость или другой аналогичный параметр договора. Зато договор страхования или опцион с непрерывной платежной функцией успешно справляются с ролью хеджа. Отделяться словесными формулировками с терминами вроде «коллапс» или «катастрофа» — все равно что прописывать в медицинской страховке щедрую сумму в случае «серьезной болезни», не оговорив характер и тяжесть заболевания, и 0 в противном случае.

Тем более некорректно со стороны исследователя вычленять платежную функцию и оценку вероятности из результатов опроса, в котором опрашиваемые оценивали угрозы<sup>1</sup>. В некоторых экспериментах, о которых рассказывает Рисунок 11.1, участников про-

<sup>1</sup> Притом что практически все экономические и информационные случайные величины, как показано еще в 1960-е годы, принадлежат классу  $D_2$  или, по крайней мере, промежуточному экспоненциальному классу (куда входит логнормальное распределение), см. [99, 162, 163, 164, 226]; туда же попадают социологические величины: размер города, число слов в языке, число связей в сети, размер фирмы, доход фирмы, макроэкономические показатели, монетарные показатели, число жертв межгосударственных конфликтов и гражданских войн [46, 199], риск операции, ущерб от землетрясений, цунами, ураганов и прочих сти-

сят оценить риск смерти от ботулизма или еще какой-то болезни, после чего исследователи приходят к выводу, что участники неправильно оценивали вероятность. Хотя в таком эксперименте участники, скорее всего, не отделяли вероятность от функции платежа.

## 11.2. МНИМАЯ ПЕРЕОЦЕНКА ХВОСТОВОЙ ВЕРОЯТНОСТИ В ПСИХОЛОГИИ

**Определение 11.5** (Подмена интегральной оценки)

Пусть  $K_p \in \mathbb{R}^+$  — порог, который случайная величина превышает с вероятностью  $p \in [0, 1]$ ; пусть при этом  $f(\cdot)$  — плотность вероятности и  $g(\cdot)$  — функция воздействия.

Обозначим  $I_1$  математическое ожидание платежной функции в случае превышения порога  $K_p$ :

$$I_1 = \int_{K_p}^{\infty} g(x) f(x) dx,$$

и обозначим  $I_2$  платеж при пороговом событии, умноженный на вероятность превысить порог:

$$I_2 = g(K_p) \int_{K_p}^{\infty} f(x) dx = g(K_p) p.$$

Можно ли оценивать ожидаемый платеж  $I_1$  как бинарный ожидаемый платеж  $I_2$ ? Эти величины тождественны лишь в той ситуации, где функция воздействия  $g(\cdot)$  постоянна выше порога  $K_p$  — например, где  $g(x) = \mathfrak{Y}(x - K_p)$ , ступенька Хевисайда в точке  $K_p$ . Если же функция воздействия  $g(\cdot)$  — меняющаяся функция с положительной производной, ожидаемый платеж  $I_1$  больше бинарного ожидаемого платежа  $I_2$  и близок к нему разве что при распределении с тонким хвостом, но никак не при жирном хвосте<sup>1</sup>.

В этом разделе примем  $g(x) = x$ , а более сложный случай нелинейной функции воздействия рассмотрим в Разделе 11.5.

### Теорема 2: Сходимость $\frac{I_1}{I_2}$

Если  $X$  следует тонкохвостому распределению из класса  $\mathcal{D}_1$ , как описано в формуле 11.2, то

$$\lim_{K_p \rightarrow \infty} \frac{I_1}{I_2} = 1. \quad (11.3)$$

Если  $X$  следует правильно меняющемуся распределению из класса  $\mathcal{D}_2$ , то

$$\lim_{K_p \rightarrow \infty} \frac{I_1}{I_2} = \lambda > 1. \quad (11.4)$$

*Доказательство.* Следует из формулы 11.2. Вот дополнительные комментарии.

хийных бедствий, неравенство доходов [40] и т. д. Напрашивается вопрос, а где же гауссовы случайные величины? Похоже, их как минимум на порядок меньше среди тех случайных величин, которые приходится учитывать в официальных прогнозах. — *Прим. автора.*

1 Это также объясняет, как мы увидим в Главе 11, почему бинарные пары не способны «рисковать собственной шкурой» при жирнохвостых распределениях. — *Прим. автора.*

### 11.2.1. Тонкие хвосты

Выше мы дали определение тонкохвостого распределения — не только гауссова, но любого распределения с хвостом тоньше, чем в субэкспоненциальном классе; будем помечать такие распределения буквой G. Распределение плотности вероятности  $f^{(G)}(\cdot)$  удовлетворяет условию:

$$\lim_{K_p \rightarrow \infty} \frac{\int_{K_p}^{\infty} x f^{(G)}(x) dx}{K_p \int_{K_p}^{\infty} f^{(G)}(x) dx} = \lim_{K_p \rightarrow \infty} \frac{I_1}{I_2} = 1. \quad (11.5)$$

В случае стандартного гауссова распределения, то есть центрального и нормализованного, интеграл в числителе

$$\int_{K_p}^{\infty} x G(x) dx = \frac{1}{\sqrt{2\pi}} \exp \frac{-K_p^2}{2}, \quad (11.6)$$

и интеграл в знаменателе  $p = \frac{1}{2} \operatorname{erfc} \frac{K_p}{\sqrt{2}}$ , где  $\operatorname{erfc}$  — дополнительная функция ошибок<sup>2</sup> и  $p$  — вероятность превысить порог  $K_p$ .

Заметим, что  $K_p \frac{I_1}{I_2}$  отвечает<sup>3</sup> обратному отношению Миллза<sup>4</sup>, используемому в страховании.

### 11.2.2. Жирные хвосты

Жирные хвосты у всех распределений в классе правильно меняющихся функций, определенном по поведению функции выживания в хвосте, при больших  $K_p$ :

$$\mathbb{P}(X > K_p) \approx L K_p^{-\alpha}, \quad \alpha > 1 \text{ и } L > 0,$$

1 В самом деле, в силу центральности коэффициент сдвига  $\mu = 0$  и в силу нормализованности коэффициент

$$\text{масштаба } \sigma = 1; \text{ поэтому плотность вероятности } G(x) = \frac{1}{\sqrt{2\pi}} \exp \frac{-(x-\mu)^2}{2\sigma^2} = \frac{1}{\sqrt{2\pi}} \exp \frac{-x^2}{2}.$$

$$\text{Поэтому ожидаемый платеж при превышении порога } \int_{K_p}^{\infty} x G(x) dx = \int_{K_p}^{\infty} x \frac{1}{\sqrt{2\pi}} \exp \frac{-x^2}{2} dx =$$

$$= -\frac{1}{\sqrt{2\pi}} \int_{x=K_p}^{x=\infty} \exp \frac{-x^2}{2} d \frac{-x^2}{2} = -\frac{1}{\sqrt{2\pi}} \exp \frac{-x^2}{2} \Big|_{x=K_p}^{x=\infty} = \frac{1}{\sqrt{2\pi}} \exp \frac{-K_p^2}{2} = G(K_p).$$

2 Функция выживания при стандартном нормальном распределении связана с дополнительной функцией

$$\text{ошибок так: } \int_{K_p}^{\infty} G(x) dx = \frac{1}{2} \operatorname{erfc} \frac{K_p}{\sqrt{2}}.$$

3 В самом деле, для гауссова распределения  $K_p \frac{I_1}{I_2} = K_p \frac{G(K_p)}{K_p \int_{K_p}^{\infty} G(x) dx} = h(K_p)$ .

4 Отношение функции выживания к плотности вероятности  $m(K) \triangleq \frac{\int_K^{\infty} f(x) dx}{f(K)}$  иногда называют отношением

Миллза, в честь таблиц Джона Миллза (*John P. Mills*, Table of the Ratio: Area to Bounding Ordinate, for Any Portion of Normal Curve. *Biometrika* 18: 395–400, 1926).

Обратное отношение называется функцией риска (*hazard function*), или интенсивностью отказов (*failure rate*). Пусть случайная величина  $X$  распределена с плотностью  $f$ . Тогда интенсивность отказов

$$h(K) \triangleq \frac{f(K)}{\int_K^{\infty} f(x) dx} = \lim_{\delta \rightarrow \infty} \frac{\mathbb{P}(K < X \leq K + \delta | X > K)}{\delta}, \text{ и если } X \text{ — это срок службы прибора до отказа или продолжи-}$$

тельность жизни пациента, то  $h(K)$  — это вероятность отказа или смерти за единицу времени в возрасте  $K$  среди доживших до возраста  $K$ .



и в котором у всякой плотности вероятности  $f^{(P)}$ <sup>1</sup>

$$\lim_{K_p \rightarrow \infty} \frac{\int_{K_p}^{\infty} x f^{(P)}(x) dx}{K_p \int_{K_p}^{\infty} f^{(P)}(x) dx} = \frac{\alpha}{\alpha - 1} > 1^2. \quad (11.7)$$

### 11.2.3. Что с чем путают

Во многих экспериментах путают  $I_1$  и  $I_2$ ; примером служит статья Канемана и Тверски 1978 года по теории перспектив<sup>3</sup> [139]. Утверждается, что участники переоценивают низкие вероятности в экспериментах, где участнику показали шансы и где результат отвечает одному-единственному платежу. Известный результат Канемана — Тверски подтвержден как робастный, однако интерпретации содержат ошибочные выводы. Практически вся последующая литература полагается на бинарный платеж  $I_2$ , путая его с  $I_1$ , — то, что автор этой книги назвал *игровым софизмом* в книге «Черный лебедь» [226], поскольку игры по необходимости отсекают размерности, действующие на практике. Робастность психологического результата всего лишь говорит о воспроизводимости в аналогичных экспериментах и не доказывает применимость выводов в других условиях, приближенных к реальным рискам, — и такое обобщение вызывает сомнения с учетом того, что реальные риски редко отвечают  $I_2$ . На самом деле даже те участники, которые переоценивали вероятность, могли все еще недооценивать математическое ожидание платежной функции.

**Стоимость путаницы:** она продолжается даже сейчас, спустя четыре десятилетия после заявлений Канемана и Тверски (1979). В своем обзоре поведенческой экономики Николас Барберис делает упор на неверную оценку вероятности (2003) [12], считая  $I_1 = I_2$ . Эрроу и др. [3] — там целый список специалистов по теории принятия решений, которые ратуют за дерегуляцию рынков пари, — тоже смешивают мир бинарных прогнозов с реальным миром и, в частности, с реальными финансовыми рынками.

Еще одним веским — и опасным — примером служит «стоимость под риском дефолта», которую определяют явным образом как  $I_2$ , а именно как вероятность дефолта  $\times (1 - \text{коэффициент погашения})$ , и эта платежная функция может кардинально отличаться от настоящего математического ожидания убытков в случае дефолта. Финансовая математика предлагает ошибочную аппроксимацию  $\text{CVaR}$ <sup>4</sup>,

<sup>1</sup> Очевидно, метка P в честь Парето.

<sup>2</sup> В самом деле, для распределения Парето с коэффициентом масштаба  $x_m > 0$  плотность вероятности  $f^{(P)}(x) = \alpha L x^{-\alpha-1}$ , где  $L = x_m^\alpha$ . Оба интеграла берутся:

$$I_1 = \int_{K_p}^{\infty} x f^{(P)}(x) dx = \alpha L \int_{K_p}^{\infty} x^{-\alpha} dx = \alpha L \frac{x^{-\alpha+1}}{-\alpha+1} \Big|_{x=K_p}^{x=\infty} = -\alpha L \frac{K_p^{-\alpha+1}}{-\alpha+1};$$

$$p = \int_{K_p}^{\infty} f^{(P)}(x) dx = \alpha L \int_{K_p}^{\infty} x^{-\alpha-1} dx = \alpha L \frac{x^{-\alpha}}{-\alpha} \Big|_{x=K_p}^{x=\infty} = L K_p^{-\alpha}.$$

<sup>3</sup> Теория перспектив (*prospect theory*) в поведенческой экономике описывает принятие решений при известной вероятности рисков.

<sup>4</sup> Математическое выражение для стоимости под риском (*Value at Risk*, VaR) для случайной величины  $X$  с функцией распределения  $F$  и порогом  $a \in [0, 1]$

$$\text{VaR}_a(X) = -\inf \{x \in \mathbb{R} \mid F_X(x) > a\},$$

и для соответствующей условной стоимости под риском (*Conditional Value at Risk*, CVaR), она же ожидаемые потери (*Expected Shortfall*, ES)

$$\text{ES}_a(X) = \mathbb{E}(-X \mid X \leq -\text{VaR}_a(X)). \text{ — Прим. автора.}$$



и эта ошибка привела к просчетам в управлении рисками — и они, возможно, стали причиной кризиса 2008 года [244].

Ложная посылка была та, что коэффициент погашения оценивался как ожидаемая стоимость залога вне ситуации дефолта. Однако ожидаемая стоимость залога в ситуации дефолта существенно ниже. В 2007 после массовых просрочек выплат по кредитам и лишения права собственности стоимость большинства залогов просела ни много ни мало на 1/3 ожидаемого значения.

**Непонимание доводов Хайека об использовании знаний.** То, что, по Хайеку, мнения консолидируются через цены, не оправдывает рынки прогнозов, как это пытаются изобразить такие публикации, как [30] или Санстейн [220]: цены принадлежат миру финансовых и коммерческих рынков; они отличаются от бинарных пари. Для Хайека [127] консолидация знаний осуществляется через цены и, как он выразился, *арбитражеров*, — и эти арбитражеры обмениваются товарами, услугами и ценными бумагами, но не бинарными пари.

**Определение 11.6** (Скорректированная вероятность в бинаризованных экспериментах)

Пусть  $p^*$  — «эффективная вероятность», при которой  $I_1 = I_2^*$ , то есть ожидаемый платеж можно оценить как бинарный ожидаемый платеж со скорректированной вероятностью превысить порог:

$$p^* = \{p^* \mid I_1 = I_2^* = K_p p^*\}.$$

Выразим из этого равенства порог  $K_p$  «в хвостах», используя полученные выше формулы для вероятности  $p$  превысить данный порог. Для гауссовой случайной величины  $K_p = \sqrt{2} \operatorname{erfc}^{-1} 2p$ ; для распределения с хвостом Парето  $K_p = p^{-1/\alpha}$ . Значит, для распределения Парето отношение эффективной вероятности, делающей бинарный платеж реалистичным, к наивной вероятности

$$\frac{p^*}{p} = \frac{\alpha}{1-\alpha},$$

благодаря чему эффективная вероятность  $p^*$  может абсурдным образом превосходить 1; это лишний раз подчеркивает, что бинарный платеж — некорректный способ задать ожидаемый платеж непрерывного распределения.

В Таблицах 11.1 и 11.2 показаны, для различных уровней вероятности  $p$ , соответствующий порог отсечения хвоста  $K_p$

$$K_p = \inf \{K \mid \mathbb{P}(X > K) > p\}$$

и эффективная вероятность  $p^*$  дебинаризации события<sup>1, 2</sup> — здесь вероятности должны быть в нижней половине, т. е.  $p < 0,5$ . Заметим, что здесь мы работаем в благоприятных условиях известных распределений вероятностей; при неопределенности параметров расхождение между  $I_1$  и  $I_2$  усугубляется<sup>3</sup>.

1 Данный анализ инвариантен к рассмотрению правого или левого хвоста. В финансовой математике принято приписывать убыткам отрицательную величину, а в других областях управления рисками в выражениях ставят знак минус перед случайной величиной, для которой изучают правый хвост распределения. — Прим. автора.

2  $K_p$  эквивалентна стоимости под риском  $\operatorname{VaR}_p$  в финансовой математике, где  $p$  — вероятность потерь. — Прим. автора.

3 Обратим внимание на закон Ван дер Вейка, см. Чирилло [44]: отношение  $\frac{I_1}{I_2}$  аналогично термину финансовой математики — ожидаемые потери (*expected shortfall*) за предельным значением  $K_p$ . — Прим. автора.

Таблица 11.1: Мнимая переоценка гауссовой случайной величины

$p$	$K_p$	$\int_{K_p}^{\infty} x f(x) dx$	$K_p \int_{K_p}^{\infty} f(x) dx$	$p^*$	$\frac{p^*}{p}$
$\frac{1}{10}$	1,28	$1,75 \times 10^{-1}$	$1,28 \times 10^{-1}$	$1,36 \times 10^{-1}$	1,36
$\frac{1}{100}$	2,32	$2,66 \times 10^{-2}$	$2,32 \times 10^{-2}$	$1,14 \times 10^{-2}$	1,14
$\frac{1}{1000}$	3,09	$3,36 \times 10^{-3}$	$3,09 \times 10^{-3}$	$1,08 \times 10^{-3}$	1,08
$\frac{1}{10000}$	3,71	$3,95 \times 10^{-4}$	$3,71 \times 10^{-4}$	$1,06 \times 10^{-4}$	1,06

Таблица 11.2: Мнимая переоценка случайной величины Парето<sup>1</sup>

$p$	$K_p$	$\int_{K_p}^{\infty} x f(x) dx$	$K_p \int_{K_p}^{\infty} f(x) dx$	$p^*$	$\frac{p^*}{p}$
$\frac{1}{10}$	8,1	8,92	0,811	1,1 (да-да, больше единицы)	11
$\frac{1}{100}$	65,7	7,23	0,65	0,11	11
$\frac{1}{1000}$	533	5,87	0,53	0,011	11
$\frac{1}{10000}$	4328	4,76	0,43	0,0011	11

В самом известном примере распределения Парето, «80 / 20» (с тех пор как Парето обнаружил, что 20 процентов итальянцев владеет 80 процентами земель), показатель хвоста  $\alpha = 1,16$ , и корректировка требует увеличивать вероятность в 7 с лишним раз.

**Пример противоположной реакции вероятности и ожидаемого платежа при росте неопределенности.** Вот пример, где благодаря асимметрии распределения ожидаемый бинарный платеж и ожидаемый платеж реагируют на изменение параметров распределения противоположным образом.

Рассмотрим логнормальное распределение, нейтральное в отношении риска,  $\mathcal{L}\left(X_0 - \frac{1}{\sigma^2}, \sigma\right)$  с плотностью  $f_L(\cdot)$ , средним  $X_0$  и дисперсией  $(e^{\sigma^2} - 1) X_0^2$ . Мы можем повышать неопределенность, увеличивая коэффициент  $\sigma$ . Математическое ожидание  $\mathbb{E}_{>X_0}$  платежа по договору, если платеж превысил  $X_0$ , можно выразить как

$$\mathbb{E}_{>X_0} = \int_{X_0}^{\infty} x f_L(x) dx = \frac{1}{2} X_0 \left( 1 + \operatorname{erf} \frac{\sigma}{2\sqrt{2}} \right),$$

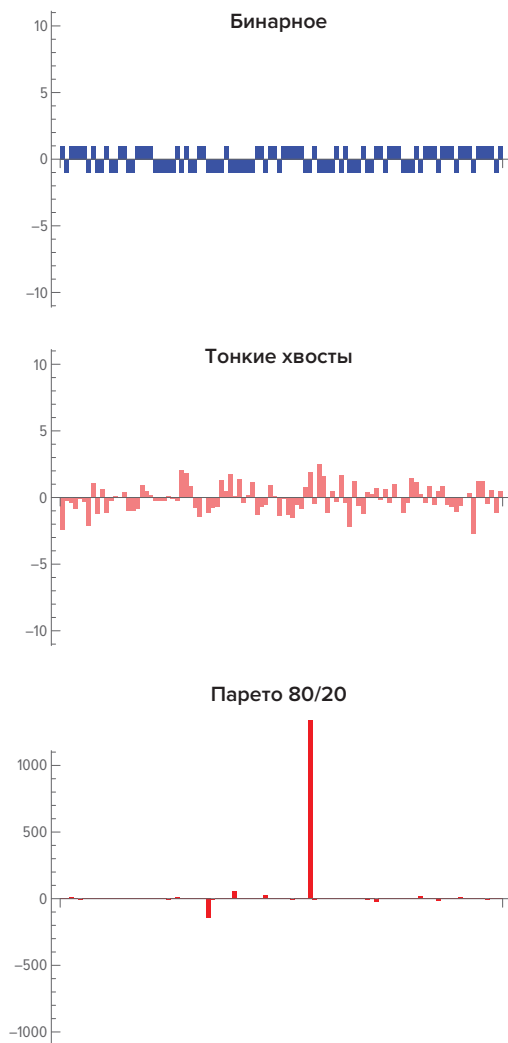
<sup>1</sup> Таблица рассчитана для показателя хвоста  $\alpha = 1,1$ .

а вероятность превышения  $X_0$  как

$$\mathbb{P}(X > X_0) = \frac{1}{2} \left( 1 - \operatorname{erf} \frac{\sigma}{2\sqrt{2}} \right),$$

где  $\operatorname{erf}$  — функция ошибок. По мере роста  $\sigma$  выражение  $\operatorname{erf} \frac{\sigma}{2\sqrt{2}} \rightarrow 1$ , поэтому  $\mathbb{E}_{>X_0} \rightarrow X_0$ .

Этот пример хорошо известен трейдерам (см. «Динамическое хеджирование» [225]): стоимость бинарного опциона со страйк-ценой  $X_0$  стремится к 0, в то время как стандартный опцион «колл» с той же страйк-ценой дорожает и приближается к стоимости актива — уже независимо от страйк-цены. Это типично для венчурного капитала: чем рискованнее проект, тем меньше шансы на успех, но зато больше награда в случае успеха. В результате математическое ожидание платежа может приблизиться к  $+\infty$ , даже когда вероятность успеха приблизится к 0.



**Рисунок 11.4:** Сравниваем три профиля платежной функции при двух распределениях — для бинарной платежной функции профиль не зависит от тонкости/жирности хвостов. Первые два графика в одном масштабе, а для Парето 80/20 ( $\alpha = 1,16$ ) вертикальную шкалу пришлось сжать на два порядка

### 11.2.4. Неопределенность распределения

#### Замечание 11: Размывание параметров распределения

Благодаря неравенству Йенсена различие между  $I_1$  и  $I_2$  усиливается при повышении эксцесса (например, если в тонкохвостом распределении сделать стохастическим коэффициент масштаба  $\sigma$  или если в распределении Парето стохастизировать показатель хвоста  $\alpha$ ).

*Доказательство.* Начнем с мира Гаусса. Посмотрим, как ведет себя разность

$$I_1 - I_2 = \int_K^\infty x f^{(G)}(x) dx - K \int_K^\infty f^{(G)}(x) dx$$

при добавлении стохастической волатильности. Пусть  $\sigma$  — коэффициент масштаба гауссова распределения. Дифференцируя при постоянном пороге  $K$ :

$$\frac{\partial^2}{\partial \sigma^2} \int_K^\infty x f^{(G)}(x) dx - K \frac{\partial^2}{\partial \sigma^2} \int_K^\infty f^{(G)}(x) dx = e^{-\frac{K^2}{2\sigma^2}} \frac{(K-1)K^3 - (K-2)K\sigma^2}{\sqrt{2\pi}\sigma^5}, \quad (11.8)$$

получаем положительное выражение при всех  $K > 0$  (поскольку при положительных  $K$  многочлен  $K^4 - K^3 - K^2 + 2K > 0$ )<sup>1</sup>.

Теперь рассмотрим чувствительность  $\frac{I_1}{I_2}$  к неопределенности показателя хвоста  $\alpha$  распределения Парето (здесь вывод выражений для отношения будет проще, чем для разности).

При  $\alpha > 1$  (это критерий конечного среднего) выражение

$$\frac{\partial^2}{\partial \alpha^2} \frac{\int_K^\infty x f^{(P)}(x) dx}{K \int_K^\infty f^{(P)}(x) dx} = \frac{2}{(\alpha-1)^3} \quad (11.9)$$

положительно и гарантирует быстрый рост при небольших  $\alpha$ . То есть чем толще хвост, тем сильнее неопределенность влияет на платеж и тем сильнее различие между  $I_1$  и  $I_2$ . ■

1 Строго говоря, при  $\sigma = \frac{1}{4}$  все-таки есть область,  $K \in (0, 4, 0, 9)$ , где многочлен  $(K-1)K^3 - (K-2)K\sigma^2$  отрицателен, однако в данном исследовании интересны только большие пороги  $K$ .

2 В самом деле, плотность распределения Парето  $f^{(P)}(x) = \alpha x_m^\alpha x^{-\alpha-1}$ , где  $x_m > 0$  — параметр масштаба. Оба

$$\begin{aligned} \text{интеграла беруться, } \int_K^\infty x f^{(P)}(x) dx &= \alpha x_m^\alpha \int_K^\infty x^{-\alpha} dx = \alpha x_m^\alpha \frac{x^{-\alpha+1}}{-\alpha+1} \Big|_{x=K}^{x=\infty} = -\alpha x_m^\alpha \frac{K^{-\alpha+1}}{-\alpha+1} \quad \text{и} \quad K \int_K^\infty f^{(P)}(x) dx = \\ &= K \alpha x_m^\alpha \int_K^\infty x^{-\alpha-1} dx = K \alpha x_m^\alpha \frac{x^{-\alpha}}{-\alpha} \Big|_{x=K}^{x=\infty} = -K \alpha x_m^\alpha \frac{K^{-\alpha}}{-\alpha}. \end{aligned}$$

$$\text{Отношение } \frac{I_1}{I_2} = \frac{\alpha}{\alpha-1} = 1 + \frac{1}{\alpha-1}, \text{ откуда } \frac{\partial}{\partial \alpha} \frac{I_1}{I_2} = -\frac{1}{(\alpha-1)^2} \text{ и } \frac{\partial^2}{\partial \alpha^2} \frac{I_1}{I_2} = \frac{2}{(\alpha-1)^3}.$$

### 11.3. КАЛИБРОВКА И МНИМАЯ КАЛИБРОВКА

В литературе по психологии изучается «калибровка» того, как участники оценили вероятности — насколько близки оказались названные шансы событий к среднему (исходя из удовлетворительной работы закона больших чисел): [152], [143]; смотрите Рисунок 3.13 в Главе 3. Эти методы, как мы показали в этом разделе, вводят в заблуждение всюду, кроме узкого круга явлений с бинарной платежной функцией (где результат либо выигрыш, либо проигрыш), причем перерасчет бинарных платежей в непрерывные невозможен или перевирает картину. В связи с этим Рисунок 11.1 демонстрирует эмпирическую бессмыслицу.

Такие калибровочные метрики, как оценка Брайера, по своей природе тонкохвостые; когда требуется измерять случайные величины с толстым хвостом, результаты не поддаются анализу.

Напомним поговорку «[Бинарными] предсказаниями сыт не будешь» — в большинстве областей для бизнеса характерно несимметричное распределение платежной функции, и попытки прокалибровать ее бинарной платежной функцией безнадежны.

#### Замечание 12: Различие между распределениями

Бинарные предсказания и калибровочные метрики на основе оценки Брайера принадлежат классу тонких хвостов.

Ниже мы приведем доказательство.

### 11.4. ПОКАЗАТЕЛИ ДЛЯ КОЛИЧЕСТВЕННОЙ ОЦЕНКИ

Резюме этого раздела приведено в Таблице 11.3 — это сравнение распределений вероятности различных показателей для измерения эффективности по явным формулам или хотя бы по оценке класса распределений вероятности. Понятно, что измерение эффективности дает ложные результаты, если случайная величина не из того класса. Для разных распределений нужны разные размеры выборки, потому что по-разному работает закон больших чисел. Ряд бинарных предсказаний быстро сходится к тонкохвостой гауссиане, даже если исходная случайная величина следует распределению с жирным хвостом, но чтобы отслеживать эффективность Р/Л в экономике при реалистичных вложениях в рискованные активы, платежная функция которых следует, например, распределению Парето, требуется много больший размер выборки [235].

Начнем с точных выражений для четырех возможных классов.

1. Практическая эффективность при условии выживания, она же оценка Р/Л или количественная кумулятивная оценка.
2. Подсчет пари — наивная сумма правильных предсказаний.
3. Оценка Брайера в игре де Финетти,  $\lambda_n^{(B)}$ .
4. Оценка  $\lambda_n^{(M4)}$  в конкурсе  $M_4$  для  $n$  наблюдений и ее развитие, предложенное для конкурса  $M_5$ .

Таблица 11.3: Показатели для оценки эффективности

Показатель	Название	Применим ли на практике
$P_T^{(R)}$	Кумулятивное отношение P/L	Адаптирован к реалистичным распределениям, особенно при фильтрации выживших
$P_n^{(P)}$	Подсчет пари	Искажает эффективность при толстых хвостах; работает только для бинарных пари и/или в областях с тонкими хвостами
$\lambda_n^{(B)}$	Оценка Брайера	Показывает минимую точность в оценке эффективности при толстых хвостах; игнорирует высшие моменты
$\lambda_n^{(M4)}$	Оценка $M_4$	Представляет точность, которая отличается от реалистичной эффективности, но отображается на реалистичное распределение исходных переменных
$\lambda_n^{(M5)}$	Предложенная оценка $M_5$	Представляет и точность, и условия выживания, предсказывая крайние значения временных рядов
$g(\cdot)$	Нелинейная платежная функция машинного обучения (не является показателем)	Представляет отдачу от рискованных вложений без словесных формулировок и правдиво отражает соотношение P/L в экономике и других областях. Аналогична спискам условий в финансовых деривативах

**P/L в пространстве платежей (с условием выживания).** Обозначение P/L расшифровывается как естественный индекс прибылей и убытков, то есть кумулятивный итог эффективности. Пусть  $X_t$  — реализация обычной одномерной случайной величины  $X$  на носителе  $\mathbb{R}$  и  $t = 1, 2, \dots, T$ . Реалистичные платежи  $P^{(R)}$  можно упрощенно представить как

$$P_T^{(R)} = P_0^{(R)} + \sum_{t \leq T} g_t(X_t), \quad (11.10)$$

где  $g_t: \mathbb{R} \rightarrow \mathbb{R}$  — измеримая функция, представляющая платежи и, возможно, зависящая от предыдущего развития, чтобы учесть условие выживания: поглощающий барьер, например, чтобы исключить банкротства. Введем зависимость от предыдущего периода  $\tau < t$  как зависимость от накопленной суммы  $\sum_{\tau \leq t} g(X_\tau)$ :

$$P_T^{(R)} = P_0^{(R)} + \sum_{t \leq T} \mathbb{1} \left( \sum_{\tau \leq t} g(X_\tau) > b \right) g(X_t), \quad (11.11)$$

где коэффициент  $b \in \mathbb{R}$  называется индексом выживаемости, а  $\mathbb{1}(\cdot)$  обозначает индикаторную функцию, принимающую значения  $\{0, 1\}$ ; условие в ее аргументе служит для обработки эргодичности или ее отсутствия [226].

### Комментарий 11.5

*P/L по определению соответствует реалистичным распределениям, а поглощающий барьер задает условие выживания.*

**Пространство частот.** В обычной литературе по психологии встречается два подхода.

**A — Подсчет предсказаний.** Доля успешных предсказаний

$$P_n^{(P)} = \frac{1}{n} \sum_{t \leq n} \mathbb{1}_{X_t \in \chi}, \quad (11.12)$$

где  $\mathbb{1}_{X_t \in \chi}$  — индикатор принадлежности случайной величины «диапазону предсказаний»  $\chi$  и  $n$  — число событий, то есть сделанных предсказаний.

**B — Обработка оценок (метод калибровки).** В отсутствие видимой чистой эффективности исследователи создают сложные показатели калибровки. Мы возьмем «золотой стандарт» в этой области — оценку по Брайеру (сформулированную де Финетти, [58]). Она хороша тем, что не позволяет спекулировать на несогласованности цен и точно калибрует вероятности: например, если участник ставит на то, что вероятность события равна 1, то он получит лучшую оценку только в том случае, если данное событие происходит все время.

$$\lambda_n^{(B)} = \frac{1}{n} \sum_{t \leq n} (f_t - \mathbb{1}_{X_t \in \chi})^2, \quad (11.13)$$

где  $f_t \in [0, 1]$  — это объявленная предсказателем вероятность события  $t$ . Лучший предсказатель получает минимальную оценку  $\lambda_n^{(B)}$ .

**Приложения. Соревнования M4 и M5.** На этих соревнованиях (Макридакис [159]) предсказатели стараются разными методами рассчитать точечную оценку и интервал возможных значений. На соревнованиях 2018 года использовалась система оценки участников M4, которая опирается на ряд оценок  $\lambda^{(M4)_j}$  и хорошо оценивает участников, которые предсказывают первый момент и дисперсию.

**Определение 11.7** (Оценки M4 для предсказаний первого момента)

Оценка точности на соревнованиях M4 (Макридакис и др. [159]) судит об участниках по следующим показателям с индексами  $j = 1, 2$

$$\lambda_n^{(M4)_j} = \frac{1}{n} \sum_i \frac{|X_{f_i} - X_{r_i}|}{s_j}, \quad (11.14)$$

где  $s_1 = \frac{1}{2} (|X_{f_i}| + |X_{r_i}|)$  и  $s_2$  (обычно) исходное среднее абсолютное отклонение для наблюдений, доступных до периода  $i$  (то есть средняя абсолютная погрешность по «наивным» прогнозам или по тестированию на выборке);  $X_{f_i}$  — предсказанная точечная оценка  $i$ -той случайной величины,  $X_{r_i}$  — реализация случайной величины, а  $n$  — число учитываемых опытов.

Другими словами, это применение средней абсолютной масштабированной погрешности (*mean absolute scaled error*, MASE) и симметричной средней абсолютной процентной погрешности (*mean absolute percentage error*, sMAPE) [133].

Предложенная оценка M5 (для соревнований 2020 года) добавляет предсказания экстремальных значений для изучаемой случайной величины и воспроизводит те же критерии, что предлагались для исходных случайных величин в Определении 11.7.

### 11.4.1. Вывод распределений

#### Распределение $P_n^{(p)}$

##### Замечание 13

Подсчет бинарных предсказаний  $P_n^{(p)}$  распределен асимптотически нормально со средним  $p$  и стандартным отклонением  $\sqrt{\frac{1}{n}(p-p^2)}$  независимо от класса распределения случайной величины  $X$ .

Это стандартные результаты, но их вывод приведен в приложении.

#### Распределение оценки по Брайеру $\lambda_n$

##### Теорема 3

Независимо от распределения случайной величины  $X$  и даже без предположения о независимости  $(f_1 - \mathbb{1}_{A_1}), \dots, (f_n - \mathbb{1}_{A_n})$ , при  $n < +\infty$  оценка  $\lambda_n$  имеет все моменты порядка  $q$ ,  $\mathbb{E}\lambda_n^q < +\infty$ .

*Доказательство.* При всех  $i$

$$(f_i - \mathbb{1}_{A_i})^2 \leq 1. \quad \blacksquare$$

На самом деле можно еще приблизиться к полному распределению оценки для разных независимых политик пары. Допустим, что бинарные предсказания  $f_i$  неза-



висимы и следуют бета-распределению<sup>1</sup>  $B(a, b)$  (которым аппроксимируются или поглощаются все одnogорбые распределения на носителе  $[0, 1]$  плюс распределение Бернулли, как две функции Дирака), и пусть  $p$  — доля успешных попыток,  $p = \mathbb{E} \mathbb{1}_{A_i}$ . Характеристическая функция для  $\lambda_n$ , где  $n$  — число применений оценки Брайера:

$$\begin{aligned} \varphi_n(t) = \pi^{n/2} & \left( 2^{-a-b+1} \Gamma(a+b) \times \right. \\ & \times \left( p {}_2\tilde{F}_2 \left( \frac{b+1}{2}, \frac{b}{2}; \frac{a+b}{2}, \frac{a+b+1}{2}; \frac{it}{n} \right) - \right. \\ & \left. \left. - (p-1) {}_2\tilde{F}_2 \left( \frac{a+1}{2}, \frac{a}{2}; \frac{a+b}{2}, \frac{a+b+1}{2}; \frac{it}{n} \right) \right) \right). \end{aligned} \quad (11.15)$$

Здесь  ${}_2\tilde{F}_2(., .; ., .; .)$  — регуляризация обобщенной гипергеометрической функции  ${}_2F_2$ :

$${}_2\tilde{F}_2(a_1, a_2; b_1, b_2; z) \triangleq \frac{{}_2F_2(a_1, a_2; b_1, b_2; z)}{\Gamma(b_1)\Gamma(b_2)}.$$

Обобщенная гипергеометрическая функция  ${}_pF_q(., .; .)$  раскладывается в ряд

$${}_pF_q(a; b; z) \triangleq \sum_{k=1}^{\infty} \frac{(a_1)_k \dots (a_p)_k}{(b_1)_k \dots (b_q)_k} \frac{z^k}{k!}$$

с использованием символа Похгаммера  $(a)_n = \prod_{i=0}^{n-1} (a+i)$ .

Таким образом, мы можем доказать, что при условии независимости слагаемых, как оговорено выше,

$$\lambda_n \xrightarrow{D} \mathcal{N}(\mu, \sigma_n), \quad (11.16)$$

где  $\mathcal{N}$  — гауссово распределение с указанным средним и среднеквадратическим отклонением. Доказательство и вывод параметров  $\mu, \sigma_n$  даны в приложении.

1 Бета-распределению  $B(a, b)$  с параметрами  $a > 0$  &  $b > 0$  точно или приблизительно следует непрерывная случайная величина  $X$  на носителе  $[0, 1]$ . По определению,  $X \sim B(a, b)$ , если плотность вероятности

$$f_X(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)},$$

где бета-функция  $B(a, b) \triangleq \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ , и форма графика зависит от параметров: так, при  $a > 1$  &  $b > 1$  плотность вероятности одnogорбая, с максимумом в точке  $\frac{a-1}{a+b-2}$ , и падает до 0 на концах носителя, тогда как при  $a < 1$  &  $b < 1$  график выпуклый вниз и на концах носителя плотность бесконечно возрастает. При любых параметрах среднее  $\frac{a}{a+b}$  и дисперсия  $\frac{ab}{(a+b)^2(a+b+1)}$ .

**Распределение P/L в экономике, или количественный показатель  $P_T^{(R)}$** **Замечание 14**

При условии выживания на момент  $T$  количественный показатель  $P_T^{(R)}$  следует распределению случайной величины  $g(x)$ .

Работает прямой подход, если нет поглощающего барьера (т. е. нет условия выживания).

**Распределение оценки М4.** Распределение абсолютного отклонения в том же классе, что сама случайная величина. Оценка Брайера является  $\ell_2$ -нормой<sup>1</sup> и основана на втором моменте, который всегда существует, поскольку де Финетти показал, что она эффективнее, чем просто вероятность, и на квадратических отклонениях. Однако при небинарных платежных функциях и жирных хвостах гораздо эффективнее полагаться на абсолютное отклонение, даже в тех случаях, когда второй момент существует [239].

**11.5. НЕВЕРБАЛЬНЫЕ ПЛАТЕЖНЫЕ ФУНКЦИИ И МАШИННОЕ ОБУЧЕНИЕ**

Предыдущие примеры фокусировались на простых платежных функциях, где иногда путаница между  $I_1$  и  $I_2$  безвредна (в ситуациях тонких хвостов). Однако учтем следующую сложность.

**Неотделимость вероятности при нелинейной платежной функции.** Когда мы начинаем рассматривать нелинейную платежную функцию  $g(\cdot)$  — то есть экономический или иной количественный отклик на случайную величину  $X$ , сила которого различна при разных уровнях  $X$ , — различие обостряется и последствия путаницы усугубляются.

**Комментарий 11.6 (Вероятность как подынтегральная величина)**

Вероятность нельзя пощупать саму по себе, она фигурирует только внутри интеграла или суммы. В мире экономики работают не индивидуальные вероятности, а количественно измеряемые платежные функции.

**Замечание 15: Неотделимость вероятности**

Пусть  $F: \rightarrow [0, 1]$  — распределение вероятностей и  $f$  — производная, и пусть  $g: \mathbb{R} \rightarrow \mathbb{R}$  — измеримая «платежная» функция. Очевидно, что для  $A'$ , подмножества множества  $A$ :

$$\int_{A'} g(x) dF(x) = \int_{A'} f(x) g(x) dx \neq \int_{A'} f(x) dx g\left(\int_{A'} dx\right)$$

или, в дискретных терминах, для функции вероятности  $\pi(\cdot)$ :

$$\sum_{x \in A'} \pi(x) g(x) \neq \sum_{x \in A'} \pi(x) g\left(\frac{1}{n} \sum_{x \in A'} x\right). \quad (11.17)$$

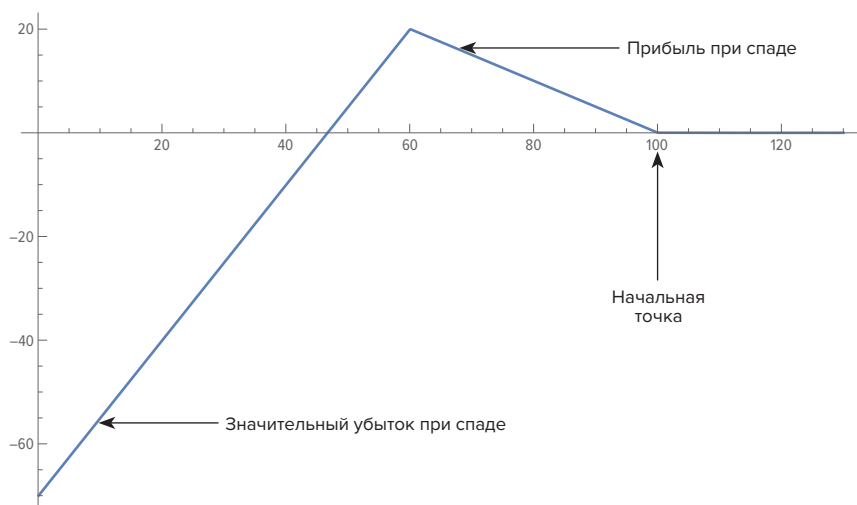
<sup>1</sup> То есть евклидовой нормой (метрикой):  $\|x\|_2 = \sqrt{\sum_i |x_i|^2}$ .

*Доказательство.* Непосредственное следует из неравенства Йенсена. ■

Другими словами, вероятность события совпадает с ожидаемым платежом, если, как мы видели ранее,  $g(x)$  — тета-функция Хевисайда.

Далее мы сосредоточимся на функциях, которые можно отслеживать математически или юридически, но которые невозможно точно выразить вербально, в таких терминах, как «мнение» или «предсказание».

**Непонимание  $g$ .** На Рисунке 11.5 показана история неправильного хеджирования *Morgan Stanley* — иллюстрация к тому, как плохо описывают нелинейный риск вложений такие вербальные концепции, как «коллапс». В 2007 игрок Уолл-стрит *Morgan Stanley* решил «хеджироваться» от «краха» рынка недвижимости, пока рынок недвижимости не начал проседать. Только в этом финансовом гиганте не понимали, что «крах» может иметь самую разную величину, в том числе хуже их худших опасений, и устроились так, чтобы получить выгоду в случае небольшого спада — ценой риска больших потерь в случае большого спада. В итоге они оказались правы в бинарном «предсказании» кризиса, но на своем мнимом хеджировании потеряли 10 миллиардов долларов.



**Рисунок 11.5:** История *Morgan Stanley*: элементарный пример нелинейной платежной функции, не поддающейся вербальному описанию. В жаргоне трейдеров на рынке деривативов такой риск называется «елочка» (*Christmas Tree*): приобретается опцион «пут» со страйк-ценой  $K$ , продается опцион «пут» с более низкой страйк-ценой  $K - \Delta_1$  и еще один с еще более низкой страйк-ценой  $K - \Delta_2$ , где  $\Delta_2 \geq \Delta_1 \geq 0$ .

На Рисунке F.6 показана более сложная платежная функция, так называемая «бабочка».

**Функция  $g$  и машинное обучение.** Заметим, что  $g$  отображается на различные функции машинного обучения, генерирующие исчерпывающее разнообразие нелинейностей на основе универсальной теоремы аппроксимации (Цыбенко [52]) или

обобщенных разложений платежной функции опциона (см. «Динамическое хеджирование» [225]).

Рассмотрим платежную функцию  $\rho: (-\infty, \infty) \rightarrow [K, \infty)$  случайной величины  $X \in \mathbb{R}$

$$\rho_{K,p}(x) = K + \frac{\ln(e^{p(x-K)} + 1)}{p}. \quad (11.18)$$

Можно выразить все нелинейные платежные функции  $g$ , используя веса  $w_i \in \mathbb{R}$ :

$$g(x) = \sum_i w_i \rho_{K_i,p}(x), \quad (11.19)$$

поскольку, используя аналогию,  $\rho_{K_i,p}$  отображается на страйк-цену  $K$  опциона «колл», если срок  $t$  нормализовать в 1, все ставки приравнять к 0 и только параметр  $\sigma$  оставить равным исходному среднеквадратическому отклонению.

Заметим, что математическое ожидание  $g(\cdot)$  представимо как сумма математических ожиданий функций ReLU:

$$\mathbb{E} g(x) = \sum_i w_i \mathbb{E} \rho_{K_i,p}(x). \quad (11.20)$$

Для дисперсии и остальных статистических параметров высших порядков получить явные формулы труднее.

### Комментарий 11.7

*Управление рисками старается изменить платежную функцию  $g(\cdot)$ , а не сделать «хорошие предсказания».*

Заметим, что  $\lambda$  — не количественный показатель, а целевой объект, к которому применяют различные количественные показатели.

### Выживание

Принятие решений — ответственное дело. Ложная калибровка может быть хорошей идеей, если снижает риск поглощения фирмы. Смотрите приложение в «Рискуя собственной шкурой» [226], где показана разница между вероятностью для статистического ансамбля и вероятностью для промежутка времени. Математическое ожидание результата для суммы  $n$  игроков за данный день отличается от математического ожидания для одного игрока за  $n$  дней, поскольку накладывается условие выживания.

В этой связи бессмысленно измерять эффективность участника, который с вероятностью 1 в конце концов обанкротится<sup>1</sup>.

## 11.6. ЗАКЛЮЧЕНИЕ

На практике в конечном счете важна итоговая эффективность (экономическая и прочая), и за некоторые «ошибки калибровки» жизнь не наказывает, а поощряет. Смещенная оценка дисперсии хорошо известна как способ повысить эффективность в машинном обучении [124], как необходимый механизм выживания при рациональном выборе (смотрите «Рискуя собственной шкурой» [226]) и как полезный механизм психологи-

<sup>1</sup> Ожидается, что на соревнованиях M5 это будет исправлено тем, что предсказателям будет предписано предсказывать минимум (или максимум) временного ряда. — Прим. автора.

ческой адаптации (Брайтон и Гигеренцер [33] приводят сильные доводы в пользу того, что если и называть данный эффект систематической ошибкой, то речь о полезной систематической ошибке). Когда ошибка не стоит вам потерь и даже улучшает ваши результаты, глупо называть ее ошибкой. Если же некоторый тип ошибок обходится вам дорого и присутствует в обществе давно, поищите в них скрытые эволюционные преимущества; например, задумайтесь о том, что принять медведя за скалу хуже, чем принять скалу за медведя.

Мы показали, что при управлении рисками не следует работать в пространстве вероятностей.

## 11.7. ПРИЛОЖЕНИЕ: ДОКАЗАТЕЛЬСТВА И ВЫВОД ФОРМУЛ

### 11.7.1. Распределение подсчета пари $P_n^{(p)}$

Мы имеем дело со средним по случайным величинам Бернулли, где есть хорошо известные результаты, но стоит повторить вывод некоторых из них. Для распределения Бернулли с параметром  $p$  характеристическая функция  $\psi(t) = 1 - p + p e^{it}$ . Для суммы  $N$  слагаемых логарифм характеристической функции

$$\psi'(\omega) = N \ln \psi \left( \frac{\omega}{N} \right)^1.$$

Обозначая  $\kappa_m$  кумулянт порядка  $m$ , порождаем кумулянты как

$$\kappa_m = -i^m \left. \frac{\partial^m \psi'(\omega)}{\partial \omega^m} \right|_{\omega \rightarrow 0}.$$

Таким образом,  $\kappa_1 = p$ ,  $\kappa_2 = \frac{(1-p)p}{N}$ ,  $\kappa_3 = \frac{(p-1)p(2p-1)}{N^2}$ ,  $\kappa_4 = \frac{(1-p)p(6(p-1)p+1)}{N^3}$ , и этим доказывается, что  $P_n^{(p)}$  сходится по закону больших чисел со скоростью  $\frac{1}{\sqrt{N}}$ , а по центральной предельной теореме сходится к гауссиане со скоростью  $\frac{1}{N}$  — потому что из приведенных выше кумулянтов эксцесс<sup>2</sup> составляет  $3 - \frac{6(p-1)p+1}{N(p-1)p}$ <sup>3</sup>.

### 11.7.2. Распределение оценки Брайера

**Базовая вероятность  $f$ .** Сначала рассмотрим распределение  $f$  базовой вероятности.

Применим бета-распределение, которое охватывает и условный, и безусловный случай (зависит от параметров  $a$  и  $b$  в формуле 11.15).

1 В самом деле, логарифмы характеристической функции складываются при сложении независимых случайных величин, так что  $\ln \psi_{X_1+\dots+X_N}(Nt) = N \ln \psi_{X_1}(t)$ ; заменяя  $t = \omega/N$ , получаем

$\ln \psi_{X_1+\dots+X_N}(\omega) = N \ln \psi_{X_1}(\omega/N)$ .

2 Эксцесс представим через кумулянты как  $3 + \kappa_4/\kappa_2^2$ . — Прим. перев.

3 В самом деле, второй кумулянт, равный дисперсии, при больших  $N$  убывает  $\sim \frac{1}{N}$  и, значит, среднеквадратическое отклонение убывает  $\sim \frac{1}{\sqrt{N}}$ . Эксцесс сходится к 3, и, значит, распределение сходится к гауссову.

**Распределение вероятностей.** Чтобы показать логику заявлений о робастности распределения *вероятностей вероятности* — другими словами, о том, что распределение вероятности случайной величины  $X$  не зависит от распределения случайной величины  $X$  (см. [68] и [143]), — напомним стандартный результат, необходимый для понимания и проверки непараметрических подходов и восходящий к Колмогорову [146].

Интеграл вероятности преобразуется так. Пусть непрерывная случайная величина  $X$  следует распределению вероятностей с кумулятивной функцией  $F_X$ , и пусть случайная величина  $U$  представляет собой преобразование случайной величины  $X$  функцией  $F_X$ , то есть  $U = F_X(X)$ . Тогда в отсутствие дополнительных условий случайная величина  $U$  следует равномерному распределению от 0 до 1. Доказывается это так. В любой точке  $u \in [0, 1]$  значение кумулятивной функции распределения вероятностей случайной величины  $U$

$$\mathbb{P}(U \leq u) = \mathbb{P}(F_X(X) \leq u) = \mathbb{P}(X \leq F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u, \quad (11.21)$$

то есть представляет собой функцию равномерного распределения.

В нашем случае, очевидно, мы имеем дело с (1) плотностью  $f$  бета-распределения (равномерного в том частном случае, когда предсказания нисколько не точны) и (2) дискретной случайной величиной  $\mathbb{1}_{A_j}$ , распределенной по Бернулли с вероятностью  $p$ .

Рассмотрим общий случай. Обозначим  $g_{a,b}$  функцию плотности бета-распределения вероятностей:

$$g_{a,b}(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, \quad 0 < x < 1.$$

Формулы для среднего и для дисперсии оценки Брайера получаются громоздкие, но все же в пределах возможностей выкладок вручную:

$$\begin{aligned} \mu &= (a(a+1)(1-p) + b(b+1)p) \frac{\Gamma(a+b)}{\Gamma(a+b+2)} \\ \sigma_n^2 &= -\frac{1}{n} \left( \frac{a(a+1)(1-p) + b(b+1)p}{(a+b)(a+b+1)} \right)^2 + \\ &+ \frac{(a+b)(a+b+1)}{(a+b+2)(a+b+3)} (p(a-b)(a+b+3)(a(a+3) + (b+1)(b+2)) - a(a+1)(a+2)(a+3)). \end{aligned}$$

Далее мы убеждаемся, что у оценки Брайера эксцесс ниже 3, то есть ее хвосты тоньше, чем у гауссианы.

*Доказательство.* Начнем с  $y_j = f - \mathbb{1}_{A_j}$ , разности между случайной величиной, которая следует бета-распределению, и дискретной случайной величиной, которая следует распределению Бернулли, где у обеих один и тот же индекс  $j$ . Для этой разности характеристическая функция  $\Psi_{(y_j)}(t) = (1 + p(-1 + e^{-it})) {}_1F_1(a; a+b; it)$ , где  ${}_1F_1(; ; \cdot)$  — функция Куммера, или вырожденная гипергеометрическая функция:

$${}_1F_1(a; b; z) = \sum_{k=0}^{\infty} \frac{a_k}{b_k} \frac{z^k}{k!}.$$

Отсюда для  $y_j^2 = (f - \mathbb{1}_{A_j})^2$  характеристическая функция

$$\begin{aligned} \psi_{(y^2)}(t) = \sqrt{2} \cdot 2^{-a-b+1} \Gamma(a+b) & \left( p {}_2\tilde{F}_2 \left( \frac{b+1}{2}, \frac{b}{2}; \frac{a+b}{2}, \frac{a+b+1}{2}; it \right) + \right. \\ & \left. + (1-p) {}_2\tilde{F}_2 \left( \frac{a+1}{2}, \frac{a}{2}; \frac{a+b}{2}, \frac{a+b+1}{2}; it \right) \right), \end{aligned} \quad (11.22)$$

где  ${}_2\tilde{F}_2(., .; ., .; .)$  — регуляризация обобщенной гипергеометрической функции  ${}_2F_2$ :

$${}_2\tilde{F}_2(a_1, a_2; b_1, b_2; z) \triangleq \frac{{}_2F_2(a_1, a_2; b_1, b_2; z)}{\Gamma(b_1)\Gamma(b_2)},$$

и обобщенная гипергеометрическая функция  ${}_pF_q(., .; .)$  раскладывается в ряд

$${}_pF_q(a; b; z) \triangleq \sum_{k=1}^{\infty} \frac{(a_1)_k \dots (a_p)_k}{(b_1)_k \dots (b_q)_k} \frac{z^k}{k!}$$

с использованием символа Похгаммера  $(a)_n = \prod_{i=0}^{n-1} (a+i)$ .

Затем используем сходимость по распределению для среднего  $\frac{1}{n} \sum_i^n y_i^2$ . Предел характеристической функции

$$\lim_{n \rightarrow \infty} \left( \psi_{(y^2)} \left( \frac{t}{n} \right) \right)^n = \exp \left( -it \frac{p(a-b)(a+b+1) - a(a+1)}{(a+b)(a+b+1)} \right), \quad (11.23)$$

то есть как у вырожденной гауссианы (функции Дирака) с коэффициентом положения

$$\frac{p(b-a) + \frac{a(a+1)}{a+b+1}}{a+b}.$$

И, наконец, оценим скорость сходимости высших моментов к гауссовым. Для

$$\text{этого рассмотрим поведение четвертого кумулянта, } \kappa_4 = -i \left. \frac{\partial^4 \Psi(.)}{\partial t^4} \right|_{t \rightarrow 0}.$$

1) В случае максимальной энтропии  $a = b = 1$ <sup>2</sup> получаем

$$\left. \kappa_4 \right|_{a=1, b=1} = -\frac{6}{7n}$$

независимо от  $p$ .

1 В самом деле, обозначим  $\frac{p(b-a) + \frac{a(a+1)}{a+b+1}}{a+b} = -\frac{p(a-b)(a+b+1) - a(a+1)}{(a+b)(a+b+1)} = x_0$ . Если случайная величина  $X$  следует вырожденному распределению с плотностью вероятности вида дельта-функция Дирака с пиком в точке  $x_0$ ,  $f_X(x) = \delta(x - x_0)$ , то соответствующая характеристическая функция  $\psi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx = \int_{-\infty}^{\infty} e^{itx} \delta(x - x_0) dx = \exp(-itx_0)$ . И наоборот, случайная величина  $X$  с характеристической функцией  $\exp(-itx_0)$  вырождена в одно-единственное значение,  $x_0$ .

2 Бета-распределение  $B(1, 1)$  — равномерное распределение на отрезке  $[0, 1]$ .

2) В случае максимальной дисперсии<sup>1</sup> по правилу Лопиталья

$$\lim_{\substack{a \rightarrow 0 \\ b \rightarrow 0}} \kappa_4 = - \frac{6(p-1)p+1}{n(p-1)p}.$$

Таким образом,  $\frac{\kappa_4}{\kappa_2^2} \rightarrow 0$  при  $n \rightarrow \infty$  со скоростью  $n^{-1}$ . ■

Далее можем извлечь функцию плотности вероятности для оценки Брайера при  $N=1$ .

При  $0 < z < 1$

$$p(z) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{(1-p)z^{a/2}(1-\sqrt{z})^b + p z^{b/2}(1-\sqrt{z})^a}{2(1-\sqrt{z})z}. \quad (11.24)$$

---

<sup>1</sup> Бета-распределение  $B(a, b)$  при  $a, b \rightarrow 0$  стремится к бесконечной плотности вблизи 0 и 1.



## ПРЕДСКАЗАНИЯ ВЫБОРОВ КАК МАРТИНГАЛ: АРБИТРАЖНЫЙ ПОДХОД<sup>1</sup>

Изучим<sup>1</sup> влияние неопределенности на бинарные прогнозы применительно к выборам. Стандартный результат в финансовой математике — что с ростом волатильности ценных бумаг давление арбитража сжимает торговлю соответствующими бинарными опционами к точке 50 %, и вариативность уменьшается на протяжении времени до истечения срока опционов. Контринтуитивным образом высокая неопределенность в стоимости ценных бумаг приводит к низкой волатильности соответствующего бинарного опциона.

Теоретически этот эффект должен присутствовать во всех доменах, где генерируются бинарные предсказания... однако мы наблюдаем нарушение данного принципа в ряде областей, где делаются бинарные предсказания, и яркий пример — выборы президента США в 2016 году. Поразительные ошибки политологов и прогнозистов, когда (1) кандидату Д. Трампу давали от 0,1 % до 3 % шансов, (2) прогнозы корректировали скачком с 48 % до 15 %, причем оба значения оправдывали высокой неопределенностью.

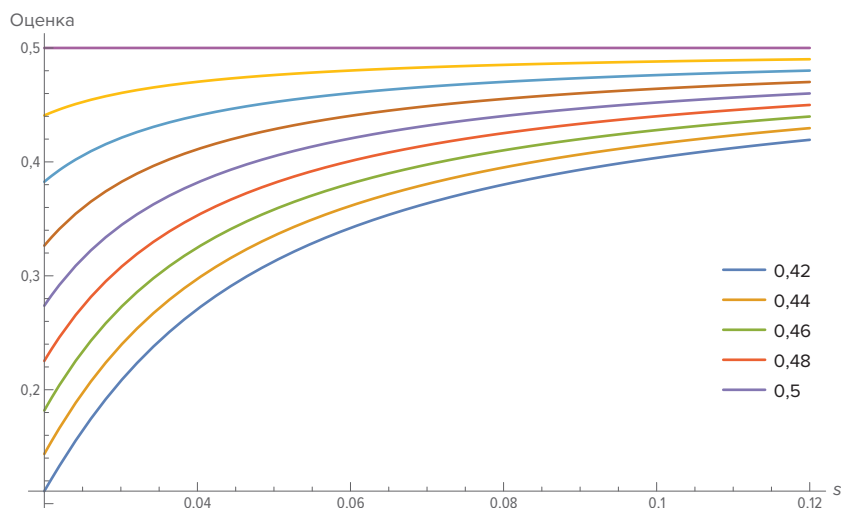
Качеству предсказания выборов принято давать статическую оценку по методу де Финетти, который состоит в минимизации оценки Брайера — количественного показателя расхождения с финальным исходом (это стандарт отслеживания точности оценщиков вероятности в самых разных областях, от выборов до погоды). По-видимому, оценивать промежуточные корректировки оценок не принято нигде за пределами

<sup>1</sup> Автор благодарен Дхруву Мадеке и Рафаэлю Дуади за подробные и обстоятельные дискуссии о статье, а также за тщательный аудит доказательств на протяжении всех итераций и, что особо мучительно, смены обозначений. Питер Карр помог при обсуждении свойств ограниченного мартингала и преобразований. Спасибо за комментарии Давиду Шимко, Анджело Ласневскому и Эндрю Папаниколау.

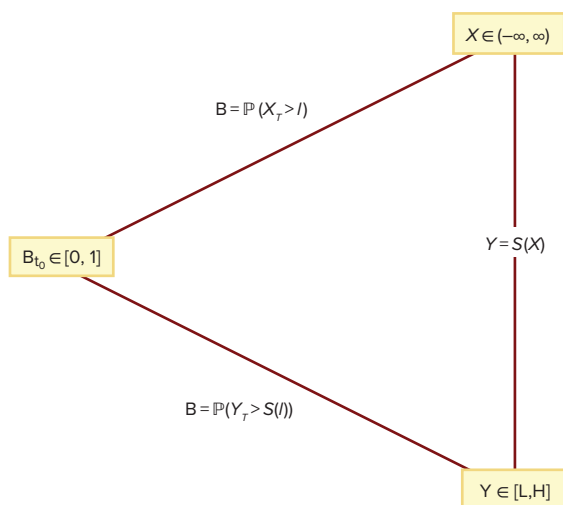
Спасибо Артуру Брайтману за подбор литературы по численным аппроксимациям ряда логистико-нормальных интегралов. Спасибо участникам Политехнического института Нью-Йоркского университета (*Tandon School of Engineering*) и семинаров агентства «Блумберг» по финансовой математике (*Bloomberg Quantitative Finance Seminars*). Я также благодарен Бруно Дюпиру, Майку Лоулеру, главным редакторам журнала *Quantitative Finance* и многочисленным доброжелателям в социальных сетях.

Дхрув Мадека в годы работы в агентстве «Блумберг» занимался аналогичной проблемой и пришел к тем же взаимосвязям между волатильностью оценки и ее границами и к тем же границам арбитража. За все ошибки отвечаю я. — *Прим. автора.*

финансовой математики, где такое встречается, как на практике, так и в литературе. Тем самым обычная практика идет вразрез с принципом самого де Финетти, считавшего вероятность двусторонней ценой «выбора».



**Рисунок 12.1:** Арбитражная «стоимость» (оценка) выборов в зависимости от ожидаемой доли голосов  $Y \in [0, 1]$  при разной ожидаемой волатильности  $s$  величины  $Y$  в период от настоящего момента до результатов выборов. Видно, что при высокой неопределенности результат оценки приближается к 0,5 независимо от оцениваемой разности голосов



**Рисунок 12.2:**  $X$  — ненаблюдаемая (теневая) случайная величина на открытом носителе  $\mathbb{R}$ ,  $Y$  — ее отображение на «голоса избирателей» или «голоса в коллегии выборщиков» взаимно-однозначной сигмодальной функцией  $S(\cdot)$ . Показана возможность оценки бинарного предсказания  $B_{t_0}$  через распределение каждой из величин  $X$ ,  $Y$  при использовании соответствующих преобразований

В этой главе мы будем придерживаться динамических моделей с непрерывным временем, основанных на принципах финансовой математики. Мы попробуем доказать, что теоретико-вероятностная оценка исхода выборов, сгенерированная «оценщиком», должна рассматриваться как торгуемая цена, то есть как стоимость бинарного опциона, подчиненного арбитражным границам (тем более что бинарные опционы и на практике используются на рынках пари). Будущие скорректированные оценки должны быть совместимы с мартингальными ценами; в противном случае на оценщике начнут спекулировать, совершая с ним интертемпоральные купли-продажи.

Математические сложности начинаются, когда мы переходим к непрерывному времени и применяем стандартный мартингальный подход: предсказание вероятности для ценных бумаг, стоимость которых пробегает  $[0, 1]$ . Согласно этому подходу, мы создаем дублирующий (теневого) мартингальный процесс  $Y$  на интервале  $[L, H]$ , преобразуя арифметическое броуновское движение  $X$  на  $(-\infty, \infty)$ , и назначаем выборам соответствующую цену. Дублирующий процесс  $Y$  можно представить, например, как число голосов, которое требуется для победы на выборах. Сложность в том, что при таком преобразовании  $X$  не может быть мартингалом, если  $Y$  — мартингал, и наоборот.

Обработкой случайной величины  $Y$  можно построить арбитражную взаимосвязь между волатильностью оценки вероятности и волатильностью соответствующей случайной величины, например числа голосов. Так удастся доказать, что при высокой неопределенности финального исхода: (1) арбитражное значение предсказания (как бинарного опциона) приближается к 50 % и (2) оценка не должна претерпевать больших изменений, даже если поллы или другие базовые величины покажут существенные вариации<sup>1</sup>.

Ценообразовательные связи строятся между (1) стоимостью бинарного опциона (то есть вероятностью предсказания), (2) оценкой  $Y$  и (3) волатильностью оценки  $Y$  на протяжении остающегося срока (смотрите Рисунки 12.1 и 12.2).

### 12.0.1. Основные результаты

Для удобства изложения начнем с обозначений.

Обозначение	
$Y_0$	наблюдаемая оценка доли голосов в диапазоне $[0, 1]$ в момент $t_0$ (это могут быть голоса избирателей или голоса в коллегии выборщиков, но нужно последовательно придерживаться чего-то одного)
$T$	дата обнародования окончательных результатов выборов или срок окончания опциона (в годах)
$t_0$	дата текущей оценки (в годах); соответственно остающееся время представляет собой разность $T - t_0$

<sup>1</sup> Центральное свойство нашей модели — она не позволяет бинарной оценке  $B(\cdot)$  варьировать больше, чем оцениваемая случайная величина  $Y$ : при соперничестве двух кандидатов оценка будет ограничена сверху или снизу отклонением  $Y$ , если будет ниже (или выше) 0,5. На практике мы можем столкнуться с вероятностью 98 % победы и 2 % поражения при гораздо более узкой разности оцениваемого числа голосов 47 % против 53 %; в случае высокой неопределенности при нашем подходе вероятности не расходятся с оцениваемым числом голосов. Но он консервативен в том смысле, что не дает более высокой доли. — Прим. автора.

$s$	среднегодовая волатильность величины $Y$ за оставшееся время. Не теряя общности, мы приняли $s$ за константу, хотя она может зависеть от времени
$B(\cdot)$	«вероятность предсказания», то есть непрерывная по времени арбитражная оценка результатов выборов, задающая арбитражные связи между $B(\cdot)$ , $Y_0$ и волатильностью $s$

## Основные результаты

$$B(Y_0, \sigma, t_0, T) = \frac{1}{2} \operatorname{erfc} \frac{l - \operatorname{erf}^{-1}(2Y_0 - 1)e^{\sigma^2(T-t_0)}}{\sqrt{e^{2\sigma^2(T-t_0)} - 1}}, \quad (12.1)$$

где

$$\sigma \approx \frac{\sqrt{\ln \left( 2\pi s^2 e^{2 \operatorname{erf}^{-1}(2Y_0 - 1)^2} + 1 \right)}}{\sqrt{2} \sqrt{T - t_0}}, \quad (12.2)$$

и  $l$  — пороговое значение, необходимое для победы (по умолчанию 0,5);  $\operatorname{erfc}$  — обычная дополнительная функция ошибок:  $\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$  и  $\operatorname{erf} z = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ .

Здесь будет уместно ответить на типичный комментарий со стороны статистиков и других специалистов вне финансовой математики: «Почему не использовать для  $Y$  обычное бета-распределение?» Потому что (1) в данной статье мы поставили задачу исследовать бинарные прогнозы, состоятельные во времени (на которых невозможно спекулировать), и (2) нам неизвестны непрерывные во времени стохастические процессы, описываемые обычным ограниченным распределением, включая бета-распределение.

### 12.0.2. Организация

Остальные части статьи организованы следующим образом. Во-первых, мы показываем процесс для  $Y$  и необходимые преобразования на основе некоторого специального броуновского движения. Во-вторых, мы выводим арбитражное соотношение, при помощи которого получается формула (12.1). Наконец, мы обсуждаем подход де Финетти и показываем, как мартингальная стоимость связана с минимизацией обычной стандартной оценки в отрасли предсказаний, а именно оценки Брайера.

**Комментарий об отсутствии аналитических решений для  $\sigma$ .** Заметим, что мы не имеем для  $Y$  аналитического представления интеграла, отражающего вариацию

$$\int_{t_0}^T \frac{\sigma}{\sqrt{\pi}} e^{-\operatorname{erf}^{-1}(2y_t - 1)^2} dt,$$

хотя для  $X$  интеграл вариации вычислимый. Методами распространения неопределенности мы сможем получить аналитические решения для плотности вероятности  $Y$ , но не моменты, потому что логистико-нормальный интеграл не допускает простых разложений [195].

**Распределения временного среза для  $X$  и  $Y$ .** Распределение временного среза — это функция плотности вероятности случайной величины  $Y$  от момента времени  $t$ , то есть представление для одного периода времени, начинающегося в момент  $t$  с  $y_0 = \frac{1}{2} + \frac{1}{2} \operatorname{erf} x_0$ . И наоборот, случайная величина  $X$  при заданном значении  $y_0$  и соответствующем значении  $x_0$ , как можно доказать, распределена нормально для периода времени  $T - t_0$  с параметрами

$$\begin{aligned}\mathbb{E}(X, T) &= x_0 e^{\sigma^2(T-t_0)}, \\ \mathbb{V}(X, T) &= \frac{e^{2\sigma^2(T-t_0)} - 1}{2}\end{aligned}$$

и эксцессом 3. Преобразованием вероятности получаем  $\varphi$ , соответствующее распределение  $Y$  с начальным значением  $y_0$ :

$$\begin{aligned}\varphi(y; y_0, t) &= \frac{1}{\sqrt{e^{2\sigma^2(T-t_0)} - 1}} \exp\left(\operatorname{erf}^{-1}(2y-1)^2 - \right. \\ &\quad \left. - \frac{\operatorname{cth} \sigma^2 t - 1}{2} \left(\operatorname{erf}^{-1}(2y-1) - \operatorname{erf}^{-1}(2y_0-1)e^{\sigma^2(T-t_0)}\right)^2\right),\end{aligned}\quad (12.3)$$

и  $\mathbb{E}Y_t = Y_0$ .

Что касается дисперсии,  $\mathbb{E}Y^2$ , то она, как отмечалось выше, не допускает аналитического вывода ни из  $\varphi(\cdot)$ , ни из стохастического интеграла; впрочем, для нее нетрудно получить оценку из заданного аналитической формулой распределения  $X$  методами распространения неопределенности для первых двух моментов (дельта-метод).

Поскольку дисперсия функции  $f$  случайной величины  $X$  с конечными моментами приближается как  $\mathbb{V}f(X) = f'(\mathbb{E}X)^2 \mathbb{V}X$ ,

$$\begin{aligned}\left. \frac{\partial S^{-1}(y)}{\partial y} \right|_{y=Y_0} s^2 &\approx \frac{e^{2\sigma^2(T-t_0)} - 1}{2}, \\ s &\approx \sqrt{e^{-2\operatorname{erf}^{-1}(2Y_0-1)^2} \frac{e^{2\sigma^2(T-t_0)} - 1}{2\pi}}.\end{aligned}\quad (12.4)$$

Аналогичными вычислениями в обратную сторону находим

$$\sigma \approx \sqrt{\frac{\ln\left(2\pi s^2 e^{-2\operatorname{erf}^{-1}(2Y_0-1)^2} + 1\right)}{2(T-t_0)}},$$

что приведено как основной результат в формуле (12.2).

Заметим, что, добавляя в наши разложения высшие моменты, не достигаем существенного повышения точности; хотя  $s$  не линейна вблизи центра, диапазон значений для волатильности (общего числа голосов или голосов в коллегии выборщиков) слишком узок, чтобы существенно повлиять на слагаемые высших порядков сверх того, что диктуется ограниченностью сигмоидных преобразований.

### 12.0.3. Обсуждение нейтральности к риску

Мы применяем стоимость, нейтральную к риску, как естественный выбор по умолчанию, когда нет мотива выбрать другой подход. Даже если случайная величина  $Y$  не является рыночным товаром, добавление рискованной премии в процесс оценки арбитражной стоимости необходимо влечет отрицательную премию для другого кандидата или кандидатов, что трудно оправдать.

Заметим далее, что стоимости опционов или бинарных пари должны удовлетворять запрету «голландской книги» (формулировка де Финетти об исключении арбитража, смотрите в [97]), т. е. рынок бинарных опционов с разумными ценами, интерпретируемыми как предсказания вероятности, исключает набор пари, с которым вы останетесь в выигрыше при всех возможных исходах. И, наконец, любое отклонение от нейтральности к риску понизит оценку Брайера (о которой будет ниже), поскольку будет отступлением от финального предсказания.

Заметим также, что не делаем никаких допущений о ставке финансирования, которые обычно делаются в финансовых дискуссиях.

### 12.1. СТОИМОСТЬ В СТИЛЕ БАШЕЛЬЕ

Пусть  $F(\cdot)$  — функция случайной величины  $X$ , удовлетворяющая критерию

$$dX_t = \sigma^2 X_t dt + \sigma dW_t. \quad (12.5)$$

Мы хотим показать, что для  $X$  существует простая цена опциона Башелье<sup>1</sup>  $B(\cdot)$ . Идея исключения арбитража заключается в том, чтобы непрерывно обновляемое предсказание мартингаловой случайной величины само было мартингалом.

Применяя формулу<sup>2</sup> Ито к  $F \triangleq B$  при условии, что  $X$  удовлетворяет (12.5), получаем

$$dF = \left( \sigma^2 X \frac{\partial F}{\partial X} + \frac{\sigma^2}{2} \frac{\partial^2 F}{\partial X^2} + \frac{\partial F}{\partial t} \right) dt + \sigma \frac{\partial F}{\partial X} dW,$$

откуда с учетом  $\frac{\partial F}{\partial t} \triangleq 0$  следует, что  $F$  должна удовлетворять дифференциальному уравнению в частных производных

$$\sigma^2 X \frac{\partial F}{\partial X} + \frac{\sigma^2}{2} \frac{\partial^2 F}{\partial X^2} + \frac{\partial F}{\partial t} = 0, \quad (12.6)$$

и это условие делает  $B$  мартингалом без дрейфа.

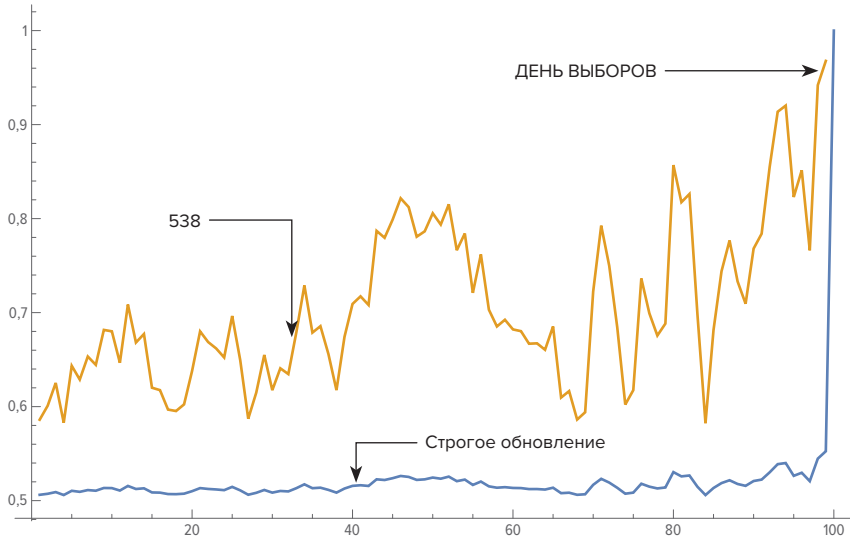
Для бинарного опциона «колл» имеем конечные условия

$$B(X, t) \triangleq F, \quad F_T = \mathfrak{H}(x - l),$$

где  $\mathfrak{H}(\cdot)$  — тета-функция Хевисайда

<sup>1</sup> С «Теории спекуляций» (1900) Башелье начинаются и моделирование броуновского процесса, и моделирование ценообразования опционов. Модель Башелье, она же нормальная модель, отличается от модели Блэка — Шоулза и логнормальной модели.

<sup>2</sup> Формула Ито (*Itô's lemma*) позволяет заменять переменную в стохастическом дифференциальном уравнении.



**Рисунок 12.3:** Теоретический подход (верхняя кривая) и практика (нижняя кривая). Графики показывают, что процесс оценки не может отслеживать волатильность оценки числа голосов (в коллегии выборщиков или другого), поскольку нарушает границы арбитража

$$\vartheta(x-l) \triangleq \begin{cases} 1, & x \geq l \\ 0, & x < l \end{cases}$$

и  $l$  — порог, при начальном условии  $x_0$  в момент времени  $t_0$  и конечном условии в момент времени  $T$

$$\frac{1}{2} \operatorname{erfc} \frac{x_0 e^{\sigma^2 t} - l}{\sqrt{e^{2\sigma^2 t} - 1}},$$

представляющем собой просто функцию выживания нормального распределения с параметрами, отвечающими процессу случайной величины  $X$ .

Аналогичным образом замечаем для аргумента взаимно-однозначного отображения  $S$  (можно использовать аргументы из борелевского множества), что

$$\vartheta(y-S(l)) \triangleq \begin{cases} 1, & y \geq S(l) \\ 0, & y < S(l) \end{cases}$$

и можно задать цены в альтернативном процессе  $B(Y, t) = \mathbb{P}(Y > \frac{1}{2})$  с порогом  $\frac{1}{2}$  или с любым порогом  $l$ , задав цены

$$B(Y_0, t_0) = \mathbb{P}(x > S^{-1}(l)).$$

Цены согласно доле голосов представимы как

$$B(Y_0, \sigma, t_0, T) = \frac{1}{2} \operatorname{erfc} \frac{l - \operatorname{erf}^{-1}(2Y_0 - 1)e^{\sigma^2(T-t_0)}}{\sqrt{e^{2\sigma^2(T-t_0)} - 1}};$$

эта формула включена в главные результаты как (12.1), что иначе выражается более громоздкой формулой

$$B(Y_0, \sigma, t_0, t) = \frac{1}{\sqrt{e^{2\sigma^2 t} - 1}} \int_{t_0}^t \exp \left( \operatorname{erf}^{-1}(2y-1)^2 - \frac{\operatorname{cth} \sigma^2 t - 1}{2} \left( \operatorname{erf}^{-1}(2y-1) - \operatorname{erf}^{-1}(2Y_0-1)e^{\sigma^2 t} \right)^2 \right) dy.$$

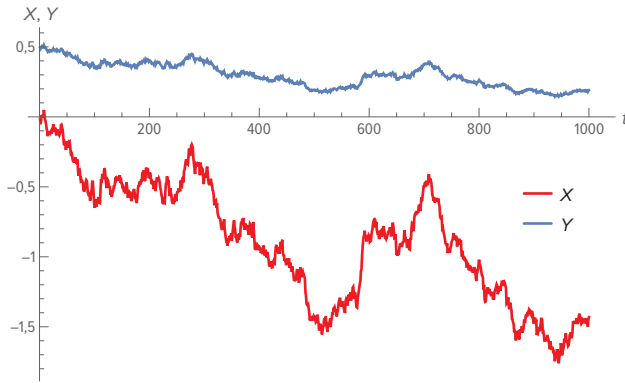


Рисунок 12.4: Процесс и двойственный процесс

## 12.2. ОГРАНИЧЕННЫЙ ДВОЙСТВЕННЫЙ МАРТИНГАЛЬНЫЙ ПРОЦЕСС

$Y_T$  — конечное значение процесса в день выборов — определено нами в промежутке  $[0, 1]$ , хотя допускает обобщение на произвольный отрезок  $[L, H]$ , где  $L, H \in [0, \infty)$ . Порог победы данного кандидата фиксируется на уровне  $l$ .  $Y$  может соответствовать числу голосов избирателей, числу голосов в коллегии выборщиков или иному количественному показателю. Мы предполагаем, что  $Y_t$  — промежуточная реализация процесса в момент времени  $t$ , синтезированная по опросам (скорректированные оценки) или иным системам.

Далее мы создаем для неограниченного арифметического стохастического процесса — ограниченный «двойственный» стохастический процесс, используя сигмоидальное преобразование. Оно бывает полезно для отображения таких процессов, как ограниченный электоральный процесс, на броуновское движение или для отображения ограниченной платежной функции на неограниченную; см. Рисунок 12.2.

### Предложение 12.1

При сигмоидальных преобразованиях  $S: x \mapsto y, \mathbb{R} \rightarrow [0, 1]$  вида (а)  $\frac{1 + \operatorname{erf} x}{2}$  или (б)  $\frac{1}{1 + \exp(-x)}$ , если  $X$  — мартингал, то  $Y$  — мартингал только при  $Y_0 = \frac{1}{2}$ , а если  $Y$  — мартингал, то  $X$  — мартингал только при  $X_0 = 0$ .

Доказательство. Схема такая. Из формулы Ито слагаемое дрейфа для  $dX_t$  превращается в (1)  $\sigma^2 X(t)$  или (2)  $\frac{1}{2} \sigma^2 \operatorname{th} \frac{X(t)}{2}$ , где  $\sigma$  — волатильность соответственно при преобразованиях вида (а)  $X_t$  и (б)  $X_t$  при мартингальном  $Y$ . Дрейф  $dY_t$  принимает вид

$$(1) \frac{\sigma^2 e^{-\operatorname{erf}^{-1}(2Y-1)^2} \operatorname{erf}^{-1}(2Y-1)}{\sqrt{\pi}} \text{ или } (2) \frac{1}{2} \sigma^2 Y(Y-1)(2Y-1), \text{ если } X \text{ — мартингал.} \quad \blacksquare$$



Поэтому мы выберем случай, где мартингалом является случайная величина  $Y$ , и представим подробности преобразования (а). Подробности процесса разработаны Карром [35]. Пусть  $X$  — арифметическое броуновское движение (12.5) с дрейфом, зависящим от  $X$ , и постоянным масштабом  $\sigma$ :

$$dX_t = \sigma^2 X_t dt + \sigma dW_t, \quad 0 < t < T < +\infty.$$

Заметим сходство с процессом Орнштейна — Уленбека, обычно записываемого как  $dX_t = \vartheta(\mu - X_t)dt + \sigma dW_t$ , только у нас  $\mu = 0$  и против правил используется отрицательный коэффициент возвращения к среднему, то есть это «коэффициент отталкивания от среднего»:  $\vartheta = -\sigma^2$ <sup>1</sup>.

Отображение случайной величины  $X \in (-\infty, \infty)$  на двойственный процесс  $Y$  построим так. При  $S: \mathbb{R} \rightarrow [0, 1]$ ,  $Y = S(X)$  и

$$S(x) = \frac{1 + \operatorname{erf} x}{2}$$

двойственный процесс, связанный с исходным взаимно-однозначным преобразованием  $S$ , так что  $y \triangleq S(x)$ , будет дифференцироваться по формуле Ито (поскольку преобразование  $S(\cdot)$  дважды дифференцируемо и  $\frac{\partial S}{\partial t} = 0$ ):

$$dS = \left( \sigma^2 X \frac{\partial S}{\partial X} + \frac{\sigma^2}{2} \frac{\partial^2 S}{\partial X^2} \right) dt + \sigma \frac{\partial S}{\partial X} dW,$$

— откуда при нулевом дрейфе процесс можно записать как

$$dY_t = s(Y_t) dW_t$$

при всех  $t > \tau$ , где  $\mathbb{E}(Y_t | Y_\tau) = Y_\tau$  и масштаб

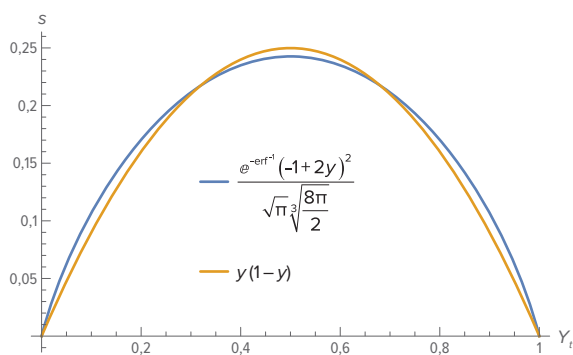
$$s(y) = \frac{\sigma}{\sqrt{\pi}} e^{-\operatorname{erf}^{-1}(2y-1)^2},$$

как видно на Рисунке 12.5, аппроксимируется квадратичной функцией вида  $y(1-y)$  с постоянным множителем<sup>2</sup>.

Мы можем восстановить формулу (12.5), обратив преобразование:  $S^{-1}(y) = \operatorname{erf}^{-1}(2y - 1)$  и еще раз применив формулу Ито. Вследствие калибровочной инвариантности цены опционов одинаковы что на основе  $X$ , что на основе  $Y$ , даже если один из процессов обладает дрейфом, а другой представляет собой мартингал. Другими словами, разными исследователями могут применяться разные оценки к электоральному порогу или более сложной случайной величине  $X$ , но результаты будут одинаковы. Выбранный нами метод вычисления цены опциона на основе  $X$  хорошо известен — это цена опциона по модели Башелье.

1 Обозначение  $\vartheta$  для коэффициента возвращения к среднему (*mean reversion coefficient*) не следует путать с обозначением  $\vartheta(\cdot)$  для функции Хевисайда.

2 Обозначение  $s(\cdot)$  для переменного масштаба случайной величины  $Y$  не следует путать с обозначением  $S(\cdot)$  для сигмоидного преобразования случайной величины  $X$  в  $Y$ .



**Рисунок 12.5:** Мгновенная волатильность случайной величины  $Y$  как функция величины  $Y$  выглядит по существу одинаково при двух преобразованиях случайной величины  $X$ . Функция масштаба, масштабированная множителем  $\frac{1}{\sqrt[3]{4\pi}}$ , приближается квадратичной функцией  $y - y^2$ . Волатильность падает при удалении от  $\frac{1}{2}$  к краям и этим удерживает случайную величину  $Y$  в интервале  $(0, 1)$ . Для простоты мы приняли  $\sigma = t = 1$

### 12.3. СВЯЗЬ С ОЦЕНЩИКОМ ВЕРОЯТНОСТИ ДЕ ФИНЕТТИ

В этом разделе кратко рассказывается об обычном подходе к оценке вероятности. Великий человек де Финетти [58] доказал, что «оценка» для «вероятности» некоторой реализации случайной величины в  $[0, 1]$  требует нелинейной функции потерь — и тем самым определение де Финетти отличается от отношения  $P/L$  для трейдера на рынке бинарных пари.



**Рисунок 12.6:** Бруно де Финетти (Bruno de Finetti; 1906–1985). Специалист по теории вероятностей, философ и исследователь математических начал страхования; сформулировал статистические оценки Брайера, которые, как мы показываем, динамически совместимы с мартингалом. © DeFinetti.org

Допустим, что участник пари в  $n$  повторях модели двух периодов,  $t_0$  и  $t_1$ , генерирует стратегию  $\mathfrak{S}$  — последовательность пари  $b_{0,i} \in [0, 1]$ , где индекс  $i = 1, 2, \dots, n$  о реализации бинарной случайной величины  $\mathbb{1}_{t_1,i}$ . Для такого участника абсолютная вариация отношения P/L за  $n$  пари будет

$$L_1(\mathfrak{S}) = \frac{1}{n} \sum_{i=1}^n |\mathbb{1}_{t_1,i} - b_{0,i}|.$$

Допустим, например, что  $\mathbb{E}(\mathbb{1}_{t_1}) = \frac{1}{2}$ . Делая ставку на вероятность, в данном случае  $\frac{1}{2}$ , участник теряет  $\frac{1}{2}$  математического ожидания, то есть столько же, как при ставке на 0 или на 1; значит, участника не поощряют ставить на точную вероятность.

Если мы работаем с одной и той же случайной величиной и не меняющимися со временем вероятностями, вариацию можно представить метрикой  $\ell_1$ :

$$L_1(\mathfrak{S}) = \frac{1}{n} \left| \mathbb{1}_{t_1} - \sum_{i=1}^n b_{0,i} \right|.$$

Де Финетти предложил использовать функцию типа «оценки Брайера», квадратичную функцию потерь, представляемую метрикой  $\ell_2$ :

$$L_1(\mathfrak{S}) = \frac{1}{n} \sum_{i=1}^n \left( \mathbb{1}_{t_1,i} - b_{0,i} \right)^2,$$

которая достигает минимума при  $b_{0,i} = \mathbb{E}(\mathbb{1}_{t_1})$ .

В нашем мире непрерывно отслеживаемой стоимости деривативов вместо двухпериодической решеточной модели нам интереснее рассмотреть при том же финальном исходе  $t_1$  стохастический процесс  $b_t$ ,  $t_0 \leq t \leq t_1$ , где арбитражная «стоимость» пари на бинарный результат должна соответствовать математическому ожиданию, и мы снова получаем отображение на оценку Брайера — в силу требования безарбитражности. Хотя никакой квадратичной функции потерь не постулируется, арбитраж исключается тем фактом, что пари является функцией мартингалла и является мартингалом, то есть условное ожидание инвариантно ко времени. При «завышенной» цене арбитражер сможет «шортить», при «заниженной» — «скупать», и так до бесконечности. Ради согласованности между пари в период  $t$  и в другие периоды  $t + \Delta t$  участникам приходится соблюдать теоретико-вероятностную дисциплину. Другими словами, если предсказатель отойдет от мартингальной стоимости, он позволит на себе спекулировать покупающим у него и затем продающим ему же обратно.

В текущей практике предсказателей выборов видно осознание необходимости минимизировать оценку Брайера, но не видно осознания той идеи, что и коррективы оценок должны иметь мартингальную стоимость.

## 12.4. ЗАКЛЮЧЕНИЕ И КОММЕНТАРИИ

Как видно на Рисунке 12.1, бинарный опцион скорее выявляет неопределенность, чем истинную оценку, и этот результат хорошо знаком трейдерам, см. [225].

Если кандидатов 3, можно обобщить процедуру, опираясь на следующую эвристику. Зададим стохастический процесс для  $Y_{1,t}$  на носителе  $[0, 1]$ , а затем для  $Y_{2,t}$  —

процесс в  $(Y_{1,t}, 1]$ ; для третьего кандидата  $Y_{3,t}$  будет остатком  $1 - Y_{2,t} - Y_{1,t}$ . Если кандидатов  $n$ , то  $Y_{n-1,t} \in (Y_{n-2,t}, 1]$  и  $n$ -й кандидат будет представлен остатком  $Y_{n,t} = 1 - \sum_{i=1}^{n-1} Y_{i,t}$ .

## ПРИЛОЖЕНИЕ: ВСЕ ДОРОГИ ВЕДУТ К ФИНАНСОВОЙ МАТЕМАТИКЕ

**Истоки.** Обри Клейтон отправил редактору письмо с жалобами на «ошибки» в описанной выше методологии. Автор совместно с Дхрувом Мадекой ответил не столько на выпады Клейтона, сколько на общий вопрос о пользе методов финансовой математики в жизни.

Мы рады отреагировать на письмо Клейтона (не прошедшее рецензирования) безотносительно содержащихся в нем выводов, воспользовавшись случаем поговорить о более глубоком непонимании роли финансовой математики в целом и арбитражных цен в частности — и с гордостью продемонстрировать, что «все дороги ведут к финансовой математике», то есть что арбитражные подходы универсальны и применимы к бинарным предсказаниям любого рода. Это также случай для второго автора комментария, Мадеки, процитировать статью (2017) [158], где он одновременно и независимо получил результаты, аналогичные Талебу (2018) [234].

### Некорректные претензии

«Талеб раскритиковал предсказанные вероятности числа голосов избирателей, в частности предсказания о выборах от Группы-538... Он [Талеб] утверждает, что предсказания 538 нарушают арбитражные границы» — это описание не соответствует фактам.

Группа-538 вообще не упомянута в [234], и Клейтон, видимо, смешивает статью в научном журнале с дебатами в твиттере. Цель статьи — строгий анализ темы выборов, без публицистики, и выборы 2016 года упомянуты как пример всего один раз<sup>1</sup>.

Однако перейдем к другим утверждениям Клейтона, не останавливаясь на путанности и запальчивости его письма.

### Неправильная арбитражная стоимость

Клейтон считает ее не то ошибкой (и вначале пишет, что «во-первых, один из так называемых стандартных результатов финансовой математики, на который опираются его оценки предсказаний выборов, ложен»), не то «лишь отчасти справедливой», как он туманно пишет ниже, давая задний ход. Мы вновь отложим в сторону ту деталь, что Талеб (2018) [234] не дает оценок результатам Группы-538, и сосредоточимся на его доводах.

Клейтон рассматривает три периода,  $t_0 = 0$ , промежуточный период  $t$  и конечный период  $T$ , где  $t_0 \leq t < T$ . Клейтон показывает специальный случай распределения вероятности, предсказанной в  $t_0$  на время от  $t = \frac{T}{2}$  до  $T$ . Это равномерное распределение для

<sup>1</sup> Кстати, проблема Группы-538 не в том, что они подняли вероятность с 0,55 до 0,85 в течение 5 месяцев, а в том, что они вносили резкие изменения в течение более коротких периодов — и это обсуждается в статье Мадеки (2017) [158]. — Прим. автора.

указанного периода. На самом деле для такой конструкции можно показать, используя интегральное преобразование вероятности, что вероятности следуют распределению, близкому к симметрическому бета-распределению,  $B(a, b)$  при  $a = b$ . Когда  $t = \frac{T}{2}$ , коэффициенты  $a = b = 1$  (это равномерное распределение). До момента  $\frac{T}{2}$  распределение имеет горб и вырождается в  $\delta$ -функцию Дирака при  $t = t_0$ . После  $\frac{T}{2}$  оно имеет яму и вырождается в два дираковских пика в точках 0 и 1 (как и положено распределению Бернулли), когда  $t$  приближается к  $T$  (и близко к распределению арксинуса<sup>1</sup>,  $B\left(\frac{1}{2}, \frac{1}{2}\right)$ , в некоторый промежуточный момент).

На самом деле построение Клейтона вводит в заблуждение, поскольку он анализирует распределение цены в момент времени  $t$  при фильтрации в момент времени  $t_0$ ; в частности, он поступает так при обсуждении арбитражных цен и арбитражного давления. Участники учитывают стоимость опционов от  $t$  до  $T$  в момент времени  $t$  (а не в период  $t_0$ ), с учетом заданной цены; при таких ограничениях бинарный опцион автоматически сходится к  $\frac{1}{2}$  при  $\sigma \rightarrow \infty$ , независимо от заданной цены, как бы далеко она ни отошла от страйк-цены (или порога). Коэффициент  $\sigma$  здесь ни в коем случае не прошлая реализованная волатильность, а только нереализованная будущая. Это можно видеть в рамках системы, представленной Талебом (2018) [234], да и в любой модели цен бинарного опциона. Цена по своей природе — не вероятность (и тем более не распределение вероятностей), а ожидание. Попросту говоря, как операторы арбитража, мы выводим цену бинарного опциона, глядя на будущую волатильность при данной цене, а не на абстрактное распределение вероятностей безо всяких условий.

При бесконечном коэффициенте  $\sigma$  остается один шум, в котором полезные сигналы тонут.

Другой путь прийти к неопределенности  $\frac{1}{2}$  — воспользоваться теорией информации и концепцией максимальной энтропии при глубокой неопределенности: энтропия ( $I$ ) распределения Бернулли с вероятностями  $p$  и  $1 - p$  представима как  $I = -((1 - p) \ln(1 - p) + p \ln p)$  и достигает максимума в точке  $\frac{1}{2}$ .

Чтобы получить выигрыш по сравнению с ценами  $\frac{1}{2}$ , нужно иметь достаточно информации, чтобы превзойти шум. Как мы увидим в следующем разделе, это нелегко.

### Арбитражные вопросы

Еще один результат в финансовой математике, накладывающий ограничения на волатильность предсказаний, состоит в следующем. Поскольку предсказания выборов

<sup>1</sup>  $B\left(\frac{1}{2}, \frac{1}{2}\right)$  называется распределением арксинуса, потому что кумулятивное распределение вероятности представимо как 
$$\begin{cases} 0, & x < 0 \\ \frac{\arcsin(2x-1)}{\pi} + \frac{1}{2}, & 0 \leq x \leq 1. \\ 1, & \text{иначе} \end{cases}$$
 Плотность 
$$\begin{cases} 0, & x < 0 \\ \frac{1}{\pi\sqrt{x(1-x)}}, & 0 \leq x \leq 1; \text{ среднее } \frac{1}{2}; \text{ дисперсия } \frac{1}{8}; \text{ коэффициент асимметрии } 0; \text{ эксцесс } \frac{3}{2}. \\ 0, & \text{иначе} \end{cases}$$

можно интерпретировать как европейский бинарный опцион, можно опереться на тот факт, что ценообразование такого опциона ограничено диапазоном от 0 до 1, и сделать утверждения о волатильности самой цены.

Суть в том, что при больших колебаниях цены на бинарный опцион простая стратегия трейдера покупать подешевле и продавать подороже гарантирует прибыль<sup>1</sup>. Вкратце обоснование сводится к тому замечанию, что при рассмотрении арифметического броуновского движения в границах  $[L, H]$

$$dB_t = \sigma dW_t \quad (12.7)$$

стохастический интеграл

$$2 \int^T (B_0 - B_t) dB_t = \sigma^2 T - (B_T - B_0)^2$$

можно повторять с нулевыми затратами, и это показывает, что значение  $B_T$  ограничено максимальным значением квадрата разности в правой части равенства. Тем самым предсказатель, генерирующий слишком волатильные вероятности — и готовый торговать согласно своим предсказаниям (т. е. рисковать собственной шкурой), — позволяет на себе спекулировать по стратегии, когда ему продают пропорционально завышению предсказания и у него покупают пропорционально занижению предсказания.

В заключение отметим, что любые численные предсказания вероятности правильно обрабатывать как цену выбора — согласно той интуитивной идее де Финетти, что предсказатель должен рисковать собственной шкурой. В этих условиях бинарные предсказания подпадают под правила цен арбитража и деривативов, которые хорошо описаны в финансовой математике. Использование подходов финансовой математики, чтобы генерировать бинарные прогнозы, не мешает байесовским методам; Талеб (2018) не утверждает, что вероятности должны быть только  $\frac{1}{2}$ ; утверждается только то, что предсказания сносит в сторону этого уровня из-за арбитражного давления и ограничений на вариативность предсказания. Структурные ограничения на корректировку предсказаний следуют из того факта, что в итоге цена окажется либо 1, либо 0<sup>2</sup>.

1 Мы заимствовали этот результат из записей Бруно Дюпира на лекциях по финансам непрерывного времени в Курантовском институте математических наук в Нью-Йорке на последнем курсе весной 2019. — *Прим. автора.*

2 Еще один способ убедиться в этом находим за пределами моделей финансовой математики: можно рассмотреть стандартную оценку вероятности. Пусть  $X_1, \dots, X_n$  — случайные величины в  $[0, 1]$  (предсказания) и  $B_T \in \{0, 1\}$  — константа (единственный исход). Нужно минимизировать оценку сделанных предсказаний

$$\lambda_n = \frac{1}{n} \sum_{i=1}^n (x_i - B_T)^2.$$

При любом заданном  $B_T$  и среднем прогнозе  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  минимальное значение  $\lambda_n$  достигается при  $x_1 = \dots = x_n$ . Предсказания с дираковским распределением  $x_1 = \dots = x_n = \frac{1}{2}$  гарантируют  $\lambda_n = \frac{1}{4}$ , и, чтобы превзойти их какой-либо стратегией с большей дисперсией, потребуется предсказание с точностью выше 75 %. (Заметим, что у предсказаний с равномерным распределением оценка  $\frac{1}{3}$ .) Это показывает нам, каково необходимое соотношение между волатильностью и полезным сигналом. — *Прим. автора.*

Возможное обстоятельство, настроившее Клейтона против финансовой математики, то, что вероятности и обосновывающие их опросы в реальной жизни могут не быть мартингалами; однако торгуемые вероятности (и, следовательно, реальные предсказания) мартингалами быть обязаны. Вот почему в статье Талеба (2018) [234] утверждается, что обработку опросов (которая может быть туманной и неторгуемой) необходимо преобразовать в обработку вероятности в промежутке  $[0, 1]$ .

### **БЛАГОДАРНОСТИ**

Рафаэлю Дуади, студентам Политехнического института Нью-Йоркского университета и участникам семинара агентства «Блумберг» по финансовой математике в Нью-Йорке.

## **Часть IV**

# **ОЦЕНОЧНЫЕ НЕРАВЕНСТВА ПРИ ЖИРНЫХ ХВОСТАХ**





## ОЦЕНКА ДЖИНИ ПРИ БЕСКОНЕЧНОЙ ДИСПЕРСИИ<sup>†</sup>

Эта глава посвящена проблемам, связанным с оценкой индекса Джини, когда процесс порождает данные с жирным хвостом, т. е. одного из классов устойчивых распределений с конечным средним, но бесконечной дисперсией — другими словами, с показателем хвоста  $\alpha \in (1, 2)$ . Мы доказываем, что в этом случае обычные непараметрические методы не способны дать надежную оценку коэффициенту Джини — они будут занижать результаты из-за плохой наблюдаемости жирных хвостов. Это обстоятельство имеет важные последствия для современных дискуссий об экономическом неравенстве.

Начнем мы с фазового перехода, который претерпевает непараметрическая оценка индекса Джини в симметричной структуре асимптотического распределения оценок, когда распределение данных смещается из домена аттракции легких хвостов в домен жирных хвостов; такой переход особенно резок, если во втором домене дисперсия становится бесконечной.

Мы также покажем, что смещение непараметрических оценок индекса Джини усугубляется при снижении  $\alpha$ .

Затем мы докажем, что оценка методом максимального правдоподобия превосходит по эффективности непараметрические методы, то есть достигает той же надежности при меньшем размере выборки.

Наконец, мы предложим простой механизм поправок к непараметрической оценке при малом размере выборки жирнохвостых данных; этот механизм основан на расстоянии между модой и средним асимптотического распределения.

### 13.1. ВВЕДЕНИЕ

Исследования имущественного неравенства — важная область для специалистов по экономике, статистике и эконофизике, сталкивающихся с процессами, в которых порождаются жирнохвостые данные, в том числе с бесконечной дисперсией [39, 144]. Что неудивительно, если вспомнить, что прототипом жирнохвостых распределений послужила модель Парето для распределения доходов домохозяйств [185]. Однако

жирнохвостость по-прежнему осложняет исследования распределений благосостояния, подрывая эффективность (и отчасти состоятельность) результатов применения ряда количественных показателей имущественного неравенства и концентрации богатств [82, 144].

В этой работе мы исследуем влияние жирных хвостов на такой знаменитый показатель экономического неравенства, как индекс Джини [78, 110, 144], который широко применяется и которым широко злоупотребляют в литературе по эконофизике и экономике как главным инструментом для описания распределения и концентрации богатств во всем мире [39, 191, 192].

При всей обширности и подробности литературы по измерению индекса Джини (см., напр., обзоры [78, 222]) ничтожно мало внимания уделялось его поведению в ситуации жирных хвостов; это тем более странно, что (1) жирные хвосты встречаются повсеместно в эмпирических распределениях доходов и благосостояния [144, 191] и (2) сам индекс Джини работает как показатель вариативности и жирнохвостости [76, 79, 80, 95].

Стандартный метод оценить индекс Джини непараметрический: доступные данные эмпирического распределения обрабатываются по приведенной ниже формуле (13.5). Но, как мы покажем в данной статье, такая оценка смещена в сторону занижения, если объект наблюдения жирнохвостый. Чтобы закрыть этот пробел, мы выводим предельное распределение непараметрической оценки Джини при наличии жирных хвостов и предлагаем возможные стратегии, как уменьшить систематическую ошибку. Мы показываем, что метод максимального правдоподобия, даже при риске ошибочной классификации модели, требует существенно меньше наблюдений, чем непараметрический метод, чтобы достичь сравнимой эффективности<sup>1</sup>.

Наши результаты существенны для дискуссии об имущественном неравенстве, недавно возобновленной Тома Пикетти в [191], поскольку оценка индекса Джини при жирных хвостах и бесконечной дисперсии может в ряде случаев сделать экономический анализ ненадежным и просто неверным. Как можно полагаться на смещенную оценку?

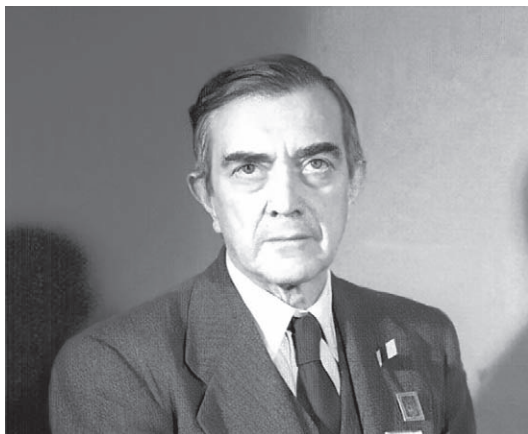
Под жирнохвостыми данными мы подразумеваем такие, которые порождаются положительной случайной величиной  $X$  с кумулятивной функцией распределения  $F(x)$ , если она правильно меняющаяся порядка  $\alpha$  [136], то есть для  $\bar{F}(x) \triangleq 1 - F(x)$  имеет место

$$\lim_{x \rightarrow \infty} x^\alpha \bar{F}(x) = L(x), \quad (13.1)$$

где  $L(x)$  — медленно меняющаяся функция:  $\lim_{x \rightarrow \infty} \frac{L(cx)}{L(x)} = 1$  при  $c > 0$  и где  $\alpha > 0$  называется показателем хвоста.

Правильно меняющиеся распределения составляют обширный класс случайных величин, свойства которых подробно изучались в контексте теории экстремальных значений [82, 116] как вероятностное поведение максимумов и минимумов. Как отмечено в [44], термины «правильно меняющиеся» и «жирнохвостые» на самом деле

<sup>1</sup> Аналогичное смещение затрагивает также непараметрическую оценку доли, или квантиля, т. е. оценку типа «верхний 1 % владеет  $x$  % всех богатств» [242]. В этой статье эта проблема распространяется на шире применяемый коэффициент Джини и исследуется глубже: прослеживаются связи с предельными теоремами. — *Прим. автора.*



**Рисунок 13.1:** *Итальянский статистик Коррадо Джини, 1884–1965. © Боккони L'Università Commerciale Luigi Bocconi*

синонимичны. Известно, что если н. о. р. наблюдения  $X_1, \dots, X_n$  следуют распределению  $F(x)$  в классе правильно меняющихся, согласно критерию (13.1), то их порождающий процесс попадает в максимальный аттрактор распределения Фреше с коэффициентом формы  $\rho$ ; пишут  $X \in \text{MDA}(\Phi(\rho))$  [116]. Это значит, что максимум по частичной выборке  $M_n = \max(X_1, \dots, X_n)$  представим как сходящийся по распределению

$$\mathbb{P}(a_n^{-1} (M_n - b_n) \leq x) \xrightarrow{D} \Phi(\rho) = e^{-x-\rho}, \rho > 0, \quad (13.2)$$

где нормализующие константы  $a_n > 0$  и  $b_n \in \mathbb{R}$ . Очевидно, что показатель  $\alpha$  хвоста правильно меняющейся функции и коэффициент формы  $\rho$  распределения Фреше связаны как  $\alpha = \frac{1}{\rho}$  [82].

Распределение Фреше, наряду с распределениями Гумбеля и Вейбулла, служит предельным распределением для максимумов в теории экстремальных значений; оно представляет предельный случай неограниченного жирного хвоста [116]. Поэтому взаимосвязь между правильно меняющимися случайными величинами и классом Фреше позволяет нам работать с обширным семейством случайных величин (и эмпирических данных) и обосновать интуитивное предположение [95, 144] о резкой зависимости индекса Джини от максимумов, т. е. экстремальных богатств, особенно при бесконечной дисперсии. Что лишний раз призывает нас к осторожности при обсуждении экономического неравенства в ситуациях жирных хвостов.

Нелишне будет напомнить, что существование (конечность) моментов для жирно-хвостой случайной величины  $X$  зависит от показателя хвоста  $\alpha$ , а именно:

$$\begin{aligned} \mathbb{E}X^\delta &< \infty, \text{ если } \delta \leq \alpha, \\ \mathbb{E}X^\delta &= \infty, \text{ если } \delta > \alpha^1. \end{aligned} \quad (13.3)$$

<sup>1</sup> Где  $\delta$  — порядок обсуждаемого момента.

В этой работе мы фокусируемся на порождающих процессах с конечным средним и бесконечной дисперсией; следовательно, согласно формуле (13.3), на классе регулярно меняющихся распределений с показателем хвоста  $\alpha \in (1, 2)$ .

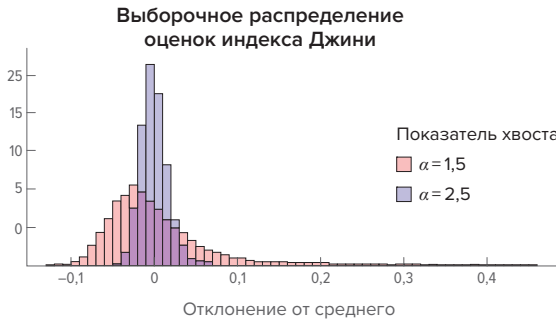
В Таблице 13.1 и на Рисунке 13.2 представлены численно и графически наши результаты на основе искусственно сгенерированных наблюдений выборки из распределения Парето с приведенной ниже формулой (13.13) и показателем хвоста  $\alpha = 1,1$ ; результаты столь красноречивы, что выводы напрашиваются сами собой.

В Таблице 13.1 сравнивается непараметрическая оценка индекса Джини по формуле (13.5) и оценка по хвосту методом максимального правдоподобия (*maximum likelihood*, ML), описанная в Разделе 13.3. Для различных размеров выборки в Таблице 13.1 мы сгенерировали по  $10^8$  выборок, чтобы вывести средние оценки по методу Монте-Карло. Как видно в первом столбце, непараметрическая оценка сходится к истинному значению индекса Джини ( $g = 0,8333$ ) чрезвычайно медленно и при этом монотонно возрастает, так что у непараметрической оценки распределение проблемное не только своими тяжелыми хвостами, но и несимметричностью.

На Рисунке 13.2 представлены численные доказательства того, что при попадании в ситуацию данных с бесконечной дисперсией предельное распределение непараметрической оценки индекса Джини теряет свои свойства нормальности и симметричности [91], сползая к несимметричному и жирнохвостому пределу. Как мы доказываем в Разделе 13.2, когда порождающий процесс принадлежит аттрактору некоторого жирнохвостого распределения, асимптотическое распределение оценки индекса Джини превращается в  $\alpha$ -устойчивый закон с перекосом вправо. Эта смена поведения ответственна за систематическую ошибку в сторону занижения оценки индекса Джини при жирных хвостах. Однако осведомленность о новом пределе позволяет нам предложить поправку к непараметрической оценке, повышающую качество и тем самым снижающую риск грубой недооценки имущественного неравенства, со всеми последствиями для экономической и социальной политики [144, 191].

**Таблица 13.1:** Сравнивается непараметрическая (NonPar) оценка индекса Джини и оценка методом максимального правдоподобия (ML). Используются данные, которые распределены по Парето с показателем хвоста  $\alpha = 1,1$  (конечное среднее, бесконечная дисперсия) при разных размерах выборки. Число моделирований методом Монте-Карло:  $10^8$

$n$ (Число набл.)	NonPar		ML		Коэффициент ошибок
	Среднее	Смещение	Среднее	Смещение	
$10^3$	0,711	−0,122	0,8333	0	1,4
$10^4$	0,750	−0,083	0,8333	0	3
$10^5$	0,775	−0,058	0,8333	0	6,6
$10^6$	0,790	−0,043	0,8333	0	156
$10^7$	0,802	−0,031	0,8333	0	$10^5+$



**Рисунок 13.2:** Гистограммы непараметрических оценок индекса Джини для двух распределений Парето (тип I) с разными показателями хвоста — одно с конечной дисперсией, другое с бесконечной. Для удобства сравнения диаграммы совмещены по горизонтали. Размер выборки  $10^3$ . Число выборок по  $10^2$  для каждого распределения

Остальная часть статьи организована так. В Разделе 13.2 мы выведем асимптотическое распределение выборочной оценки индекса Джини в ситуации данных с бесконечной дисперсией. В Разделе 13.3 мы рассмотрим оценку методом максимального правдоподобия; в Разделе 13.4 создадим иллюстрацию с наблюдениями распределения Парето; в Разделе 13.5 предложим простую поправку на основе расстояния мода — среднее асимптотического распределения непараметрической оценки, устраняющую систематическую ошибку малой выборки. Раздел 13.6 подводит итоги. В Техническое приложение в конце Раздела 13 вынесены длинные доказательства главных результатов этой работы.

### 13.2. АСИМПТОТИКИ НЕПАРАМЕТРИЧЕСКОЙ ОЦЕНКИ ПРИ БЕСКОНЕЧНОЙ ДИСПЕРСИИ

Сейчас мы выведем асимптотическое распределение непараметрической оценки индекса Джини в условиях жирнохвостого порождающего процесса с конечным средним, но бесконечной дисперсией.

Так называемая стохастическая репрезентация Джини — это

$$g = \frac{1}{2} \frac{\mathbb{E}|X' - X''|}{\mu} \in [0, 1], \quad (13.4)$$

где  $X'$  и  $X''$  — н. о. р. экземпляры случайной величины  $X$  с носителем  $[c, \infty)$ ,  $c > 0$  и конечным средним  $\mathbb{E}X = \mu$ . Величина  $\mathbb{E}|X' - X''|$  известна как «средняя разность по Джини» (*Gini Mean Difference*,  $\text{GMD}^1$ ) [222]. Для удобства дальнейшего изложения представим  $g = \frac{\vartheta}{\mu}$ , где  $\vartheta = \frac{\mathbb{E}|X' - X''|}{2}$ .

Таким образом, индекс Джини случайной величины  $X$  есть среднее ожидаемое отклонение реализации  $X$  от другой независимой реализации  $X$ , масштабированное с коэффициентом удвоенное среднее [81].

Наиболее распространенная непараметрическая оценка индекса Джини для выборки  $X_1, \dots, X_n$  определяется как

$$G^{\text{NP}}(X_n) = \frac{\sum_{1 \leq i, j \leq n} |X_i - X_j|}{(n-1) \sum_{i=1}^n X_i}, \quad (13.5)$$

<sup>1</sup> Для той же величины  $\mathbb{E}|X' - X''|$  в литературе встречается также термин средняя абсолютная разность (*mean absolute difference*) и обозначение  $\text{MD}(X)$ .

что можно иначе представить как

$$G^{NP}(X_n) = \frac{\sum_{i=1}^n \left(2 \frac{i-1}{n-1} - 1\right) X_{(i)}}{\sum_{i=1}^n X_{(i)}} = \frac{\frac{1}{n} \sum_{i=1}^n Z_{(i)}}{\frac{1}{n} \sum_{i=1}^n X_{(i)}}, \quad (13.6)$$

где  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  — упорядоченные данные  $X_1, \dots, X_n$ , так что  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  и  $Z_{(i)} = \left(2 \frac{i-1}{n-1} - 1\right) X_{(i)}$ <sup>1</sup>. Известна асимптотическая нормальность оценки по формуле (13.6) в предположении конечной дисперсии для порождающего процесса [144, 222]. Такой результат непосредственно следует из свойств  $U$ -данных<sup>2</sup> и  $L$ -оценок<sup>3</sup>, фигурирующих в формуле (13.6).

Стандартным подходом к выводу предельного распределения для оценки в формуле (13.6) и любой другой линейной комбинации упорядоченных статистических данных было бы исследовать в пределе при  $n \rightarrow \infty$  аппроксимацию последовательности упорядоченных статистических данных последовательностью н. о. р. случайных величин [56, 151]. Однако при этом обычно требуется принадлежность порождающего процесса классу  $L^2$  — интегрируемость квадрата процесса, — и такого допущения мы здесь не делаем.

Лемма 13.1 (доказана в Приложении) показывает, как справиться со случаем, когда последовательность упорядоченных статистических данных порождена случайными величинами, следующими жирнохвостому распределению класса  $L^1$ , то есть с интегрируемостью только первой степени.

### Лемма 13.1

Рассмотрим последовательность  $R_n = \frac{1}{n} \sum_{i=1}^n \left(\frac{i}{n} - U_{(i)}\right) F^{-1}(U_{(i)})$ , где  $U_{(i)}$  — упорядоченные статистические данные выборки н. о. р. случайных величин, следующих равномерному распределению. Пусть  $F^{-1}(U_{(i)}) \in L^1$ . Тогда

$$R_n \xrightarrow{L^1} 0 \quad (13.7)$$

и

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} R_n \xrightarrow{L^1} 0, \quad (13.8)$$

где  $\alpha \in (1, 2)$  и где  $L_0(n)$  — медленно меняющаяся функция.

1 В самом деле,  $\sum_{1 \leq i, j \leq n} |X_i - X_j| = \sum_{1 \leq i, j \leq n} |X_{(i)} - X_{(j)}| = \sum_{i=1}^n \sum_{j=1}^n \begin{cases} X_{(i)} - X_{(j)}, & i \geq j \\ X_{(j)} - X_{(i)}, & \text{иначе} \end{cases} = \sum_{i=1}^n ((i-1)X_{(i)} - (n-i)X_{(i)}) =$   
 $= \sum_{i=1}^n (2i - n - 1)X_{(i)} = (n-1) \sum_{i=1}^n \left(2 \frac{i-1}{n-1} - 1\right) X_{(i)}.$

2 Так в английской литературе обозначают класс несмещенных статистических данных (от *unbiased class of statistics*).

3 Так в английской литературе обозначают класс оценок, вычисляемых как линейная комбинация статистических данных (от *linear combination of order statistics*).

### 13.2.1. Краткое повторение $\alpha$ -устойчивых случайных величин

Введем ряд обозначений для  $\alpha$ -устойчивых распределений, которые понадобятся нам при изучении асимптотического предела оценки индекса Джини. Случайная величина  $X$  следует  $\alpha$ -устойчивому распределению — пишут:  $X \sim S(\alpha, \beta, \gamma, \delta)$ , — если ее характеристическая функция

$$\mathbb{E} e^{i\gamma X} = \begin{cases} \exp\left(-\gamma^\alpha |t|^\alpha \left(1 - i\beta(\operatorname{sgn} t) \operatorname{tg} \frac{\pi\alpha}{2}\right) + i\delta t\right), & \alpha \neq 1 \\ \exp\left(-\gamma |t| \left(1 + i\beta \frac{2}{\pi} (\operatorname{sgn} t) \ln |t|\right) + i\delta t\right), & \alpha = 1 \end{cases},$$

где  $\alpha \in (0, 2)$  управляет хвостом,  $\beta \in [-1, 1]$  — коэффициент асимметрии,  $\gamma \in \mathbb{R}^+$  — коэффициент масштаба и  $\delta \in \mathbb{R}$  — коэффициент положения. Эта система параметров  $\alpha$ -устойчивых распределений называется S 1 [181, 209<sup>1</sup>].

Любопытно, что  $\alpha$  — индекс  $\alpha$ -устойчивой случайной величины соответствует  $\alpha$  — показателю хвоста правильно меняющейся функции согласно формуле (13.1). Как показано в [91, 181], правильно меняющаяся случайная величина порядка  $\alpha$  является  $\alpha$ -устойчивой. Поэтому мы здесь не вводим для этих параметров разных обозначений. Поскольку мы работаем с распределениями, характеризующимися конечным средним, но бесконечной дисперсией, мы ограничиваем диапазон  $\alpha \in (0, 2)$ , и при таком ограничении две альфы совпадают.

Напомним, что при  $\alpha \in (1, 2]$  математическое ожидание  $\alpha$ -устойчивой случайной величины  $X$  совпадает с коэффициентом положения  $\delta$ , т. е.  $\mathbb{E}X = \delta$ . За подробностями отсылаем к [181, 209].

Стандартная  $\alpha$ -устойчивая случайная величина кратко записывается

$$S_{\alpha, \beta} \sim S(\alpha, \beta, 1, 0). \quad (13.9)$$

Заметим, что  $\alpha$ -устойчивые распределения составляют подкласс бесконечно делимых распределений. Благодаря замкнутости по свертке их можно использовать для описания предельного поведения (масштабированных) частичных сумм  $S_n = \sum_{i=1}^n X_i$  в контексте обобщенной центральной предельной теоремы (*Generalized central limit theorem*, GCLT) [91]<sup>2</sup>. В особом случае  $\alpha = 2$  получаем нормальное распределение — предельное распределение классических вариантов ЦПТ, предполагающих конечную дисперсию.

1 Именно в этой, второй ссылке указана книга Геннадия Пенхосовича Самородниченко и Мюрада Такку (1994) с параметрами, которые в дальнейшем стали называться системой параметров S 1 (только там распределение и параметры обозначены  $S_\alpha(\sigma, \beta, \mu)$ , где  $\sigma \in \mathbb{R}^+$  — коэффициент масштаба и  $\mu \in \mathbb{R}$  — коэффициент положения).

2 Уильям Феллер в книге по ссылке формулирует такую «обобщенную ЦПТ»: пусть  $X_k$  следуют симметричным распределениям  $F_k$  и пусть при всех  $t > 0$

$$\sum_{k=1}^n \int_{|x|>t} a_n F_k \{dx\} \rightarrow 0, \quad a_n^{-2} \sum_{k=1}^n \int_{|x|>t} a_n x^2 F_k \{dx\} \rightarrow 1.$$

Тогда можно доказать, что  $\frac{S_n}{a_n} \xrightarrow{D} \mathcal{N}(a)$ .



Далее о случайной величине в аттракторе  $\alpha$ -устойчивого распределения будем писать  $X \in \text{DA}(S_\alpha)$ . Заметим только, что данное условие для предела частичных сумм равносильно условию в формуле (13.2) для предела частичных максимумов [82, 91].

### 13.2.2. $\alpha$ -устойчивый асимптотический предел индекса Джини

Рассмотрим выборку н. о. р. наблюдений  $X_1, \dots, X_n$  с непрерывной кумулятивной функцией  $F(x)$  в классе правильно меняющихся функций, определенных формулой (13.1), с показателем хвоста  $\alpha \in (1, 2)$ . Порождающий процесс этой выборки принадлежит аттрактору распределения Фреше с  $\rho \in \left(\frac{1}{2}, 1\right)$ , поскольку  $\rho = \frac{1}{\alpha}$ .

Следующие две теоремы помогут нам получить асимптотическое распределение для оценки индекса Джини, представленного формулой (13.6), когда порождающий процесс характеризуется бесконечной дисперсией: Теорема 1 касается предельного распределения «средней разности по Джини», числителя в формуле (13.6), а Теорема 2 распространяет этот результат на весь индекс Джини. Доказательства обеих теорем вынесены в Приложение.

#### Теорема 1

Рассмотрим последовательность  $(X_i)_{1 \leq i \leq n}$  н. о. р. случайных величин на  $[c, +\infty)$  при  $c > 0$ , чтобы распределение  $X$  принадлежало аттрактору  $\alpha$ -устойчивого распределения,  $X \in \text{DA}(S_\alpha)$  при  $\alpha \in (1, 2)$ . Тогда среднее отклонение Джини (Gini mean deviation, Gini Mean Difference, GMD) по выборке  $\frac{1}{n} \sum_{i=1}^n Z_{(i)}$  сходится по распределению

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \left( \frac{1}{n} \sum_{i=1}^n Z_{(i)} - \vartheta \right) \xrightarrow{D} S_{\alpha,1}, \quad (13.10)$$

где  $Z_i = (2F(X_i) - 1)X_i$ ,  $\mathbb{E}Z_i = \vartheta$ ,  $L_0(n)$  — медленно меняющаяся функция, удовлетворяющая формуле (13.37) (см. Приложение) и  $S_{\alpha,1}$  — стандартное  $\alpha$ -устойчивое распределение с перекосом вправо согласно формуле (13.9).

Более того, статистический показатель  $\frac{1}{n} \sum_{i=1}^n Z_{(i)}$  — асимптотически состоятельная оценка для GMD, т. е.  $\frac{1}{n} \sum_{i=1}^n Z_{(i)} \xrightarrow{P} \vartheta$ .

Заметим, что Теорему 1 можно переформулировать в терминах максимального аттрактора  $\text{MDA}(\Phi(\rho))$  согласно формуле (13.2).

#### Теорема 2

При посылках Теоремы 1 оценка индекса Джини  $G^{\text{NP}}(X_n) = \frac{\sum_{i=1}^n Z_{(i)}}{\sum_{i=1}^n X_{(i)}}$  сходится по распределению

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \left( G^{\text{NP}}(X_n) - \frac{\vartheta}{\mu} \right) \xrightarrow{D} Q, \quad (13.11)$$

где  $\mathbb{E}Z_i = \vartheta$ ,  $\mathbb{E}X_i = \mu$ ,  $L_0(n)$  — медленно меняющаяся функция, заданная в Теореме 1, и случайная величина  $Q$  следует  $\alpha$ -устойчивому распределению с перекосом вправо

$$S\left(\alpha, 1, \frac{1}{\mu}, 0\right).$$

Более того, статистический показатель  $\frac{\sum_{i=1}^n Z_{(i)}}{\sum_{i=1}^n X_{(i)}}$  — асимптотически состоя-  
 тельная оценка для индекса Джини, т. е.  $\frac{\sum_{i=1}^n Z_{(i)}}{\sum_{i=1}^n X_{(i)}} \xrightarrow{P} \frac{g}{\mu} = g$ .

В ситуации жирных хвостов порождающего процесса, при  $\alpha \in (1, 2)$ , Теорема 2 говорит нам, что асимптотическое распределение оценки Джини перекошено вправо, независимо от других подробностей порождающего процесса. Поэтому жирнохвостые данные не просто делают предельную оценку Джини жирнохвостой, они изменяют форму предельного закона, отклоняя его от обычной симметричной гауссианы. В результате оценка Джини, хотя ей по-прежнему гарантирована асимптотическая состоятельность [151], приближается к истинному значению медленно и снизу. Мы уже видели подобное поведение в Таблице 13.1.

### 13.3. ОЦЕНКА МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Теорема 2 показывает, что обычная непараметрическая оценка индекса Джини — не лучший выбор в ситуации распределений с бесконечной дисперсией, так как ее распределение в асимптотическом пределе перекошено и жирнохвостое. Наша цель — найти оценки, распределение которых остается асимптотически нормальным даже после перехода к порождающим процессам с жирными хвостами, и эта цель недостижима для непараметрических методов, поскольку они все подпадают под центральную предельную теорему для  $\alpha$ -устойчивых распределений [82, 91]. Значит, придется прибегнуть к параметрическим техникам.

Как показывает Теорема 3, если известно параметрическое семейство порождающего процесса, то можно оценить индекс Джини при помощи ММП. Полученная оценка не просто асимптотически нормальная, она асимптотически эффективная.

В Теореме 3 мы рассматриваем случайные величины  $X$ , распределение которых принадлежит обширному и гибкому экспоненциальному семейству [211], т. е. плотность представима как

$$f_{\vartheta}(x) = h(x) e^{\eta(\vartheta)T(x) - A(\vartheta)},$$

где  $\vartheta \in \mathbb{R}$  и  $T(x)$ ,  $\eta(\vartheta)$ ,  $h(x)$  и  $A(\vartheta)$  — известные функции.

#### Теорема 3

Пусть  $X \sim F_{\vartheta}$  из экспоненциального семейства. Тогда оценка  $\vartheta$  методом максимального правдоподобия дает асимптотически нормальную и эффективную оценку индекса Джини,  $G^{\text{ML}}(X_n)_{\vartheta}$ . А именно:

$$\sqrt{n} (G^{\text{ML}}(X_n)_{\vartheta} - g_{\vartheta}) \xrightarrow{D} \mathcal{N}(0, g_{\vartheta}^2 I^{-1}(\vartheta)), \quad (13.12)$$

где  $g'_{\vartheta} = \frac{dg_{\vartheta}}{d\vartheta}$  и  $I$  — информация Фишера.

*Доказательство.* Результат легко получить из асимптотической эффективности оценок максимального правдоподобия для экспоненциального семейства и принципа инвариантности оценок ММП. В случае индекса Джини принцип инвариантности гарантирован непрерывностью и монотонностью  $g_{\vartheta}$  по  $\vartheta$ . Асимптотическая дисперсия выводится дельта-методом [211]. ■

### 13.4. ИЛЛЮСТРАЦИЯ С РАСПРЕДЕЛЕНИЕМ ПАРЕТО

Для иллюстрации используем искусственно сгенерированные жирнохвостые данные, распределенные по Парето I [185] с плотностью

$$f(x) = \alpha c^\alpha x^{-\alpha-1}, x \geq c. \quad (13.13)$$

Нетрудно убедиться, что соответствующая функция выживания  $\bar{F}(x)$  принадлежит правильно меняющемуся классу с коэффициентом хвоста  $\alpha$  и медленно меняющейся функцией  $L(x) = c^{\alpha 1}$ . Поэтому мы можем применить результаты Раздела 13.2 и получить приведенные ниже следствия.

#### Следствие 13.1

Пусть  $X_1, \dots, X_n$  — последовательность н. о. р. наблюдений, распределенных по Парето с показателем хвоста  $\alpha \in (1, 2)$ . Непараметрическая оценка Джини характеризуется следующим пределом:

$$D_n^{\text{NP}} = G^{\text{NP}}(X_n) - g \sim S \left( \alpha, 1, \frac{C_\alpha^{-\frac{1}{\alpha}}}{n^{\frac{1}{\alpha}}} \frac{\alpha-1}{\alpha}, 0 \right)^2. \quad (13.14)$$

*Доказательство.* Без потери общности примем  $c = 1$  в формуле (13.13). Нужный результат получим простым применением Теоремы 2, помня, что распределение Парето — в аттракторе  $\alpha$ -устойчивых случайных величин с медленно меняющейся функцией  $L(x) = 1$ . Последовательность  $c_n$ , удовлетворяющих формуле (13.37), превращается в  $c_n = n^{\frac{1}{\alpha}} C_\alpha^{-\frac{1}{\alpha}}$ ; поэтому  $L_0(n) = C_\alpha^{-\frac{1}{\alpha}}$  и не зависит от  $n$ . Кроме того, среднее распределения также является функцией  $\alpha$ , то есть  $\mu = \frac{\alpha}{\alpha-1}$ .

#### Следствие 13.2

Пусть выборка  $X_1, \dots, X_n$  распределена как в Следствии 13.1 и пусть  $G_g^{\text{ML}}$  — оценка индекса Джини максимального правдоподобия, как указано в Теореме 3. Тогда оценка Джини по ММП, масштабированная по истинному среднему  $g$ , имеет предел

$$D_n^{\text{ML}} = G_\alpha^{\text{ML}}(X_n) - g \sim \mathcal{N} \left( 0, \frac{4\alpha^2}{n(2\alpha-1)^4} \right), \quad (13.15)$$

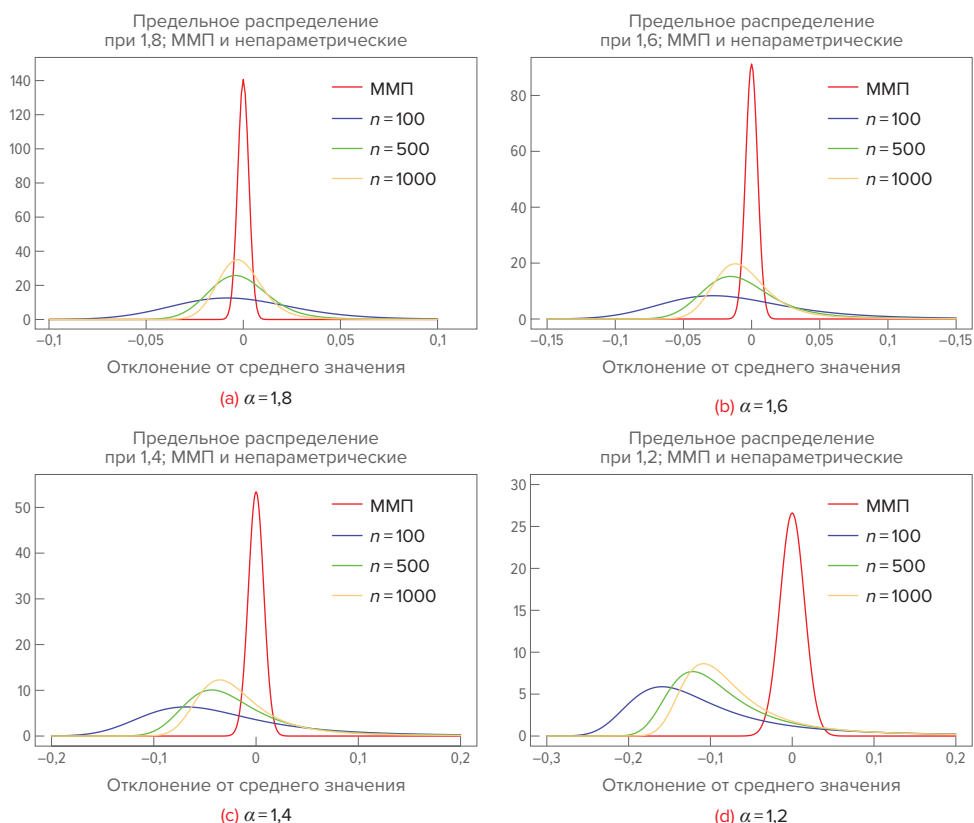
где  $\mathcal{N}$  — гауссово распределение.

*Доказательство.* Известно функциональное представление оценки индекса Джини максимального правдоподобия:  $G_g^{\text{ML}} = \frac{1}{2\alpha^{\text{ML}} - 1}$  [144]. Поэтому нужный результат следует из того факта, что распределение Парето (с известным минимальным значением  $x_m$ ) принадлежит экспоненциальному семейству, а значит, удовлетворяет условиям регулярности для асимптотической нормальности и эффективности оценки максимального правдоподобия. Заметим также, что информация Фишера для распределения Парето имеет вид  $\frac{1}{\alpha^2}$ . ■

1 То есть константой.

2 Обозначение  $C_\alpha$  объясняется в формуле (13.37).

Теперь, выведя оба асимптотических распределения, можем сравнить качество сходимости оценки ММП и непараметрической оценки, когда данные распределены по Парето — прототипического примера жирнохвостых наблюдений. В частности, можем аппроксимировать распределение отклонений оценки от истинного значения  $g$  индекса Джини при конечных размерах выборки, используя формулы (13.14) и (13.15).



**Рисунок 13.3:** Сравниваем асимптотические распределения оценок максимального правдоподобия и непараметрических оценок при различных показателях хвоста  $\alpha$ . Число наблюдений ММП вступо постоянно,  $n = 100$ . Заметим, что, даже если у всех распределений среднее ноль, мода распределений непараметрической оценки отличается от нуля в силу их асимметрии

На Рисунке 13.3 показано, как распределены отклонения от среднего для двух типов оценки и как эти распределения сжимаются с ростом числа наблюдений. Для наглядности мы зафиксировали число наблюдений для оценок ММП, варьируя его только для непараметрических оценок. Мы провели данное исследование для разных показателей хвоста порождающего процесса, чтобы показать, каково воздействие жирнохвостости на состоятельность оценки. Следует заметить, что при уменьшении показателя хвоста вблизи 1 (порог, за которым среднее станет бесконечным) мода распределения непараметрической оценки существенно смещается от среднего того же распределения (которое по определению 0: ведь речь об отклонении от среднего). Этот

эффект ответственен за наблюдаемую на практике систематическую ошибку малых выборок. И такого не происходит с оценкой ММП благодаря нормальности предела при любых показателях хвоста.

Сформулируем эту мысль в более строгих терминах, оценив число наблюдений  $\tilde{n}$ , при котором точность непараметрической оценки сравнивается с ММП в сценариях с различными хвостами. Рассмотрим функцию отношения правдоподобия

$$r(c, n) = \frac{\mathbb{P}_S(|D_n^{\text{NP}}| > c)}{\mathbb{P}_N(|D_{100}^{\text{ML}}| > c)}, \quad (13.16)$$

где  $\mathbb{P}_S(|D_n^{\text{NP}}| > c)$  и  $\mathbb{P}_N(|D_{100}^{\text{ML}}| > c)$  — вероятности (соответственно  $\alpha$ -устойчивая и гауссова) того, что центральные оценки (непараметрическая или ММП) превысят порог  $c$ , фигурирующий в формулах (13.15) и (13.14). В непараметрическом случае числу наблюдений  $n$  позволено изменяться, а в случае ММП оно зафиксировано на ста. Затем найдем значение  $\tilde{n}$ , при котором  $r(c, \tilde{n}) = 1$  для данного  $c$ .

В Таблице 13.2 показаны результаты при разных порогах  $c$  и показателях хвоста  $\alpha$ . Видно, в частности, что показатель ММП эффективнее, чем непараметрический, для которого требуется намного больше наблюдений, чтобы достичь той же хвостовой вероятности, как у ММП при  $n = 100$ . Так, при  $\alpha = 1,2$  непараметрической оценке требуется  $80 \times 10^6$  наблюдений, чтобы вероятность нарушения порога  $\pm 0,02$  стала такой же, как у оценки ММП.

**Таблица 13.2:** Число наблюдений  $\tilde{n}$ , при котором непараметрическая оценка сравнивается по хвостовым вероятностям с оценкой ММП из фиксированного числа наблюдений  $n = 100$  — для различных порогов  $c$  и различных показателей хвоста  $\alpha$

Порог  $c$  из формулы (13.16)

$\alpha$	0,005	0,01	0,015	0,02
1,8	$27 \times 10^3$	$12 \times 10^5$	$12 \times 10^6$	$63 \times 10^5$
1,5	$21 \times 10^4$	$21 \times 10^4$	$46 \times 10^5$	$81 \times 10^7$
1,2	$33 \times 10^8$	$67 \times 10^7$	$20 \times 10^7$	$80 \times 10^6$

Интересно заметить, что число наблюдений, при котором совпадают хвостовые вероятности в формуле (13.16), зависит от порога неравномерным образом. Это можно было предвидеть, ведь порог пробегает от нуля до бесконечности, а хвостовые вероятности остаются неизменными для данного значения  $n$ . С учетом того, что предельные распределения одногорбые, можно ожидать, что необходимое число наблюдений для совпадения хвостовых вероятностей будет взлетать до больших значений лишь после пересечения некоторого порога.

Вывод: при жирнохвостых данных с бесконечной дисперсией оценка на основе обычного ММП лучше, чем непараметрическая.

### 13.5. ПОПРАВКА НА МАЛУЮ ВЫБОРКУ

Теорему 2 можно также использовать для корректировки смещения непараметрической оценки при малых размерах выборки. Идея в том, чтобы осознать роль симметрии для одногорбых распределений, где большинство наблюдений близки к моде.

При симметричных распределениях мода совпадает со средним, и большинство наблюдений близки к среднему, но при асимметричных распределениях это не так: у непрерывного одностороннего распределения с перекосом вправо мода меньше среднего. Учитывая, что асимптотическое распределение непараметрической оценки индекса Джини перекошено вправо, ожидается, что наблюдаемое значение индекса Джини будет обычно занижаться (по сравнению со средним). Это расхождение (т. е. систематическая ошибка) можно количественно охарактеризовать расстоянием между модой и средним, и при известном расстоянии можно скорректировать оценку Джини, прибавив к ней эту величину, недооценку<sup>1</sup>.

На языке формул задача в том, чтобы вывести такую скорректированную непараметрическую оценку  $G^C(X_n)$ , что

$$G^C(X_n) = G^{NP}(X_n) + \|m(G^{NP}(X_n)) - \mathbb{E}G^{NP}(X_n)\|, \quad (13.17)$$

где  $\|m(G^{NP}(X_n)) - \mathbb{E}G^{NP}(X_n)\|$  — расстояние между модой  $m$  и средним распределения непараметрической оценки Джини  $G^{NP}(X_n)$ .

Корректировка по формуле (13.17) равносильна сдвигу распределения  $G^{NP}(X_n)$ , совмещающему моду с истинным значением индекса Джини.

Идеально было бы измерить расстояние мода — среднее  $\|m(G^{NP}(X_n)) - \mathbb{E}G^{NP}(X_n)\|$ , пользуясь точным распределением индекса Джини; тогда бы поправка получилась максимально точной. Однако распределение конечной выборки не всегда легко вывести: для него требуются предположения о структуре параметров порождающего процесса (и для жирнохвостых данных она в большинстве случаев неизвестна [144]). Поэтому мы предлагаем приблизить распределение конечной выборки предельным распределением непараметрического Джини, полученным в Разделе 13.2, и оценивать расстояние мода — среднее по предельному распределению. Такой порядок действий оставляет больше свободы при выборе допущений модели и потенциально уменьшает число оцениваемых параметров — учитывая, что предельное распределение зависит только от показателя хвоста и среднего данных и среднее обычно можно считать функцией показателя хвоста; например, в случае Парето  $\mu = \frac{\alpha}{\alpha - 1}$ .

Воспользовавшись этой зависимостью, положение — масштаб  $\alpha$ -устойчивых распределений и формулой (13.11), аппроксимируем распределение  $G^{NP}(X_n)$  для конечных выборок

$$G^{NP}(X_n) \sim S(\alpha, 1, \gamma(n), g), \quad (13.18)$$

где  $\gamma(n) = \frac{1}{n^{\frac{\alpha-1}{\alpha}}} \frac{L_0(n)}{\mu}$  — коэффициент масштаба предельного распределения.

Как следствие, благодаря линейности моды для  $\alpha$ -устойчивых распределений имеем

$$\|m(G^{NP}(X_n)) - \mathbb{E}G^{NP}(X_n)\| \approx \|m(\alpha, \gamma(n)) + g - g\| = \|m(\alpha, \gamma(n))\|,$$

где  $m(\alpha, \gamma(n))$  — функция моды  $\alpha$ -устойчивого распределения с нулевым средним.

<sup>1</sup> Во время написания статьи мы проверили и другую идею — использовать расстояние между медианой и средним. Эффективность оказалась сравнимая. — *Прим. автора.*

Таким образом, искомая поправка не требует знания истинного индекса Джини — ведь  $m(\alpha, \gamma(n))$  не зависит от  $g$ . Тогда поправку можно оценить как

$$\hat{m}(\alpha, \gamma(n)) = \operatorname{argmax}_x s(x)^1, \quad (13.19)$$

где обрабатываемая численными методами функция  $s(x)$  аналогична плотности соответствующего  $\alpha$ -устойчивого распределения в формуле (13.18), только сцентрированная в 0. Это связано с тем, что для  $\alpha$ -устойчивых распределений мода недоступна в виде аналитических выражений, но легко находится численно [181], пользуясь единственностью горба.

Таким образом, скорректированная непараметрическая оценка

$$G^C(X_n) = G^{NP}(X_n) + \hat{m}(\alpha, \gamma(n)) \quad (13.20)$$

имеет асимптотическое распределение

$$G^C(X_n) \sim S(\alpha, 1, \gamma(n), g + \hat{m}(\alpha, \gamma(n))). \quad (13.21)$$

Заметим, что поправка  $\hat{m}(\alpha, \gamma(n))$  зависит от показателя хвоста  $\alpha$  порождающего процесса и через коэффициент масштаба  $\gamma(n)$  соответствующего предельного распределения — от размера выборки  $n$ , причем зависимость  $\hat{m}(\alpha, \gamma(n))$  от  $n$  убывающая, и  $\lim_{n \rightarrow \infty} \hat{m}(\alpha, \gamma(n)) = 0$ . Дело в том, что с ростом  $n$  распределение, описанное формулой (13.18), центрируется вокруг среднего, сжимая расстояние между модой и средним до нуля. Тем самым гарантируется асимптотическая эквивалентность скорректированной оценки и непараметрической. Достаточно учесть, что

$$\begin{aligned} \lim_{n \rightarrow \infty} |G^C(X_n) - G^{NP}(X_n)| &= \lim_{n \rightarrow \infty} |G^{NP}(X_n) + \hat{m}(\alpha, \gamma(n)) - G^{NP}(X_n)| = \\ &= \lim_{n \rightarrow \infty} |\hat{m}(\alpha, \gamma(n))| = 0. \end{aligned}$$

Естественно, поправка улучшает  $G^C(X_n)$  при малых выборках. Также учтем, что, согласно формуле (13.21), у распределения скорректированной оценки среднее равно сумме  $g + \hat{m}(\alpha, \gamma(n))$ , которая сходится к истинному индексу Джини  $g$  при  $n \rightarrow \infty$ .

Теоретически качество данной поправки можно измерить и улучшить, поскольку оно зависит от расстояния между точным распределением  $G^{NP}(X_n)$  и его  $\alpha$ -устойчивым пределом: чем ближе эти два распределения друг к другу, тем лучше наша приближенная оценка. Однако в большинстве случаев точное распределение  $G^{NP}(X_n)$  неизвестно и подробности недоступны.

Из уже изложенного выше ясно, что поправка зависит от показателя хвоста данных и, возможно, от их среднего. Если данные параметры неизвестны априори, их

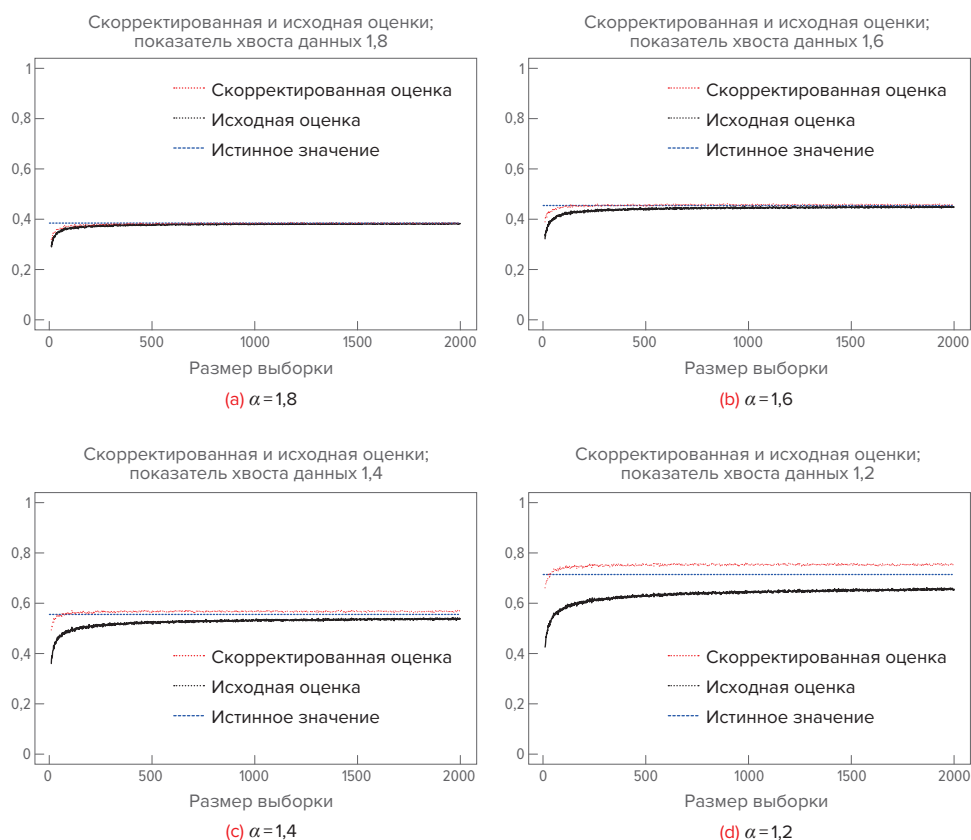
1 Аргумент максимизации:  $\operatorname{argmax}_x f(x)$  — значение аргумента, при котором  $f(x)$  достигает максимума.

В англоязычной литературе пишут с пробелом  $\operatorname{arg max}_x f(x)$ .

придется оценить. Это внесет дополнительную неопределенность, которая отразится на качестве поправки.

В заключении этого раздела обсудим влияние корректирующей процедуры на простом примере. В эксперименте Монте-Карло мы моделируем 1000 выборок из данных по Парето — увеличивая размер выборки от  $n = 10$  до  $n = 2000$ , — и для каждого размера выборки вычисляем исходную непараметрическую оценку  $G^{\text{NP}}(X_n)$  и скорректированную  $G^{\text{C}}(X_n)$ . Мы повторяем эксперимент при различных  $\alpha$ . Результаты показаны на Рисунке 13.4.

Очевидно, что скорректированные оценки всегда лучше нескорректированных — имеют меньшее абсолютное отклонение от истинного значения Джини. В нашем конкретном эксперименте при малых размерах выборки  $n \leq 1000$  видно разительное улучшение при всех показателях хвоста  $\alpha \in (1, 2)$ . При больших размерах выборки, как и следовало ожидать, различие между оценками сокращается; наша поправка убывающим образом зависит от  $n$  и от  $\alpha$ . Заметим, что при показателе хвоста 2 получим симметричное гауссово распределение и совпадение обеих оценок, потому что при конечной дисперсии непараметрическая оценка не смещена.



**Рисунок 13.4:** Сравниваем скорректированную непараметрическую оценку (красная, верхняя кривая) и обычную непараметрическую оценку (черная, нижняя кривая). При малых размерах выборки поправка явно улучшает качество оценки



### 13.6. ВЫВОДЫ

В этой главе мы изучили асимптотическое поведение непараметрической оценки индекса Джини в ситуации бесконечной дисперсии порождающего процесса, которая странным образом игнорируется в литературе. Коренная ошибка широко применяемых непараметрических методов состоит в предположении, что асимптотическая состоятельность гарантирует хорошее доасимптотическое поведение.

Мы показали, что параметрический подход улучшает асимптотические результаты благодаря свойствам оценки максимального правдоподобия. Поэтому мы настаиваем, что при подозрении в жирнохвостости данные предпочтительно обрабатывать параметрическими методами.

В ситуациях, где полностью параметрический подход недоступен, мы предлагаем простой механизм, корректирующий непараметрическую оценку с учетом расстояния между модой и средним ее асимптотического распределения. Но даже при успешном вычислении поправки мы предлагаем пользоваться скорректированной оценкой с осторожностью, поскольку поправка вносит дополнительную неопределенность.

### ТЕХНИЧЕСКОЕ ПРИЛОЖЕНИЕ

#### Доказательство Леммы 13.1

Пусть  $U = F(X)$  — равномерно распределенная случайная величина, полученная из случайной величины  $X$  стандартным интегральным преобразованием. Для упорядоченных статистических данных имеем [56]:  $X_{(i)} \stackrel{\text{a.s.}}{=} F^{-1}(U_{(i)})$ . Поэтому

$$R_n = \frac{1}{n} \sum_{i=1}^n \left( \frac{i}{n} - U_{(i)} \right) F^{-1}(U_{(i)}). \quad (13.22)$$

Из определения эмпирического распределения вероятностей следует, что

$$R_n = \frac{1}{n} \sum_{i=1}^n (F_n(U_{(i)}) - U_{(i)}) F^{-1}(U_{(i)}), \quad (13.23)$$

где  $F_n(u) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{U_{(i)} \leq u}$  — эмпирическое кумулятивное распределение случайных величин, следующих равномерному распределению.

Чтобы доказать сходимость  $R_n \xrightarrow{L^1} 0$ , мы ограничим  $R_n$  сверху последовательно, сходящейся к нулю. Начнем с того, что

$$\mathbb{E}|R_n| \leq \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left| (F_n(U_{(i)}) - U_{(i)}) F^{-1}(U_{(i)}) \right|. \quad (13.24)$$

Чтобы построить ограничивающую последовательность для выражения в правой части формулы (13.24), можно использовать тот факт, что даже при  $L^1$ -интегрируемой  $F^{-1}(U_{(i)})$  разность  $F_n(U_{(i)}) - U_{(i)}$  интегрируема бесконечно и можно применить неравенство Гёльдера при  $q = \infty$  и  $p = 1$ . Из него следует, что

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left| (F_n(U_{(i)}) - U_{(i)}) F^{-1}(U_{(i)}) \right| \leq \frac{1}{n} \sum_{i=1}^n \mathbb{E} \sup_{U_{(i)}} |F_n(U_{(i)}) - U_{(i)}| \mathbb{E} |F^{-1}(U_{(i)})|. \quad (13.25)$$

Далее по неравенству Коши — Шварца<sup>1</sup>

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \mathbb{E} \sup_{U_{(i)}} |F_n(U_{(i)}) - U_{(i)}| \mathbb{E} |F^{-1}(U_{(i)})| \leq \\ & \leq \left( \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E} \sup_{U_{(i)}} |F_n(U_{(i)}) - U_{(i)}| \right)^2 \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E} |F^{-1}(U_{(i)})| \right)^2 \right)^{\frac{1}{2}}. \end{aligned} \quad (13.26)$$

Теперь вспомним, что независимо от упорядочивания последовательности н. о. р.  $U_i$ ,  $i = 1, \dots, n$ , сумма  $\sum_{i=1}^n F^{-1}(U_{(i)}) \stackrel{\text{a.s.}}{=} \sum_{i=1}^n F^{-1}(U_i)$  и что  $\mathbb{E} F^{-1}(U_i) = \mu$ . Поэтому вторую часть формулы (13.26) можно представить в виде

$$\mu \left( \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E} \sup_{U_{(i)}} |F_n(U_{(i)}) - U_{(i)}| \right)^2 \right)^{\frac{1}{2}}. \quad (13.27)$$

Достаточно будет доказать, что это выражение стремится к нулю при  $n \rightarrow \infty$ .

Мы знаем, что  $F_n$  — эмпирическая функция распределения равномерно распределенных случайных величин. Применяя неравенство треугольника, можем ограничить выражение, которое в формуле (13.27) возводится в степень  $\frac{1}{2}$ :

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E} \sup_{U_{(i)}} |F_n(U_{(i)}) - U_{(i)}| \right)^2 \leq \\ & \leq \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E} \sup_{U_{(i)}} |F_n(U_{(i)}) - F(U_{(i)})| \right)^2 + \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E} \sup_{U_{(i)}} |F(U_{(i)}) - U_{(i)}| \right)^2. \end{aligned} \quad (13.28)$$

Поскольку мы работаем с равномерно распределенными случайными величинами,  $F(U) = U$ , и второе слагаемое в правой части исчезает.

Далее мы можем ограничить  $\mathbb{E} \sup_{U_{(i)}} |F_n(U_{(i)}) - F(U_{(i)})|$  при помощи так называемого неравенства<sup>2</sup> Вапника — Червоненкиса, равномерного ограничения для эмпирических процессов [29, 55, 258], и получить

$$\mathbb{E} \sup_{U_{(i)}} |F_n(U_{(i)}) - U_{(i)}| \leq \sqrt{\frac{\ln(n+1) + \ln 2}{n}}. \quad (13.29)$$

<sup>1</sup> *Cauchy-Schwarz inequality* в русской литературе также называется неравенством Коши — Буняковского.  
<sup>2</sup> Теория Вапника — Червоненкиса (*Vapnik-Chervonenkis theory*, VC) изучает эффективность статистического обучения. В ней в качестве функции потерь классификатора  $f$  рассматривается  $R(f) \triangleq \mathbb{P}(f(X) \neq Y)$  и оценивается эмпирический риск  $\hat{R}_n(f) \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{f(X_i) \neq Y_i}$ , который согласно теореме VC сходится к функции потерь:  $\mathbb{E} \sup_f |\hat{R}_n(f) - R(f)| \leq 2 \sqrt{\frac{\ln S_n + \ln 2}{n}}$ , если  $S_n \triangleq \max\{f(X_1), \dots, f(X_n)\}$  не растет быстрее, чем какая-нибудь степень  $n$ .

Комбинируя формулу (13.29) с формулой (13.27), получаем ограничивающую последовательность

$$\mu \left( \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E} \sup_{U_{(i)}} |F_n(U_{(i)}) - U_{(i)}| \right)^2 \right)^{\frac{1}{2}} \leq \mu \sqrt{\frac{\ln(n+1) + \ln 2}{n}}, \quad (13.30)$$

которая стремится к нулю при  $n \rightarrow \infty$ , что и требовалось для первого утверждения леммы.

Чтобы доказать второе утверждение леммы, достаточно заметить, что правая часть формулы (13.30) продолжит сходиться к нулю даже после умножения на  $\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)}$ , если  $\alpha \in (1, 2)$ .

### Доказательство Теоремы 1

Первая часть доказательства будет состоять в том, что мы перепишем формулу (13.10) как функцию н. о. р. случайных величин, а не упорядоченных данных статистики, чтобы стало возможным сослаться на центральную предельную теорему (ЦПТ).

Начнем с последовательности

$$\frac{1}{n} \sum_{i=1}^n Z_{(i)} = \frac{1}{n} \sum_{i=1}^n \left( 2 \frac{i-1}{n-1} - 1 \right) F^{-1}(U_{(i)}). \quad (13.31)$$

Используя интегральное преобразование вероятности  $X = F^{-1}(U)$ , где случайная величина  $U$  следует стандартному равномерному распределению<sup>1</sup>, и прибавляя и вычитая  $\frac{1}{n} \sum_{i=1}^n (2U_{(i)} - 1)F^{-1}(U_{(i)})$ , можем в формуле (13.31) переписать правую часть как

$$\frac{1}{n} \sum_{i=1}^n Z_{(i)} = \frac{1}{n} \sum_{i=1}^n (2U_{(i)} - 1)F^{-1}(U_{(i)}) + \frac{1}{n} \sum_{i=1}^n 2 \left( \frac{i-1}{n-1} - U_{(i)} \right) F^{-1}(U_{(i)}). \quad (13.32)$$

Затем, пользуясь свойствами упорядоченных статистических данных [56], получаем следующую эквивалентность почти наверное:

$$\frac{1}{n} \sum_{i=1}^n Z_{(i)} \stackrel{\text{a.s.}}{=} \frac{1}{n} \sum_{i=1}^n (2U_i - 1)F^{-1}(U_i) + \frac{1}{n} \sum_{i=1}^n 2 \left( \frac{i-1}{n-1} - U_{(i)} \right) F^{-1}(U_{(i)}). \quad (13.33)$$

Заметим, что первое слагаемое в правой части формулы (13.33) является нужной нам функцией н. о. р. случайных величин, тогда как второе слагаемое представляет собой остаток; поэтому введем обозначения

1 *Standard uniform distribution* — распределение вероятностей  $F_U(u) = \begin{cases} 0, & u < 0 \\ u, & 0 \leq u \leq 1. \end{cases}$  Любую непрерывную

случайную величину  $X$  можно преобразовать в стандартную равномерную  $F(X) \stackrel{D}{=} U$ , используя в качестве преобразования  $F$  распределение вероятностей этой случайной величины,  $F(x) = F_X(x)$ . И наоборот, произвольную непрерывную случайную величину  $X$  можно представить как результат преобразования стандартной равномерной случайной величины  $X = F^{-1}(U)$ .

$$\frac{1}{n} \sum_{i=1}^n Z_{(i)} \stackrel{\text{a.s.}}{=} \frac{1}{n} \sum_{i=1}^n Z_i + R_n,$$

где  $Z_i = (2U_i - 1)F^{-1}(U_i)$  и  $R_n = \frac{1}{n} \sum_{i=1}^n 2\left(\frac{i-1}{n-1} - U_{(i)}\right) F^{-1}(U_{(i)})$ .

Учитывая формулу (13.10) и пользуясь разложением согласно (13.33), можем переписать наше утверждение как

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \left( \frac{1}{n} \sum_{i=1}^n Z_{(i)} - \vartheta \right) = \frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \left( \frac{1}{n} \sum_{i=1}^n Z_i - \vartheta \right) + \frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} R_n. \quad (13.34)$$

Пользуясь вторым утверждением Леммы 13.1 и теоремой<sup>1</sup> Слуцкого, можно доказать сходимость в формуле (13.10), обратив внимание на поведение последовательности

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \left( \frac{1}{n} \sum_{i=1}^n Z_{(i)} - \vartheta \right), \quad (13.35)$$

где  $Z_i = (2U_i - 1)F^{-1}(U_i) = (2F(X_i) - 1)X_i$ . Поэтому для доказательства теоремы достаточно доказать, что  $Z_i$  принадлежит жирнохвостому аттрактору.

Напомним, что, согласно посылке,  $X \in \text{DA}(S_\alpha)$ , где  $\alpha \in (1, 2)$ . Данное допущение позволяет использовать вариант ЦПТ, утверждающий сходимость суммы толстохвостых случайных величин. Однако нужно еще доказать, что и  $Z \in \text{DA}(S_\alpha)$ , т. е. что  $\mathbb{P}(|Z| > z) \sim L(z)z^{-\alpha}$ , где  $\alpha \in (1, 2)$  и  $L(z)$  медленно меняющаяся функция.

Заметим, что

$$\mathbb{P}(|\tilde{Z}| > z) \leq \mathbb{P}(|Z| > z) \leq \mathbb{P}(2X > z),$$

где  $\tilde{Z} = (2U - 1)X$  и  $U \perp X$ . Первое ограничение опирается на положительную зависимость между  $X$  и  $F(X)$ , и его удастся строго доказать после того, как мы заметим, что  $2UX \leq 2F(X)X$  в силу так называемого перестановочного неравенства<sup>2</sup> [122]. Ограничение сверху, в свою очередь, тривиально.

Пользуясь свойствами медленно меняющихся функций, представим  $\mathbb{P}(2X > z) \sim 2^\alpha L(z)z^{-\alpha}$ . Чтобы доказать  $\tilde{Z} \in \text{DA}(S_\alpha)$ , воспользуемся теоремой<sup>3</sup> Бреймана, которая гарантирует замкнутость класса  $\alpha$ -устойчивых распределений по умножению, если у второго сомножителя не слишком жирный хвост [267].

1 Теорема Слуцкого, она же теорема Крамера, гласит следующее. Пусть две последовательности случайных величин сходятся: одна по распределению к случайной величине и другая по вероятности к константе,  $X_n \xrightarrow{D} X$  и  $Y_n \xrightarrow{P} c$ . Тогда  $X_n + Y_n \xrightarrow{D} X + c$  и  $X_n Y_n \xrightarrow{D} X c$ .

2 Оно же неравенство об одномонотонных последовательностях, или транснеравенство. Если две числовые последовательности одинаковой длины  $x_1, \dots, x_n$  и  $y_1, \dots, y_n$  заданы с точностью до перестановок, то упорядочение  $x_{(1)} \leq \dots \leq x_{(n)}$  и  $y_{(1)} \leq \dots \leq y_{(n)}$  обеспечивает максимальное скалярное произведение  $x_{(1)}y_{(1)} + \dots + x_{(n)}y_{(n)}$ .

3 Теорема Бреймана (1965) доказывает, что если  $X \in \text{MDA}(\Phi)$  при некотором  $\gamma > 0$  и  $Y$  удовлетворяет  $\mathbb{E} Y^{Y+\varepsilon} < \infty$  при некотором  $\varepsilon > 0$ , то  $\lim_{u \rightarrow \infty} \frac{\mathbb{P}(XY > u)}{\mathbb{P}(X > u)} = \mathbb{E} Y^\gamma$ . Здесь  $\Phi$  — распределение Фреше,  $\Phi_\gamma(x) = \exp(-x^{-\gamma})$  при  $x > 0$  и  $\text{MDA}$  — максимальный аттрактор.

Чтобы применить эту теорему, перепишем  $\mathbb{P}(|\tilde{Z}| > z)$  как

$$\mathbb{P}(|\tilde{Z}| > z) = \mathbb{P}(\tilde{Z} > z) + \mathbb{P}(-\tilde{Z} > z) = \mathbb{P}(\tilde{U}X > z) + \mathbb{P}(-\tilde{U}X > z),$$

где  $\tilde{U}$  — стандартная равномерная случайная величина и  $\tilde{U} \perp X$ .

Сосредоточимся на  $\mathbb{P}(\tilde{U}X > z)$ , поскольку процедура для  $\mathbb{P}(-\tilde{U}X > z)$  такая же.

$$\mathbb{P}(\tilde{U}X > z) = \mathbb{P}(\tilde{U}X > z \mid \tilde{U} > 0) \mathbb{P}(\tilde{U} > 0) + \mathbb{P}(\tilde{U}X > z \mid \tilde{U} \leq 0) \mathbb{P}(\tilde{U} \leq 0), \text{ при } z \rightarrow +\infty.$$

Вероятность  $\mathbb{P}(\tilde{U}X > z \mid \tilde{U} \leq 0) \rightarrow 0$ . В то же время, применяя теорему Бреймана, для  $\mathbb{P}(\tilde{U}X > z \mid \tilde{U} > 0)$  можно получить

$$\mathbb{P}(\tilde{U}X > z \mid \tilde{U} > 0) \rightarrow \mathbb{E}(\tilde{U}^\alpha \mid U > 0) \mathbb{P}(X > z) \mathbb{P}(U > 0).$$

Следовательно,

$$\mathbb{P}(|\tilde{Z}| > z) \rightarrow \frac{1}{2} \mathbb{E}(\tilde{U}^\alpha \mid U > 0) \mathbb{P}(X > z) + \frac{1}{2} \mathbb{E}((-\tilde{U})^\alpha \mid U \leq 0) \mathbb{P}(X > z).$$

Таким образом,

$$\begin{aligned} \mathbb{P}(|\tilde{Z}| > z) &\rightarrow \frac{1}{2} \mathbb{P}(X > z) \left( \mathbb{E}(\tilde{U}^\alpha \mid U > 0) + \mathbb{E}((-\tilde{U})^\alpha \mid U \leq 0) \right) = \\ &= \frac{2\alpha}{1-\alpha} \mathbb{P}(X > z) \sim \frac{2^\alpha}{1-\alpha} L(z)z^{-\alpha}, \end{aligned}$$

откуда по теореме сжатия<sup>1</sup> [91] можно заключить, что

$$\mathbb{P}(|\tilde{Z}| > z) \sim L(z)z^{-\alpha}$$

при  $z \rightarrow \infty$ . Следовательно,  $Z \in \text{DA}(S_\alpha)$ .

Теперь мы готовы применить обобщенную центральную предельную теорему [82] к последовательности  $Z_i$ :

$$n c_n^{-1} \left( \frac{1}{n} \sum_{i=1}^n Z_i - \mathbb{E} Z_i \right) \xrightarrow{D} S_{\alpha, \beta}, \quad (13.36)$$

где  $S_{\alpha, \beta}$  — стандартная  $\alpha$ -устойчивая случайная величина,  $\mathbb{E} Z_i = \vartheta$  и  $c_n$  — последовательность, которая должна удовлетворять требованию

$$\lim_{n \rightarrow \infty} \frac{n L(c_n)}{c_n^\alpha} = \frac{\Gamma(2-\alpha)}{\alpha-1} \left| \cos \frac{\pi\alpha}{2} \right| = C_\alpha. \quad (13.37)$$

Заметим, что  $c_n$  можно представить как  $c_n = n^{\frac{1}{\alpha}} L_0(n)$ , где  $L_0(n)$  — тоже медленно меняющаяся функция, не обязательно совпадающая с  $L(n)$ .

<sup>1</sup> Она же «теорема о двух милиционерах» (карабинерах, жандармах, городских), она же «теорема о сэндвиче» (*sandwich theorem*): функция, которая ограничена снизу и сверху функциями, сходящимися к общему пределу, сходится к тому же пределу.

Коэффициент асимметрии  $\beta$  таков, что

$$\frac{\mathbb{P}(Z > z)}{\mathbb{P}(|Z| > z)} \rightarrow \frac{1 + \beta}{2}.$$

Вспоминая, что, по построению,  $Z \in [-c, +\infty)$ , можем данное выражение свести к

$$\frac{\mathbb{P}(Z > z)}{\mathbb{P}(Z > z) + \mathbb{P}(-Z > z)} \rightarrow \frac{\mathbb{P}(Z > z)}{\mathbb{P}(Z > z)} = 1 \rightarrow \frac{1 + \beta}{2} \quad (13.38)$$

и, следовательно,  $\beta = 1$ . Это, в сочетании с формулой (13.34), результатом для остатка  $R_n$  в Лемме 13.1 и теоремой Слуцкого, позволяет нам заключить, что те же слабые пределы остаются в силе и для упорядоченной последовательности  $Z_{(i)}$  в формуле (13.10).

## Доказательство Теоремы 2

В качестве первого шага докажем, что оценка индекса Джини  $\frac{\sum_{i=1}^n Z_{(i)}}{\sum_{i=1}^n X_i}$ , где фигурирует<sup>1</sup> упорядочение последовательности н. о. р., эквивалентна по распределению отношению  $\frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n X_i}$ <sup>2</sup>. Для этого достаточно применить разложение из формулы (13.33) к формуле (13.11), получив

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \left( G^{\text{NP}}(X_n) - \frac{g}{\mu} \right) = \frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \left( \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n X_i} - \frac{g}{\mu} \right) + \frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} R_n. \quad (13.39)$$

Опираясь на Лемму 13.1 и применяя теоремы о непрерывном отображении и Слуцкого, удастся доказать, что второе слагаемое в (13.39) сходится по вероятности к нулю. Таким образом, для доказательства утверждения теоремы достаточно доказать слабый предел для последовательности

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \left( \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n X_i} - \frac{g}{\mu} \right). \quad (13.40)$$

Раскрывая формулу (13.40) и вспоминая, что  $Z_i = (2F(X_i) - 1)X_i$ , получаем

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \frac{n}{\sum_{i=1}^n X_i} \left( \frac{1}{n} \sum_{i=1}^n X_i \left( 2F(X_i) - 1 - \frac{g}{\mu} \right) \right). \quad (13.41)$$

Сомножитель  $\frac{n}{\sum_{i=1}^n X_i}$  в формуле (13.41) сходится по вероятности к  $\frac{1}{\mu}$ ; это доказывается по теореме о непрерывном отображении с учетом того факта, что наши случайные величины  $X$  положительные. Следовательно, этот сомножитель можно учесть в пределе по теореме Слуцкого.

<sup>1</sup> Через  $Z_{(i)} = \left( 2 \frac{i-1}{n-1} - 1 \right) X_{(i)}$ .

<sup>2</sup> Где  $Z_i = (2F(X_i) - 1)X_i$ , то есть определяется через  $X_i$  без упорядочения.

Сначала сосредоточимся на исследовании предельного закона для сомножителя

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \frac{1}{n} \sum_{i=1}^n X_i \left( 2F(X_i) - 1 - \frac{\vartheta}{\mu} \right). \quad (13.42)$$

Обозначим  $\hat{Z}_i = X_i \left( 2F(X_i) - 1 - \frac{\vartheta}{\mu} \right)$  и заметим, что  $\mathbb{E}\hat{Z}_i = 0$ , поскольку  $\mathbb{E}Z_i = \vartheta$  и  $\mathbb{E}X_i = \mu$ .

Чтобы характеризовать предельное распределение последовательности  $\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \frac{1}{n} \sum_{i=1}^n \hat{Z}_i$  согласно ОЦПТ, нужно, чтобы  $\hat{Z} \in \text{DA}(S_\alpha)$ . Если это верно, то ОЦПТ применима

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \left( \frac{\sum_{i=1}^n \hat{Z}_i}{n} - \mathbb{E}\hat{Z}_i \right). \quad (13.43)$$

Заметим, что в силу  $\mathbb{E}\hat{Z}_i = 0$  формула (13.43) равносильна формуле (13.42).

Чтобы доказать  $\hat{Z} \in \text{DA}(S_\alpha)$ , вспомним, что последовательность  $\hat{Z}_i = X_i \left( 2F(X_i) - 1 - \frac{\vartheta}{\mu} \right)$  есть не что иное, как сдвинутая на  $\frac{\vartheta}{\mu}$  последовательность  $Z_i = (2F(X_i) - 1)X_i$ .

Рассуждая аналогично тому, как в Теореме 1 рассуждали о  $Z$ , доказываем  $\hat{Z} \in \text{DA}(S_\alpha)$ . В частности, укажем, что  $\hat{Z}$  и  $Z$  (и, следовательно,  $X$ ) характеризуются одним и тем же показателем  $\alpha$  и медленно меняющейся функцией  $L(n)$ .

Заметим, что согласно посылке теоремы  $X \in [c, \infty)$  и  $c > 0$  и что мы имеем дело с непрерывными распределениями, так что  $\hat{Z} \in \left[ -c \left( 1 + \frac{\vartheta}{\mu} \right), \infty \right)$ . Поэтому левый хвост  $\hat{Z}$  не вносит вклада в предельный коэффициент асимметрии  $b$ , который остается равным 1 (как для  $Z$ ) согласно уравнению (13.38).

Следовательно, по ОЦПТ мы, наконец, получаем

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \left( \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n X_i} - \frac{\vartheta}{\mu} \right) \xrightarrow{D} \frac{1}{\mu} S(\alpha, 1, \frac{1}{\mu}, 0). \quad (13.44)$$

В завершение доказательства заметим, что, как доказывается формулой (13.39), слабый предел для индекса Джини характеризуется последовательностью н. о. р.

$\frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n X_i}$ , без упорядочивания и что  $\alpha$ -устойчивая случайная величина замкнута по мас-

штабированию с постоянным коэффициентом [209].

## СУПЕРАДДИТИВНОСТЬ И СМЕЩЕННЫЕ ОЦЕНКИ ВКЛАДА КВАНТИЛЕЙ<sup>†a</sup>

Сделанная по выборке количественная оценка вклада, внесенного верхним процентилем (концентрация), смещена вниз, неустойчива, крайне чувствительна к размеру выборки и выпукла вверх по большим отклонениям. Она дополнительно портится в доменах с хвостами степенного закона, особенно при низких значениях показателя. Как показано в настоящей статье, такие оценки могут варьировать с течением времени и повышаться с увеличением выборки, создавая иллюзию структурных изменений, отслеживаемых изменениями концентрации. Кроме того, такие оценки несостоятельны при суммировании и комбинировании распределений, поскольку взвешенное среднее оценок концентрации в  $A$  и  $B$  имеет тенденцию оказываться ниже оценки концентрации в  $A \cup B$ . И можно показать, что при толстых хвостах рост общей суммы должен сопровождаться ростом размера выборки для количественной оценки концентрации. Мы исследуем супераддитивность этой оценки и смещение при однородных и смешанных распределениях.

### 14.1. ВВЕДЕНИЕ

Вильфредо Парето в свое время заметил, что 80 % земель в Италии принадлежит 20 % населения и наоборот, и от этого наблюдения начинается изучение класса распределений по степенному закону и распространяется популярная фраза «80 на 20». Самоподобие — фундаментальное свойство степенных законов [162] и [163] — допускает рекурсию: правило «80 на 20» применимо к оставшимся 20 % и так далее; для верхнего процента населения получаем владение 53 % всего богатства.

Похоже, данный способ количественно оценить концентрацию подвержен существенной систематической ошибке, зависящей от способа измерения, и вполне вероятно, что реальное отношение концентрации, наблюдавшееся Парето, было выше: верхний процентиль мог быть ближе к 70 %, а если бы в последующем выборка росла, то оценка, приближаясь к истинному уровню, также росла бы с годами. Мы покажем, что чем полнее выборка начинает представлять, например, благосостояние жите-



лей — благодаря технологическому прогрессу, — а также чем больше становится само население и экономика, тем больше растет данный показатель просто в силу расширения пространства выборки и роста суммарного значения.



**Рисунок 14.1:** Вильфредо Парето в молодости, до того, как открыл степенные законы

Суть проблемы в том, что в классе случайных величин с одним жирным хвостом — ограниченных слева и неограниченных справа,  $X \in [x_{\min}, \infty)$ , — вклад квантиля в выборку оказывается смещенной оценкой для истинного вклада этого квантиля.

Определим вклад  $q$ -го квантиля

$$\kappa_q = q \frac{\mathbb{E}(X \mid X > h(q))}{\mathbb{E}X},$$

где  $h(q) = \inf \{h \in [x_{\min}, \infty) \mid \mathbb{P}(X > h) \leq q\}$  — порог, который случайная величина  $X$  превышает с вероятностью  $q$ .

Для данной выборки  $(X_k)_{1 \leq k \leq n}$  «естественной» оценкой, используемой в большинстве академических исследований, служит отношение  $\hat{\kappa}_q \triangleq \frac{q\text{-й квантиль}}{\text{сумма}}$ , которое можно выразить как

$$\hat{\kappa}_q \triangleq \frac{\sum_{i=1}^n \mathbb{1}_{X_i > \hat{h}_q} X_i}{\sum_{i=1}^n X_i},$$

где  $\hat{h}_q$  — оценка порога для превышения с вероятностью  $q$ :

$$\hat{h}(q) = \inf \left\{ h \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i > h} \leq q \right. \right\}.$$

Мы увидим, что наблюдаемая величина  $\hat{\kappa}_q$  смещена вниз относительно истинного отношения  $\kappa_q$ , такого, которое должно оставаться в силе за пределами выборки; что эта систематическая ошибка находится в прямой зависимости от жирности хвоста распределения; что при большой жирности недооценка остается существенной даже при огромных размерах выборки.

## 14.2. ОЦЕНКА ДЛЯ НЕСМЕШАННЫХ РАСПРЕДЕЛЕНИЙ С ХВОСТОМ ПАРЕТО

Пусть  $X$  — случайная величина из класса распределений с правым хвостом степенного закона, то есть:

$$\mathbb{P}(X > x) = L(x) x^{-\alpha}, \quad (14.1)$$

где  $L: [x_{\min}, +\infty) \rightarrow (0, +\infty)$  — медленно меняющаяся функция, определяемая критерием  $\lim_{x \rightarrow \infty} \frac{L(kx)}{L(x)} = 1$  при всех  $k > 0$ .

При низких порогах квантиля ( $< 50\%$ ) различия между распределениями слабы, будь то  $t$ -распределение Стьюдента,  $\alpha$ -устойчивое распределение по Леви, распределение Дагума<sup>1</sup> [53], [54], распределение Сингха — Маддалы<sup>2</sup> [213] или само распределение Парето.

При показателях хвоста  $1 \leq \alpha \leq 2$ , как замечено в [235] (Глава 8 в этой книге), закон больших чисел работает, но крайне медленно. Проблема обостряется при приближении сверху к 1, а при  $\alpha = 1$  сходимость пропадает вовсе.

### 14.2.1. Смещение и сходимость

**Простое распределение Парето.** Рассмотрим вначале плотность  $\varphi_\alpha(x)$  распределения  $\alpha$ -Парето с нижней границей  $x_{\min} > 0$ , то есть  $\varphi_\alpha(x) = \alpha x_{\min}^\alpha x^{-\alpha-1} \mathbb{1}_{x \geq x_{\min}}$ , и кумулятивное распределение  $\mathbb{P}(X > x) = \left(\frac{x_{\min}}{x}\right)^\alpha$ . При таких допущениях отсечение  $q$ -квантиля происходит в точке  $h(q) = x_{\min} q^{-1/\alpha}$ , и доля этого квантиля получается

$$\kappa_q = \frac{\int_{h(q)}^{\infty} x \varphi_\alpha(x) dx}{\int_{x_{\min}}^{\infty} x \varphi_\alpha(x) dx} = \left(\frac{h(q)}{x_{\min}}\right)^{1-\alpha} = q^{\frac{\alpha-1}{\alpha}}. \quad (14.2)$$

Даже если распределение случайной величины  $X$  следует  $\alpha$ -Парето только после точки отсечения  $x_{\text{cut}}$  и эта точка ниже границы квантиля  $h(q)$ , так что  $\mathbb{P}(X > x) = \left(\frac{\lambda}{x}\right)^\alpha$  при некотором  $\lambda > 0$ , мы все еще знаем границу  $h(q) = \lambda q^{-1/\alpha}$ , и

$$\kappa_q = \frac{\alpha}{\alpha-1} \frac{\lambda}{\mathbb{E}X} q^{\frac{\alpha-1}{\alpha}}.$$

1 Случайная величина  $X$  распределена по Дагуму,  $X \sim D(a, b, p, \delta)$ , где коэффициенты формы  $a, p > 0$ , коэффициент масштаба  $b > 0$  и масса в начальной точке  $\delta \in [0, 1]$ , если распределение вероятностей  $F_X(x) = \delta + (1-\delta)(1+(x/b)^{-a})^{-p}$ . При этом индекс Джини  $g = \frac{\Gamma(p)\Gamma(2p+1/a)}{\Gamma(2p)\Gamma(p+1/a)} - 1$ . Среднее существует при  $a > 1$ , дисперсия существует при  $a > 2$ .

2 Распределению Сингха — Маддалы (*S. K. Singh, G. S. Maddala*), оно же распределение Бэрра (*Irving W. Burr*), следует случайная величина вида  $1/X$ , где случайная величина  $X$  следует дагумову распределению.

Таким образом, для оценки  $\kappa_q$  требуется знать показатель  $\alpha$  и коэффициент масштаба  $\lambda$  или хотя бы его отношение к математическому ожиданию случайной величины  $X$ .

В Таблице 14.1 показано смещение  $\hat{\kappa}_q$  от  $\kappa_q$  в случае распределения  $\alpha$ -Парето с  $\alpha = 1,1$  — это значение выбрано как возможный на практике экономический показатель, например для распределения богатств в мире или в отдельной стране, включая развитые страны<sup>1</sup>. При такой альфе оценка крайне чувствительна к «малым» выборкам, а именно к выборкам меньше  $10^8$ . Мы выполнили триллион моделирований с различными размерами выборки. Истинное значение  $\kappa_{0,01} \approx 0,657933$ , но даже при размере выборки 100 миллионов оценка серьезно смещена, как видно в таблице.

Естественно, смещение стремительно (и нелинейно) сокращается при удалении показателя  $\alpha$  от 1, становясь слабым вблизи 2, хотя при смешанном распределении с переменным  $\alpha$  влияние сохраняется дольше, как мы увидим позже. И смещение слабее при спуске от верхнего процентиля к низким квантилям; таким образом, проблема острая только для пресловутого «одного процента» и только при низких показателях.

**Таблица 14.1:** Смещение оценки от  $\kappa = 0,657933$  (по  $10^{12}$  запусков Монте-Карло)

$\hat{\kappa}(n)$	Среднее	Медиана	Среднеквадратическая ошибка по запускам Монте-Карло
$\hat{\kappa}(10^3)$	0,405235	0,367698	0,160244
$\hat{\kappa}(10^4)$	0,485916	0,458449	0,117917
$\hat{\kappa}(10^5)$	0,539028	0,516415	0,0931362
$\hat{\kappa}(10^6)$	0,581384	0,555997	0,0853593
$\hat{\kappa}(10^7)$	0,591506	0,575262	0,0601528
$\hat{\kappa}(10^8)$	0,606513	0,593667	0,0461397

Глядя на такие результаты и число выполненных тестов, выдвигаем гипотезу: смещение  $\kappa_q - \hat{\kappa}_q(n)$  «по порядку величины» составляет  $c(\alpha, q)n^{-b(q)(\alpha-1)}$ , где константы  $b(q)$  и  $c(\alpha, q)$  еще надо найти. Результаты моделирования совместимы с  $b(q) = 1$  независимо от  $\alpha$  и  $q$ , но медленная сходимость оценки и стандартного отклонения к 0 затрудняют точное суждение об этих константах.

**Общий случай.** Чтобы рассмотреть произвольное распределение, обозначим вклад квантиля при заданном пороге  $h$

$$\kappa_h = \mathbb{P}(X > x) \frac{\mathbb{E}(X | X > h)}{\mathbb{E}X} = \frac{\mathbb{E}(X \mathbb{1}_{X>h})}{\mathbb{E}X},$$

то есть  $\kappa_q = \kappa_{h(q)}$ . Обозначим также оценку по выборке размером  $n$

$$\hat{\kappa}_h \triangleq \frac{\sum_{i=1}^n \mathbb{1}_{X_i > h} X_i}{\sum_{i=1}^n X_i},$$

где  $X_i$  — независимые экземпляры случайной величины  $X$ . Интуитивно смещение оценки  $\hat{\kappa}_q$  от  $\kappa_q$  происходит от разной выпуклости вверх оценки концентрации как функции от новости: окажется ли очередное значение выборки выше порога. Пусть

<sup>1</sup> Такое значение ниже типичного в литературе по экономике, где оно в районе 2, но, следуя [86], мы не можем нашими наблюдениями исключить и более низкие значения. — Прим. автора.

$A_h(n) = \sum_{i=1}^n \mathbb{1}_{X_i > h} X_i$  и  $S(n) = \sum_{i=1}^n X_i$ , так что  $\hat{\kappa}_h(n) = \frac{A_h(n)}{S(n)}$  и пусть порог  $h$  зафиксирован. Если очередное значение в выборке  $X_{n+1} < h$ , то новое значение  $\hat{\kappa}_h(n+1) = \frac{A_h(n)}{S(n) + X_{n+1}}$ .

Это значение выпукло вниз по  $X_{n+1}$ , так что неопределенность из-за  $X_{n+1}$  увеличивается по математическому ожиданию. В случае же, если очередное значение в выборке  $X_{n+1} > h$ , новое значение  $\hat{\kappa}_h(n+1) \approx \frac{A_h(n) + X_{n+1} - h}{S(n) + X_{n+1} - h} = 1 - \frac{S(n) - A_h(n)}{S(n) + X_{n+1} - h}$ , и теперь значение выпукло вверх по  $X_{n+1}$ , так что неопределенность из-за  $X_{n+1}$  уменьшается. Соревнование между этими двумя противоположными эффектами заканчивается в пользу последнего из-за большей выпуклости вверх по случайной величине, а также более высокой вариативности (по какому количественному показателю ее ни оценивать) случайной величины при условии, что она превысила порог, чем при условии, что она оказалась ниже порога. Эффект тем сильнее, чем жирнее правый хвост распределения. В итоге находим  $\mathbb{E} \hat{\kappa}_h(n) \leq \frac{\mathbb{E} A_h(n)}{\mathbb{E} S(n)} = \kappa_h$  (заметим, что эффект недооценки уровня концентрации дополнительно усиливается, если перестать фиксировать порог  $\hat{h}(q)$  при добавлении очередной выборки, из-за небольшого увеличения математического ожидания оценки  $\hat{h}(q)$ ; впрочем, этим эффектом можно пренебречь). Приходим к следующему.

#### Предложение 14.1

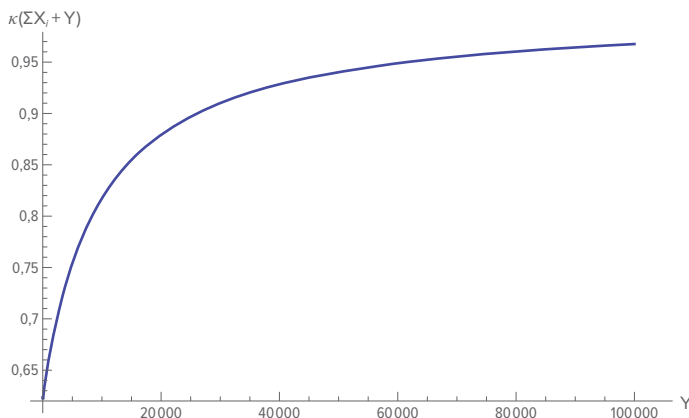
Пусть  $X = (X)_{i=1}^n$  — случайная выборка размером  $n > \frac{1}{q}$ ,  $Y = X_{n+1}$  — одно дополнительное наблюдение и пусть  $\hat{\kappa}_h(X \sqcup Y) \triangleq \frac{\sum_{i=1}^n \mathbb{1}_{X_i > h} X_i + \mathbb{1}_{Y > h} Y}{\sum_{i=1}^n X_i + Y}$ . Заметим, что всякий раз, когда  $Y > h$ , вторая частная производная по  $Y$

$$\frac{\partial^2 \hat{\kappa}_h(X \sqcup Y)}{\partial Y^2} \leq 0.$$

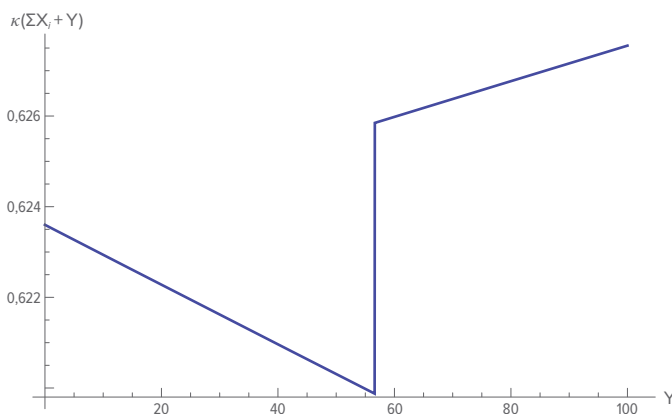
Это неравенство остается в силе и для  $\hat{\kappa}_q$ , поскольку значение  $\hat{h}(q, X \sqcup Y)$  не зависит от конкретного значения  $Y > \hat{h}(q, X)$ .

Ситуация, с которой мы столкнулись, отличается от обычного эффекта малой выборки — возникающего в результате сильного влияния редкого наблюдения из хвостов распределения, поскольку у такого наблюдения мало шансов попасть в малую выборку. Обычный эффект малой выборки устраняется при повторениях выборки. Однако теперь мы рассматриваем оценку с выпуклостью вверх, создающую верхний предел для измерений при конечном  $n$ , которое отсекает большие отклонения, и возникают проблемы с агрегацией, которые мы изложим ниже в Теореме 1.

На практике даже при огромных выборках вклад редких событий в  $\kappa_q$  замедляет сходимость выборочной оценки к истинному значению. Чтобы получить лучшую, несмещенную оценку, нужно идти другим путем: оценить параметры распределения  $(\hat{\alpha}, \hat{\lambda})$  и затем вычислить теоретический вклад хвоста  $\kappa_q(\hat{\alpha}, \hat{\lambda})$ . Фальк [86] замечает, что даже при использовании оценок  $\alpha$  и  $\lambda$  сходимость крайне медленная, а именно порядка  $n^{-\delta} / \ln n$ , где показатель  $\delta$  зависит от  $\alpha$  и дивергенции между фактическим распределением и теоретическим Парето (оценивается расстоянием Хеллингера). В частности,  $\delta \rightarrow 0$  при  $\alpha \rightarrow 1$ , что весьма замедляет сходимость при низких  $\alpha$ .



**Рисунок 14.2:** Влияние дополнительных наблюдений на  $\kappa$



**Рисунок 14.3:** Влияние дополнительных наблюдений на  $\kappa$  (фрагмент). Выпуклость вниз слева от порога  $h$  (здесь это  $Y < 57$ ) занимает ничтожную площадь порядка  $1/n$

### 14.3. НЕРАВЕНСТВО АГРЕГАЦИИ НЕРАВЕНСТВ

При оценке среднего жирного хвоста случайной величины  $(X)_i^j$  по  $m$  подвыборкам размера  $n_i$ , при общем числе наблюдений  $n = \sum_{i=1}^m n_i$ , двухуровневая нумерация наблюдений вида  $j$ -е наблюдение в  $i$ -й подвыборке не влияет на итог по всем  $n$  наблюдениям. Несущественность разбиения  $n$ -выборки на  $m$  подвыборки демонстрирует выпуклость вверх количественного показателя  $\kappa^1$ . Докажем далее, что глобальная концентрация, найденная как  $\hat{\kappa}_q$  по широкому набору данных, окажется выше локальных концентраций, так что, например, агрегация данных по Европе даст  $\hat{\kappa}_q$  выше усредненной концентрации по странам — «неравенство между неравенствами». Другими словами, мы утверждаем, что смещение оценки при использовании  $\hat{\kappa}_q(n)$  дополнительно усиливается, когда выборку разбивают на подвыборки и берут взвешенное среднее оценок  $\hat{\kappa}_q(n_i)$ .

<sup>1</sup> Та же выпуклость вверх и то же общее смещение применимы к логнормальному распределению, и усиливаются при высокой дисперсии. — Прим. автора.

**Теорема 4**

Разрежем  $n$  значений данных на  $m$  подвыборок  $N = N_1 \cup \dots \cup N_m$  размером  $n_1, \dots, n_m$ , где  $\sum_{i=1}^m n_i = n$ , и обозначим как  $S_1, \dots, S_m$  суммы величин в подвыборках и как  $S = \sum_{i=1}^m S_i$  сумму по всей выборке. Тогда

$$\mathbb{E} \hat{\kappa}_q(N) \geq \sum_{i=1}^m \mathbb{E} \frac{S_i}{S} \mathbb{E} \hat{\kappa}_q(N_i).$$

Если дополнительно предположить, что случайные величины  $X_j$  одинаково распределены во всех подвыборках, то

$$\mathbb{E} \hat{\kappa}_q(N) \geq \sum_{i=1}^m \frac{n_i}{n} \mathbb{E} \hat{\kappa}_q(N_i).$$

Другими словами, агрегация оценок концентрации по подвыборкам, взвешенная согласно размерам той или иной подвыборки, недооценивает концентрацию всей выборки.

*Доказательство.* Элементарной индукцией вопрос сводится к случаю двух подвыборок. Пусть  $q \in (0, 1)$  и выборки  $(X_1, \dots, X_m)$  и  $(X'_1, \dots, X'_n)$  состоят каждая из положительных н. о. р. случайных величин:  $X_i$  следуют распределению  $p(dx)$  и  $X'_j$  следуют распределению  $p'(dx')$ . Примем для простоты, что  $qm$  и  $qn$  — целые числа. Обозначим  $S = \sum_{i=1}^m X_i$  и  $S' = \sum_{i=1}^n X'_i$ . Определим  $A = \sum_{i=1}^{mq} X_{[i]}$ , где  $X_{[i]}$  —  $i$ -е наибольшее значение

в  $(X_1, \dots, X_m)$ , и  $A' = \sum_{i=1}^{nq} X'_{[i]}$ , где  $X'_{[i]}$  —  $i$ -е наибольшее значение в  $(X'_1, \dots, X'_n)$ . Кроме

того, обозначим  $S'' = S + S'$  и  $A'' = \sum_{i=1}^{(m+n)q} X''_{[i]}$ , где  $X''_{[i]}$  —  $i$ -е наибольшее значение в объединенной выборке  $(X_1, \dots, X_m, X'_1, \dots, X'_n)$ .

Оценки  $q$ -концентрации для выборок  $X = (X_1, \dots, X_m)$ ,  $X' = (X'_1, \dots, X'_n)$  и  $X'' = (X_1, \dots, X_m, X'_1, \dots, X'_n)$  будут

$$\kappa = \frac{A}{S} \quad \kappa' = \frac{A'}{S'} \quad \kappa'' = \frac{A''}{S''}.$$

Требуется доказать, что

$$\mathbb{E} \kappa'' \geq \mathbb{E} \frac{S}{S''} \mathbb{E} \kappa + \mathbb{E} \frac{S'}{S''} \mathbb{E} \kappa'.$$

Заметим, что

$$A = \max_{\substack{J \subset \{1, \dots, m\} \\ |J|=qm}} \sum_{i \in J} X_i$$

и, аналогичным образом,  $A' = \max_{J' \subset \{1, \dots, n\}, |J'|=qn} \sum_{i \in J'} X'_i$  и  $A'' = \max_{J'' \subset \{1, \dots, m+n\}, |J''|=q(m+n)} \sum_{i \in J''} X_i$ , где большие номера из  $J''$  используются в смысле  $X_{m+i} = X'_i$ , при  $i = 1, \dots, n$ . Если  $J \subset \{1, \dots, m\}$ ,  $|J| = qm$  и  $J' \subset \{1, \dots, n\}$ ,  $|J'| = qn$ , то множество  $J'' = J \cup J'$  имеет мощность  $q(m+n)$ . Значит, независимо от конкретной выборки сумма  $A + A' = \sum_{i \in J''} X_i \leq A''$ . Следовательно,  $\kappa'' \geq \frac{S}{S''} \kappa + \frac{S'}{S''} \kappa'$ , и мы можем записать:

$$\mathbb{E} \kappa'' \geq \mathbb{E} \frac{S}{S''} \kappa + \mathbb{E} \frac{S'}{S''} \kappa'.$$

Покажем теперь, что

$$\mathbb{E} \frac{S}{S''} \kappa = \mathbb{E} \frac{A}{S''} \geq \mathbb{E} \frac{S}{S''} \mathbb{E} \frac{A}{S}.$$

Если это так, то для  $\kappa'$  будет аналогичным образом верно, что

$$\mathbb{E} \frac{S'}{S''} \kappa' = \mathbb{E} \frac{A'}{S''} \geq \mathbb{E} \frac{S'}{S''} \mathbb{E} \frac{A'}{S'},$$

и мы сможем получить требуемое

$$\mathbb{E} \kappa'' \geq \mathbb{E} \frac{S}{S''} \mathbb{E} \kappa + \mathbb{E} \frac{S'}{S''} \mathbb{E} \kappa'.$$

Пусть  $T = X_{[mq]}$  — точка отсечения (где  $[mq]$  — целая часть  $mq$ ), так что  $A = \sum_{i=1}^m X_i \mathbb{1}_{X_i \geq T}$ , и пусть  $B = S - A = \sum_{i=1}^m X_i \mathbb{1}_{X_i < T}$ . Условно к  $T$  величины  $A$  и  $B$  независимы:  $A$  представляет собой частичную сумму, в которую вошли  $mq$  величин по критерию быть выше  $T$ , тогда как  $B$  представляет собой независимую частичную сумму, в которую вошли  $m(1-q)$  величин по критерию быть ниже  $T$ . Кроме того,  $A$  и  $B$  независимы от  $S'$ . Пусть  $p_A(t, da)$  и  $p_B(t, db)$  — распределения величин  $A$  и  $B$  соответственно, считая  $T = t$ . Вспомним, что  $p'(ds')$  — распределение  $S'$ , и обозначим  $q(dt)$  распределение  $T$ . В этих обозначениях:

$$\mathbb{E} \frac{S}{S''} \kappa = \int \int \frac{a+b}{a+b+s'} \frac{a}{a+b} p_A(t, da) p_B(t, db) q(dt) p'(ds').$$

При заданных  $b$ ,  $t$  и  $s'$  зависимости  $a \rightarrow \frac{a+b}{a+b+s'}$  и  $a \rightarrow \frac{a}{a+b}$  представляют собой две возрастающие функции одной и той же переменной  $a$ , и поэтому условно к  $T$ ,  $B$  и  $S'$

$$\mathbb{E} \left( \frac{S}{S''} \kappa \mid T, B, S' \right) = \mathbb{E} \left( \frac{A}{A+B+S'} \mid T, B, S' \right) \geq \mathbb{E} \left( \frac{A+B}{A+B+S'} \mid T, B, S' \right) \mathbb{E} \left( \frac{A}{A+B} \mid T, B, S' \right).$$

Поскольку данное неравенство выполнено при всех значениях  $T$ ,  $B$  и  $S'$ , оно выполнено для безусловного ожидания, и мы можем записать:

$$\mathbb{E} \frac{S}{S''} \kappa \geq \mathbb{E} \frac{S}{S''} \mathbb{E} \frac{A}{S}.$$

Если две выборки следуют одному и тому же распределению, можем записать:

$$\mathbb{E} \kappa'' \geq \frac{m}{m+n} \mathbb{E} \kappa + \frac{n}{m+n} \mathbb{E} \kappa'.$$

В самом деле, в этом случае можно заметить, что  $\mathbb{E} \frac{S}{S''} = \frac{m}{m+n}$ . В самом деле,  $S = \sum_{i=1}^m X_i$  и  $X_i$  одинаково распределенные; значит,  $\mathbb{E} \frac{S}{S''} = m \mathbb{E} \frac{X}{S''}$ . Но в то же время  $\mathbb{E} \frac{S''}{S''} = 1 = (m+n) \mathbb{E} \frac{X}{S''}$ ; значит,  $\mathbb{E} \frac{X}{S''} = \frac{1}{m+n}$ . Аналогичным образом  $\mathbb{E} \frac{S'}{S''} = \frac{n}{m+n}$ , что и требовалось для последнего неравенства.

На этом теорема доказана. ■

Пусть  $X$  — положительная случайная величина и  $h \in (0, 1)$ . Вспомним, что мы определяли теоретическую  $h$ -оценку концентрации как:

$$\kappa_h = \mathbb{P}(X > h) \frac{\mathbb{E}(X | X > h)}{\mathbb{E}X},$$

тогда как по  $n$ -выборке  $h$ -оценка концентрации —  $\hat{\kappa}_h(n) = \frac{A(n)}{S(n)}$ , где  $A(n)$  и  $S(n)$  определены, как выше, для  $n$ -выборки  $X = (X_1, \dots, X_n)$  н. о. р. случайных величин, распределение которых такое же, как у случайной величины  $X$ .

### Теорема 5

Для любого  $n \in \mathbb{N}$

$$\mathbb{E}\hat{\kappa}_h(n) < \kappa_h$$

и  $\lim_{n \rightarrow \infty} \hat{\kappa}_h(n) = \kappa_h$  в смысле сходимости почти наверное и по вероятности.

*Доказательство.* Приведенное выше следствие показывает, что последовательность  $n \mathbb{E}\hat{\kappa}_h(n)$  супераддитивна; поэтому последовательность  $\mathbb{E}\hat{\kappa}_h(n)$  возрастает. Более того, благодаря закону больших чисел последовательность  $\frac{1}{n}S(n)$  сходится почти наверное и по вероятности к  $\mathbb{E}X$ , а последовательность  $\frac{1}{n}A(n)$  сходится почти наверное и по вероятности к  $\mathbb{E}(X \mathbb{1}_{X>h}) = \mathbb{P}(X > h) \mathbb{E}(X | X > h)$ ; значит, их отношение также сходится почти наверное к  $\kappa_h$ . С другой стороны, это отношение ограничено 1. Теорема Лебега о мажорируемой сходимости<sup>1</sup> позволяет вывести требуемую сходимость по вероятности. ■

## 14.4. СМЕШАННЫЕ РАСПРЕДЕЛЕНИЯ ПО ПОКАЗАТЕЛЮ ХВОСТА

Рассмотрим теперь случайную величину  $X$ , распределение которой  $p(dx)$  представляет собой смесь параметрических распределений с различными значениями параметра:  $p(dx) = \sum_{i=1}^m \omega_i p_{\alpha_i}(dx)$ . Типичную  $n$ -выборку случайной величины  $X$  можно сгенерировать из выборок размером  $n_i = \omega_i n$  из случайных величин  $X_{\alpha_i}$  с распределением  $p_{\alpha_i}$ . Приведенная выше теорема показывает, что в этом случае

$$\mathbb{E}\hat{\kappa}_q(n, X) \geq \sum_{i=1}^m \mathbb{E} \frac{S(\omega_i n, X_{\alpha_i})}{S(n, X)} \mathbb{E}\hat{\kappa}_q(\omega_i n, X_{\alpha_i}).$$

<sup>1</sup> Применительно к теории вероятности теорема Лебега о мажорируемой сходимости утверждает, что если последовательность случайных величин  $X_n \xrightarrow{\text{a.s.}} X$  и интегрируемая случайная величина  $Y \geq X_n$  почти наверное, то случайные величины  $X_n$  и  $X$  интегрируемы, и  $\lim_{n \rightarrow \infty} \mathbb{E}X_n = \mathbb{E}X$ .





**Рисунок 14.4:** Пьер-Симон, маркиз де Лаплас (вверху). Его имя носит одно распределение и горстка результатов. Между тем он стоял за открытием как распределения Коши, так и распределения Гаусса (см. закон Стиглера об эпонимии [218]). Портрет Жан-Батист Полен Герен написал в 1838, после смерти Лапласа. Карл Фридрих Гаусс (внизу), портрет кисти Христиана Альбрехта Йенсена. Гауссовым называют распределение, которое все историки приписывают Лапласу



При  $n \rightarrow +\infty$  каждое из отношений  $\frac{S(\omega_i n, X_{\alpha_i})}{S(n, X)}$  сходится почти наверное к соответствующему весу  $\omega_i$ ; следовательно, имеет место следующее неравенство выпуклости вниз:

$$\kappa_q(X) \geq \sum_{i=1}^m \omega_i \kappa_q(X_{\alpha_i}).$$

Особо интересен случай распределения Парето. Здесь параметр  $\alpha$  представляет показатель хвоста распределения. Если нормализовать ожидания к 1, распределение  $X_\alpha$  будет  $F_\alpha(x) = 1 - \left(\frac{x}{x_{\min}}\right)^{-\alpha}$ , и тогда

$$\kappa_q(X_\alpha) = q^{\frac{\alpha-1}{\alpha}}$$

и

$$\frac{\partial^2}{\partial \alpha^2} \kappa_q(X_\alpha) = q^{\frac{\alpha-1}{\alpha}} \frac{(\ln q)^2}{\alpha^3} > 0.$$

Следовательно,  $\kappa_q(X_\alpha)$  — выпуклая вниз функция показателя  $\alpha$ , и мы можем записать

$$\kappa_q(X) \geq \sum_{i=1}^m \omega_i \kappa_q(X_{\alpha_i}) \geq \kappa_q(X_{\bar{\alpha}}),$$

где  $\bar{\alpha} = \sum_{i=1}^m \omega_i \alpha_i$ .

Теперь допустим, что  $X$  — положительная случайная величина, распределение которой неизвестно, однако известно, что его хвост убывает по степенному закону с неизвестным показателем степени. Несмещенная оценка показателя степени, с некоторой величиной неопределенности (т. е. с распределением возможных истинных значений вокруг некоторого среднего), приведет к смещению  $\kappa_q$  в сторону недооценки.

Поскольку уровень концентрации зависит только от хвоста распределения, это неравенство применимо также в случае комбинации распределений со степенным спадом, как в формуле (23.1):

$$\mathbb{P}(X > h) = \sum_{j=1}^N \omega_j L_j(x) x^{-\alpha_j}. \quad (14.3)$$

Малейшая неопределенность с показателем степени повышает индекс концентрации. Эту систематическую ошибку можно оценить, рассмотрев среднее  $\bar{\alpha} > 1$  между двумя значениями  $\alpha^+ = \alpha + \delta$  и  $\alpha^- = \alpha - \delta$ . Неравенство выпуклости вниз записывается так:

$$\kappa_q(\bar{\alpha}) = q^{1-\frac{1}{\bar{\alpha}}} < \frac{1}{2} \left( q^{1-\frac{1}{\alpha+\delta}} + q^{1-\frac{1}{\alpha-\delta}} \right).$$

Таким образом, на практике оценка  $\bar{\alpha}$  в районе 3/2 (т. н. полукубический степенной закон) даст результаты, аналогичные результатам при  $\alpha$  ближе к 1, которые мы использовали в предыдущем разделе. Просто зависимость  $\kappa_q(\alpha)$  выпукла вниз и в ней

доминирует эффект второго порядка  $\frac{(\ln q) q^{1-\frac{1}{\alpha+\delta}} (\ln q - 2(\alpha+\delta))}{(\alpha+\delta)^4}$ , который усугубляется при низких  $\alpha$ .

Чтобы показать, сколь ненадежны оценки неравенства и концентрации по вкладу того или иного квантиля, укажем, что стандартная ошибка 0,3 в оценке  $\alpha$  приведет к росту  $\kappa_q(\alpha)$  на 0,25.

### 14.5. ЧЕМ БОЛЬШЕ СУММА, ТЕМ ВЫШЕ $\hat{\kappa}_q$

Есть сильная зависимость между оценкой  $\hat{\kappa}_q$  и суммой  $S = \sum_{j=1}^m X_j$ : среди случаев повышения  $\hat{\kappa}_q$  ожидаемая сумма больше. Действительно, согласно Теореме 4 оценка  $\hat{\kappa}_q$  положительно коррелирует с суммой  $S$ .

Пусть рассматриваемые случайные величины касаются благосостояния; на Рисунке 14.5 наблюдаем такой условный рост; другими словами, когда изучаемое распределение принадлежит классу жирных хвостов, максимальные значения одного порядка с суммой, и случаи прибавки богатства связаны с прибавкой оценки неравенства. При такой динамике нелепо предполагать, что прибавка благосостояния произойдет благодаря нижней части или хотя бы средней. (Тот же довод можно применить к войнам, пандемиям, размеру компаний и т. д.)

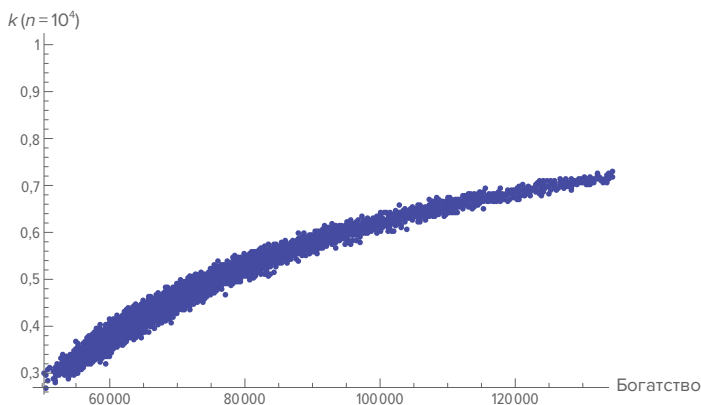


Рисунок 14.5: Влияние прибавки богатства на  $\hat{\kappa}$

### 14.6. ЗАКЛЮЧЕНИЕ И ТОЧНАЯ ОЦЕНКА КОНЦЕНТРАЦИИ

Возможно, что концентрация высока на уровне порождающего процесса, но наблюдения по малым подобластям дали низкий уровень  $\kappa_q$ . Изучая временные ряды, легко создать историческую иллюзию роста исследуемой величины, например концентрации богатств, тогда как на уровне порождающего процесса уровень был высок с самого начала; может случиться и так, что рост исследуемой величины реально происходит, но наблюдаемый рост отчасти объясняется расширением области наблюдения<sup>1</sup>.

Даже оценка показателя  $\alpha$  может быть смещена в некоторых областях, где не видна полная картина: можно доказать, что при неопределенности с «истинным» показателем  $\alpha$  не следует, как обычно делается, брать взвешенное среднее доступных оценок, а лучше взять минимальную из них.

Не следует анализировать изменения  $\hat{\kappa}_q$  по годам без такой поправки. Известно, что ряд теорий был построен на утверждениях о «росте» неравенства, как в [191], без учета природы  $\kappa_q$ . Выдвигались теории о «вариации» неравенства без ссылки на стохастичность оценки и несогласованность  $\kappa_q$  в разное время по разным подобластям. Хуже того, опровергались эти теории опять-таки не глядя на эффект размера,

<sup>1</sup> Обычно у накопленного богатства хвост распределения жирнее, чем у дохода; см. [99]. — Прим. автора.

с привлечением контрданных по выборке другого масштаба, и диалог о неравенстве оказывался статистически неинформативен<sup>1</sup>.

Похоже, что это обычная ошибка при статистических выводах в научной литературе о жирнохвостых данных. Вопросы вызывает сама методология оценок концентрации и изменений концентрации. Например, в тезисе Стивена Пинкера [194] о том, что насилие в мире снижается, замечаем несостоятельный вывод о концентрации ущерба от войн на основании  $\hat{\kappa}_q$ , сделанных по ничтожно малым генеральным совокупностям для данной жирности хвостов<sup>2</sup>. В силу жирнохвостости числа погибших во время войн и острых конфликтов поправка быстро сводит на нет утверждения о статистически значимом спаде военного насилия.

#### 14.6.1. Робастные методы и использование полных данных

Часто приходится встречать довод типа «метод количественной оценки концентрации по вкладу квантиля  $\hat{\kappa}$  робастен и учитывает полный набор данных». Увы, робастные методы обычно терпят неудачу при жирнохвостых данных, см. Главу 8. Хуже того, даже прошедшие тест на «робастность» методы прямой оценки центовых квантилей останутся привязаны к статической и специфической генеральной совокупности, не допускающей агрегации. Эта технология не позволяет взять пробу и сделать научные статистические суждения об истинных свойствах целого.

Возьмем для примера страховую компанию или, еще лучше, компанию перестраховки. Если в ее годовом бухгалтерском отчете оказалось мало исков, это еще не значит, что ее экономическое положение благополучное. На основании выборки за год нелепо было бы судить о концентрации убытков на один страховой случай. Годовые бухгалтерские отчеты не используют для предсказания вариаций год за годом; такие отчеты служат экономистам только для выявления хвостовых и прочих событий, чтобы далее анализировать риски для эффективности, учитывая ее стохастическую природу. Вырабатывая политику, особенно при жирных хвостах, важно отличать «бухгалтерские» (детерминистские) величины от «экономических» (стохастических). То же относится к ущербу от войн: мы не можем оценивать серьезность будущего риска по данным о прошлой истории.

#### 14.6.2. Как правильно оценивать концентрацию?

В наше время специалисты по управлению риском стали применять CVaR и другие количественные показатели, которые экстраполятивны и не выпуклы вверх; примером служит показатель степени хвоста вблизи нижней границы диапазона, как мы убедились при обобщении Теоремы 2 и повторении вывода соответствующей  $\kappa$  или, более строго, при интегрировании функций по возможным состояниям. Указанные методы уточнения обеспечивают меньшую систематическую ошибку и оберегают от потери связи с реальностью при агрегации; в финансовой математике аналогичным

1 Chris Giles. Piketty findings undercut by errors // Financial Times, May 23, 2014. — *Прим. автора.*

2 По данным Ричардсона [194]: «[Войны] следуют правилу 80 на 2: почти восемьдесят процентов смертей вызвано *двумя процентами* [выделил автор цитаты] войн». Очевидно, и Пинкер, и цитируемая им литература следуют в количественной оценке вооруженных конфликтов несостоятельной методологии, поскольку при жирнохвостости войн оценка по центовым квантилям дает огромное смещение. Дальнейшие утверждения о среднем при низких показателях степени хвоста — лженаука. — *Прим. автора.*

образом методами «стохастической волатильности» корректируются цены опционов, когда к стандартному отклонению прибавляется так называемая «улыбка», пропорциональная изменчивости параметра, который представляет волатильность и ошибки измерения. В нашем случае эту роль играет «стохастическая альфа», или «стохастический показатель степени хвоста»<sup>1</sup>. Под экстраполятивностью мы подразумеваем перенос массы в хвост при учете реализаций вне выборки, превышающих наблюдавшиеся крайние значения<sup>2</sup>.

## БЛАГОДАРНОСТИ

Покойному Бенуа Бандельброту, Бранко Милановичу, Доминик Гега, Феликсу Салмону, Бруно Дюпиру, покойному Марку Йору, Альберту Ширяеву, анонимному судье, персоналу ресторана *Luciano* в Бруклине и *Naya* в Манхэттене.

---

1 Заметим также, в дополнение к проблеме оценок центовых квантилей, что некоторые авторы, например [192], работая с цензурированными данными, восполняют нехватку данных в хвостах интерполяцией Парето (на основе параметра хвоста); отсутствующий поддиапазон заполняется условным средним вкладом поддиапазона, но это не полноценная замена расширению степенного закона, и существенное смещение остается. — *Прим. автора.*

2 Даже логнормальное распределение с подгонкой параметра масштаба работает до некоторой степени, поскольку повышение стандартного отклонения перегоняет массу вероятности в правый хвост. Заметим также, что теоремы должны быть применимы и к скачкам Пуассона, но мы сосредоточились на применении к степенному закону, поскольку методы подгонки скачков Пуассона интерполятивны, а подгонка внутренних пробелов в выборке легче, чем подгонка в стороне от выборки. — *Прим. автора.*

## **Часть V**

### **СТАТЬИ О ТЕНЕВЫХ МОМЕНТАХ**



## ТЕНЕВЫЕ МОМЕНТЫ ЯВЛЕНИЙ С МНИМО БЕСКОНЕЧНЫМ СРЕДНИМ<sup>†</sup>

В этой главе предлагается подход к вычислению условных моментов жирнохвостых явлений, которые на первый взгляд кажутся имеющими бесконечное среднее. Проблемы этого рода характерны для случайной величины  $Y$  с тяжелым хвостом и протяженным, но ограниченным носителем.

Мы вводим концепцию двойственного распределения, получаемого логарифмическим преобразованием, которое гладким образом удаляет верхнюю границу. У такого двойственного распределения хвост можно исследовать согласно теории экстремальных значений, без лишних допущений о параметрах, а полученные оценки можно затем использовать для исследования исходного распределения, и его моменты можно вычислить обратным преобразованием.

Центральное отличие нашего подхода от простого усечения состоит в гладкости преобразования исходного распределения в двойственное, обеспечивающее применение теории экстремальных значений.

Возможными областями применения будут военные потери, операционный риск, разрушение окружающей среды, сложные сети и многие другие явления эконофизики.

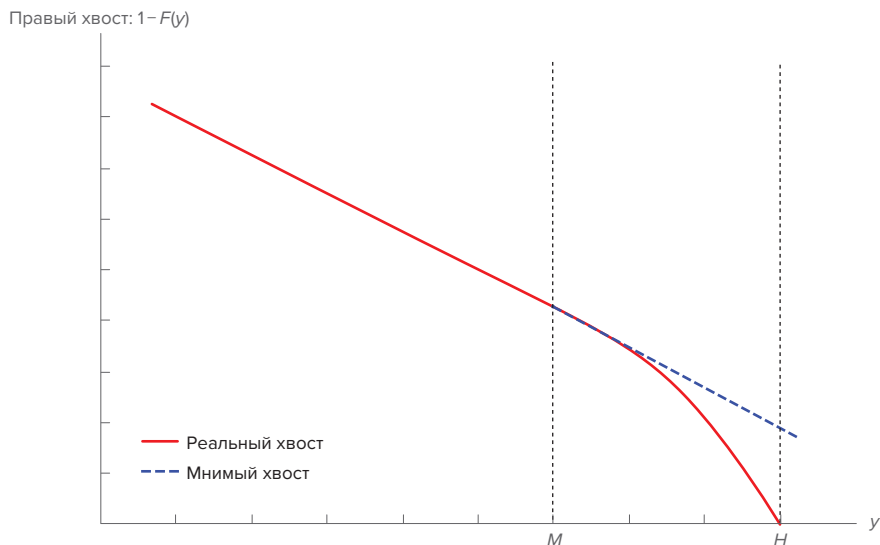
### 15.1. ВВЕДЕНИЕ

Рассмотрим случайную величину  $Y$  с тяжелым хвостом и с конечным носителем  $[L, H]$ . Без потери общности положим, что нижняя граница  $L \gg 0$  и верхняя граница  $H$  конечная, но далекая; соответственно вероятность наблюдений вблизи верхней границы крайне низка, и доступные наблюдения все ниже некоторого  $M \ll H < \infty$ .

На Рисунке 15.1 проблема представлена на графике<sup>1</sup>. Сплошной линией показан реальный хвост нашей случайной величины  $Y$  с далекой верхней границей  $H$ . Однако все наблюдавшиеся значения не превосходили некоторого  $M \ll H$ , и нам невольно представляется мнимый хвост, показанный пунктирной линией. В большинстве случаев различие между двумя хвостами останется невидимо, поскольку очевидное расхождение произойдет лишь вблизи  $H$ .

<sup>†</sup> Это график функции выживания,  $\bar{F}_Y(y) = 1 - F_Y(y)$ . Случайная величина  $Y$  далее трактуется как величина убытков, положительная в случае убытков, так что стоимость под 5-процентным риском  $\text{VaR}_{0.05}(-Y) = F_Y^{-1}(0,95) = \bar{F}_Y^{-1}(0,05)$  и ожидаемые потери с тем же доверительным порогом  $\text{ES}_{0.05} Y = \mathbb{E}(Y | Y > \bar{F}_Y^{-1}(0,05))$ .





**Рисунок 15.1:** Схематичный график возможного заблуждения, если игнорировать существование конечного верхнего предела  $H$  и опираться только на наблюдения, которые все были ниже  $M$

Теперь допустим, что нужно исследовать хвост случайной величины  $Y$  и, поскольку она жирнохвостая, удобно пренебречь ограничением  $H < \infty$  и использовать распределение из класса Фреше<sup>1</sup>. В теории крайних значений [184] распределение  $F$  случайной величины  $Y$  относят к классу Фреше, если  $\bar{F}(y) = 1 - F(y) = y^{-\alpha}L(y)$ , где  $L(y)$  — медленно меняющаяся функция. Другими словами, класс Фреше — это класс распределений, у которых правый хвост следует степенному закону.

Глядя на данные, можно подумать, что правый хвост ведет себя, как пунктирная линия на Рисунке 15.1, и оценить  $\alpha$  ниже 1. Согласно свойствам степенных законов, это значит, что  $EY$  бесконечно (как и все высшие моменты). Кроме того, это значит, что выборочное среднее статистически незначимо и никакой робастности ждать не приходится [168]. Но если  $H$  конечно, такого быть не может: все моменты случайной величины на конечном носителе конечны.

В качестве выхода можно обратиться к параметрической модели, сочетающей жирный хвост и конечный носитель, например усеченное распределение Парето [1]. Но что делать, если  $Y$  ведет себя по Парето не всюду, а только в верхней части хвоста? Подгонять комбинацию моделей со многими параметрами?

В следующем разделе мы предложим простое общее решение, не требующее сильных допущений о параметрах.

## 15.2. ДВОЙСТВЕННОЕ РАСПРЕДЕЛЕНИЕ

Мы находим более удобным не заменять хвосты распределения, а преобразовать данные так, что станут применимы хорошо изученные распределения. На Рисунке 15.1

<sup>1</sup> Заметим, что отнести  $Y$  к классу Фреше нельзя. Если случайная величина ограничена сверху, она не может принадлежать классу Фреше, только классу Вейбулла [116]. — Прим. автора.

реальный и мнимый хвосты остаются неразличимы на большом протяжении. Мы предлагаем воспользоваться этим обстоятельством и преобразовать  $Y$ , устранив верхнюю границу таким образом, что новая, «двойственная» случайная величина  $Z$  будет иметь тот же хвост, что кажущийся хвост для  $Y$ . Затем можно оценить коэффициент формы  $\alpha$  хвоста  $Z$ , вернуться к  $Y$  и вычислить ее моменты или, точнее, ее избыточные моменты, то есть условные моменты выше заданного порога, опираясь на добытые сведения о хвосте  $Z$ .

Пусть  $Y$  определена на носителе  $[L, H]$ . Определим функцию преобразования

$$\varphi(Y) = L - H \ln \frac{H - Y}{H - L}. \quad (15.1)$$

Можно убедиться, что  $\varphi$  — гладкая:  $\varphi \in C^\infty$ ,  $\varphi^{-1}(\infty) = H$  и  $\varphi^{-1}(L) = \varphi(L) = L$ . Новая случайная величина  $Z = \varphi(Y)$  имеет прежнюю нижнюю<sup>1</sup> границу  $L$  и бесконечную верхнюю границу. Заметим, что вид преобразования  $\varphi(\cdot)$  не зависит ни от каких параметров распределения  $Y$ .

По построению  $z = \varphi(y) \approx y$  при больших  $H$ <sup>2</sup>. Это значит, что, если верхняя граница далека и приближение к ней маловероятно, то результаты, которые мы, скорее всего, увидим для хвоста  $Y$ , будут неотличимы от хвоста  $Z = \varphi(Y)$  — если не пытаться распространить их в область ближе к  $H$ . При этом, в отличие от  $Y$ , двойственная случайная величина  $Z$  не ограничена. Следовательно, ее распределение можно спокойно смоделировать в классе Фреше, изучить хвост и затем вернуться к случайной величине  $Y$  и узнать ее моменты (которые у самого двойственного распределения  $Z$  существовать не могут)<sup>3</sup>.

Есть разные способы изучить хвост  $Z$ , см., например, [184] и [87]. Мы предлагаем опереться на так называемую теорему Пикэндса — Болкемы — де Гаана [116]. Эта теорема позволяет сфокусироваться на правом хвосте распределения, не уделяя много внимания поведению ниже заданного порога  $u$ . В нашем случае  $u \geq L$ .

Рассмотрим случайную величину  $Z$  с функцией распределения  $G$  и обозначим  $G_u$  условную функцию распределения величины превышения порога  $W = Z - u$  при условии, что  $Z$  выше или равна порогу<sup>4</sup>. Масштабируя по условию, получаем

$$G_u(w) = \mathbb{P}(Z - u \leq w | Z > u) = \frac{G(u + w) - G(u)}{1 - G(u)},$$

причем  $0 \leq w \leq z_G - u$ , где  $z_G$  — правая конечная точка  $G$ .

1 Поскольку функция  $\varphi(Y)$  возрастает. В самом деле, выражение  $\frac{Y-L}{H-L}$  как функция  $Y$  убывает, поэтому  $-H \ln \frac{Y-L}{H-L}$  возрастает.

2 В самом деле, при  $H \gg 1$  имеем  $\frac{y-L}{H-L} \ll 1$ , откуда  $\varphi(y) = L - H \ln \frac{y-L}{H-L} = L - H \ln \left(1 - \frac{y-L}{H-L}\right) \approx L + H \frac{y-L}{H-L} = y + L \frac{y-L}{H-L} \approx y$ .

3 Заметим, что использование логарифмического преобразования вполне естественно в контексте функции полезности (*utility function*).

4 Пишут так:  $W = Z - u |_{Z \geq u}$ . Наблюдения, в которых  $Z < u$ , не считаются наблюдениями  $W$ , так что случайная величина  $W$  неотрицательна.

Теорема Пикэндса, Болкема и де Гаана гласит, что для широкого класса функций распределения  $G$  при большом  $u$  можно приблизить обобщенным распределением Парето  $G_u$ , т. е.  $G_u(w) \rightarrow \text{GPD}(w; \xi, \sigma)$  при  $u \rightarrow \infty$ , где

$$\text{GPD}(w; \xi, \sigma) = \begin{cases} 1 - \left(1 + \xi \frac{w}{\sigma}\right)^{-1/\xi}, & \text{если } \xi \neq 0, \\ 1 - e^{-\frac{w}{\sigma}}, & \text{если } \xi = 0 \end{cases}, w \geq 0. \quad (15.2)$$

Так называемый коэффициент формы  $\xi$  соответствует  $1/\alpha$  и управляет жирностью хвостов и существованием моментов. У случайной величины, следующей обобщенному распределению Парето, момент порядка  $p$  существует тогда и только тогда, когда  $\xi < 1/p$  или  $\alpha > p$  [184]. Оба коэффициента  $\xi$  и  $\sigma$  можно оценить при помощи ММП или метода моментов [116]<sup>1</sup>.

### 15.3. ВЕРНЕМСЯ К Y: ТЕНЕВОЕ СРЕДНЕЕ, ИЛИ СРЕДНЕЕ ПО ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Обозначим плотности случайных величин  $Y$  и  $Z$  как  $f$  и  $g$ .

Мы помним, что  $Z = \varphi(Y)$ , так что  $Y = \varphi^{-1}(Z) = (L - H)e^{\frac{L-Z}{H}} + H$ <sup>2</sup>.

Предположим теперь, что мы нашли такой порог  $u = L^* \geq L$ , что  $G_u(w) \approx \text{GPD}(w; \xi, \sigma)$ . Тогда и хвост  $Y$  выше  $L^*$  представим этой функцией.

Итак, мы можем, во-первых, приравнять массу вероятности в этих хвостах:

$$\int_{L^*}^{\infty} g(z) dz = \int_{L^*}^{\varphi^{-1}(\infty)} f(y) dy, \quad (15.3)$$

и, во-вторых, имеем формулу для плотности случайной величины  $Z$ :

$$g(z; \xi, \sigma) = \frac{1}{\sigma} \left(1 + \xi \frac{z}{\sigma}\right)^{-\frac{1}{\xi}-1}, z \in [L^*, \infty)^3. \quad (15.4)$$

Обозначая  $\alpha = \xi^{-1}$ , получаем

$$f(y; \alpha, \sigma) = \frac{H}{\sigma(H-y)} \left(1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - \ln(H-y)))\right)^{-\alpha-1}, y \in [L^*, H)^4. \quad (15.5)$$

1 Есть и другие методы работы с конечными (или выпуклыми вверх) верхними границами, как то: замедленные степенные законы (с экспоненциальным затуханием) [197] или растянутые экспоненты [149]; по своей природе они близки к нашему изложению, но не допускают немедленного применения теории экстремальных значений или аналогичных методов оценки параметров. — Прим. автора.

2 В самом деле, по определению  $\varphi$  из  $Z = \varphi(Y)$  сразу следует, что  $\frac{L-Z}{H} = \ln \frac{H-Y}{H-L}$ , откуда  $e^{\frac{L-Z}{H}} = \frac{H-Y}{H-L}$  и, следовательно,  $Y = H - (H-L)e^{\frac{L-Z}{H}}$ .

3 В самом деле,  $g$  — это плотность, или производная функции распределения неограниченной случайной величины  $Z$ , которая выше порога следует  $\text{GPD}(z; \xi, \sigma) = 1 - \left(1 + \xi \frac{z}{\sigma}\right)^{-1/\xi}$  согласно (15.2), заменяя  $w$  на  $z$ .

Производная  $\frac{\partial}{\partial z} \text{GPD}(z; \xi, \sigma) = \frac{\partial}{\partial z} \left(1 - \left(1 + \xi \frac{z}{\sigma}\right)^{-1/\xi}\right) = \frac{1}{\sigma} \left(1 + \xi \frac{z}{\sigma}\right)^{-\frac{1}{\xi}-1}$ .

4 В самом деле, преобразованию случайной величины  $Z = \varphi(Y) = L - H \ln \frac{H-Y}{H-L} = L + H \ln \frac{H-L}{H-Y}$  отвечает преобразование плотности  $g_Z(z) = \frac{1}{\sigma} \left(1 + \frac{z}{\alpha\sigma}\right)^{-\alpha-1}$  в плотность  $f_Y(y) = g_Z(\varphi(y)) \left|\frac{d\varphi(y)}{dy}\right|$ . Первый множитель

или, в терминах функции распределения,

$$F(y; \alpha, \sigma) = 1 - \left( 1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - \ln(H-y))) \right)^{-\alpha}. \quad (15.6)$$

В силу взаимной однозначности преобразования  $\varphi$  очевидно, что параметры  $f$  и  $g$ , полученные методом максимального правдоподобия, совпадут — функции правдоподобия для  $f$  и  $g$  различаются только масштабным коэффициентом.

Мы можем вывести теневое среднее<sup>2</sup> для  $Y$  при условии  $Y > L^*$  как

$$\mathbb{E}(Y | Y > L^*) = \int_{L^*}^H y f(y; \alpha, \sigma) dy, \quad (15.7)$$

откуда

$$\mathbb{E}(Y | Y > L^*) = (H - L^*) e^{\frac{\alpha\sigma}{H}} \left( \frac{\alpha\sigma}{H} \right)^{\alpha} \Gamma\left(1 - \alpha, \frac{\alpha\sigma}{H}\right) + L^*. \quad (15.8)$$

Теперь условное среднее  $Y$  выше  $L^*$  можно оценить по стандартным оценкам  $\hat{\alpha}$  и  $\hat{\sigma}$ , полученным из аппроксимации GPD для хвоста  $Z$ . Стоит отметить, что, если  $L^* = L$ , то  $\mathbb{E}(Y | Y > L^*) = \mathbb{E} Y$ , условное среднее  $Y$  выше  $L$  — это в точности среднее  $Y$ .

Естественно, аналогичным образом можно получить и другие моменты, даже если для них потребуется прибегать к численным методам. Наш метод применим в общем случае, но особенно полезен, когда по имеющимся данным хвост  $Y$  кажется слишком жирным для существования каких-либо конечных моментов, как нередко случается при исследовании риска операционных убытков, распределения степени для больших сложных сетей и прочих экономическxих явлений.

Пусть, например, для некоторой величины  $Z$  коэффициент  $\xi > 1$ . Тогда бесконечны и  $\mathbb{E}(Z | Z > L^*)$ , и  $\mathbb{E} Z^3$ . Как показано на Рисунке 15.1, нам захочется предположить, что и  $\mathbb{E} Y$  бесконечно — и данные, скорее всего, будут подсказывать такой же вывод, если оценивать  $\hat{\xi}$  по хвосту<sup>4</sup> величины  $Y$ . Но это невозможно, ведь  $H < \infty$ , и даже при  $\xi > 1$  мы сможем рассчитать  $\mathbb{E}(Y | Y > L^*)$  по формуле (15.8).

находим подстановкой  $g_z(\varphi(y)) = \frac{1}{\sigma} \left( 1 + \frac{1}{\alpha\sigma} \left( L + H \ln \frac{H-L}{H-y} \right) \right)^{-\alpha-1}$ , второй сомножитель находим дифференцированием  $\frac{\partial \varphi(y)}{\partial y} = \frac{d}{dy} \left( L - H \ln \frac{H-L}{H-y} \right) = \frac{H}{H-y}$ .

1 В самом деле, функцию распределения  $F_Y(y)$  однозначно найдем из двух условий: граничного  $F_Y(H) = 1$  и дифференциального  $\frac{dF_Y(y)}{dy} = f_Y(y)$ . Формула (15.6) удовлетворяет граничному условию, потому что  $F_Y(H) = 1 - \left( 1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - \ln(H-H))) \right)^{-\alpha} = 1 - \left( 1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - (-\infty))) \right)^{-\alpha} = 1 - \left( 1 + \frac{1}{\alpha\sigma} \infty \right)^{-\alpha} = 1 - \infty^{-\alpha} = 1$ ; она также удовлетворяет дифференциальному условию, потому что  $\frac{dF_Y(y)}{dy} = -\frac{d}{dy} \left( 1 - \left( 1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - \ln(H-y))) \right)^{-\alpha} \right) = -\frac{d}{dy} \left( 1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - \ln(H-y))) \right)^{-\alpha} = -(-\alpha) \left( 1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - \ln(H-y))) \right)^{-\alpha-1} \frac{d}{dy} \left( 1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - \ln(H-y))) \right) = \alpha \left( 1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - \ln(H-y))) \right)^{-\alpha-1} \frac{1}{\alpha\sigma} \frac{d}{dy} (L + H(\ln(H-L) - \ln(H-y))) = \frac{1}{\sigma} \left( 1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - \ln(H-y))) \right)^{-\alpha-1} \cdot H \times \frac{d}{dy} (-\ln(H-y)) = \frac{H}{\sigma} \left( 1 + \frac{1}{\alpha\sigma} (L + H(\ln(H-L) - \ln(H-y))) \right)^{-\alpha-1} \frac{1}{H-y}.$

2 Мы называем среднее по генеральной совокупности — в отличие от выборочного среднего — «теневым», потому что непосредственно из данных оно невидимо. — Прим. автора.

3 Вспомним, что для случайной величины  $Z$ , следующей GPD,  $\mathbb{E} Z^p < \infty \Leftrightarrow \xi < 1/p$ . — Прим. автора.

4 Из-за сходимости  $1 - F(y)$  и  $1 - G(z)$ , по крайней мере в области ниже  $M$ , аппроксимация GPD даст для их хвостов статистически неразличимые оценки  $\hat{\xi}$  [184]. — Прим. автора.

### Стоимость под риском и ожидаемые потери

Обращая формулу (15.6), мы можем рассчитать функцию квантилей случайной величины  $Y$  при  $Y \geq L^*$ :

$$Q(p; \alpha, \sigma, H, L) = e^{-y(p)} \left( L e^{\frac{L-\alpha\sigma}{H}} + H e^{y(p)} - H e^{\frac{L-\alpha\sigma}{H}} \right), \quad (15.9)$$

где  $y(p) = \frac{\alpha\sigma}{(1-p)^{1/\alpha} H}$  и  $p \in [0, 1]$ <sup>1</sup>. Функцию квантилей мы опять-таки рассматриваем условную по  $Y$  выше  $L^*$ .

По формуле (15.9) легко посчитать стоимость под риском (*Value-at-Risk*, VaR) для  $Y \geq L^*$  при любом доверительном уровне. Например, 5 %-я VaR операционных потерь  $Y$  в горизонте 1 года<sup>2</sup> — это просто  $\text{VaR}_{0.05} Y = Q(0.95; \alpha, \sigma, H, L)$ .

Нас также может интересовать еще одна характеристика хвостового риска  $Y$  — так называемые ожидаемые потери (*expected shortfall*, ES), определяемые как  $\mathbb{E}(Y | Y > u \geq L^*)$ <sup>3</sup>. Это просто обобщение формулы (15.8).

Мы можем получить ожидаемые потери, вычислив сначала функцию избыточного среднего условной величины  $Y | Y \geq u$ , которая определяется как

$$e_u(Y) = \mathbb{E}(Y - u | Y > u) = \frac{\int_u^\infty (u - y) f(y; \alpha, \sigma) dy}{1 - F(u)}$$

и которую мы найдем при  $u \geq L^*$ . Пользуясь формулой (15.5), получаем

$$\begin{aligned} e_u(Y) &= (H - L) e^{\frac{\alpha\sigma}{H}} \left( \frac{\alpha\sigma}{H} \right)^\alpha \left( 1 + \frac{1}{\alpha\sigma} \left( L + H \ln \frac{H - L}{H - y} \right) \right)^\alpha \times \\ &\times \Gamma \left( 1 - \alpha, \frac{\alpha\sigma}{H} + \frac{L}{H} + \ln \frac{H - L}{H - y} \right). \end{aligned} \quad (15.10)$$

После этого ожидаемые потери вычисляются просто как

$$\mathbb{E}(Y | Y > u \geq L^*) = e_u(Y) + u.$$

В финансовой математике и в управлении рисками ES и VaR могут комбинироваться. Например, пусть кроме стоимости под 5 %-м риском нам еще интересно найти ожидаемые потери с доверительной вероятностью 5 % и пусть известно, что эти потери находятся в хвосте распределения величины потерь,  $Y \geq L^*$ . Тогда нам достаточно сложить  $\text{VaR}_{0.05} Y + e_{\text{VaR}_{0.05} Y}(Y)$ .

1 В самом деле,  $p = F_y(y) = 1 - \left( 1 + \frac{L + H(\ln(H - L) - \ln(H - y))}{\alpha\sigma} \right)^{-\alpha} \Leftrightarrow L + H(\ln(H - L) - \ln(H - y)) = \alpha\sigma((1 - p)^{-1/\alpha} - 1) \Leftrightarrow \ln(H - L) - \frac{1}{H}(\alpha\sigma((1 - p)^{-1/\alpha} - 1) - L) = \ln(H - y) \Leftrightarrow y = H - (H - L) \exp\left(\frac{\alpha\sigma}{(1 - p)^{1/\alpha} H} + \frac{L - \alpha\sigma}{H}\right) = Q_y(p)$ .

2 Имеется в виду, что  $Y$  положительна в случае убытков, так что тяжелые хвостовые убытки расположены в правом хвосте  $Y$ . В определении стоимости под риском, приведенном в Разделе 2, фигурировала величина платежа  $X = -Y$ .

3 Если речь о потерях, ожидаемых с вероятностью 5 %,  $\text{ES}_{0.05} Y$ , то  $u = \text{VaR}_{0.05} Y$ .

## 15.4. СРАВНЕНИЕ С ДРУГИМИ МЕТОДАМИ

Существует три способа явно отрезать хвосты распределений Парето (не считая методов растянуть, или «замедлить рост», функции распределения).

1) Первый состоит в резком отсечении части распределения по заданной конечной точке и далее необходимой перенормировки. Можно отрезать оба хвоста и нормализовать функцию распределения на отрезке от  $L$  до  $H$ , передав массу вероятности на все точки отрезка поровну.

2) При втором предполагается, что точка  $H$  служит поглощающим барьером, то есть что все реализации исходной случайной величины, превышающие  $H$ , сжимаются в дельта-функцию Дирака в точке  $H$ ; такое практикуется в моделировании деривативов. В этом случае новое распределение совпадает с исходным, например с обычным Парето, вплоть до точки  $H$ .

3) Третий способ был представлен в этом разделе.

Вопрос уже прорабатывался в финансовой математике, где сопоставляли усечение нормального распределения (для уточнения нормальной модели Башелье, где использовалось просто гауссово распределение) и логарифмическое преобразование (Sprenkle, 1961 [216]); для стандартной модели выбрали логарифмическое преобразование и соответствующее однохвостое логнормальное распределение. Помимо преимуществ, зависящих от распределения, таких как аддитивность логарифмов платежных функций, отмечались и преимущества общего рода: из моделей устранялся «уступ», резкое изменение в функции плотности, порождающее неустойчивость оценок риска по негладким функциям.

В теории экстремальных значений (см. Beirlant, 2014) для усечения распределения решили подвергать превышение в хвосте преобразованию вида  $Y^{-\alpha} \rightarrow Y^{-\alpha} - H^{-\alpha}$  и применять ТЭЗ к результату. Преобразование зависит от параметра  $\alpha$ , который приходится оценивать по ММП. Мы видим проблемы с выбором недостаточно гладкого преобразования. Это та же проблема, что возникает с моделями финансовых активов, а именно резкий «уступ», ниже которого есть плотность вероятности, а выше нет. Из-за этого оценка математического ожидания оказывается выше, чем наша, особенно при  $\alpha < 1$ , как показано на Рисунке 15.2.

Мы можем продемонстрировать этот эффект так. Пусть мы наблюдаем распределение, которое похоже на Парето, но усечено, и пусть мы обрабатываем его как неко-

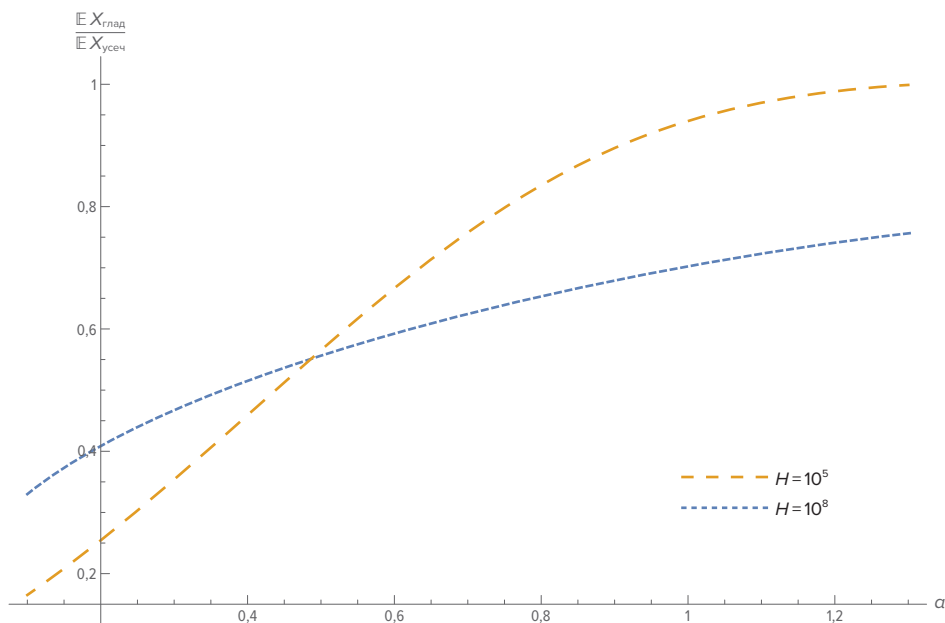
торое распределение Парето. Плотность  $f(x) = \frac{1}{\sigma} \left( \frac{x-L}{\alpha\sigma} + 1 \right)^{-\alpha-1}$ ,  $x \in [L, \infty)$  после усече-

ния превратится в  $g(x) = \frac{\left( \frac{x-L}{\alpha\sigma} + 1 \right)^{-\alpha-1}}{\sigma(1-\alpha^\alpha \sigma^\alpha (\alpha\sigma + H-L)^{-\alpha})}$ ,  $x \in [L, H]$ .

Моменты порядка  $p$  усеченного Парето (наблюдаемых реализаций порождающего процесса и усечения) такие:

$$M(p) = \alpha e^{-\pi\rho} (\alpha\sigma)^\alpha (\alpha\sigma - L)^{p-\alpha} \frac{B_{\frac{H}{L-\alpha\sigma}}(p+1, -\alpha) - B_{\frac{L}{L-\alpha\sigma}}(p+1, -\alpha)}{\left( \frac{\alpha\sigma}{\alpha\sigma + H-L} \right)^\alpha - 1}, \quad (15.11)$$

где  $B_x(a, b)$  — неполная бета-функция Эйлера  $B_x(a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt$ .



**Рисунок 15.2:** Отношение математического ожидания после гладкого преобразования и после усечения

В итоге отношение среднего по распределению Парето после мягкого усечения к среднему по усеченному Парето

$$r(H, \alpha) = e^{-\frac{\alpha}{H}} \left( \frac{\alpha}{H} \right)^{\alpha} \frac{-\left( \frac{\alpha + H}{\alpha} \right)^{\alpha} + H + 1}{(\alpha - 1) \left( \left( \frac{\alpha}{H} \right)^{\alpha} - \left( \frac{\alpha + H}{H} \right)^{\alpha} \right) E_{\alpha} \frac{\alpha}{H}}, \quad (15.12)$$

где  $E_{\alpha} \frac{\alpha}{H}$  — интегральная показательная функция,  $E_{\alpha} z = \int_1^{\infty} \frac{e^{t(-z)}}{t^{\alpha}} dt$ .

## 15.5. ОБЛАСТИ ПРИМЕНЕНИЯ

**Операционный риск.** Убытки фирмы ограничены ее капитализацией, то есть существуют известные максимумы потерь.

**Лимитированные договоры перестрахования.** Практически во всех договорах перестрахования прописаны лимиты (максимальные иски), но перестраховщик может заключить много договоров на общий источник риска, и тогда сложение договоров поднимет верхнюю границу потенциального суммарного ущерба.

**Острые конфликты.** Хотя войны имеют крайне жирные хвосты, все же максимальное число жертв военного конфликта не может превысить население мира.

**Кредитный риск.** Как и у договоров перестрахования, у займов определены конечные максимальные потери.

**Размер города.** Хотя продемонстрировано, что города по своему населению следуют распределению Ципфа, население отдельного города не может превысить население мира.

**Ущерб окружающей среде.** Экологические катастрофы исключительно жирно-хвосты, и все же зона бедствия ограничена нашей планетой (в некоторых задачах — одним континентом), то есть верхняя граница заведомо существует.

**Сложные сети.** Число соединений в сети конечно<sup>1</sup>.

**Размер компании.** Объем продаж компании не может превысить валовый внутренний продукт (*gross domestic product*, GDP).

**Землетрясения.** Максимальный ущерб от землетрясения ограничен энергией землетрясения.

**Гидрология.** Максимальный уровень наводнения поддается оценке из физических соображений.

---

<sup>1</sup> И в сети с  $N$  узлами число соединений одного узла максимум  $N-1$ .



## О ХВОСТОВОМ РИСКЕ ОСТРОГО КОНФЛИКТА<sup>†</sup>

Мы исследуем различные статистические картины острых конфликтов современной исторической эры, фокусируясь на недостаточности и ненадежности данных. Мы применяем методы из теории экстремальных значений, для чего логарифмически преобразуем данные, устраняя конечность носителя; затем, пользуясь ограниченностью максимально возможного числа жертв, преобразуем данные обратно и узнаем ожидаемые средние. Мы находим, что выборочное среднее недооценивает ожидаемое среднее минимум втрое, то есть при наивном наблюдении происходит грубая недооценка серьезности конфликтов. Мы проверяем робастность, применяя метод складного ножа к выборкам из высоких и низких оценок. Мы изучаем продолжительность времени между поступлением хвостовых событий и подтверждаем отсутствие памяти (первого порядка). Полученная статистическая картина расходится с утверждениями о «долгом мире».

### 16.1. ВВЕДЕНИЕ И РЕЗЮМЕ

Это исследование во многом посвящено новым статистическим методологиям работы с жирнохвостыми (и ненадежными) данными, а также со случайными величинами, следующими ограниченному распределению, которое локально ведет себя как Парето — эти свойства присущи, например, статистике острых конфликтов<sup>1</sup>.

---

<sup>1</sup> Благодарности: капитан Марк Вейзенборн выполнил неблагодарную и жуткую работу по компиляции данных из разных источников и привязке всех конфликтов к нарративам в Википедии (см. Приложение 1). Мы также получили щедрую помощь в социальных сетях, где публиковали данные для ознакомления и критики, а также советы от историков, которых благодарим в том же приложении. И мы благодарны покойному Бенуа Мандельброту за идеи насчет хвостовых свойств войн и конфликтов, — а еще Яниру Бар-Яму, Рафаэлю Дуади...

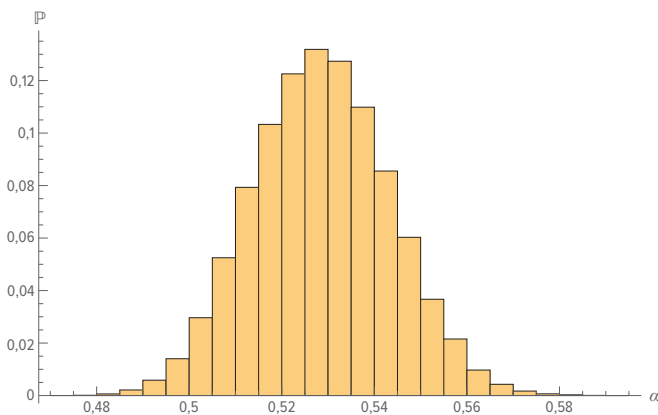
Уровень насилия гораздо суровее, чем представляется в обычной аналитике и в господствующей теории «долгого мира», утверждающей, будто уровень насилия снизился. Адаптируя методы теории экстремальных значений и внося поправки на ошибки в репортажах о конфликтах и исторических оценках жертв войны, мы строим статистические картины крупных острых конфликтов, начиная с 50 тысяч жертв в наше время или с эквивалентного числа по отношению к населению мира в прошлом (например, в XVIII веке это 5 тыс.). Вразрез с текущим дискурсом полученные нами статистические картины показывают, что (1) риск острого конфликта не спадает, а начинает недооцениваться из-за наивных методик, опирающихся на тенденцию среднегодовой статистики или на выборочное среднее как оценку истинного среднего явления с крайне жирным хвостом; (2) судя по продолжительности времени между поступлением событий, у вооруженных конфликтов отсутствует память, что несовместимо с концепцией временного тренда. В нашем анализе рассматриваются: (1) необработанные данные, собранные и оцененные историками; (2) наивное преобразование, используемое некоторыми историками и социологами, которое изменяет масштаб прошлых конфликтов и числа жертв по отношению к численности населения; (3) более важное логарифмическое преобразование, благодаря которому можно учесть, что число жертв конфликта не может превысить население мира. (Последнее аналогично преобразованию данных в логарифм платежной функции в финансовой математике, чтобы корректно использовать функции распределения, определенные на всей оси действительных чисел.)

В общем и целом, если говорить о классе данных (необработанных или перемасштабированных), то мы наблюдаем, что (1) число жертв следует распределению в классе степенного закона<sup>1</sup>. В случае логарифмически преобразованных данных мы наблюдаем  $0,4 \leq \alpha \leq 0,7$ , то есть крайне жирнохвостое явление с несуществующим средним (и этот результат получен робастным образом); (2) судя по продолжительности времени между поступлением событий, конфликты выше порога 50 тысяч жертв следуют однородному процессу Пуассона, без определенного тренда, что противоречит популярному нарративу о спаде насилия; (3) истинное среднее, которое следует ожидать в будущем и которое наиболее совместимо с данными, при всей высокой стохастичности можно оценить: оно  $\approx 3 \times$  прошлое среднее, то есть вдвое хуже прошлого.

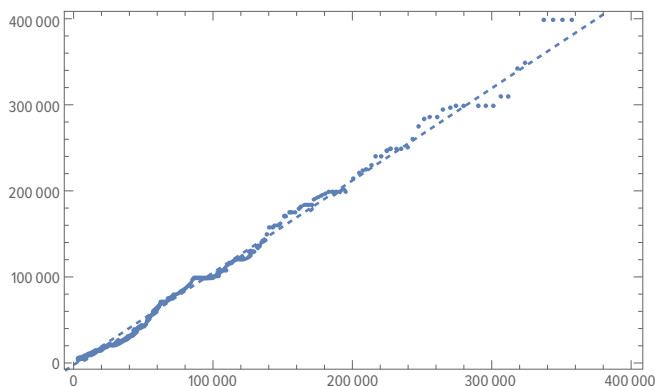
Далее мы объясним: (1) как возникает грубая недооценка среднего (то есть математического ожидания числа жертв) при обычных методах анализа данных, то есть по среднему имеющихся наблюдений, а не по специальной оценке истинного среднего. Здесь ситуация отличается от экспоненциального хвоста, где уровень шума ниже и наивная картина точнее; (2) как обманчиво долгие (и волатильные) интервалы между большими конфликтами порождают ложные концепции.

Чтобы скорректировать неточности в числовых оценках историков, мы проводим обычный бутстрэп наших оценок, а также оцениваем методом Монте-Карло недостоверность статистики войн и бессобытийных периодов в письменной истории.

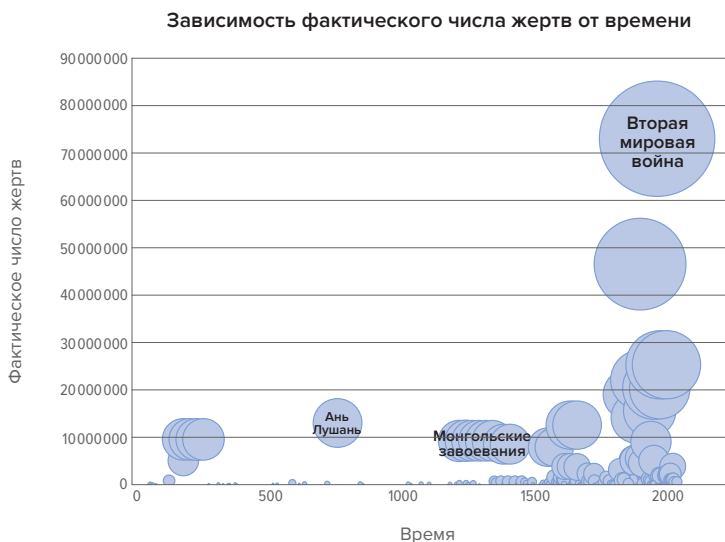
<sup>1</sup> Ряд прошлых исследований обнаружил в этих данных следование распределению Парето (например, [38]). Наше исследование, отличаясь использованием методов теории экстремальных значений, бутстрэпа по надежности и преобразований компактного носителя, приводит к другой калибровке и интерпретации.



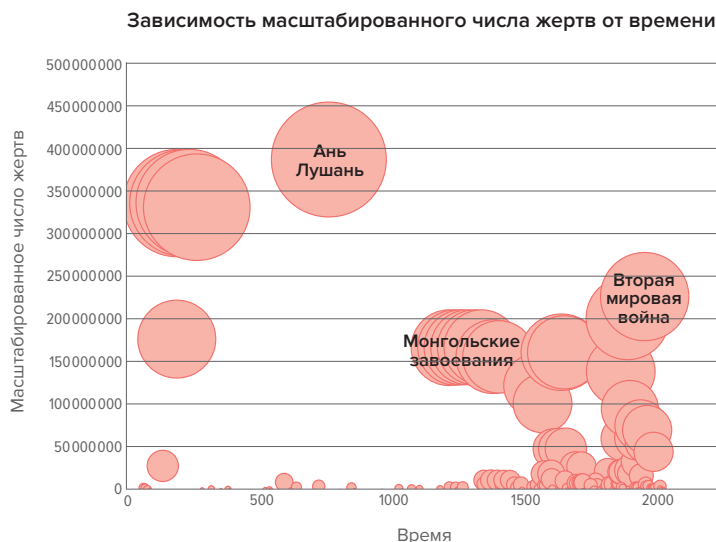
**Рисунок 16.1:** Значения показателя хвоста  $\alpha$  по оценке Хилла, полученной из 100 000 масштабированных оценок числа жертв, равномерно выбираемых между нижними и верхними оценками каждого конфликта. Показатель хвоста слегка отличается (без статистической значимости) от полученного методом максимального правдоподобия для всех данных, если сосредоточиться на верхних 100 отклонениях



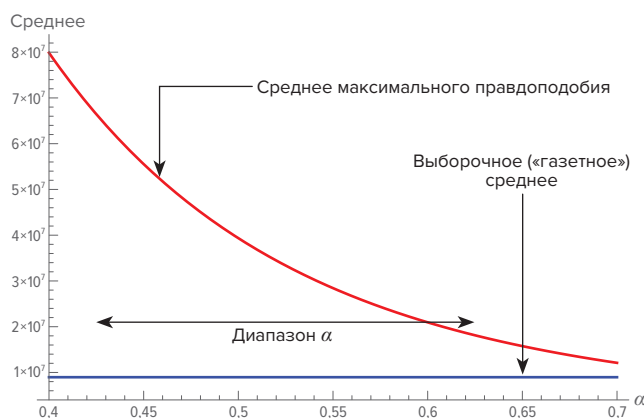
**Рисунок 16.2:** Квантиль-квантильный график зависимости масштабированных данных в ближнем хвосте от распределения в стиле Парето II — Lomax



**Рисунок 16.3:** Число погибших в наиболее известных исторических конфликтах в зависимости от времени. Если конфликт длился дольше 25 лет, он представлен последовательностью одинаковых конфликтов с соответствующей суммой жертв и с шагом по времени 25 лет



**Рисунок 16.4:** Масштабированное число погибших в вооруженных конфликтах и репрессиях в зависимости от времени. Данные масштабированы по отношению к населению мира. Если конфликт длился дольше 25 лет, он представлен последовательностью одинаковых конфликтов с соответствующей суммой жертв и с шагом по времени 25 лет



**Рисунок 16.5:** Наблюдаемое «газетное» среднее и среднее по ММП (полученное после обратного преобразования данных к распределению на компактном носителе) при различных значениях  $\alpha$  (и, соответственно, различных сочетаний  $\sigma_\alpha$ ,  $\alpha$ ). Возможный «диапазон  $\alpha$ » получен из возможных вариаций данных бутстрэпом и моделированием надежности

## 16.2. ОБЗОР СТАТИСТИЧЕСКОЙ ДИСКУССИИ

### 16.2.1. Результаты

**Хвосты Парето.** Метод превышений над порогом<sup>1</sup> показывает для этих данных (как исходных, так и масштабированных) сильные признаки распределения Парето, с вероятностью выживания  $\mathbb{P}(X > x) = \lambda(x)x^{-\alpha}$ , где  $\lambda: [L, +\infty) \rightarrow (0, +\infty)$  — медленно меняющаяся функция согласно критерию  $\lim_{x \rightarrow \infty} \frac{\lambda(kx)}{\lambda(x)} = 1$  при всех  $k > 0$ .

<sup>1</sup> *Peaks-over-threshold*, POT (дословно «пики над порогом») — метод в теории экстремальных значений. В изучаемом распределении выделяется хвост, который используется для нахождения стоимости под риском и ожидаемых потерь.

Мы находим параметры  $G(\cdot)$ , обобщенного распределения Парето (*Generalized Pareto Distribution*, GPD), приведенные в Таблице 16.1, используя представление  $G(x) = 1 - (1 + \xi x/\beta)^{-1/\xi}$ ; так что полученная для масштабированных данных оценка  $\xi \approx 1,88 \pm 0,14$  соответствует показателю хвоста  $\alpha = \frac{1}{\xi} = 0,53 \pm 0,04$ .

**Таблица 16.1:** Оценки и стандартные ошибки параметров обобщенного распределения Парето для числа жертв конфликтов свыше порога 50 тыс. И для фактического числа жертв, и для масштабированного мы приводим число событий выше порога

Данные	Число превышений	$\xi$	$\beta$
Исходные данные	307	1,5886 (0,1467)	3,6254 (0,8191)
Наивное масштабирование	524	1,8718 (0,1259)	14,3254 (2,1111)
Логарифмическое масштабирование	524	1,8717 (0,1277)	14,3261 (2,1422)

**Отсутствие памяти, влияющей на начало очередного конфликта.** Таблицы 16.2 и 16.3 показывают длительности интервалов между моментами событий, и видно, что, если событий масштаба Второй мировой войны не будет сто лет, это не повлияет на ожидания. Не заметно никакой автокорреляции, никакой статистически значимой временной структуры (никаких следов процесса с самовозбуждением<sup>1</sup>); см. Рисунок 16.8.

**Таблица 16.2:** Средний интервал между событиями и среднее абсолютное отклонение интервала между событиями крупнее 1, 2, 5 и 10 млн жертв (по фактической оценке)

Порог	Среднее	MAD
1	26,71	31,66
2	42,19	47,31
5	57,74	68,60
10	101,58	144,47

**Таблица 16.3:** Средний интервал между событиями и среднее абсолютное отклонение интервала между событиями крупнее 1, 2, 5, 10, 20 и 50 млн жертв (по масштабированной оценке)

Порог	Среднее	MAD
1	11,27	12,59
2	16,84	18,13
5	26,31	27,29
10	37,39	41,30
20	48,47	52,14
50	67,88	78,57

<sup>1</sup> Оптимистическая гипотеза была бы скорее о том, что возникновение острых конфликтов — процесс с самоторможением. Однако и самовозбуждение, и самоторможение отвечают одному и тому же предположению о модели: положительной корреляции между интенсивностью кровопролития сегодня и четверть века спустя.

**Полные распределения.** Масштабированные данные соответствуют распределению в стиле Ломакса с тем же хвостом, как полученный методом превышений над порогом, с сильной адекватностью. Для событий с числом жертв  $> L = 10K, 25K, 50K$  и т. д. мы подгоняем разные распределения Парето II, или распределения Ломакса, используя  $\alpha$  хвоста согласно GPD с коэффициентом масштаба  $\sigma = 84\,360$ , т. е. с плотностью  $\frac{\alpha}{\sigma} \left( \frac{-L + \sigma + x}{\sigma} \right)^{-\alpha-1}$ ,  $x \geq L$ .

Кроме того, мы рассматриваем более обширный массив статистических картин, отвечающих парам  $\alpha, \sigma_\alpha$  при разных данных на основе возможных альтернативных значений  $\alpha$ , с перекалибровкой  $\sigma$  максимального правдоподобия; см. Рисунок 16.5.

**Различие между выборочным средним и средним максимального правдоподобия.** В Таблице 16.4 показаны значения среднего по параметрам распределения Парето, приведенным выше, после обратного преобразования к конечному носителю. «Истинное» среднее, оно же в данном моделировании «статистическое» среднее максимального правдоподобия, в 3–4 раза выше наивно наблюдаемого «газетного» среднего. Это значит, что «газетное» наблюдаемое среднее, с его порочным принципом полагаться на выборочное среднее, недооценит истинное среднее втрое, если не хуже, а потом, когда дальнейшие наблюдения приблизят оценку к реальности, «газетный» вывод о возросшем уровне острых конфликтов будет статистически недопустимым.

**Таблица 16.4:** Выборочное среднее и среднее по ММП при разных минимальных значениях  $L$ . Данные масштабированные

$L$	Выборочное среднее	Среднее по ММП	Отношение
10 тыс	$9,079 \times 10^6$	$3,11 \times 10^7$	3,43
25 тыс	$9,82 \times 10^6$	$3,62 \times 10^7$	3,69
50 тыс	$1,12 \times 10^7$	$4,11 \times 10^7$	3,67
100 тыс	$1,34 \times 10^7$	$4,74 \times 10^7$	3,53
200 тыс	$1,66 \times 10^7$	$6,31 \times 10^7$	3,79
500 тыс	$2,48 \times 10^7$	$8,26 \times 10^7$	3,31

### 16.2.2. Заключение

С позиций аналитики хвостов прошлое выглядит намного опаснее, с намного большим числом острых конфликтов, чем с позиций наивного наблюдения среднего по историческим временным последовательностям.

## 16.3. ОБСУЖДЕНИЕ МЕТОДОВ

### 16.3.1. Метод масштабирования

Нам нужно избавиться от ограниченного носителя, чтобы применить степенные законы (см. предыдущие главы), и мы делаем это так. Случайную величину  $X_t$ , число жертв в конфликте в период времени  $t$ , мы сначала масштабируем по формуле  $X'_t = \frac{X_t}{H_t}$ , где  $H_t$  — численность человечества в период  $t$ . О методах оценки  $H_t$  смотрите в приложении.

Затем, обозначив  $H$  нынешнее максимальное население и  $L$  наивно масштабированный минимум числа жертв конфликта по принятому нами определению, мы вводим функцию гладкого преобразования  $\varphi: [L, H] \rightarrow [L, \infty)$  такую, что:

- i)  $\varphi$  «гладкая»:  $\varphi \in C^\infty$ ,
- ii)  $\varphi^{-1}(\infty) = H$ ,
- iii)  $\varphi^{-1}(L) = \varphi(L) = L$ .

А именно такую:

$$\varphi(x) = L - H \ln \frac{H-x}{H-L}. \quad (16.1)$$

Пользуясь неограниченностью  $X_r = \varphi(X)$ , мы проанализируем эту величину и подберем показатель степенного закона. Затем выполним обратное преобразование и получим свойства  $X$ . Заметим при этом, что при больших  $H$  наше преобразование  $\varphi(x) \approx x$ . Это значит, что при далекой верхней границе параметры хвоста, которые мы подберем для  $x$  или для  $\varphi(x)$ , будут по существу одинаковы. Фундаментальное различие будет только философское и методическое: мы удаляем верхнюю границу (которая все равно вряд ли будет достигнута).

Далее мы используем наивно масштабированное число жертв как аргумент  $\varphi(\cdot)$ .

В качестве упражнения выберем  $H = P_{i_0}$ .

Распределение  $X$  можно вывести из распределения  $X_r$  так:

$$\int_L^\infty f(x_r) dx_r = \int_L^{\varphi^{-1}(\infty)} g(x) dx, \quad (16.2)$$

где  $\varphi^{-1}(u) = (L-H)e^{\frac{L-u}{H}} + H$ .

В данном случае, выбрав распределение Парето — Ломакса:

$$f(x_r) = \frac{\alpha}{\sigma} \left( \frac{-L + \sigma + x_r}{\sigma} \right)^{-\alpha-1}, \quad x_r \in [L, \infty), \quad (16.3)$$

получаем

$$g(x) = \frac{\alpha H}{\sigma(H-x)} \left( \frac{\sigma - H \ln \frac{H-x}{H-L}}{\sigma} \right)^{-\alpha-1}, \quad x_r \in [L, H),$$

откуда убеждаемся, что  $\int_L^H x g(x) dx = 1$ . Следовательно, математическое ожидание

$$\mathbb{E}g(X; L, H, \sigma, \alpha) = \int_L^H x g(x) dx, \quad (16.4)$$

$$\mathbb{E}g(X; L, H, \sigma, \alpha) = \alpha H \left( \frac{1}{\alpha} - \frac{(H-L)e^{\sigma/H} E_{\alpha+1} \frac{\alpha}{H}}{\alpha} \right), \quad (16.5)$$

где  $E_{(\cdot)}(\cdot)$  — интегральная показательная функция  $E_n z = \int_1^\infty \frac{e^{t(-z)}}{t^n} dt$ .

Заметим, что мы опирались на свойство инвариантности:

**Замечание 16**

Если  $\hat{\vartheta}$  — оценка максимального правдоподобия (*maximum likelihood estimator*, MLE) для величины  $\vartheta$ , то абсолютно непрерывное преобразование  $\varphi$  отображит ее на оценку максимального правдоподобия преобразованной величины,  $\varphi(\hat{\vartheta}) = \hat{\varphi}(\vartheta)$ .

Дополнительные подробности смотрите в [211].

**16.3.2. Условное математическое ожидание (в нестрогом изложении)**

Мы будем заменять гладкую функцию в  $C^\infty$  на ступеньку Хевисайда, то есть на индикаторную функцию  $\mathbb{1}: \mathbb{R} \rightarrow \{0, 1\}$ ; преобразованная величина запишется как  $\mathbb{1}_{X \in [L, H]}$ :

$$\mathbb{E} \mathbb{1}_{X \in [L, H]} = \frac{\int_L^H x f(x) dx}{\int_L^H f(x) dx},$$

что в случае распределения Парето — Ломакса даст

$$\mathbb{E} \mathbb{1}_{X \in [L, H]} = \frac{\frac{\alpha \sigma^\alpha (H-L)}{\sigma^\alpha - (H-L+\sigma)^\alpha} + (\alpha-1)L + \sigma}{\alpha-1}. \quad (16.6)$$

**16.3.3. Надежность данных и влияние на хвостовые оценки**

Данные об острых конфликтах во многом отрывочны, часто представляют собой некритичный пересказ, основываются на туманных оценках свидетелей, не дают возможности проверить факты по другим источникам того же периода. Такое событие седьмого века, как восстание Лушаня, привело, как считается, к гибели 26 миллионов человек, но нет надежных и точных методов, чтобы проверить эту оценку. О войне за независимость Алжира есть данные как со стороны Франции, так и со стороны повстанцев, но нет профессионального научного источника.

Как уже говорилось выше в этой главе, мы используем данные по-разному: в исходном виде, с наивным масштабированием по текущему населению мира и с логарифмическим преобразованием, обеспечивающим удобную теорию без верхней границы.

Для некоторых наблюдений на основе исторических источников сообщается не просто оценка числа жертв, а нижняя и верхняя граница. Пусть  $X_t$  — число жертв некоторого конфликта, произошедшего в момент  $t$ . В принципе, мы могли бы определить тройки вида

- $\{X_t, X_t^l, X_t^u\}$  для фактических оценок (исходных данных), где  $X_t^l$  и  $X_t^u$  представляют нижнюю и верхнюю границу, если эти данные доступны;
- $\left\{Y_t = X_t \frac{P_{2015}}{P_t}, Y_t^l = X_t^l \frac{P_{2015}}{P_t}, Y_t^u = X_t^u \frac{P_{2015}}{P_t}\right\}$  для наивно масштабированных данных, где  $P_{2015}$  — население мира в 2015 году и  $P_t$  — население мира в моменты времени  $t = 1, \dots, 2014$ ;
- $\{Z_t = \varphi(Y_t), Z_t^l = \varphi(Y_t^l), Z_t^u = \varphi(Y_t^u)\}$  для логарифмически преобразованных данных.



Для очистки совести мы не стали просто использовать среднее между нижней и верхней границей, а выполнили процедуру Монте-Карло (подробности смотрите в [201]) и не обнаружили значительных различий по интересующим нас величинам (таким, как показатель хвоста  $\alpha = 1/\xi$ ):

1) для каждого события  $X$ , где есть границы, мы предположили для простоты равномерное распределение числа жертв от нижней границы до верхней, т. е.  $X \sim U(X^l, X^u)$ . На самом деле любое ограниченное распределение дало бы в пределе те же результаты благодаря центральной предельной теореме;

2) далее мы сгенерировали большое число экземпляров Монте-Карло и в каждом экземпляре назначили событию  $X$  случайное значение, следуя распределению  $U(X^l, X^u)$ ;

3) для каждого экземпляра мы вычислили интересующие нас статистические показатели, обычно показатель степени хвоста, и полученные так значения далее усреднили.

Проведенная процедура показала, что точность оценок не влияет на хвост распределения числа жертв; напротив, показатель хвоста весьма устойчив.

Для событий, у которых границы не заданы, был вариант оставлять их как есть или вносить возмущения, задав фиктивные границы вокруг заданных значений (и далее обработать в экземплярах Монте-Карло наравне с имевшими заданные границы значениями). Мы выбрали второй подход.

То же относится к обработке  $Y_t$  и  $Z_t$ .

Заметим, что  $\alpha$  хвоста, полученная из среднего, отличается от среднего по разным оценкам альфы, и это заставило нас выполнить отдельные анализы по разным оценкам.

**Технический комментарий.** В таком моделировании часто ищут смещение «стохастической альфы» из-за ошибок и ненадежности данных (Глава 18). При размере выборки  $n$  параметр  $\hat{\theta}_m$  — средний параметр, полученный при большом числе моделирований Монте-Карло. Пусть  $X_i$  — данный смоделированный вектор Монте-Карло с индексом  $i$ , и  $X_\mu$  — средняя оценка между нижней и верхней границами. Поскольку

по всем моделированиям Монте-Карло  $\frac{1}{m} \sum_{j \leq m} \|X_j\|_1 = \|X_\mu\|_1$ , кроме  $\forall j$ , постольку  $\|X_j\|_1 \neq \|X_\mu\|_1$ ,  $\hat{\theta}_m = \frac{1}{m} \sum_{j \leq m} \hat{\theta}(X_j) \neq \hat{\theta}(X_\mu)$ . Рассмотрим, например, оценку максимального правдоподобия для альфы хвоста Парето,  $\hat{\alpha}(X_i) \triangleq n \left( \sum_{1 \leq i \leq m} \ln \frac{x_i}{L} \right)^{-1}$ . При  $\Delta \geq x_m$  определим

$$\hat{\alpha}(X_i \sqcup \Delta) \triangleq \frac{1}{2} \left( \frac{n}{\sum_{i=1}^n \ln \frac{x_i}{L} - \ln \frac{\Delta}{L}} + \frac{n}{\sum_{i=1}^n \ln \frac{x_i}{L} + \ln \frac{\Delta}{L}} \right),$$

которая, благодаря выпуклости вверх логарифмической функции, дает неравенство

$$\forall \Delta \geq x_m, \hat{\alpha}(X_i \sqcup \Delta) \geq \hat{\alpha}(X_i).$$

#### 16.3.4. Определение «события»

Даже «знатные» конфликты часто определяются произвольно, что затрудняет учет статистики: то один и тот же конфликт носит несколько имен, то два и более конфликтов проходят под одним именем. И мы не нашли удовлетворительной классификации, отличающей войну от конфликта.

Ради единообразия мы считали одним событием то, что длилось менее 25 лет, и делили более длинный конфликт на последовательность событий, которым приписывали по 25 лет или меньше. Соответственно, монгольские завоевания, которые продолжались больше века с четвертью, мы считали за несколько событий. При другом подходе получается нелепость; например, придется считать эпоху от Франко-прусской войны до Второй мировой «германскими завоеваниями», а не последовательностью событий, которые современники называли разными именами. Собственно, фундаментальные источники, такие как *Encyclopedia of War* [189], вместо «монгольских завоеваний» пишут о множестве конфликтов; чем глубже погружается историк в свою область, тем более он склонен разбивать конфликты на отдельные «знатные» события. У разных историков монгольские завоевания насчитывают от 12 до 55 конфликтов.

Опять-таки возможные споры о том, что считать «известным событием», разрешаются бутстрэппингом. Кстати, наши выводы инвариантны к укрупнению или дроблению монгольских завоеваний.

Еще одна трудность — нет четкого протокола, как отделить гибель людей непосредственно на войне и сокращение численности населения из-за войны (например, от голода при блокаде). Известны споры историков вокруг Первой Иудейской войны: число смертей оценивается в 30 тысяч на войне и много больше (от 350 тыс. до 1 млн, согласно Иосифу Флавия) от голода и от гибели гражданских лиц.

#### 16.3.5. Пропущенные события

Мы предполагаем, что многие войны не попали в нашу выборку; правда, мы сомневаемся, что пропали события из «хвостов» распределения, поскольку у крупных конфликтов больше шансов попасть в историю. И мы предполагаем, что пропущенные события распределены между данными случайно, без кластеризации.

Но мы учитываем систематическую ошибку из-за различий в точности и дотошности истории в разные времена: события ближе к современности попадают в анналы надежнее, чем в далеком прошлом. Если подняли минимальное значение  $L$ , рассчитывая, что число пропущенных событий и их влияние должны резко уменьшиться. И действительно, когда при проверке робастности мы поднимали планку еще выше, до минимума  $L = 500$  тыс., то результаты анализа не менялись.

Простая оценка методом складного ножа, когда анализ повторяют, удалив из выборки некоторую часть событий, показывает нам зависимость анализа от пропущенных событий, и мы находим эту зависимость несущественной, если фокусироваться на хвосте распределения числа жертв. Другими словами, если исследуются экстремальные значения, и удаление 30 % событий не приводит к расхождению параметров с исходными результатами, можно не беспокоиться о том, что 30 % событий, возможно, было пропущено в исходных данных — это вряд ли сделало хвосты тоньше<sup>1</sup>.

<sup>1</sup> Обратное неверно. В этом суть несимметричности черного лебедя: такая процедура не компенсирует отсутствие в данных хвостового события, «черного лебедя». Одно-единственное событие «черного

### 16.3.6. Систематическая ошибка выжившего

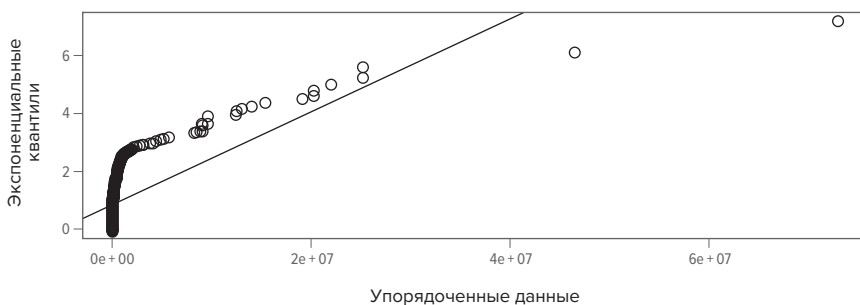
Мы не учитывали в данном анализе систематическую ошибку выжившего, полагая ее несущественной, до 1960-го, пока вероятность конфликта, уничтожающего все человечество, была пренебрежимо мала. После этого данный риск повысился, прежде всего в связи с ядерным и другим оружием массового поражения.

## 16.4. АНАЛИЗ ДАННЫХ

Наши данные: число жертв в зависимости от времени, — графически представлены на Рисунках 16.3 и 16.4. На Рисунке 16.3 показаны оценки фактического числа жертв, а на Рисунке 16.4 это число масштабировано, чтобы привести к населению мира в 2015-м (около 7,2 миллиарда человек)<sup>1</sup>. Глядя на Рисунок 16.3, можно предположить, что с течением истории число смертей в вооруженных конфликтах росло; если так, то это довод в пользу предположения о росте военного насилия. Глядя на Рисунок 16.4, можно предположить прямо противоположное, что (масштабированное) число жертв со временем уменьшалось, особенно в последние сто лет; это довод в пользу предположения о спаде военного насилия. Мы далее покажем, что обе интерпретации безнадёжно наивны, потому что не учитывают тот факт, что мы имеем дело с экстремальными событиями.

### 16.4.1. Превышения над порогом

Поскольку данные жирнохвостые, что заметно уже в таких простейших проверках, как гистограммы логарифмов и графики квантиль-квантиль (на Рисунке 16.6 показан график квантиль-квантиль фактического числа жертв и экспоненциального распределения: очевидная выпуклость вверх сигнализирует о жирнохвостом распределении), — постольку представляется уместным для моделирования зависимости числа жертв войн от времени прибегнуть к апробированному методу теории экстремальных значений: метод превышений над порогом (*Peaks-over-Threshold*, POT) [184].



**Рисунок 16.6:** Квантиль-квантильный график фактического числа жертв и стандартного экспоненциального квантиля. Точки данных расположены с выпуклостью вверх — четкий сигнал жирных хвостов

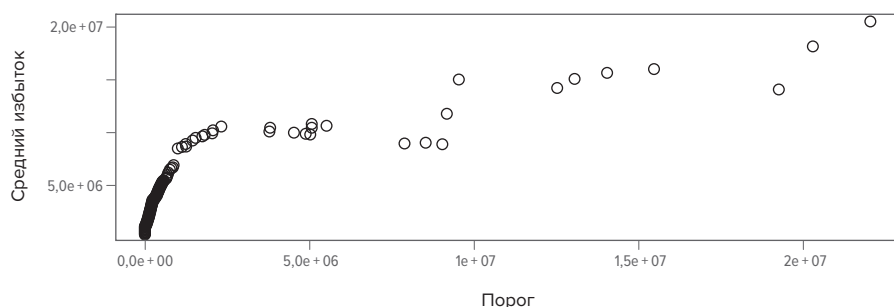
лебеда» способно существенно умирить хвост. В данном же случае хвост так и так жирный, и если часть информации была пропущена, это вряд ли сделало хвост тоньше. — Прим. автора.

<sup>1</sup> Заметим, что в формуле (16.1) при  $N = 7,2$  миллиарда  $\phi(x) \approx x$ . Поэтому Рисунок 16.4 также хорошо представляет данные после логарифмического преобразования. — Прим. автора.

По методу POT для последовательности н. о. р. случайных величин случаи превышений над некоторым высоким порогом  $u$  (который придется идентифицировать) происходят в моменты времени согласно однородному процессу Пуассона, тогда как величину превышений можно моделировать обобщенным распределением Парето (*Generalized Pareto Distribution*, GPD). Моменты наступления и величины разных событий предполагаются независимыми друг от друга.

В нашем случае предположение о независимости войн друг от друга не кажется слишком смелым, если учесть их разнесенность по эпохам и континентам. Но остальные предположения, напротив, придется проверять.

Начнем с того, что узнаем порог  $u$ , выше которого есть шанс успешной аппроксимации распределением GPD. Известен ряд эвристических инструментов для такой работы, от графика Ципфа до графиков функции среднего избытка, когда ищется линейность, характерная для жирнохвостых явлений [44, 82]. На Рисунке 16.7 показан график функции среднего избытка для фактического числа жертв:<sup>1</sup> очевиден тренд на повышение, начинающийся с порога 5 тысяч жертв. Ради лучшей подгонки разумно взять порог с небольшим запасом,  $u = 50$  тысяч<sup>2</sup>.



**Рисунок 16.7:** График функции среднего избытка (*mean excess function plot*, MEPLLOT) по фактическому числу жертв. Виден тренд на повышение, почти линейный в первой части графика; это признак жирного правого хвоста. При высоком пороге изменчивость повышается, но это объясняется тем, что в этой области мало наблюдений выше порога; подробно рассматривать такое поведение бесполезно

#### 16.4.2. Интервалы во временных рядах и автокорреляция

Чтобы проверить, что по времени события следуют равномерному процессу Пуассона — это основное допущение для метода POT, — мы можем изучить время между последовательными свершениями, или интервал, и убедиться, что он следует экспоненциальному распределению. Кроме того, нужно убедиться, что нет автокорреляции между интервалами.

Что автокорреляции нет, ясно видно на Рисунке 16.8. В совместимости данных с экспоненциальным распределением интервала между событиями убеждаемся при

1 Аналогичные результаты получаются и для масштабированных величин (наивно и логарифмически). Для краткости мы всегда, когда нет серьезных различий, показываем на графиках только одну из двух величин. — Прим. автора.

2 Далее удачность этого выбора подтверждается проверками адекватности модели по критериям согласования. — Прим. автора.

помощи ряда эвристических и аналитических инструментов; для краткости мы здесь опустим эти положительные результаты проверок.

Однако обратим внимание, поскольку это полезно и в других отношениях, на базовые статистические показатели, которые мы приводим в Таблицах 16.2 и 16.3 и которые касаются интервалов между катастрофическими, крупнейшими по числу жертв событиями<sup>1</sup>. Уже из этих простых сведений ясно, сколь ненадежны утверждения об историческом снижении уровня острых конфликтов. Для событий крупнее 10 миллионов жертв по фактическим оценкам видим средний интервал 101,58 года при среднем абсолютном отклонении 144,47<sup>2</sup>. Это значит, что нет ничего необычного в том, что ни одного события такой величины не свершилось за последние годы; согласно такому распределению данное событие с равной уверенностью ожидается и завтра, и через сто лет. Это также значит, что нечего и пытаться экстраполировать тренд для экстремальных событий данного типа. Напоследок заметим, что событие масштаба Второй мировой войны придется ждать еще дольше: оно происходит раз в 2014 лет, если считать по фактическим величинам (по масштабированному числу жертв придется учесть восстание Ань Лушаня).

### 16.4.3. Анализ хвоста

Убедившись, что данные, очевидно, удовлетворяют предпосылкам РОТ в отношении пуассоновского процесса, можем, наконец, подогнать обобщенное распределение Парето к наблюдаемым избыткам.

Рассмотрим случайную величину  $X$  с функцией распределения  $F$  и обозначим  $F_u$  функцию условного распределения  $X$  выше заданного порога  $u$ . Далее определим (согласно [184]) случайную величину  $Y$ , масштабированное превышение порога  $u$  величиной  $X$ , как следующую к распределению с функцией

$$F_u(y) = \mathbb{P}(X - u \leq y \mid X > u) = \frac{F(u+y) - F(u)}{1 - F(u)}$$

при  $0 \leq y \leq x_F - u$ , где  $x_F$  — это правая конечная точка исходного распределения  $F$ . Пикэндс [190], Болкема и де Ганн [8], [9] и [10] показали, что для широкого класса исходных функций распределения  $F$  (в так называемом аттракторе распределения GEV<sup>3</sup> [184]) при большом  $u$  можно аппроксимировать  $F_u$  обобщенным распределением Парето:  $F_u(y) \rightarrow G(y)$  при  $u \rightarrow \infty$ , где

$$G(y) = \begin{cases} 1 - (1 + \xi y / \beta)^{-1/\xi}, & \text{если } \xi \neq 0, \\ 1 - e^{-y/\beta}, & \text{если } \xi = 0 \end{cases}. \quad (16.7)$$

1 В Таблице 16.2 не приведены средние интервалы между событиями с числом жертв 20 млн (50 млн и более). Потому что число таких наблюдений в фактических, немасштабированных данных ограничено. В частности, 20 млн жертв превысили только два события, и оба свершились в последние 150 лет, с интервалом короче 20 лет. Действительно ли мы живем в мирную эпоху? — Прим. автора.

2 По масштабированным величинам интервалы короче, но интерпретация та же. — Прим. автора.

3 Обобщенное распределение экстремальных значений (*generalized extreme value distribution*, GEV) охватывает семейства распределений Гумбеля, Фреше и Вейбула.

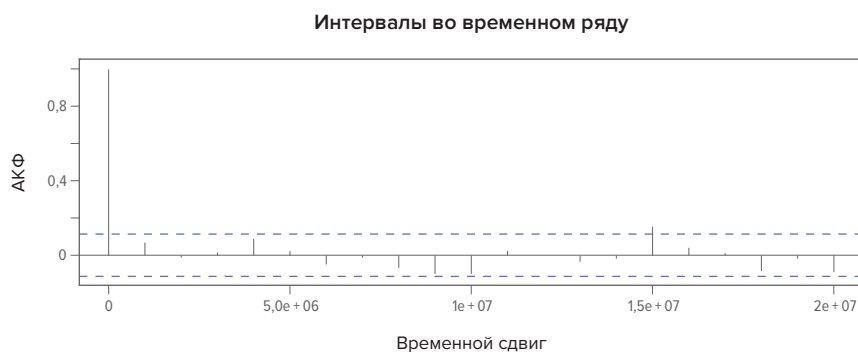
Можно доказать, что распределение GPD охватывает промежуточные случаи между экспоненциальным распределением (при  $\zeta=0$ ) и классом распределений Парето. Подробности есть в [184].

Параметры в формуле (16.7) можно оценить методом максимального правдоподобия или методом взвешенных моментов вероятности [184]. Оценить адекватность модели затем можно бутстрэп-тестами [262].

В Таблице 16.1 приведены наши оценки по ММП для фактического и масштабированного числа жертв выше порога 50 тыс. Этот порог оказался лучшим компромиссом между стремлениями к высокой адекватности модели и к использованию большого числа наблюдений ради надежности стандартных ошибок. Фактические и масштабированные данные дали два разных набора оценок, но интерпретация тех и других согласуется между собой. По этой причине мы сосредоточимся здесь на обсуждении фактического числа жертв.

Параметр  $\zeta$  для нас самый важный: он управляет жирностью правого хвоста. Если  $\zeta$  больше 1 (а мы получили 1,5886), то для нашего обобщенного распределения Парето не определены никакие моменты — ситуация крайней жирнохвостости. Естественно, что для выборки можно рассчитать все моменты, но они будут принципиально ненадежными и их теоретическая интерпретация будет некорректной (хотя данная некорректность встречается сплошь и рядом). Согласно нашей модели, отнюдь не исключены самые катастрофические события. Стоит отметить, что результаты моделирования статистически значимы, со стандартной ошибкой 0,1467.

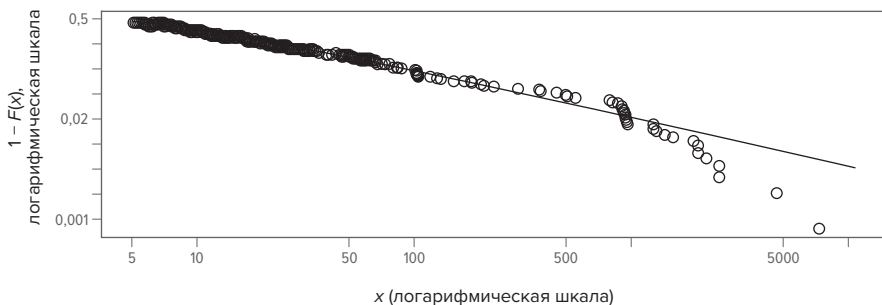
На Рисунках 16.9 и 16.10 наша подогнанная модель сравнивается с фактическими данными. На обоих можно видеть адекватность GPD для большинства наблюдений выше порога 50 тыс. жертв. Есть расхождение для самых крупных событий, таких как Вторая мировая война и восстание Ань Лушаня:<sup>1</sup> по нашей модели ожидалось более крупные события. Это известная проблема с изучением экстремальных данных [184]: бывает, что особо крупные значения поджидают вас в ближайшем будущем.



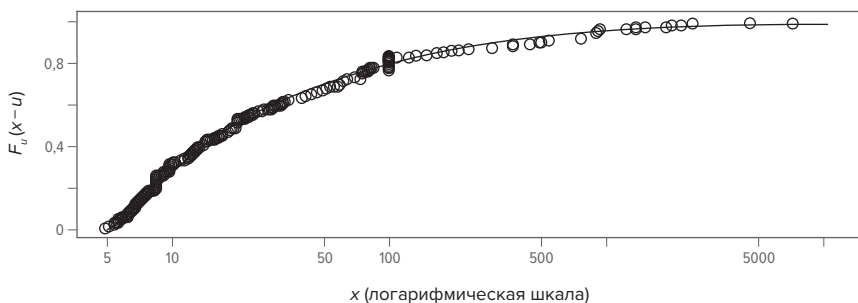
**Рисунок 16.8:** График автокорреляционной функции (АКФ) промежутков между событиями с одинаковым фактическим числом жертв; никакой статистически значимой автокорреляции не просматривается

<sup>1</sup> Если удалить из данных два самых крупных события, гипотезу GPD станет невозможно отвергнуть при уровне значимости 5 %. — Прим. автора.

Другое расхождение — для событий от 5 до 10 млн жертв (опять-таки немалых!), которые случались немного чаще, чем ожидается по нашей модели GPD. Это еще раз напоминает нам о характерном «экстремальном» поведении числа жертв войны, не позволяющем экстраполиацию упрощенных трендов.



**Рисунок 16.9:** Подгонка хвоста GPD под данные фактического числа жертв (в десятках тысяч). Параметры согласно Таблице 16.1, первая строка



**Рисунок 16.10:** Подгонка кумулятивного распределения GPD под данные фактического числа жертв (в десятках тысяч). Параметры согласно Таблице 16.1, первая строка

#### 16.4.4. Альтернативный взгляд на максимумы

Другой метод в теории экстремальных значений — блочные максимумы: данные разбивают на блоки, в каждом выбирают максимальное значение и только эти выбранные максимумы рассматривают далее. Теорема Фишера — Типпета [184] гарантирует, что нормализованные максимумы сходятся по распределению к обобщенному распределению крайних значений, GEV.

$$\text{GEV}(x, \xi) = \begin{cases} \exp\left(-\left(1 + \xi x\right)^{\frac{1}{\xi}}\right), & \xi \neq 0, 1 + \xi x > 0. \\ \exp(-\exp(-x)), & \xi = 0 \end{cases}$$

У этого распределения есть естественная связь с GPD, и мы отсылаем к [184] за подробностями.

Если мы разделим наши данные на 100-летние блоки, получим 21 наблюдение (последний блок будет остатком истории с 2001 по 2014 год). Метод максимального правдоподобия дает  $\xi$  больше 2, относя распределение к максимальному аттрактору Фреше, что совместимо с оценкой хвоста явления как тяжелого. Значение  $\xi$  больше 2 в распределении GEV дополнительно подтверждает предположение о том, что моментов нет и правый хвост очень тяжелый.

#### 16.4.5. Анализ полных данных

Помня об ограничениях, мы можем попробовать подогнать модель по всем данным, соблюдая при этом для числа жертв больше 10 000 распределение Парето согласно формуле (16.3) с  $\alpha \approx 0,53$ . Адекватность модели для «близкого хвоста» ( $L = 10$  тыс.) можно видеть на Рисунке 16.2. Аналогичные Рисунку 16.2 результаты показаны для различных значений в приведенной ниже таблице, и у всех одинаковая оценка согласия.

$L$	$\sigma$
10 000	84 260
25 000	899 953
50 000	116 794
100 000	172 733
200 000	232 358
500 000	598 292

Различные возможные значения среднего в формуле (16.4) можно вычислить для разных наборов  $\alpha$ , при одной степени свободы: соответствующая  $\sigma$  — это оценка по ММП с использованием  $\alpha$  как фиксированного параметра. Если размер выборки  $n$

$$\text{и } x_i \text{ — наблюдения выше } L, \sigma_{\alpha} = \left\{ \sigma \left| \frac{\alpha n}{\sigma} - (\alpha + 1) \sum_{i=1}^n \frac{1}{x_i - L + \sigma} = 0, \sigma > 0 \right. \right\}.$$

Выборочное среднее при  $L = 10$  тыс. составляет  $9,12 \times 10^6$  по 100 тыс. моделированных, и разброс значений показан на Рисунке 16.15.

«Истинное» среднее из формулы (16.4) составляет  $3,1 \times 10^7$ ; мы повторили вычисления при  $L = 10$  тыс., 20 тыс., 50 тыс., 100 тыс., 200 тыс. и 500 тыс. и нашли отношение истинной оценки среднего к наблюдаемой строго между 3 и 4, смотрите Таблицу 16.4. Заметим, что оценка среднего как  $\approx$  в 3,5 раза больше наблюдаемого среднего по выборке — это только общая прикидка; величина эта стохастическая и не выявляет точной информации; она лишь показывает нам, сколь нелепо ориентироваться на наивную оценку среднего.

Потому что при жирных хвостах среднее, выведенное из оценок  $\alpha$ , строже и имеет меньшую погрешность. В самом деле, оценка  $\alpha$  асимптотически гауссова, тогда как среднее распределения по степенному закону, если вообще существует, гораздо стохастичнее. По этому вопросу смотрите обсуждение «медленного закона больших чисел» в Разделе 8.



Мы находим, что среднее с усечением по  $L = 10$  тыс. заметно ниже, если вычислять по формуле (16.6); около  $1,8835 \times 10^7$ .

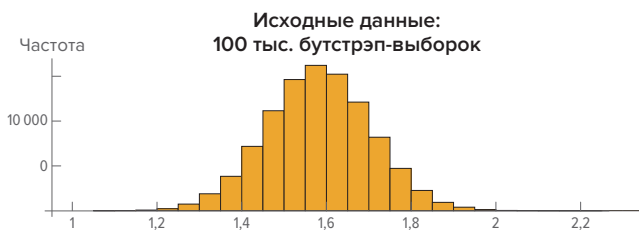
В заключение заметим, что среди конфликтов с числом жертв 10 тыс. и выше доля конфликтов с числом жертв ниже среднего составляет 96 %. Обозначая среднее  $m$ ,

$$\mathbb{P}(X < m) = 1 - \left( 1 - \frac{H}{\sigma} \ln \left( \alpha e^{\sigma/H} E_{\alpha+1} \frac{\sigma}{H} \right) \right)^{-\alpha}.$$

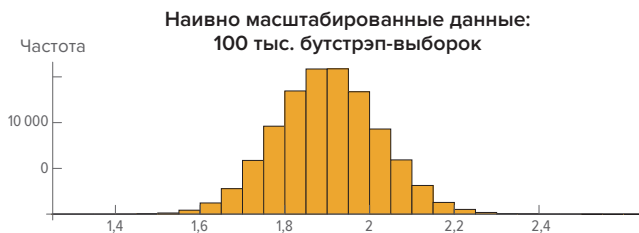
## 16.5. ДОПОЛНИТЕЛЬНЫЕ ТЕСТЫ РОБАСТНОСТИ И НАДЕЖНОСТИ

### 16.5.1. Бутстрэп для GPD

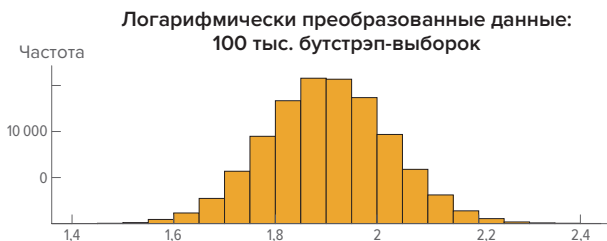
Чтобы проверить чувствительность к качеству и точности данных, мы решили выполнить бутстрэп-анализ. И для исходных данных, и для масштабированных мы сгенерировали 100 тыс. новых выборок путем случайного выбора 90 % наблюдений, применяя замену. На Рисунках 16.11, 16.12 и 16.13 показана устойчивость наших оценок  $\xi$ . В частности,  $\xi > 0$  во всех выборках, что показывает крайнюю жирнохвостость числа жертв вооруженных конфликтов. Оценки  $\xi$  в Таблице 16.1, по-видимому, хорошая аппроксимация для наших действительных параметров формы GPD, невзирая на неточности и пропуски наблюдений.



**Рисунок 16.11:** Распределение параметра  $\xi$  по 100 тысячам бутстрэп-выборок фактических данных. Каждая выборка сгенерирована случайным выбором с заменой при использовании 90 % первоначальных наблюдений



**Рисунок 16.12:** Распределение параметра  $\xi$  по 100 тысячам бутстрэп-выборок наивно масштабированных данных. Каждая выборка сгенерирована случайным выбором с заменой при использовании 90 % первоначальных наблюдений

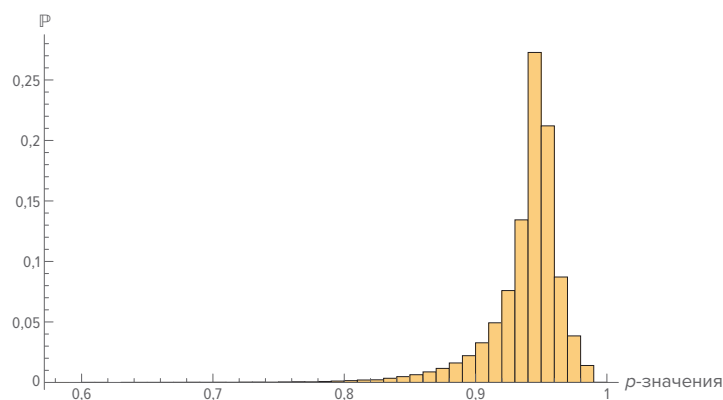


**Рисунок 16.13:** Распределение параметра  $\xi$  по 100 тысячам бутстрэп-выборок логарифмически преобразованных данных. Каждая выборка сгенерирована случайным выбором с заменой при использовании 90 % первоначальных наблюдений

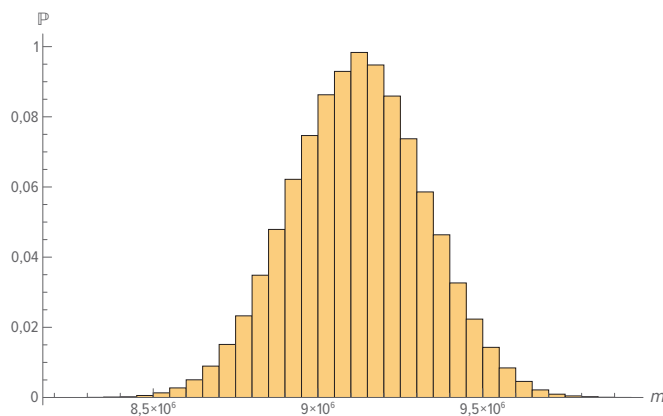
### 16.5.2. Внесение возмущений в границы оценок

Мы выполнили анализ для «близкого хвоста» методами Монте-Карло, которые обсудили в Разделе 16.3.3. Посмотрим на  $p$ -значения второго порядка, то есть на чувствительность  $p$ -значений при различных оценках на Рисунке 16.14: практически все результаты удовлетворяют одним и тем же уровням статистической значимости и адекватности модели.

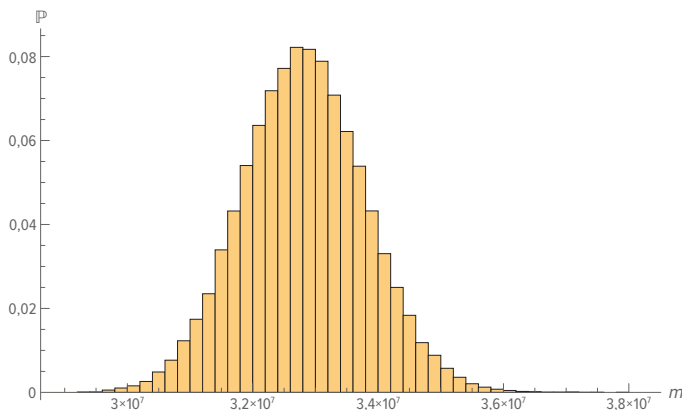
Кроме того, посмотрим на значения выборочных средних и средних по ММП на основе альфы при различных перестановках (Рисунки 16.15 и 16.16).



**Рисунок 16.14:**  $p$ -значения модели Парето — Ломакса при 100 тыс. сочетаний. Цель была не узнать  $p$ -значение, а проверить робастность, посмотрев на изменчивость при перестановке оценок



**Рисунок 16.15:** Масштабированное выборочное среднее при 100 тыс. оценок между нижней и верхней границами



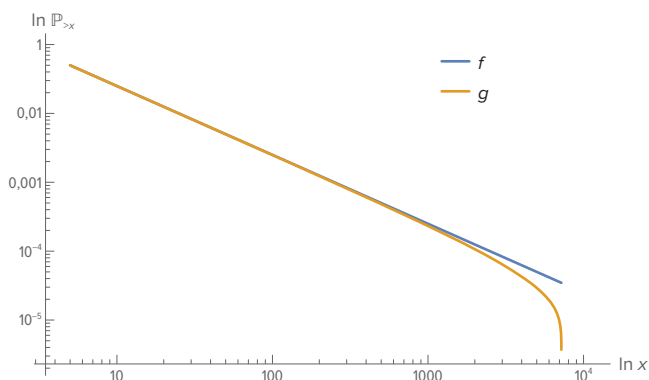
**Рисунок 16.16:** Масштабированное среднее ММП при 100 тыс. оценок между нижней и верхней границами

## 16.6. ЗАКЛЮЧЕНИЕ: МИР ОПАСНЕЕ, ЧЕМ КАЖЕТСЯ?

Говоря самыми простыми словами, наши выводы таковы, что, даже если дальнейшие события в 3 раза поднимут средний уровень насилия, нам не придется переписать эту главу или изменить калибровку параметров.

- В самом деле, если опираться на анализ статистики, то мир опаснее, чем по легкомысленным расчетам. Уровень острых конфликтов недооценивается при статистически неграмотном, в газетном стиле разговоре о среднем без понимания стохастичности интервалов между событиями.
- Преобразование к компактному носителю позволило нам выполнить анализ и калибровку этой недооценки, получив, пусть и с заметным шумом, представление о величине недооценки и границах.
- Другими словами, крупное событие и даже рост наблюдаемого среднего уровня острых конфликтов не нарушит согласия со статистическими свойствами модели, и можно будет сказать, что в наших взглядах «ничего не изменилось».
- Мы не можем распространить наши выводы на статистику убийств, поскольку ограничились  $L > 10\,000$ , но вряд ли их уровень существенно влияет на хвосты; скорее там его влияние как капля в море. Динамика числа убийств отличается от военного насилия; случалось, что в одних и тех же обществах наблюдался низкий уровень убийств и высочайший риск погибнуть в результате острого конфликта. (Число жертв убийства при масштабировании из 70 в год на 100 тыс. жителей на современное население мира дает нам  $5,04 \times 10^6$ . Даже падение до минимальных уровней не выйдет за пределы погрешности среднего по острым конфликтам крупнее 10 000 жертв.)

- Мы проигнорировали при анализе данных ошибку выжившего (то есть тот факт, что, *будь мир более опасным местом, нас бы здесь не было и мы бы его не обсуждали*). Если учесть ошибку выжившего, оценка риска повысится. Появление соответствующего хвоста требует начать такой учет в анализе последующего периода. С 1960 года один-единственный конфликт — который чуть было не случился — способен достичь максимального числа жертв; раньше такого не было. (Мы можем переписать модель, раздробив мир на множество изолированных друг от друга миров и введя  $n$  независимых случайных величин  $X_i$ , каждая с максимальным значением  $H_i$ , так что  $\sum_{i=1}^n w_i H_i = H$ , при всех  $w_i > 0$ ,  $\sum_{i=1}^n w_i = 1$ . В таком случае максимум (то есть наихудший конфликт) потребовал бы совместной вероятности того, что все  $X_1, X_2, \dots, X_n$  окажутся вблизи своих максимальных значений, что при субэкспоненциальном распределении будет много менее вероятным, чем достижение максимального значения одной случайной величиной.)<sup>1</sup>



**Рисунок 16.17:** Сравнение на логарифмическом графике функций  $f$  и  $g$ . Видно гладкое закругление вблизи  $H$ , аналогичное границе американского опциона

## 16.7. БЛАГОДАРНОСТИ

Данные компилировал капитан Марк Вейзенборн. Благодарим Бена Кирнана за комментарии о восточноазиатских конфликтах.

<sup>1</sup> Сколько же нам ждать, чтобы научно обоснованно объявить, что войны некоторой величины стали случаться реже? Все просто: интервалы между войнами следуют экспоненциальному распределению без памяти, поэтому функция выживания для отклонения втрое больше среднего составляет  $e^{-3} \approx 0,05$ . Это значит, что перед научно обоснованными заявлениями надо ждать втрое дольше, чем средний интервал. Говоря о таких больших войнах, как Первая мировая и Вторая мировая, ждать надо 300 лет. *Печально, но факт.* — Прим. автора.

## G

### КАКОВА ВЕРОЯТНОСТЬ ТРЕТЬЕЙ МИРОВОЙ ВОЙНЫ?<sup>\*\*</sup>

Этот раздел взят из статьи в рамках дискуссий среди публицистов-интеллектуалов, заявивших о снижении уровня насилия «согласно данным» без понимания научной сложности вопроса; для значимых утверждений в условиях жирных хвостов требуется больше данных и более аккуратные подходы. Наш ответ (автора и П. Чирилло) можно считать подведением итогов спора и формулировкой морали по поводу наивного эмпиризма при жирных хвостах.

В одном из последних номеров *Significance*<sup>1</sup> мистер Питер Макинтайр спросил, какова вероятность третьей мировой войны в текущем столетии. Профессор Майкл Шпагат написал, что никто не знает ответа — и здесь мы с ним совершенно согласны. Затем он добавил, что «большая война возможна, но, на мой взгляд, крайне маловероятна». И в качестве довода сослался на газетные публикации и научно-популярную книгу профессора Стивена Пинкера «Лучшее в нас». Профессор Пинкер утверждает, что в мире наблюдается долговременный спад уровня насилия, предполагающий, что воинственность человечества претерпевает структурные изменения.

И далее в своем ответе профессор Шпагат имел неосторожность упомянуть нашу статью (в этой книге это Глава 16), часть нашего большого исследования жирнохвостых случайных величин.

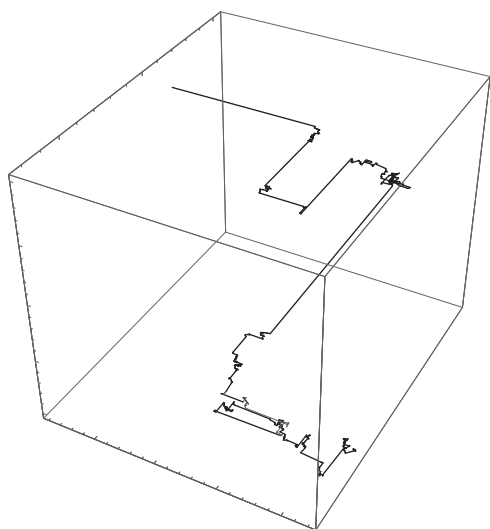
В чем особенность жирнохвостых величин? Основное влияние на их параметры (начиная со среднего значения) оказывают события экстремальные, «в хвостах». Расхожий пример — распределение Парето «80 / 20».

---

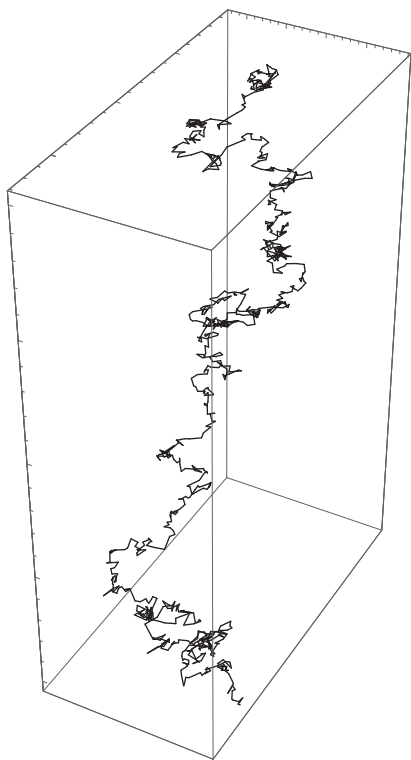
<sup>1</sup> Речь идет о редакционной статье 5 января 2016, *Ask a statistician: What are the chances of World War III?*



**Рисунок G.1:** После Наполеона в Европе наступило затишье. Пока не появился национализм



**Рисунок G.2:** История движется скачками. Жирнохвостый исторический процесс, в котором события распределены по степенному закону, отвечающему принципу «80/20», то есть  $\alpha \approx 1,13$ , представленный в 3D как процесс с независимыми приращениями [имеется в виду последовательность значений случайной величины в дискретные моменты времени, в которой приращение этой величины от одного момента до следующего не зависит от предшествующих приращений]



**Рисунок G.3:** Как это себе представляют заявители «Великой стабилизации» [предположение, что США с середины 1980-х вступили в эпоху без экономических кризисов] и «Долгого мира» [предположение, что мир с 1945 вступил в эпоху без мировых войн], история как тонкохвостый процесс

Мы показываем, говоря прямо, что данные не поддерживают предположение о структурных изменениях в воинственности человечества. И первая ошибка профессора Шпагата в том, что он неправильно прочитал наше утверждение; мы не делаем ни пессимистических, ни оптимистических заявлений, мы всего лишь полагаем, что статистик должен следовать принципам статистической теории и не указывать данным, в чью пользу говорить.

Вернемся к первым принципам.

### **Фундаментальные принципы**

В своей основе статистика — наука о том, как убедить людей не строить научные теории на пустом месте, то есть там, где нет значимого отличия от чистой случайности. Иначе происходит пресловутое «одурачивание случайностью».

Уточним далее, что, когда случайные величины жирнохвостые, обычные механизмы закона больших чисел существенно замедлены, и для выводов требуется больше данных за более длительные периоды наблюдений. Коварным образом это так не для всех утверждений, и для некоторых утверждений достаточно немногих данных: статистические выводы в домене жирных хвостов асимметричны. Требуется гораздо больше данных для того, чтобы подтвердить отсутствие черных лебедей, чем для того,

чтобы подтвердить наличие черных лебедей, и нам нужно гораздо больше данных для заявлений о спаде насилия, чем для заявлений о его подъеме.

И, наконец, заметим, что те заявления, которые не являются статистически значимыми и для которых незначимость доказана, не следует использовать как опору для построения научных теорий.

Эти фундаментальные принципы часто упускают, потому что обучение в общественных науках заимствует примитивный инструментарий тонкохвостых доменов [2]. В физике обычны ситуации, где можно делать утверждения по малым наборам наблюдений и можно обойтись даже без стандартной статистической методологии, потому что у изучаемых величин низкая дисперсия. Чем выше дисперсия, тем больше требуется данных для статистических заявлений. При жирных хвостах дисперсия обычно высока и недооценивается по прошлым данным.

Вторая — и более серьезная — ошибка Шпагата и Пинкера в том, что они полагают хвостовые события как нечто далекое от среднего и не влияющее на него, тогда как в действительности хвостовые события включены в среднее.

*Для жирнохвостых величин среднее почти полностью определяется экстремальными значениями. Если у вас большая неопределенность с хвостами, то у вас большая неопределенность и со средним.*

Поэтому некорректно говорить, будто частота острых конфликтов спала, не считая риска хвостовых событий; это все равно что сказать про человека, что он добрейшая душа, не считая одного случая, когда он принес в школу автомат и убил 30 учеников.

### **Робастность**

Наше исследование старается построить максимально робастную статистическую картину острых конфликтов, полагаясь на методы из теории экстремальных значений и статистические методы работы с жирными хвостами. Кроме того, мы выполняем проверки робастности, чтобы оценить влияние несовершенства данных, собранных тысячи лет назад: наши результаты должны оставаться в силе, даже если треть (а то и больше) данных ошибочна.

### **Интервалы между событиями**

Мы показываем, что интервалы между крупными конфликтами крайне велики и совместимы с гипотезой о равномерном процессе Пуассона; поэтому никакой тренд установить невозможно, как и невозможно объявить современных людей менее воинственными, чем прежде. Для конфликтов, вызывающих минимум 10 миллионов жертв (включая конфликты и менее кровавые, чем Первая мировая война и Вторая мировая война), время ожидания в среднем 136 лет при среднем абсолютном отклонении 267 лет (или чаще, если масштабировать данные к сегодняшнему населению: в среднем 52 года при отклонении 61 год). Очевидно, что семидесяти лет так называемого «долгого мира» недостаточно для сильных утверждений о возможности третьей мировой войны в близком будущем.



### Недооценка среднего

Кроме того, мы находим, что среднее по прошлым конфликтам недооценивает истинное статистическое среднее минимум наполовину. Почему? Учтите, что 90–97 % наблюдений попадает ниже среднего, и станет ясно, что поправка согласно теории экстремальных значений будет существенной. (При крайне жирных хвостах статистическое среднее может быть ближе к максимальному прошлому наблюдению, чем к среднему по выборке.)

### Обычная ошибка

Похожие ошибки делались и в прошлом. В 1860 некий Г. Т. Бокль<sup>1</sup> рассуждал столь же антинаучно в отношении статистики, как Пинкер и Шпагат.

Что это варварское занятие вместе с развитием общества быстро падает, известно каждому, даже поверхностно знакомому с историей Европы. Если мы сравним одно столетие с другим, то увидим, что уже довольно давно войны сделались менее частыми; теперь это движение так ясно видно, что до начала последних военных действий мы жили в мире почти сорок лет, обстоятельство беспрецедентное... Возникает вопрос, насколько нравственное чувство способствовало этому важному успеху<sup>2</sup>.

Способствовало чему-то нравственное чувство или нет, но столетие после сочинений мистера Бокля оказалось самым большим душегубством за всю историю человечества.

В заключение скажем, что считаем делом чести указать на фундаментальные статистические ошибки в журнале под названием *Significance*, ибо вопрос стоит именно о значимости и донесении принципов статистической строгости до широкой публики.

---

1 Buckle H. T. (1858) *History of Civilization in England, Vol. 1*, London: John W. Parker and Son. — Прим. автора.

2 Перевод К. Бестужева-Рюмина, 1863.

## **Часть VI**

### **СТАТЬИ О МЕТАВЕРОЯТНОСТИ**



## КАК ТОЛСТЫЕ ХВОСТЫ ВОЗНИКАЮТ ИЗ РЕКУРСИВНОЙ ЭПИСТЕМОЛОГИЧЕСКОЙ НЕОПРЕДЕЛЕННОСТИ\*

**Противоположность центральной предельной теоремы.** В рамках центральной предельной теоремы мы начинаем с некоторого распределения и приходим к гауссову. Но в этой области скорее случится нечто противоположное. Помните, как мы ужирняли хвосты гауссианы, стохастизируя дисперсию? Давайте теперь используем тот же метод для метавероятности, добавляя уровни неопределенности<sup>1</sup>.

**Регрессия<sup>2</sup> (погрешность погрешности).** Главная проблема, порождающая черного лебедя, — ограниченное понимание погрешности модели (или репрезентации), а если это понимание достигнуто, то недостаток понимания погрешностей второго порядка (в методах вычисления погрешностей); согласно логике регрессии, повторением данного способа размышления нельзя дойти до предела (*особенно если нет причин остановиться*). Но нет проблемы остановить эту рекурсию в заранее объявленной точке, не подпадающей под количественные и статистические методы.

**Новый вывод степенных законов: не статистический, а эпистемологический.** Заметим, что ранее мы выводили степенные законы из статистических соображений, таких как аккумуляция преимуществ<sup>3</sup>, предпочтительное связывание<sup>4</sup>, эффекты игры

1 Ранняя версия этой главы была представлена на семинаре памяти Бенуа Мандельброта 29 апреля 2011 в Нью-Хейвене, штат Коннектикут. — *Прим. автора.*

2 Англ. *Regress Argument* — один из доводов скептицизма, т. н. второе Пирроново положение: *Любое доказательство требует доказать то, на что опирается, и так до бесконечности.*

3 Англ. *cumulative advantage* — то же, что Юлов процесс и принцип Матфея: феномен неравномерного распределения преимуществ, в котором сторона, уже ими обладающая, продолжает их накапливать и умножать. Ср. *...ибо всякому имеющему дастся и приумножится, а у неимеющего отнимется и то, что имеет* (Мф. 25:29).

4 Англ. *Preferential attachment process* — процесс роста безмасштабной случайной сети, когда вероятность добавления связи к тому или иному узлу пропорциональна текущему числу связей узла.

по принципу «победитель получает все», критичность<sup>1</sup> и свойства, выведенные Юлом, Мандельбротом, Ципфом, Саймоном, Баком и другими порождаются структурными условиями или нарушением той независимости слагаемых в сумме случайных величин, которая предполагается для применения центральной предельной теоремы; см. [90], [212], [100], [162], [161]. Напротив, эта статья полностью эпистемологическая, основанная на стандартных философских проблемах сомнения и регрессии.



**Рисунок 17.1:** Ранняя версия этой главы была представлена на семинаре памяти Бенуа Мандельброта

## 17.1. МЕТОДЫ И ВЫКЛАДКИ

### 17.1.1. Уровни неопределенности

Возьмем стандартное распределение вероятностей, скажем гауссово. Оценивая его среднеквадратическое отклонение  $\sigma$ , нужно и самой этой оценке приписать некоторое среднеквадратическое отклонение. Эта неопределенность в оценке неопределенности на жаргоне торговцев опционами называется «волатильность волатильности» (смотрите у Талеба, 1997, Дермана, 1994, Дюпира, 1994, Халла и Уайта, 1997), а в нашем рассказе здесь это будет «степень неопределенности степени неопределенности». И нет причины останавливаться на этом уровне: мы можем вводить в рас-

<sup>1</sup> Англ. *criticality* или *self-organized criticality* — то же, что Баков процесс: поведение вблизи аттрактора.

смотрение все новые неопределенности высоких порядков, степень неопределенности степени неопределенности и так далее. Нет никаких оснований полагать, что на каком-то шаге процедуры должна возникнуть полная определенность.

### 17.1.2. Интегралы высоких порядков в стандартном гауссовом случае

Начнем с гауссова случая и сосредоточимся на неопределенности предполагаемого среднеквадратического отклонения. Определим  $\varphi(\mu, \sigma; x)$  как гауссову функцию плотности для значения  $x$  при среднем  $\mu$  и среднеквадратическом отклонении  $\sigma$ .

Стохастическое стандартное отклонение второго порядка — это интеграл  $\varphi$  по всем значениям  $\sigma \in \mathbb{R}^+$  согласно функции плотности  $f(\bar{\sigma}, \sigma_1; \sigma)$ , где  $\sigma_1$  — ее коэффициент масштаба (согласно нашему намерению отслеживать погрешность погрешности), который не обязательно представляет собой среднеквадратическое отклонение; математическое ожидание  $\bar{\sigma}$  обозначим  $\bar{\sigma}_1$ .

$$f_1(x) = \int_0^{\infty} \varphi(\mu, \sigma; x) f(\bar{\sigma}, \sigma_1; \sigma) d\sigma.$$

Обобщая до  $N$ -го порядка, находим, что функция плотности  $f(x)$  принимает вид

$$f_N(x) = \int_0^{\infty} \dots \int_0^{\infty} \varphi(\mu, \sigma; x) f(\bar{\sigma}, \sigma_1; \sigma) f(\bar{\sigma}_1, \sigma_2; \sigma_1) \dots f(\bar{\sigma}_{N-1}, \sigma_N; \sigma_{N-1}) d\sigma d\sigma_1 d\sigma_2 \dots d\sigma_{N-1}. \quad (17.1)$$

Чтобы так исследовать итоговое распределение  $f_N(x)$ , нужно знать множество параметров и вид множества распределений. Например, в финансовой математике, чтобы  $\sigma$  не принимала отрицательных значений, для этого параметра вместо гауссова распределения традиционно используют логнормальное распределение  $\sigma^2$ , или гауссово распределение для логарифма отношения  $\ln \frac{\sigma_1^2}{\sigma_2^2}$ . Для каждого уровня погрешности придется задавать меру  $f$ . Вместо всего этого, как мы увидим далее, можно использовать приближение средним отклонением  $\sigma$ .

Дискретизировать последовательностью вложенных бинарных ветвлений по  $\sigma$  — простая мультипликативная процедура.

Мы видели в прошлой главе эффективное упрощение для захвата выпуклости вниз — отношения или разности  $\varphi(\mu, \sigma; x)$  и  $\int_0^{\infty} \varphi(\mu, \sigma; x) f(\bar{\sigma}, \sigma_1; \sigma) d\sigma$  (отклонение первого порядка), — путем использования взвешенного среднего значений  $\sigma$ ; например, для простого случая стохастической волатильности первого порядка это будет

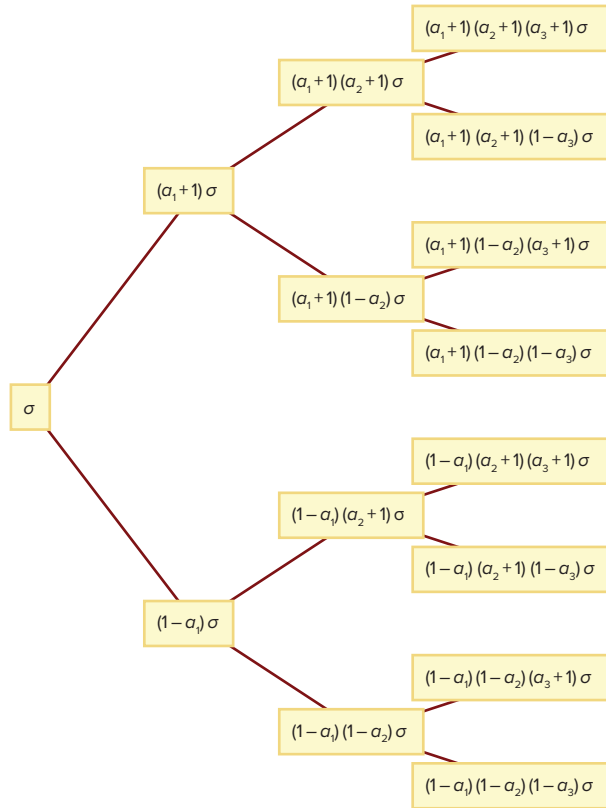
$$\sigma(1 \pm a_1),$$

при  $0 \leq a_1 < 1$ , где  $a_1$  — пропорциональное среднее абсолютное отклонение для  $\sigma$ , другими словами — мера абсолютной погрешности для  $\sigma$ . Мы используем  $\frac{1}{2}$  как вероятность каждого из двух состояний. В отличие от рассмотренной выше ситуации мы

сохраняем не дисперсию, а среднеквадратическое отклонение. Поэтому распределение с использованием стохастического среднеквадратического отклонения первого порядка можно выразить так:

$$f_1(x) = \frac{1}{2}(\varphi(\mu, \sigma(1+a_1), x) + \varphi(\mu, \sigma(1-a_1), x)). \quad (17.2)$$

Теперь припишем погрешности  $a_1$  неопределенность, выраженную через  $a_2$  аналогичным образом. Соответственно вместо  $a_1$  начнем использовать  $\frac{1}{2}a_1(1 \pm a_2)$ .



**Рисунок 17.2:** Три уровня погрешности, если  $\sigma$  следует мультипликативной процедуре

Стохастическое среднеквадратическое отклонение второго порядка:

$$f_2(x) = \frac{1}{4}(\varphi(\mu, \sigma(1+a_1(1+a_2)), x) + \varphi(\mu, \sigma(1-a_1(1+a_2)), x) + \\ + \varphi(\mu, \sigma(1+a_1(1-a_2)), x) + \varphi(\mu, \sigma(1-a_1(1-a_2)), x)) \quad (17.3)$$

и  $N$ -го порядка

$$f_N(x) = \frac{1}{2^N} \sum_{i=1}^{2^N} \varphi(\mu, \sigma M_i^N, x),$$

где  $M_i^N$  —  $i$ -я строка (скаляр) матрицы  $M^N(2^N \times 1)$

$$M^N = \left( \prod_{j=1}^N (a_j \mathbf{T}_{i,j} + 1) \right)_{i=1}^{2^N}$$

и  $\mathbf{T}_{i,j}$  — элемент  $i$ -й строки  $j$ -го столбца матрицы всех возможных цепочек длины  $N$  в алфавите  $\{-1, 1\}$ , то есть последовательностей вида  $(1, 1, 1, \dots)$  всевозможных сочетаний 1 и  $-1$ .

При  $N=3$

$$\mathbf{T} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{pmatrix}$$

и

$$M^3 = \begin{pmatrix} (1-a_1)(1-a_2)(1-a_3) \\ (1-a_1)(1-a_2)(1+a_3) \\ (1-a_1)(1+a_2)(1-a_3) \\ (1-a_1)(1+a_2)(1+a_3) \\ (1+a_1)(1-a_2)(1-a_3) \\ (1+a_1)(1-a_2)(1+a_3) \\ (1+a_1)(1+a_2)(1-a_3) \\ (1+a_1)(1+a_2)(1+a_3) \end{pmatrix}.$$

Таким образом,  $M_1^3 = (1-a_1)(1-a_2)(1-a_3)$  и т. п.

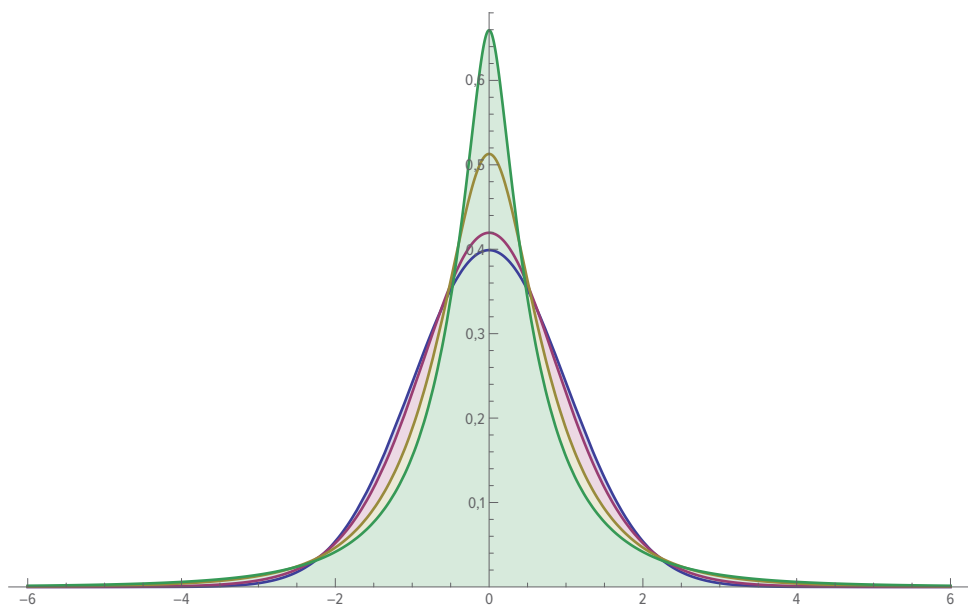
Заметим, что разные погрешности  $a_i$  не аналогичны ошибкам выборочных оценок; скорее их можно сравнить с проекцией уровня погрешности в будущее. Повторим: эти параметры — эпистемологические.

**Итоговое смешанное распределение.** Взвешенное среднее распределение примет вид (вспомним, что  $\varphi$  — обычная гауссова функция плотности, зависящая от среднего  $\mu$ , среднеквадратического отклонения  $\sigma$  и значения случайной величины  $x$ ):

$$f(x | \mu, \sigma, M, N) = 2^{-N} \varphi(\mu, \sigma M_i^N, x).$$

Ее можно аппроксимировать логнормальным распределением для  $\sigma$  и соответствующей дисперсии  $V$ . Но нас интересует именно  $V$ , и она зависит от поведения погрешностей высшего порядка.





**Рисунок 17.3:** Утолщение хвостов (и повышение пиков) при росте  $N$ ; здесь  $N = 0, 5, 10, 25, 50$ ; все  $a = \frac{1}{10}$

Рассмотрим далее различные режимы для погрешностей высшего порядка.

### РЕЖИМ 1 (ВЗРЫВНОЙ): СЛУЧАЙ ПОСТОЯННОГО ПАРАМЕТРА $a$

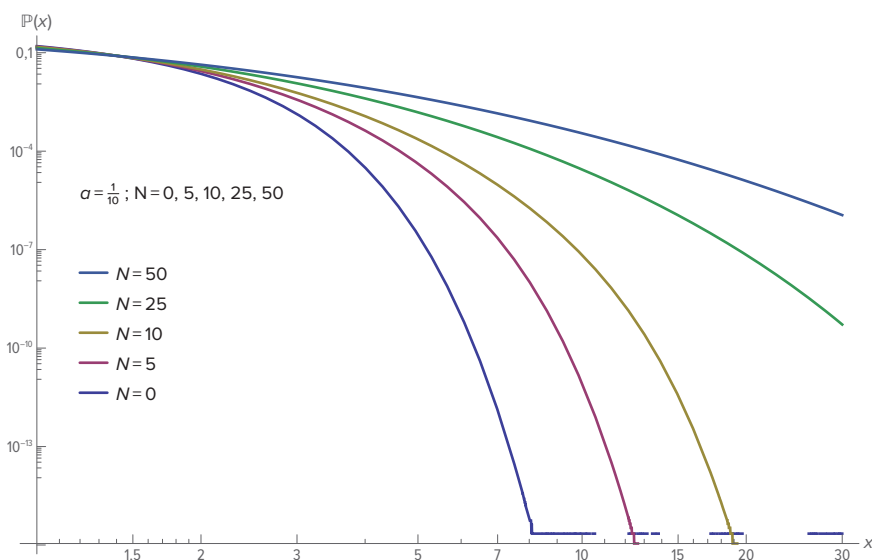
**Особый случай постоянного  $a$ .** Пусть  $a_1 = a_2 = \dots = a_N = a$ , т. е. это случай плоского графика пропорциональной погрешности  $a$ . Матрица  $M$  коллапсирует в обычное биномиальное дерево для рассеяния на уровне  $N$ .

$$f(x|\mu, \sigma, M, N) = 2^{-N} \sum_{j=0}^N \binom{N}{j} \varphi\left(\mu, \sigma(a+1)^j (1-a)^{N-j}, x\right). \quad (17.4)$$

Благодаря линейности сумм при постоянном  $a$  можно использовать биномиальное распределение для весовых коэффициентов моментов (снова заметим, что действует искусственное ограничение этого анализа: первый момент  $\mu$  жестко задан и известен априори).

Порядок	Момент
1	$\mu$
2	$\sigma^2(a^2+1)^N + \mu^2$
3	$3\mu\sigma^2(a^2+1)^N + \mu^3$
4	$6\mu^2\sigma^2(a^2+1)^N + \mu^4 + 3(a^4+6a^2+1)^N\sigma^4$

Снова заметим, что, странным образом, несмотря на взрывной характер высших моментов, математическое ожидание абсолютной величины  $x$  не зависит ни от  $a$ , ни



**Рисунок 17.4:** Вероятность превысить  $x$  на графике в дважды логарифмическом масштабе; видим аналогичное степенному закону приближение к горизонтали при росте  $N$ . Здесь все  $a = \frac{1}{10}$

от  $N$ , поскольку возмущения  $\sigma$  не влияют на первый абсолютный момент  $= \sqrt{\frac{2}{\pi}} \sigma$  (где фигурирует исходное значение  $\sigma$ ). Ситуация была бы другой при сложении  $x$ .

С каждой рекурсией дисперсия процесса умножается на  $(1 + a^2)$ . Процедура аналогична стохастизации в модели волатильности, где среднеквадратическое отклонение (а не дисперсия) следует логнормальному распределению, волатильность которого растет вместе с  $N$  и, следовательно, в пределе дисперсия бесконечна.

### Следствия

При постоянном  $a > 0$ , как и в более общем случае переменного  $a$ , если  $a_n \geq a_{n-1}$ , моменты взрываются.

**А.** При самом малом значении  $a > 0$  величина  $(1 + a^2)^N$  неограниченно возрастает, из-за чего второй момент (но не первый) уходит в бесконечность при  $N \rightarrow \infty$ . Даже незначительная, казалось бы, погрешность 0,001 % приведет к взрыву моментов и непригодности класса распределений  $\mathcal{L}^2$ .

**В.** В этой ситуации необходимость степенных законов или иных распределений вне нормы  $\mathcal{L}^2$  диктуется эпистемологическими причинами и не зависит от наблюдений данных в прошлом.

Заметим, что нужна априорная причина (в философском смысле), чтобы превратиться при некотором  $N$  и положить предел безудержному росту второго момента.

**Сходимость к свойствам, аналогичным степенным законам.** На примере следующего графика в дважды логарифмическом масштабе (Рисунок 17.4) можно видеть, как при высших порядках статистической волатильности при постоянном пропорцио-

нальном стохастическом коэффициенте  $a_1 = a_2 = \dots = a_N = \frac{1}{10}$  плотность приближается к степенному закону (подобно логнормальному распределению при высокой дисперсии); это проявляется как более плоский график плотности на дважды логарифмическом графике. Вероятности продолжают расти в хвостах по мере добавления уровней неопределенности, пока не приближаются к границам степенного закона, хотя первый момент коварным образом остается инвариантным.

Тот же эффект имеет место, когда  $a$  возрастает, приближаясь к 1 и предел показателя хвоста  $P_{>x}^{-1}$  приближается к 1, хотя остается  $> 1$ .

### 17.1.3. Влияние на малые вероятности

Далее найдем оценки для эффектов толстого хвоста. Очевидным эффектом является рост маловероятных событий.

Возьмем вероятность превышения порога  $K$  при заданном  $N$  и постоянном параметре  $a$ :

$$\mathbb{P}_{>K} | N = 2^{-N-1} \sum_{j=0}^N \binom{N}{j} \operatorname{erfc} \frac{K}{\sqrt{2\sigma(a+1)^j(1-a)^{N-j}}}, \quad (17.5)$$

где  $\operatorname{erfc}(\cdot)$  — дополнительная функция ошибок,  $1 - \operatorname{erf}(\cdot)$ ,  $\operatorname{erf} z = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ .

**Эффект выпуклости вниз.** В следующих таблицах показано, при разных значениях  $N$ , отношение вероятности превысить заданный высокий порог к вероятности такого же события в случае стандартного гауссова распределения.

Таблица 17.1: Случай  $a = \frac{1}{10}$

$N$	$\frac{\mathbb{P}_{>3} N}{\mathbb{P}_{>3} N=0}$	$\frac{\mathbb{P}_{>5} N}{\mathbb{P}_{>5} N=0}$	$\frac{\mathbb{P}_{>10} N}{\mathbb{P}_{>10} N=0}$
5	1,01724	1,155	7
10	1,0345	1,326	45
15	1,05178	1,514	221
20	1,06908	1,720	922
25	1,0864	1,943	3347

Таблица 17.2: Случай  $a = \frac{1}{100}$

$N$	$\frac{\mathbb{P}_{>3} N}{\mathbb{P}_{>3} N=0}$	$\frac{\mathbb{P}_{>5} N}{\mathbb{P}_{>5} N=0}$	$\frac{\mathbb{P}_{>10} N}{\mathbb{P}_{>10} N=0}$
5	2,74	146	$1,09 \times 10^{12}$
10	4,43	805	$8,99 \times 10^{15}$
15	5,98	1980	$2,21 \times 10^{17}$
20	7,38	3529	$1,20 \times 10^{18}$
25	8,64	5321	$3,62 \times 10^{18}$

1 То есть правого хвоста функции выживания  $\bar{F}(x) = 1 - F_X(x) = \mathbb{P}(X > x)$ .

## 17.2. РЕЖИМ 2: СЛУЧАИ ЗАТУХАЮЩИХ ПАРАМЕТРОВ $a_n$

Как мы говорили, для затухания параметра  $a$  или полного прерывания последовательности при некотором  $N$  мы можем задать априорные причины (и, собственно, никакие другие причины не помогут). Если  $a_i$  высоких порядков снижаются, моменты в общем и целом станут ограничены (наследия хвосты от логнормального распределения  $\sigma$ ).

### 17.2.1. Режим 2-а; «потери» погрешности высокого порядка

Пусть погрешности высокого порядка «теряют энергию» в геометрической прогрессии со знаменателем  $\lambda$ ,  $0 \leq \lambda < 1$ , так что  $a_N = \lambda a_{N-1}$ , и поэтому  $a_N = \lambda^{N-1} a_1$ , где  $a_1$  — обычная сила стохастического среднеквадратического отклонения. Пусть  $\mu = 0$ .

При  $N = 2$  второй момент принимает вид:

$$M_2(2) = (a_1^2 + 1) \sigma^2 (a_1^2 \lambda^2 + 1).$$

При  $N = 3$  второй момент принимает вид:

$$M_2(3) = (a_1^2 + 1) \sigma^2 (a_1^2 \lambda^2 + 1)(a_1^2 \lambda^4 + 1).$$

При произвольном  $N$  второй момент принимает вид:

$$M_2(N) = (a_1^2 + 1) \sigma^2 (a_1^2 \lambda^2 + 1) \prod_{i=1}^{N-1} (a_1^2 \lambda^{2i} + 1). \quad (17.6)$$

Формулу (17.6) можно переписать с использованием символа Похгаммера  $(a; q)_N = \prod_{i=1}^{N-1} (1 - aq^i)$  так:

$$M_2(N) = \sigma^2 (-a_1^2; \lambda^2)_N,$$

и тогда предел можно записать как

$$\lim_{N \rightarrow \infty} M_2(N) = \sigma^2 \frac{(\lambda^2; \lambda^2)_2 (a_1^2; \lambda^2)_\infty}{(\lambda^2 - 1)^2 (\lambda^2 + 1)}.$$

Четвертый момент получаем рекурсией:

$$M_4(N) = 3\sigma^4 \prod_{i=1}^{N-1} (6a_1^2 \lambda^{2i} + a_1^4 \lambda^{4i} + 1), \text{ или}$$

$$M_4(N) = 3\sigma^4 \left( -(3 - 2\sqrt{2})a_1^2; \lambda^2 \right)_N \left( -(3 + 2\sqrt{2})a_1^2; \lambda^2 \right)_N, \quad (17.7)$$

в пределе

$$\lim_{N \rightarrow \infty} M_4(N) = 3\sigma^4 \left( -(3 - 2\sqrt{2})a_1^2; \lambda^2 \right)_\infty \left( -(3 + 2\sqrt{2})a_1^2; \lambda^2 \right)_\infty. \quad (17.8)$$

Таким образом, предельный второй момент при  $\lambda = 0,9$  и  $a_1 = 0,2$  составляет всего-навсего  $1,28 \sigma^2$ , существенное, но не злокачественное смещение выпуклости вниз. Предельный четвертый момент составляет  $9,88 \sigma^4$ , втрое больше гауссова ( $3 \sigma^4$ ), но конечный. При малых  $a$  и при  $\lambda$  близко к 1 четвертый момент падает до гауссова.

### 17.2.2. Режим 2-b; второй метод, немультимпликативная погрешность

По второму методу при  $N$  рекурсиях среднееквадратическое отклонение становится

$$\sigma(1 \pm (a_1(1 \pm (a_2(1 \pm a_3(\dots))))))$$

$$\mathbb{P}(X, \mu, \sigma, N) = \frac{1}{L} \sum_{i=1}^L f\left(x, \mu, \sigma\left(1 + (\mathbf{T}^N \cdot \mathbf{A}^N)_i\right)\right),$$

где  $(\mathbf{T}^N \cdot \mathbf{A}^N)_i$  — скалярное произведение  $i$ -й строки  $\mathbf{T}^N$ , матрицы всех цепочек длины  $N$  в алфавите  $\{1, -1\}$ , на  $\mathbf{A}^N$ , вектор из степеней параметра  $a$

$$\mathbf{A}^N = (a^j)_{j=1, \dots, N};$$

$L$  — длина матрицы  $\mathbf{T}^{N1}$ .

Например, при  $N=3$  имеем  $\mathbf{A}^3 = \begin{pmatrix} a \\ a^2 \\ a^3 \end{pmatrix}$  и

$$\mathbf{T}^3 \cdot \mathbf{A}^3 = \begin{pmatrix} a + a^2 + a^3 \\ a + a^2 - a^3 \\ a - a^2 + a^3 \\ a - a^2 - a^3 \\ -a + a^2 + a^3 \\ -a + a^2 - a^3 \\ -a - a^2 + a^3 \\ -a - a^2 - a^3 \end{pmatrix}.$$

Моменты имеют вид:

$$M_1(N) = \mu,$$

$$M_2(N) = \mu^2 + 2\sigma,$$

$$M_4(N) = \mu^4 + 12\mu^2\sigma + 12\sigma^2 \sum_{i=0}^N a^{2i}$$

и в пределе

$$\lim_{N \rightarrow \infty} M_4(N) = \frac{12\sigma^2}{1-a^2} + \mu^4 + 12\mu^2\sigma,$$

и это весьма умеренное значение.

### 17.3. ПРЕДЕЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Смотрите у Талеба и Чирилло [241], как исследуется предельное распределение, которое при правильных условиях оказывается логнормальным. Собственно, логнормальные аппроксимации работают хорошо, если погрешность погрешности составляет постоянную долю.

<sup>1</sup> То есть  $L = 2^N$ .

## СТОХАСТИЧЕСКИЙ ПОКАЗАТЕЛЬ ХВОСТА ПРИ АСИММЕТРИЧНЫХ СТЕПЕННЫХ ЗАКОНАХ<sup>1</sup>

Мы исследуем случайные величины в классе степенного закона / медленно меняющихся функций при стохастическом показателе степени хвоста  $\alpha$ , имеющем собственное распределение<sup>1</sup>. Мы показываем, какое влияние оказывает стохастичность на математическое ожидание и высшие моменты случайной величины. В частности, моменты асимметричной случайной величины с правым хвостом или перекосом в правую сторону, если существуют, растут с увеличением дисперсии  $\alpha$ , а для случайной величины с перекосом влево — убывают. То же относится к условным потерям (*conditional shortfall*, они же *conditional value at risk*, CVaR)<sup>2</sup>, или функциям среднего избытка (*mean excess*, ME).

Мы доказываем утверждение в общем случае и исследуем особую ситуацию логнормального распределения  $\alpha \in [b, \infty)$ ,  $b > 1$ .

Стохастичность показателя степени порождает существенное смещение оценки среднего и высших моментов; таков эффект неопределенности данных. Он сказывается на погрешности выборки: неопределенность относительно  $\alpha$  повышает ожидаемое среднее.

Это смещение сохраняется при накоплении наблюдений, даже когда достигается сходимость к устойчивому распределению. Мы выводим неравенства, связанные с такой асимметрией.

Кроме того, мы рассматриваем ситуацию усеченного степенного закона (т. е. с компактным носителем) и применяем результаты к изучению статистики острых конфликтов Чирилло и Талебом (2016). Мы показываем, что с учетом неопределенности исторических данных необходимо увеличивать истинное среднее.

<sup>1</sup> Первое сообщение на конференции *Extremes and Risks in Higher Dimensions* в Lorentz Center, Лейден, Нидерланды, сентябрь 2016.

<sup>2</sup> Они же ожидаемые потери (*expected shortfall*, ES).

## 18.1. ИСТОРИЯ ВОПРОСА

Стохастическая волатильность появилась как эвристический прием в финансовой математике, изобретенный трейдерами, которые искали смещение в оценке опционов; суть приема в том, чтобы приписать гауссову распределению несколько возможных дисперсий, действующих на коротких участках или начиная с некоторой даты в будущем. Опционы «далеко от денег»<sup>1</sup> (т. е. связанные с хвостовыми событиями) приносят тем большую прибыль, чем больше неопределенность в дисперсии распределения, поскольку имеют выпуклость вниз по среднеквадратическому отклонению.

Идея стохастической волатильности породила семейство моделей броуновского движения со стохастической дисперсией (смотрите обзор Гезерала [102]) и оказалась полезной для изучения поведения негауссовых случайных величин и их функций (например, цен опционов).

Подобно опционам, которые имеют выпуклость вниз по масштабу распределения, во многих ситуациях ожидания имеют выпуклость вниз по показателю степени хвоста степенного закона. В данном тезисе исследуется два случая:

- Стандартные степенные законы — с одним хвостом или асимметричные.
- Псевдостепенные законы — где поведение случайной величины сходно со степенным законом, однако носитель компактный, как в исследовании частоты острых конфликтов [46], где число жертв войны не может превысить некоторый максимум.

## 18.2. ОДНОХВОСТЫЕ РАСПРЕДЕЛЕНИЯ СО СТОХАСТИЧЕСКОЙ АЛЬФОЙ

### 18.2.1. Общие случаи

#### Определение 18.1

Пусть  $X$  — случайная величина из класса правых хвостов степенного закона, и тем самым ее носитель  $[x_0, +\infty)$ ,  $x_0 \in \mathbb{R}$ , и пусть она принадлежит следующему подклассу.

Подкласс  $\mathfrak{P}_1$ :

$$\left\{ X \mid \mathbb{P}(X > x) = L(x)x^{-\alpha}, \frac{\partial^q L(x)}{\partial x^q} = 0 \text{ при } q \geq 1 \right\}. \quad (18.1)$$

Заметим, что в этом подклассе допустим сдвиг функции распределения и любые отрицательные  $x_0$ , лишь бы  $x_0 > -\infty$ . Весь класс следующий.

Класс  $\mathfrak{P}$ :

$$\left\{ X \mid \mathbb{P}(X > x) \sim L(x)x^{-\alpha} \right\}, \quad (18.2)$$

<sup>1</sup> Англ. *far from the money* — опционы, страйк-цена которых сильно отличается от текущей рыночной цены базового актива. Владелец опциона колл может воспользоваться правом купить по страйк-цене, владелец опциона пут может воспользоваться правом продать по страйк-цене; так или иначе, нахождение опциона далеко от денег означает для владельца опциона возможность большой прибыли, если опцион «в деньгах» (*in the money option*), то есть если разница между ценами в ту сторону, в какую выгодно владельцу опциона. Если разница между ценами не в ту сторону, владелец увидит, что опцион «вне денег» (*out of the money option*), и не станет им пользоваться.

где  $\sim$  значит, что в пределе отношение правой части к левой стремится к 1 при  $x \rightarrow \infty$ , и где  $L: [x_{\min}, +\infty) \rightarrow (0, +\infty)$  — медленно меняющаяся функция, определяемая по критерию  $\lim_{x \rightarrow \infty} \frac{L(kx)}{L(x)} = 1$  при любых  $k > 0$ .  $L'(x)$  монотонна. Константа  $\alpha > 0$ .

Допустим также, что:

$$\lim_{x \rightarrow \infty} L'(x)x = 0, \quad (18.3)$$

$$\lim_{x \rightarrow \infty} L''(x)x = 0. \quad (18.4)$$

Имеем

$$\mathfrak{P}_1 \subset \mathfrak{P}.$$

Заметим, что первый класс соответствует распределениям Парето (при должном сдвиге и масштабировании), если  $L$  константа, а класс  $\mathfrak{P}$  соответствует более общему одностороннему степенному закону.

### 18.2.2. Неравенство стохастической альфы

В дальнейшем в этой статье будем обозначать  $X'$  версию случайной величины  $X$  со стохастической  $\alpha$ , константой в случае  $X$ .

#### Утверждение 18.1

Пусть  $p = 1, 2, \dots$ , и  $X'$  аналогична описанной выше случайной величине  $X$  из  $\mathfrak{P}_1$  (класс однохвостых правильно меняющихся распределений) с  $x_0 \geq 0$ , но со стохастической  $\alpha$ , причем во всех реализациях  $\alpha_i > p$  и среднее по всем реализациям сохраняется,  $\bar{\alpha} = \alpha$ . Тогда

$$\mathbb{E}X'^p \geq \mathbb{E}X^p.$$

#### Утверждение 18.2

Пусть  $K$  — некоторый порог. Для случайной величины  $X$  в классе  $\mathfrak{P}$  ожидаемые условные потери (CVar):

$$\lim_{K \rightarrow \infty} \mathbb{E}(X' | X' > K) \geq \lim_{K \rightarrow \infty} \mathbb{E}(X | X > K).$$

Схема доказательства следующая.

Заметим, что  $\mathbb{E}X^p$  имеет выпуклость вниз по  $\alpha$  в следующем смысле. Пусть  $X_{\alpha_i}$  — случайная величина, следующая распределению с постоянным показателем хвоста  $\alpha_i$ , где  $\alpha_i > p$ ,  $\forall i$ , и пусть  $\omega_i$  — положительные нормированные веса:  $\sum_i \omega_i = 1$ ,  $0 \leq |\omega_i| \leq 1$ ,  $\sum_i \omega_i \alpha_i = \bar{\alpha}$ . Согласно неравенству Йенсена:

$$\sum_i \omega_i \mathbb{E}X_{\alpha_i}^p \geq \mathbb{E}X_{\bar{\alpha}}^p.$$

Поскольку классы определяются по функциям выживания, нам сначала нужно найти соответствующие функции плотности,  $\varphi(x) = \alpha x^{-\alpha-1} L(x, \alpha) - x^{-\alpha} L^{(1,0)}(x, \alpha)$ , и нормирующую константу



$$L(x_0, \alpha) = x_0^\alpha \frac{2x_0 L^{(1,0)}(x_0, \alpha)}{\alpha - 1} - \frac{2x_0^2 L^{(2,0)}(x_0, \alpha)}{(\alpha - 1)(\alpha - 2)} \quad (18.5)$$

$\alpha \neq 1, 2$ , когда существует соответственно первая и вторая производная. Запись частной производной как  $L^{(p,0)}(x_0, \alpha)$  — сокращение<sup>1</sup> вместо  $\left. \frac{\partial^p L(x, \alpha)}{\partial x^p} \right|_{x=x_0}$ .

Согласно теореме представления Караматы, [22], [248], функция  $L$  на  $[x_0, +\infty)$  медленно меняющаяся тогда и только тогда, когда представима в виде

$$L(x) = \exp\left(\int_{x_0}^x \frac{\varepsilon(t)}{t} dt\right) + \eta(x),$$

где  $\eta(\cdot)$  — ограниченная измеримая функция, сходящаяся к конечному числу при  $x \rightarrow +\infty$ , и  $\varepsilon(x)$  — ограниченная измеримая функция, сходящаяся к нулю при  $x \rightarrow +\infty$ .

Соответственно  $L'(x)$  уходит в 0 при  $x \rightarrow \infty$ . Мы дополнительно предполагали в (18.3) и (18.4), что  $L'(x)$  стремится к 0 быстрее, чем  $x$ , и что  $L''(x)$  стремится к 0 быстрее, чем  $x^2$ . Интегрируя по частям, получаем

$$\mathbb{E}X^p = x_0^p + p \int_{x_0}^{\infty} x^{p-1} d\bar{F}(x),$$

где  $\bar{F}$  — функция выживания в формулах (23.1) и (18.2). Интегрируя по частям еще три раза и устроя производные  $L(\cdot)$  порядка выше 2, получаем

$$\mathbb{E}X^p = -\frac{x_0^{p-\alpha} L(x_0, \alpha)}{p-\alpha} - \frac{x_0^{p-\alpha+1} L^{(1,0)}(x_0, \alpha)}{(p-\alpha)(p-\alpha+1)} + \frac{x_0^{p-\alpha+2} L^{(2,0)}(x_0, \alpha)}{(p-\alpha)(p-\alpha+1)(p-\alpha+2)}, \quad (18.6)$$

что в особом случае  $X$  из  $\mathfrak{P}_1$  сводится к:

$$\mathbb{E}X^p = x_0^{p-\alpha} \frac{\alpha}{\alpha - p}. \quad (18.7)$$

Что же касается Предложения 2, мы можем подойти к доказательству со стороны того свойства, что  $\lim_{x \rightarrow \infty} L'(x) = 0$ . Это позволяет доказать закон Ван дер Вейка об инвариантности неравенства Парето к порогу хвоста, то есть что  $\frac{\mathbb{E}(X | X > K)}{K}$  сходится к константе при  $K \rightarrow +\infty$ .

Формула (18.6) представляет точные требования к функциональной форме  $L(x)$ , при которых выпуклость вниз распространяется на подклассы между  $\mathfrak{P}_1$  и  $\mathfrak{P}$ .

Наши результаты сохраняют силу для распределений, преобразованных путем сдвига и масштабирования, вида  $x \mapsto x - \mu + x_0$  (Парето II), или путем дальнейшего преобразования к типам Парето II и IV.

Заметим, что для простоты наше представление  $\mathfrak{P}_1$  использует один и тот же параметр  $x_0$  и как коэффициент масштаба, и как минимальное значение.

Мы можем убедиться, что математическое ожидание из формулы (18.7) выпукло вниз по  $\alpha$ :  $\frac{\partial^2}{\partial \alpha^2} \mathbb{E}X^p = x_0^{p-\alpha} \frac{2p}{(\alpha - p)^3}$ .

<sup>1</sup> Две позиции для указания порядка производной отвечают двум аргументам, по которым возможно дифференцирование.

### 18.2.3. Аппроксимации для класса $\mathfrak{P}$

Для остальных распределений класса,  $\mathfrak{P} \setminus \mathfrak{P}_1$ , наши результаты остаются в силе, когда мы записываем аппроксимацию математического ожидания случайной величины  $X$  как константу, умноженную на интеграл  $x^{-\alpha}$ , а именно как

$$\mathbb{E}X \approx k \frac{\nu(\alpha)}{\alpha - 1}, \quad (18.8)$$

где  $k$  — положительная константа, не зависящая от  $\alpha$ , и  $\nu(\cdot)$  приближается линейной функцией  $\alpha$  (плюс порог). Это математическое ожидание будет выпукло вниз по  $\alpha$ .

**Пример: t-распределение Стьюдента.** Для t-распределения Стьюдента с хвостовым  $\alpha$  — это «сложная» медленно меняющаяся функция, которая широко используется для симметричных степенных хвостов в финансовой математике, — половинное среднее, или среднее одностороннего распределения (т. е. с носителем  $\mathbb{R}^+$ ) принимает вид

$$2\nu(\alpha) = 2 \frac{\sqrt{\alpha} \Gamma \frac{\alpha+1}{2}}{\sqrt{\pi} \Gamma \frac{\alpha}{2}} \approx \alpha \frac{1 + \ln 4}{\pi},$$

где  $\Gamma(\cdot)$  — гамма-функция.

### 18.3. СУММЫ СТЕПЕННЫХ ЗАКОНОВ

Поскольку мы изучаем здесь сходимость к устойчивому распределению, мы рассматриваем ситуации  $1 < \alpha < 2$ , так что  $p = 1$ , и мы будем заботиться только о среднем<sup>1</sup>.

Мы видим, что выпуклость вниз среднего инвариантна к суммированию случайных величин, следующих распределению Парето, как рассмотренная выше случайная величина  $X$ . Устойчивое распределение имеет среднее, и при стандартном наборе параметров оно выглядит не зависящим от  $\alpha$ , тогда как на самом деле зависимость есть.

Пусть случайная величина  $Y$  следует распределению Парето с плотностью  $f(y) \triangleq \alpha \lambda^\alpha y^{-\alpha-1}$ ,  $y \geq \lambda > 0$  и показатель хвоста  $1 < \alpha < 2$ . Пусть далее  $Y_1, Y_2, \dots, Y_n$  — идентичные независимые копии  $Y$ . Пусть  $\chi(t)$  — характеристическая функция для  $f(y)$ ; она имеет вид  $\chi(t) = \alpha(-it)^\alpha \Gamma(-\alpha, -it)$ , где  $\Gamma(\cdot, \cdot)$  — неполная гамма-функция. Из характеристической функции можно получить среднее  $n$  слагаемых  $\frac{1}{n}(Y_1 + Y_2 + \dots + Y_n)$ , а именно как  $\left( \chi\left(\frac{t}{n}\right) \right)^n$ . Взяв первую производную

$$\begin{aligned} -i \frac{\partial}{\partial t} \left( \chi\left(\frac{t}{n}\right) \right)^n &= \\ &= (-i)^{\alpha(n-1)} n^{1-\alpha n} \alpha^n \lambda^{\alpha(n-1)} t^{\alpha(n-1)-1} \left( \Gamma\left(-\alpha, -\frac{it\lambda}{n}\right) \right)^{(n-1)} \left( (-i)^\alpha \alpha \lambda^\alpha t^\alpha \Gamma\left(-\alpha, -\frac{it\lambda}{n}\right) - n^\alpha e^{\frac{it\lambda}{n}} \right) \end{aligned} \quad (18.9)$$

<sup>1</sup> Поскольку моменты порядка  $p = 2, 3, \dots$ , т. е. дисперсия и высшие моменты, заведомо бесконечны.

и предел

$$\lim_{n \rightarrow \infty} -i \frac{\partial}{\partial t} \left( \chi \left( \frac{t}{n} \right) \right)^n \Big|_{t=0} = \lambda \frac{\alpha}{\alpha - 1}, \quad (18.10)$$

видим, что сходящееся асимптотическое распределение для среднего будет иметь средним коэффициент масштаба, умноженный на  $\frac{\alpha}{\alpha - 1}$ , и не будет зависеть от  $n$ .

Пусть  $\chi^S(t)$  — характеристическая функция соответствующего устойчивого распределения  $S_{\alpha, \beta, \mu, \sigma}$ , которому следует сумма бесконечного числа копий случайной величины  $Y$ . Согласно теореме Леви о непрерывности следующие два утверждения равносильны:

- $\frac{1}{n} \sum_{i \leq n} Y_i \xrightarrow{D} S$ , которая следует распределению  $S_{\alpha, \beta, \mu, \sigma}$  (где  $\xrightarrow{D}$  значит сходимость по распределению) и
- $\chi^S(t) = \lim_{n \rightarrow \infty} \left( \chi \left( \frac{t}{n} \right) \right)^n$ .

Таким образом, мы имеем дело со стандартным результатом [272], [209] точного выражения для сумм Парето [269], заменяя стандартное  $\mu$  на среднее, полученное выше:

$$\chi^S(t) = \exp \left( i \left( \lambda \frac{\alpha t}{\alpha - 1} + |t|^\alpha \left( i + \beta \operatorname{tg} \frac{\pi \alpha}{2} \operatorname{sgn} t \right) \right) \right).$$

#### 8.4. АСИММЕТРИЧНЫЕ УСТОЙЧИВЫЕ РАСПРЕДЕЛЕНИЯ

Пользуясь соображениями симметрии, мы можем убедиться, что отражение распределения из подклассов  $\mathfrak{P}_1$  и  $\mathfrak{P}_2$  вокруг  $y_g$  дает случайную величину с отрицательным средним и, следовательно, деградацию от стохастической  $\alpha$ .

Центральными вопросами становятся следующие.

##### Замечание 17: Сохранение асимметрии

Нормированная сумма в классе однохвостых распределений  $\mathfrak{P}_1$  с математическим ожиданием, зависящим от формы в (18.8), должна сходиться по распределению к асимметричному устойчивому распределению  $S_{\alpha, \beta, \mu, 1}$  с  $\beta \neq 0$ .

##### Замечание 18

Пусть  $Y'$  — это  $Y$  при стохастической  $\alpha$ , сохраняющей среднее. Влияние выпуклости вниз принимает вид

$$\operatorname{sgn}(\mathbb{E}Y' - \mathbb{E}Y) = \operatorname{sgn}(\beta).$$

Схема доказательства такая. Рассмотрим две медленно меняющиеся функции из (23.1), по одной с каждой стороны от хвостов. Их комбинация

$$L(y) = 1_{y < y_g} L^-(y) + 1_{y \geq y_g} L^+(y) = \begin{cases} L^+(y), y \geq y_g \\ L^-(y), y < y_g \end{cases}$$

позволяет задать разные пределы

$$\begin{cases} \lim_{y \rightarrow \infty} L^+(y) = c \\ \lim_{y \rightarrow -\infty} L^-(y) = d \end{cases}.$$

Как показано в [209],

$$\text{если } \begin{cases} \mathbb{P}(Y > y) \sim cy^{-\alpha}, y \rightarrow +\infty \\ \mathbb{P}(Y < y) \sim d|y|^{-\alpha}, y \rightarrow -\infty \end{cases},$$

то  $Y$  сходится по распределению к  $S_{\alpha, \beta, \mu, 1}$  с коэффициентом  $\beta = \frac{c-d}{c+d}$ .

Мы можем показать, что среднее представимо как  $(\lambda_+ - \lambda_-) \frac{\alpha}{\alpha-1}$ , где

$$\lambda_+ \geq \lambda_-, \text{ если } \int_{y_g}^{\infty} L^+(y) dy \geq \int_{-\infty}^{y_g} L^-(y) dy.$$

## 18.5. РАСПРЕДЕЛЕНИЕ ПАРЕТО С ЛОГНОРМАЛЬНЫМ РАСПРЕДЕЛЕНИЕМ АЛЬФЫ

Теперь допустим, что  $\alpha$  следует сдвинутому логнормальному распределению со средним  $\alpha_0$  и минимальным значением  $b$ , то есть  $\alpha - b$  следует логнормальному распределению  $\mathcal{L}(\ln \alpha_0 - \frac{\sigma^2}{2}, \sigma)$ . Параметр  $b$  позволяет нам работать с нижним пределом показателя хвоста, чтобы удовлетворить требуемому конечному среднему. Мы знаем, что показатель хвоста в конечном итоге сойдется к  $b$ , но этот процесс может оказаться медленным.

### Предложение 18.3

Предполагая для стохастизированной случайной величины  $X'$  конечное математическое ожидание и для сдвинутого показателя хвоста  $\alpha - b$  логнормальное распределение по закону  $\mathcal{L}(\ln \alpha_0 - \frac{\sigma^2}{2}, \sigma)$ , используя  $b \geq 1$  как минимальное значение для  $\alpha$  и коэффициент масштаба  $\lambda$ , получаем

$$\mathbb{E}Y' = \mathbb{E}Y + \lambda \frac{e^{\sigma^2} - b}{\alpha_0 - b}. \quad (18.11)$$

Нам нужно  $b \geq 1$ , чтобы избежать проблем с бесконечным математическим ожиданием.

Пусть  $\varphi(y; \alpha)$  — плотность при стохастическом показателе хвоста. При  $\alpha > 0$ ,  $\alpha_0 > b$ ,  $b \geq 1$ ,  $\sigma > 0$ ,  $Y \geq \lambda > 0$

$$\begin{aligned} \mathbb{E}Y &= \int_b^{\infty} \int_L^{\infty} y \varphi(y; \alpha) dy d\alpha = \\ &= \int_b^{\infty} \lambda \frac{\alpha}{\alpha-1} \frac{1}{\sqrt{2\pi\sigma}(\alpha-b)} \exp \frac{-(\ln(\alpha-b) - \ln(\alpha_0-b) + \sigma^2/2)^2}{2\sigma^2} d\alpha = \lambda \frac{\alpha_0 + e^{\sigma^2} - b}{\alpha_0 - b}. \end{aligned} \quad (18.12)$$

### Аппроксимация плотности

При  $b = 1$  (это нижняя граница для  $b$ ) получаем плотность при стохастической  $\alpha$ :

$$\varphi(y; \alpha_0, \sigma) = \lim_{k \rightarrow \infty} \frac{1}{y^2} \sum_{i=0}^k \frac{1}{i!} L(\alpha_0 - 1)^i e^{\frac{1}{2}i(i-1)\sigma^2} (\ln \lambda - \ln y)^{i-1} (i + \ln \lambda - \ln y). \quad (18.13)$$

Этот результат получен разложением  $\alpha$  вблизи ее нижнего предела,  $b$  (для простоты мы взяли  $b = 1$ ), с последующим интегрированием каждого слагаемого.

## 18.6. РАСПРЕДЕЛЕНИЕ ПАРЕТО С ГАММА-РАСПРЕДЕЛЕНИЕМ АЛЬФЫ

### Предложение 18.4

Предполагая конечное математическое ожидание для случайной величины  $X'$  с коэффициентом масштаба  $\lambda$  и стохастизированным показателем  $\alpha$  таким, что все значения  $\alpha$  превышают 1 и сдвинутая случайная величина  $\alpha - 1$  следует гамма-распределению с плотностью  $\varphi(\cdot)$ , средним  $\alpha_0 - 1$  и дисперсией  $s^2$ , получаем

$$\mathbb{E}X' = \mathbb{E}X + \lambda \frac{s^2}{(\alpha_0 - 1)(\alpha_0 - s - 1)(\alpha_0 + s - 1)}. \quad (18.14)$$

*Доказательство.*

$$\varphi(\alpha - 1) = \frac{e^{\frac{(\alpha-1)(\alpha_0-1)}{s^2}} \left( \frac{s^2}{(\alpha-1)(\alpha_0-1)} \right)^{-\frac{(\alpha_0-1)^2}{s^2}}}{(\alpha-1)\Gamma\left(\frac{(\alpha_0-1)^2}{s^2}\right)}, \quad \alpha > 1^1. \quad (18.15)$$

$$\begin{aligned} \mathbb{E}X' &= \int_{\lambda}^{\infty} dx \, x \int_1^{\infty} \alpha \lambda^{\alpha} x^{-\alpha-1} \varphi(\alpha) d\alpha^2 = \lambda \int_1^{\infty} \frac{\alpha}{(\alpha-1)} = \frac{e^{\frac{(\alpha-1)(\alpha_0-1)}{s^2}} \left( \frac{s^2}{(\alpha-1)(\alpha_0-1)} \right)^{-\frac{(\alpha_0-1)^2}{s^2}}}{(\alpha-1)\Gamma\left(\frac{(\alpha_0-1)^2}{s^2}\right)} d\alpha^3 = \\ &= \frac{\lambda}{2} \left( \frac{1}{\alpha_0 + s - 1} + \frac{1}{\alpha_0 - s - 1} + 2 \right)^4. \end{aligned} \quad (18.16)$$

1 В самом деле, в стандартных обозначениях, если  $X \sim \text{Gamma}(a, b^{-1})$ , т. е. случайная величина  $X \in \mathbb{R}^+$  следует гамма-распределению с коэффициентом формы  $a > 0$  и коэффициентом масштаба  $b > 0$ , то плотность вероятности  $\varphi_X(X) = \frac{e^{-bx} (xb)^a}{x\Gamma(a)}$ , среднее  $\frac{a}{b}$ , дисперсия  $\frac{a}{b^2}$ . В нашем случае альфа минус 1  $\sim \text{Gamma}\left(\frac{(\alpha_0-1)^2}{s^2}, \frac{s^2}{\alpha_0-1}\right)$ ,

случайная величина альфа минус 1 следует гамма-распределению с коэффициентом формы  $\frac{(\alpha_0-1)^2}{s^2}$  и коэф-

фициентом масштаба  $\frac{\alpha_0-1}{s^2}$ , и плотность вероятности  $\varphi(\alpha - 1) = \frac{e^{-\frac{\alpha_0-1}{s^2}(\alpha-1)} \left( \frac{s^2}{(\alpha-1)\frac{\alpha_0-1}{s^2}} \right)^{\frac{(\alpha_0-1)^2}{s^2}}}{(\alpha-1)\Gamma\left(\frac{(\alpha_0-1)^2}{s^2}\right)}$ , среднее  $\alpha_0 - 1$ ,

дисперсия  $s^2$ .

2 В самом деле, случайная величина  $X'[\lambda, \infty)$  следует распределению Парето и имеет плотность  $f_{X'}(x; \alpha) = \alpha \lambda^{\alpha} x^{-\alpha-1} \varphi(\alpha)$ .

3 В самом деле,  $\int_{\lambda}^{\infty} dx \, x \alpha \lambda^{\alpha} x^{-\alpha-1} = \lambda^{\alpha} \frac{x^{-\alpha+1}}{-\alpha+1} \Big|_{x=\lambda}^{x=\infty} = \lambda \frac{\alpha}{\alpha-1}$ .

4 В самом деле, без стохастизации показателя хвоста случайная величина  $X \in [\lambda, \infty)$  следует распределению Парето с плотностью  $f_X(x) = \alpha_0 \lambda^{\alpha_0} x^{-\alpha_0-1}$  и математическим ожиданием  $\mathbb{E}X = \int_{\lambda}^{\infty} dx \, x f_X(x) =$

## 18.7. ОГРАНИЧЕННЫЙ СТЕПЕННОЙ ЗАКОН В РАБОТЕ ЧИРИЛЛО И ТАЛЕБА (2016)

В таких исследованиях, как [46] и [45], используются ограниченные степенные законы в применении соответственно к уровню острых конфликтов и к операционному риску. Благодаря верхней границе случайная величина  $Z$  имеет конечное ожидание даже при  $\alpha < 1$ .

Предложенные методы предусматривали следующее гладкое преобразование случайной величины: мы начинаем с  $Z \in [L, H)$ ,  $L > 0$  и растягиваем до  $X \in [L, \infty)$ , для которой будет корректно предположить распределение по степенному закону.

Для этого используем гладкое логарифмическое преобразование случайной величины:

$$X = \varphi(Z) = L - H \ln \frac{H-Z}{H-L}$$

и распределение плотности случайной величины  $X$  по степенному закону

$$f(x) = \frac{\left(\frac{x-L}{\alpha\sigma} + 1\right)^{-\alpha-1}}{\sigma}.$$

Так мы получаем распределение случайной величины  $Z$ , имеющее конечное математическое ожидание при любой положительной  $\alpha$ , по которой вторая производная

$$\begin{aligned} \frac{\partial^2 \mathbb{E}Z}{\partial \alpha^2} = & \frac{H-L}{H^3} \left( e^{\frac{\alpha\sigma}{H}} \left( 2H^3 G_{3,4}^{4,0} \left( \frac{\alpha\sigma}{H} \middle| \alpha+1, \alpha+1, \alpha+1 \right) - \right. \right. \\ & - 2H^2 (H+\sigma) G_{2,3}^{3,0} \left( \frac{\alpha\sigma}{H} \middle| \alpha+1, \alpha+1 \right) + \\ & \left. \left. + \sigma(\alpha\sigma^2 + (\alpha+1)H^2 + 2\alpha H\sigma) E_\alpha \frac{\alpha\sigma}{H} \right) - H\sigma(H+\sigma) \right) \end{aligned} \quad (18.17)$$

---

$= \int_L^\infty dx \alpha_0 \lambda^{\alpha_0} x^{-\alpha_0} = \lambda^{\alpha_0} \frac{x^{-\alpha_0+1}}{-\alpha_0+1} \Big|_{x=L}^{x=\infty} = \frac{\lambda \alpha_0}{\alpha_0+1}$ . Итак, стохастическая прибавка к математическому ожиданию

$$\begin{aligned} \mathbb{E}X - \mathbb{E}X &= \frac{\lambda}{2} \left( \frac{1}{\alpha_0+s-1} + \frac{1}{\alpha_0-s-1} + 2 \right) - \frac{\lambda \alpha_0}{\alpha_0-1} = \lambda \left( \frac{\alpha_0-1}{(\alpha_0+s-1)(\alpha_0-s-1)} + 1 - \frac{\alpha_0}{\alpha_0-1} \right) = \lambda \left( \frac{\alpha_0-1}{(\alpha_0+s+1)(\alpha_0-s-1)} - \frac{1}{\alpha_0-1} \right) = \\ &= \lambda \frac{(\alpha_0-1)^2 - (\alpha_0+s-1)(\alpha_0-s-1)}{(\alpha_0-1)(\alpha_0+s-1)(\alpha_0-s-1)} = \lambda \frac{s^2}{(\alpha_0-1)(\alpha_0+s-1)(\alpha_0-s-1)}. \end{aligned}$$

- 1 Случайная величина  $X \in [L, \infty)$  следует обобщенному распределению Парето (*generalized Pareto distribution*, GPD) с коэффициентом масштаба  $\sigma > 0$ , коэффициентом формы, он же показатель степени хвоста  $\alpha > 0$  и коэффициентом положения  $L \in \mathbb{R}$ , если функция распределения имеет вид степенного закона  $F_X(X) = 1 - \left(\frac{x-L}{\alpha\sigma} + 1\right)^{-\alpha}$ .

Плотность вероятности  $f_X(X) = \frac{1}{\sigma} \left(\frac{x-L}{\alpha\sigma} + 1\right)^{-\alpha-1}$  односторонняя с максимумом на краю  $x=L$ ; среднее  $\infty$  при

$\alpha \leq 1$ , иначе  $L + \frac{\alpha\sigma}{\alpha-1}$ ; дисперсия  $\infty$  при  $\alpha \leq 2$ , иначе  $\frac{\alpha\sigma^2}{(\alpha-1)^2(\alpha-2)}$ .

очевидно положительна в диапазоне возмущений, который рассматривался в [46]<sup>1,2</sup>. При тех значениях  $\alpha$ , в районе  $\frac{1}{2}$ , выпуклость вниз математического ожидания по  $\alpha$  велика и, соответственно, смещение оценки ярко выражено.

Такая выпуклость вниз имеет практические следствия. Исторические данные о числе жертв острых конфликтов за два тысячелетия в принципе ненадежны [46]. Из погрешности данных вытекает неточность в оценке показателя хвоста, которую надо как-то учесть в вычислениях. Изложенные выше результаты показывают, что неопределенность в  $\alpha$  сместит «истинное» статистическое среднее (среднее согласно процедуре, а не наивное выборочное среднее) скорее вверх, чем вниз; тем самым подтверждается, что учет еще и этой неопределенности заставляет повысить оценку числа жертв<sup>3</sup>.

### 18.8. ДОПОЛНИТЕЛЬНЫЕ КОММЕНТАРИИ

Оценка среднего и ожидаемых потерь может смещаться из-за неопределенности в показателе хвоста при анализе недостаточных, ненадежных или даже незащищенных от подделки данных.

Эти результаты можно расширить со статистических выводов на моделирование процессов, таких как сложный пуассоновский процесс с иерархией степенных законов [217] (по которым распределены время наступления пуассонова события и скачок) или процесс Леви. Последний можно анализировать как последовательность распределений с так называемой «выборкой по уровням», или дискретизацией процесса [50]. Поскольку математическое ожидание суммы скачков равно сумме математических ожиданий, появляется та же выпуклость вниз, которую мы получили в формуле (18.8).

### 18.9. БЛАГОДАРНОСТИ

Марко Авельянеде, Роберту Фрею, Рафаэлю Дуади, Паскуале Чирилло.

<sup>1</sup>  $G_{3,4}^{4,0}\left(\frac{\alpha\sigma}{H}\left|\alpha+1, \alpha+1, \alpha+1, \right.\right)_{1, \alpha, \alpha, \alpha,}$  — это G-функция Мейера. — Прим. автора.

<sup>2</sup> G-функция Мейера определяется через интеграл Меллина — Барнса на комплексной плоскости как

$$G_{p,q}^{m,n}z \left| \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \right. \triangleq \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^m \Gamma(b_j - s) \prod_{j=1}^n \Gamma(1 - a_j + s)}{\prod_{j=m+1}^q \Gamma(1 - b_j + s) \prod_{j=n+1}^p \Gamma(a_j - s)} z^s ds.$$

<sup>3</sup> Особенно в далеком прошлом, если в нем неопределенность особенно велика. Таким образом, максимально правдоподобное число жертв в далеком прошлом надо повысить; это прольет воду на мельницу гипотезы Пинкера и пр. о большом числе жертв в прошлом и тенденции снижения в наше время.

## МЕТА-РАСПРЕДЕЛЕНИЕ $p$ -ЗНАЧЕНИЙ И $p$ -ХАКИНГ<sup>†</sup>

Мы представляем точное распределение вероятностей  $p$ -значений (мета-распределение) для ансамблей статистически одинаковых явлений, а также распределение минимального  $p$ -значения среди  $m$  независимых тестов. Мы выводим распределение для выборок малого размера  $2 < n \leq n^* \approx 30$ , а также предельное распределение для большого  $n$ . Мы рассматриваем свойства «силы» теста через распределение его инверсии при данном  $p$ -значении и параметрах.

Доказано, что  $p$ -значения крайне асимметричны и волатильны, независимо от размера выборки  $n$ , и сильно варьируют при повторении с теми же протоколами в отношении идентичных копий данного стохастического процесса; из-за этой волатильности минимальное  $p$ -значение существенно смещено от «истинного». Доказано, что задание силы мало помогает без существенного увеличения размера выборки или уменьшения  $p$ -значения на порядок величины.

Формулы позволяют исследовать устойчивость воспроизведения результатов, « $p$ -хакинг»<sup>1</sup> и другие аспекты метаанализа — включая метараспределение результатов  $p$ -хакинга.

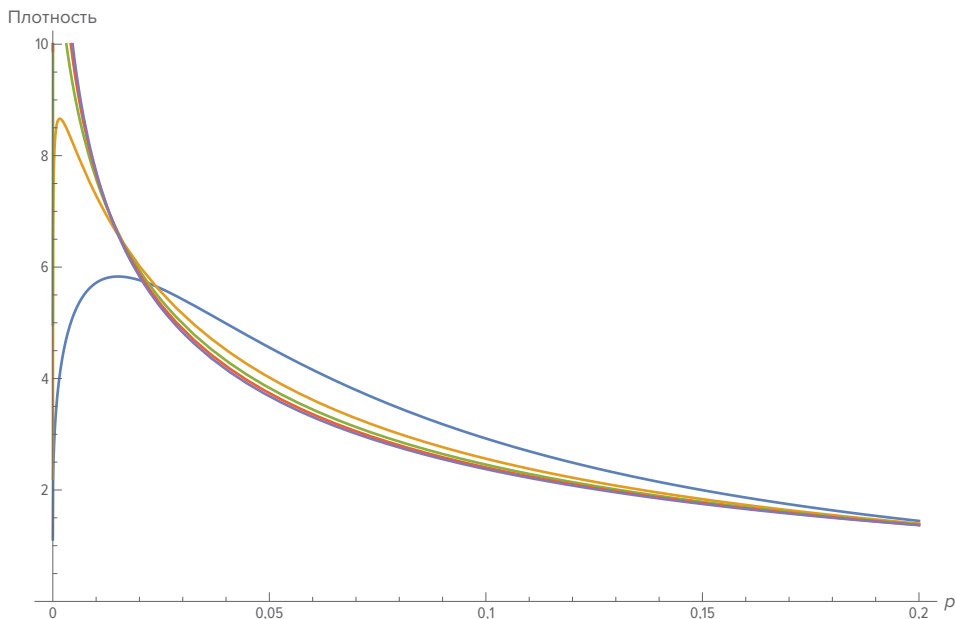
С теоретико-вероятностной точки зрения ни  $p$ -значение 0,05, ни «статистическая мощность»<sup>2</sup> 0,9 ничего не дают.

Если мы знаем «истинное»  $p$ -значение  $p_s$ , как будут выглядеть его реализации в различных попытках на идентичных копиях одного и того же статистического явления? Под истинным значением  $p_s$  мы подразумеваем ожидаемое значение согласно закону

1 Тестирование втайне от публики большого числа маловероятных гипотез в надежде на удачу и попытка затем создать впечатление статистической значимости данных, подтверждающих удачную гипотезу.

2 Англ. *statistical power* — вероятность, с которой некоторая процедура тестирования некоторой гипотезы защищена от ложноотрицательного заключения (от т. н. ошибки второго рода). Когда требуют мощность  $\geq 0,9$  от научного исследования, проверяющего некоторую гипотезу, тем самым требуют, чтобы при данном методе исследования риск отвергнуть верную гипотезу не превышал 0,1.





**Рисунок 19.1:** Различные значения по формуле 19.1, показывающие сходимость к предельному распределению  $p$ -значения

больших чисел на ансамбле  $m$  возможных выборок исследуемого явления, то есть  $\frac{1}{m} \sum_{i=1}^m p_i \xrightarrow{P} p_s$  (где  $\xrightarrow{P}$  означает сходимость по вероятности). Аналогичный анализ сходимости можно предложить в отношении соответствующей «истинной медианы»  $p_m$ . Главный результат статьи в том, что распределение выборок малого  $n$  можно представить более-менее в явном виде (с использованием обратных специальных функций), как и простое предельное распределение для большого  $n$ , используя единственный параметр — медианное  $p$ -значение,  $p_m$ . Нам не удалось получить явное выражение для  $p_s$ , но мы обошлись медианой. В итоге доступна простая формула для распределения минимального  $p$ -значения, позволяющая понять смещения в научных исследованиях.

Оказалось, как видно на Рисунке 19.2, что распределение крайне асимметрично, а именно перекошено вправо, так что 75 % реализаций «истинного»  $p$ -значения 0,05 будут оценены  $< 0,05$  (и если по протоколу положено признавать успешное наблюдение некоторого явления на основании полученных данных, когда вероятность, что они всего лишь игра случая, ниже 0,05, то пограничные по надежности попытки пронаблюдать явление будут признаны вдвое чаще, чем отвергнуты). Что еще хуже, 60 % случаев с истинным  $p$ -значением 0,12 будут оценены ниже 0,05.

Хотя носитель распределения компактный, оно ведет себя как крайне жирнохвостое. Так, при наблюдаемом  $p$ -значении 0,02 «истинное»  $p$ -значение, скорее всего,  $> 0,1$  (и вполне может приблизиться к 0,2), притом что среднеква-

дратическое отклонение  $> 0,2$  (то есть получился плюс-минус слон) и среднее отклонение около  $0,35$  (то есть получился слон плюс-минус два слона). Из-за суровой асимметрии оценка дисперсии в  $\mathcal{L}^1$  и  $\mathcal{L}^2$  (и в нормах высших порядков) сильно варьирует в зависимости от  $p_s$ ; таким образом, среднеквадратическое отклонение не является пропорциональным: когда по выборке  $p$ -значение  $0,01$ , есть существенная вероятность, что истинное значение  $> 0,3$ .

**Итак, рассуждая о  $p$ -значениях, мы не знаем, о чем говорим.**

Прошлые попытки построить явное метараспределение можно видеть в публикациях [132] и [208]; там рассматривается иерархия гауссовых распределений и система параметров не столь лаконична. Серьезность проблемы со значимостью так называемой «статистической значимости» обсуждалась в [105]; вносились байесовские поправки, [138], и по результатам рекомендовалось такое же ужесточение стандарта, до  $p$ -значений  $\approx 0,01$ . Но серьезность асимметрии распределения  $p$ -значений становится очевидна, только если увидеть метараспределение.

Ради удобства обозначений мы используем  $n$  как размер выборки в заданном исследовании и  $m$  как число испытаний, приводящих к  $p$ -значению.

## 19.1. ДОКАЗАТЕЛЬСТВА И ВЫВОД ФОРМУЛ

### Предложение 19.1

Пусть случайная величина  $P \in [0, 1]$  отвечает полученному из выборки однохвостому  $p$ -значению согласно паре  $t$ -статистика (с неизвестной дисперсией) — медианное значение  $\mathbb{M}(P) = p_M \in [0, 1]$ , полученное по выборке размером  $n$ . Распределение по ансамблю копий выборки с одинаковой статистикой имеет функцию плотности

$$\varphi(p; p_M) = \begin{cases} \varphi(p; p_M)_L, & p < \frac{1}{2}, \\ \varphi(p; p_M)_H, & p > \frac{1}{2} \end{cases},$$

$$\varphi(p; p_M)_L = \lambda_p^{-\frac{n+1}{2}} \left( \frac{\lambda_p(1-\lambda_{p_M})}{1-(1-\lambda_p)\lambda_{p_M} - 2\sqrt{(1-\lambda_p)\lambda_p}\sqrt{(1-\lambda_{p_M})\lambda_{p_M}}} \right)^{\frac{1}{2}}$$

$$\left( \frac{1}{\lambda_p} - \frac{2\sqrt{(1-\lambda_p)}\sqrt{\lambda_{p_M}}}{\sqrt{\lambda_p}\sqrt{(1-\lambda_{p_M})}} + \frac{1}{1-\lambda_{p_M}} - 1 \right)^{-\frac{n}{2}},$$

$$\varphi(p; p_M)_H = (1-\lambda'_p)^{-\frac{n+1}{2}} \left( \frac{(1-\lambda'_p)(1-\lambda_{p_M})}{1-\lambda'_p\lambda_{p_M} + 2\sqrt{(1-\lambda'_p)\lambda_p}\sqrt{(1-\lambda_{p_M})\lambda_{p_M}}} \right)^{\frac{n+1}{2}}, \quad (19.1)$$

где  $\lambda_p = \Gamma_{2p}^{-1}\left(\frac{n}{2}, \frac{1}{2}\right)$ ,  $\lambda_{p_M} = \Gamma_{1-2p_M}^{-1}\left(\frac{1}{2}, \frac{n}{2}\right)$ ,  $\lambda'_p = \Gamma_{2p-1}^{-1}\left(\frac{1}{2}, \frac{n}{2}\right)$  и  $\Gamma_{(\cdot)}^{-1}(\cdot, \cdot)$  — обратная регуляризованная неполная бета-функция.

**Замечание 19**

При  $p = \frac{1}{2}$  распределение не существует в теории, хотя на практике возможно, и эту трудность можно обойти с помощью последовательности  $p_{M_k} = \frac{1}{2} \pm \frac{1}{k}$ , как показано на Рисунке 19.3, где график сходится к равномерному распределению на  $[0, 1]$ . Заметим также, что «нулевой гипотезой» называют множество меры 0.

*Доказательство.* Пусть  $Z$  — случайная нормированная величина с реализациями  $\zeta$  из вектора  $\vec{v}$ , содержащего  $n$  реализаций и имеющего выборочное среднее  $m_v$  и выборочное среднееквадратичное отклонение  $s_v$ ,  $\bar{\zeta} = \frac{m_v - m_h}{\frac{s_v}{\sqrt{n}}}$ , где  $m_h$  — это уровень,

по которому выполняется тестирование; поэтому, предполагая следование t-распределению Стьюдента с  $n$  степенями свободы и, что важно, предположительно обеспечивающему среднее  $\bar{\zeta}$ , имеем

$$f(\zeta; \bar{\zeta}) = \frac{\left( \frac{n}{(\bar{\zeta} - \zeta)^2 + n} \right)^{\frac{n+1}{2}}}{\sqrt{n} B\left(\frac{n}{2}, \frac{1}{2}\right)},$$

где  $B(., .)$  — стандартная бета-функция. Пусть  $g(.)$  — односторонняя функция выживания t-распределения Стьюдента с нулевым средним и  $n$  степенями свободы:

$$g(\zeta) = \mathbb{P}(Z > \zeta) = \begin{cases} \frac{1}{2} I_{\frac{n}{\zeta^2 + n}}\left(\frac{n}{2}, \frac{1}{2}\right), & \zeta \geq 0 \\ \frac{1}{2} \left( I_{\frac{\zeta^2}{\zeta^2 + n}}\left(\frac{1}{2}, \frac{n}{2}\right) + 1 \right), & \zeta < 0 \end{cases},$$

где  $I_{(.)}(., .)$  — неполная бета-функция.

Рассмотрим теперь распределение  $g \circ f(\zeta)$ . Пользуясь тем, что функция  $g(.)$  — заведомо борелевская, и обозначая  $p$  случайную величину вероятность, стандартным преобразованием получаем:

$$\varphi(p; \zeta) = \frac{f(g^{(-1)}(p))}{|g'(g^{(-1)}(p))|}.$$

Благодаря симметричности  $Z$  мы можем преобразовать  $\zeta$  в соответствующую медианную вероятность выживания. Поскольку ровно по половине наблюдений попадает с каждой стороны от  $\bar{\zeta}$ , мы можем гарантировать, что преобразование сохранит медиану:  $g(\bar{\zeta}) = \frac{1}{2}$ , значит,  $\varphi(p_M, .) = \frac{1}{2}$ . Поэтому в итоге случаи разбиваются на поло-

жительные  $\left\{ \zeta \left| \frac{1}{2} I_{\frac{n}{\zeta^2 + n}}\left(\frac{1}{2}, \frac{n}{2}\right) = p_M \right. \right\}$  и отрицательные  $\left\{ \zeta \left| \frac{1}{2} \left( I_{\frac{\zeta^2}{\zeta^2 + n}}\left(\frac{1}{2}, \frac{n}{2}\right) + 1 \right) = p_M \right. \right\}$ . Заме-

ной получаем формулу (19.1), и Предложение 19.1 доказано. ■

Заметим, что  $n$  не повышает значимость:  $p$ -значения вычисляются из нормированных величин (из чего следует универсальность метараспределения), и большое  $n$  соответствует повышенной сходимости к гауссиане. При больших  $n$  мы можем доказать следующее предложение.

### Предложение 19.2

При тех же предположениях, что выше, для  $\varphi(\cdot)$  предельное распределение

$$\lim_{n \rightarrow \infty} \varphi(p; p_M) = \exp(-\operatorname{erfc}^{-1} 2p_M (\operatorname{erfc}^{-1} 2p_M - 2\operatorname{erfc}^{-1} 2p)), \quad (19.2)$$

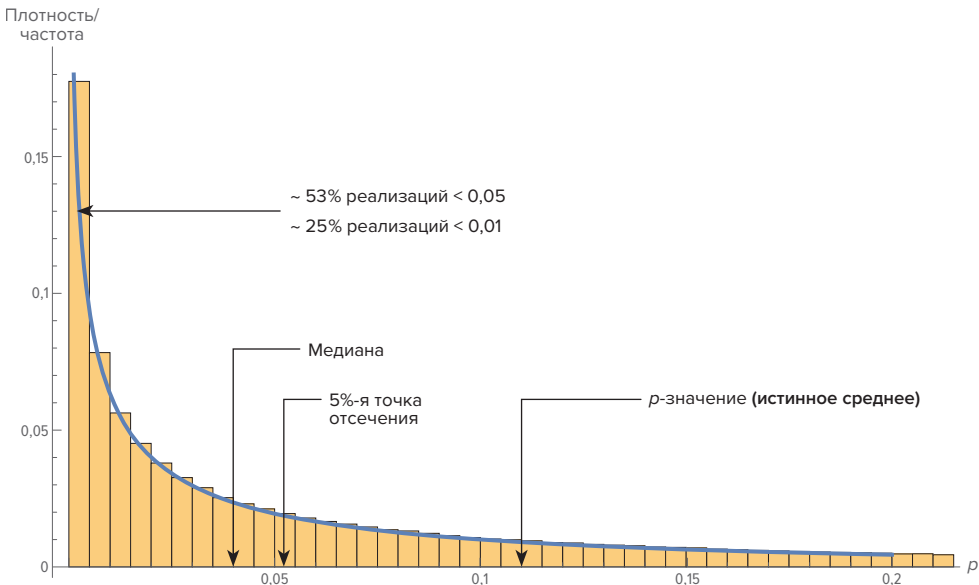
где  $\operatorname{erfc}(\cdot)$  — дополнительная функция ошибок и  $\operatorname{erfc}^{-1}(\cdot)$  — обратная функция.

Предельная кумулятивная функция распределения  $\Phi(\cdot)$ :

$$\Phi(k; p_M) = \frac{1}{2} \operatorname{erfc}(\operatorname{erf}^{-1}(1-2k) - \operatorname{erf}^{-1}(1-2p_M)). \quad (19.3)$$

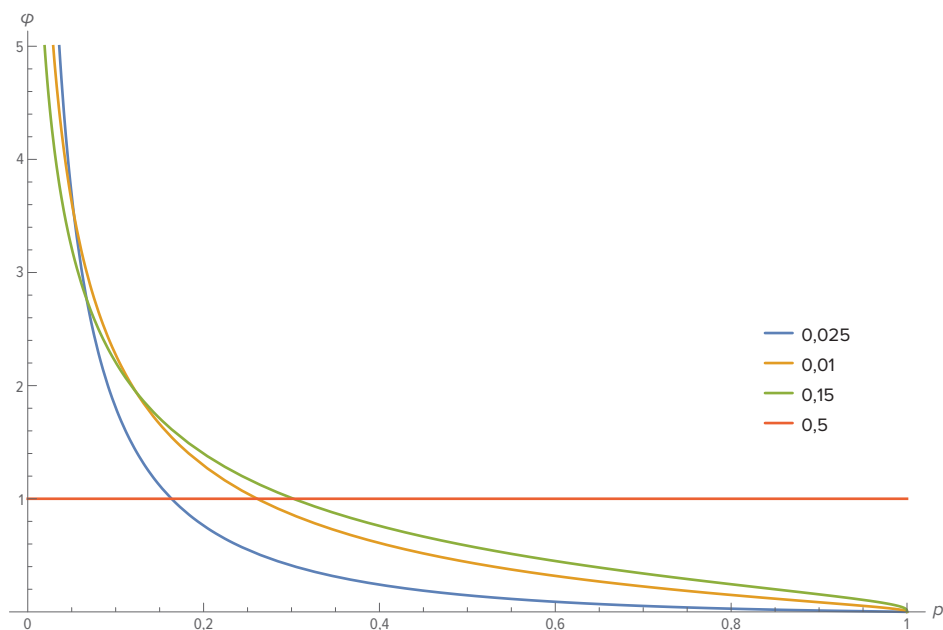
*Доказательство.* При большом  $n$  распределение  $Z = \frac{m_v}{\frac{s_v}{\sqrt{n}}}$  становится гауссовым,

и однохвостая функция выживания  $g(\cdot) = \frac{1}{2} \operatorname{erfc} \frac{\zeta}{\sqrt{2}}$  стремится  $\zeta(p) \rightarrow \sqrt{2} \operatorname{erfc}^{-1} p$ . ■



**Рисунок 19.2:** Распределение вероятностей однохвостого  $p$ -значения при математическом ожидании 0,11, сгенерированное как методом Монте-Карло (гистограмма), так и аналитически (график  $\varphi(\cdot)$  непрерывной линией). Мы извлекли все возможные подвыборки из ансамбля с заданными свойствами. Грубая асимметрия распределения делает среднее значение существенно выше большинства наблюдений, создавая многочисленные иллюзии «статистической значимости»

Это предельное распределение применимо к парам с тестами, имеющими известную или предположительную выборочную дисперсию, поскольку тест становится гауссовой случайной величиной, что эквивалентно сходимости  $t$ -теста Стьюдента к гауссиане при большом  $n$ .



**Рисунок 19.3:** Распределение вероятностей  $p$ -значения при различных значениях  $p_M$ . Видно, что при  $p_M = \frac{1}{2}$  возникает равномерное распределение

### Замечание 20

При значениях  $p$ , близких к 0,  $\phi$  в формуле (19.2) удобно вычислять так:

$$\begin{aligned} \phi(p; p_M) &= \sqrt{2\pi} p_M \sqrt{-\ln 2\pi p_M^2} \\ &e^{\sqrt{-\ln(-2\pi \ln 2\pi p^2) - 2\ln p}} \sqrt{-\ln(-2\pi \ln 2\pi p_M^2) - 2\ln p_M} + O(p^2). \end{aligned} \quad (19.4)$$

Эта аппроксимация работает точнее в диапазоне значений  $0 < p < \frac{1}{2\pi}$  — то есть самых употребительных.

Отсюда можем получить численные результаты для свертки  $\phi$  при преобразовании Фурье и аналогичных методах.

Можем получить распределение минимального  $p$ -значения при  $m$  испытаниях статистически одинаковых ситуаций и так получить представление о « $p$ -хакинге»,

который определяют как попытку исследователя получить как можно более низкие  $p$ -значения для большого числа экспериментов или повторять попытки, пока один из тестов не покажет статистическую значимость.

### Предложение 19.3

Распределение минимума по  $m$  наблюдениям статистически одинаковых  $p$ -значений становится (при предельном распределении из Предложения 19.2) таким:

$$\varphi_m(p; p_M) = m e^{(\operatorname{erfc}^{-1} 2 p_M)(2 \operatorname{erfc}^{-1} 2 p - \operatorname{erfc}^{-1} 2 p_M)} \left( 1 - \frac{1}{2} \operatorname{erfc}(\operatorname{erfc}^{-1} 2 p - \operatorname{erfc}^{-1} 2 p_M) \right)^{m-1}. \quad (19.5)$$

*Доказательство.*  $\mathbb{P}(p_1 > p, p_2 > p, \dots, p_m > p) = \prod_{i=1}^m \bar{\Phi}(p_i) = (\bar{\Phi}(p))^m$ . Взяв первую производную, получаем результат. ■

Не пользуясь предельным распределением, мы выполняем численное интегрирование при разных значениях  $m$ , как показано на Рисунке 19.4. Для  $m$  испытаний математическое ожидание с большей точностью вычисляется как

$$\mathbb{E} p_{\min} = \int_0^1 -m \varphi(p; p_M) \left( \int_0^p \varphi(u; \cdot) du \right)^{m-1} dp.$$

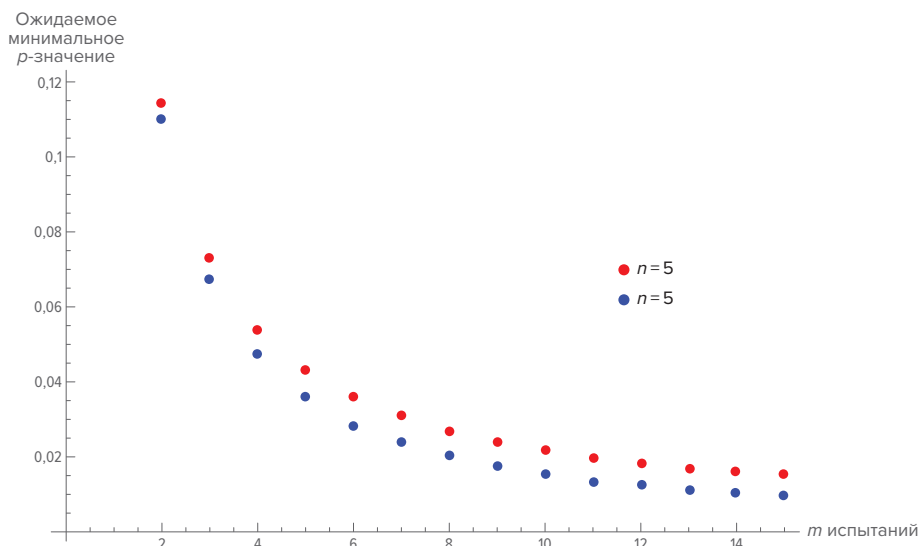


Рисунок 19.4: Значение « $p$ -хакинга» при различном числе испытаний  $m$  и при  $p_M = 0,15$  и  $p_s = 0,22$

## 19.2. ОБРАТНАЯ МОЩНОСТЬ ТЕСТА

Пусть  $\beta$  — мощность теста и  $p$  — его  $p$ -значение применительно к наблюдаемому значению  $X$ , полученному по случайным выборкам размера  $n$ , притом что ненаблю-

даемый параметр был  $\vartheta$ . Чтобы перекалибровать надежность  $\beta$  как истинную меру мощности, решим обратную задачу:

$$\begin{aligned}\beta &\rightarrow X_{\vartheta, p, n} \\ \Delta \updownarrow &\checkmark \\ \beta^{-1}(X)\end{aligned}$$

#### Предложение 19.4

Пусть  $\beta_c$  — проекция мощности теста из реализаций, которые предположительно следуют t-распределению Стьюдента и оцениваются при параметре  $\vartheta$ . Тогда

$$\Phi(\beta_c) = \begin{cases} \Phi(\beta_c)_L, & \beta_c < \frac{1}{2} \\ \Phi(\beta_c)_H, & \beta_c > \frac{1}{2} \end{cases},$$

где

$$\Phi(\beta_c)_L = \left( \frac{Y_3}{Y_3 - 2\sqrt{(1-Y_3)Y_3}\sqrt{(1-Y_1)Y_1} + 2Y_3\sqrt{(1-Y_1)Y_1} - Y_1(2\sqrt{(1-Y_3)Y_3} - 1)} \right)^{\frac{n+1}{2}}, \quad (19.6)$$

$$\Phi(\beta_c)_H = \frac{B\left(\frac{1}{2}, \frac{n}{2}\right)}{B\left(\frac{n}{2}, \frac{1}{2}\right)} (1-Y_2)^{-\frac{n+1}{2}}$$

$$\left( \left( 2(1 - \sqrt{(1-Y_2)Y_2} - Y_2) \sqrt{\frac{1}{Y_3} - 1} + 2\sqrt{(1-Y_2)Y_2} - 1 \right) Y_2 + \frac{1}{Y_3} \right)^{-\frac{n+1}{2}}, \quad (19.7)$$

$$\text{где } Y_1 = I_{2\beta_c}^{-1}\left(\frac{n}{2}, \frac{1}{2}\right), \quad Y_2 = I_{2\beta_c-1}^{-1}\left(\frac{1}{2}, \frac{n}{2}\right) \text{ и } Y_3 = I_{(1, 2\beta_c-1)}^{-1}\left(\frac{n}{2}, \frac{1}{2}\right).$$

### 19.3. ПРИЛОЖЕНИЕ И ВЫВОДЫ

- Можно смело утверждать, что при такой стохастичности реализаций  $p$ -значений и распределения их минимума желающие иметь пресловутую 5 %-ю достоверность — и с опорой на нее делать статистические выводы — должны требовать  $p$ -значения на добрый порядок ниже.
- Авторам, пытающимся исследовать воспроизводимость, например, в статье Центра открытой науки<sup>1</sup> в «Сайенс» [49], следовало бы учесть предел погрешности в собственной процедуре и существенное смещение в сторону

<sup>1</sup> Некоммерческая организация в Шарлотсвилле, штат Виргиния; основана в 2013 г. с миссией «повысить открытость, честность и воспроизводимость научных исследований». Вначале исследовала воспроизводимость в психологии.

желаемых ими результатов (ошибка первого типа). Неудивительно, что тесты, исходно показавшие значимость, не достигают ее при воспроизведении; на самом деле удивительным было бы как раз воспроизведение результатов, исходно показавших значимость вблизи пограничной.

- «Мощность» теста имеет ту же проблему, если не снизить  $p$ -значения или не переставить планку теста на более высокий уровень, такой как 0,99.

## **БЛАГОДАРНОСТИ**

Марко Авельянеде, Паскуале Чирилло, Яниру Бар-Яму, дружественным собеседникам в твиттере...



## Н

### НЕКОТОРЫЕ НЕДОРАЗУМЕНИЯ В ПОВЕДЕНЧЕСКОЙ ЭКОНОМИКЕ

Мы видели ранее (Главы 3 и 11), что проблема «переоценки хвостов» участниками психологического эксперимента скорее должна относиться на счет использования ложной «нормативной» модели психологами и специалистами по принятию решений, ничего не понимающими в жирных хвостах. Здесь мы используем два случая как иллюстрацию такого неправильного использования вероятности, разоблачаемого нашей простой эвристикой — ввести эффект второго порядка и посмотреть, как повлияет неравенство Йенсена на оператор математического ожидания.

Одно из упомянутых вольных использований вероятности (загадка премии по акциям) связано со сторонниками «подталкивания», навязчивого и нечестного метода, придуманного теми психологами, которые ищут способы манипулировать решениями, принимаемыми гражданами.

#### **Н.1. ПРИМЕР ИССЛЕДОВАНИЯ: ЛОЖНАЯ СПЕЦИФИКАЦИЯ БЛИЗОРУКОЙ БОЯЗНИ ПОТЕРЬ**

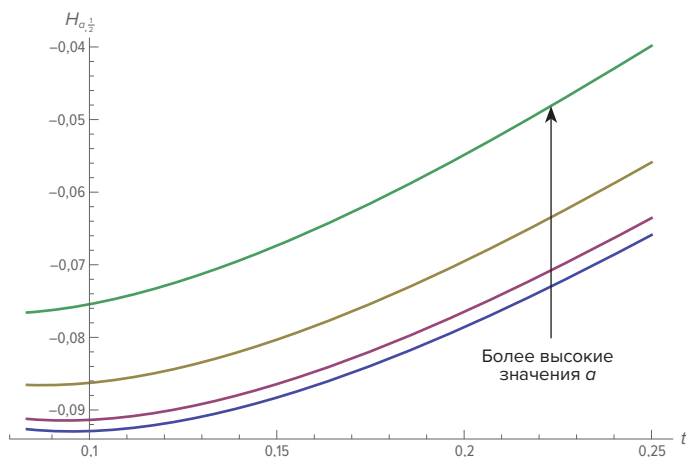
Пресловутая «загадка премии по акциям», первоначально обнаруженная Мехрой и Прескоттом [169], названа так потому, что исторически доходность акций превосходила инвестиции с фиксированным доходом, хотя, казалось бы, такая разница должна была привлечь спекулянтов и исчезнуть.

Мы сразу обнаруживаем, что тот анализ упускал из виду неэргодичность в данной области, что мы видели в Главе 3: на самом деле участники не стремились к безусловной оценке рыночной отдачи; со стороны индивидуального инвестора было бы глупо использовать вероятности ансамбля и закон больших чисел, имея всего одну жизнь. Кроме того, «положительное математическое ожидание рыночной отдачи» не достаточное условие для положительного математического ожидания инвестора; требуется стратегия масштабирования пути в стиле Келли, или динамическое хеджирование, зависящее от пути.

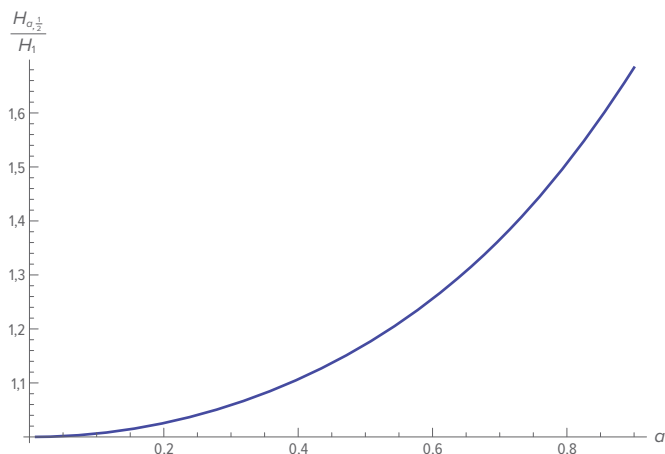
Бенарци и Талер [17] утверждают, что теория перспектив Канемана — Тверски [139] объясняет такое поведение, порожденное близорукостью. Даже если это верно, такой анализ рассыпается при толстых хвостах.

Давайте ужирним хвосты распределения, стохастизировав, например, коэффициент масштаба, и посмотрим, что останется от приводимых в литературе результатов, на первый взгляд абсурдных. Как мы сейчас увидим, при более строгом теоретико-вероятностном анализе эта абсурдность подтверждается.

### Близорукая боязнь потерь



**Рисунок Н.1:** На «полезность»  $H_{a,p}(t)$  из теории перспектив влияет неопределенность второго порядка — неопределенность дисперсии. Здесь  $\sigma = 1$ ,  $\mu = 1$  и  $t$  — переменная [как поясняется в тексте, здесь выполняется стохастизация дисперсии в  $1 \pm a$  раз: с вероятностью  $\frac{1}{2}$  в  $1 + a$  раз и с вероятностью  $\frac{1}{2}$  в  $1 - a$  раз]



**Рисунок Н.2:** Отношение  $\frac{H_{a, \frac{1}{2}}(t)}{H_1(t)}$ , или спад «полезности» под действием эффектов второго порядка

Возьмем из теории перспектив оценивающую функцию  $w$  изменения финансового положения  $x$  с параметрами  $\lambda$  и  $\alpha$ :

$$w_{\lambda, \alpha}(x) = x^\alpha \mathbb{1}_{x \geq 0} - \lambda(-x^\alpha) \mathbb{1}_{x < 0}.$$

Пусть  $\varphi_{\mu, \sigma\sqrt{t}}(x)$  — плотность нормального распределения с соответствующими средним и среднеквадратическим отклонением (масштабированными с коэффициентом  $t$ ).

Ожидаемая «полезность» (в смысле теории перспектив):

$$H_0(t) = \int_{-\infty}^{\infty} w_{\lambda, \alpha}(x) \varphi_{\mu, \sigma\sqrt{t}}(x) dx = \quad (\text{H.1})$$

$$= \frac{1}{4\sqrt{\pi}} \sigma^{\alpha} (2t)^{\frac{\alpha}{2}} \left( \left( \Gamma \frac{\alpha+1}{2} \right) (1-\lambda) {}_1F_1 \left( -\frac{\alpha}{2}; \frac{1}{2}; -\frac{t\mu^2}{2\sigma^2} \right) + \right. \\ \left. + \frac{\mu}{\sigma} \left( \Gamma \frac{\alpha+2}{2} \right) \sqrt{2t} (1+\lambda) {}_1F_1 \left( \frac{1-\alpha}{2}; \frac{3}{2}; -\frac{t\mu^2}{2\sigma^2} \right) \right). \quad (\text{H.2})$$

Из формулы (H.2) видим, что более частая выборка эффективности ведет к снижению полезности. Таким образом, Бенарци и Талер на самом деле пытались найти близорукость при такой периодичности выборки и, соответственно, частоте выборки, при которой возникает «премия»; ошибка их в том, что они упустили эффекты второго порядка.

Мы же, пробуя разные стохастизации  $\sigma$ , эвристически открываем другую картину: что, если, с малой вероятностью, дисперсия может многократно возрастать при неизменной общей дисперсии? Важно, что мы не изменяем дисперсию; мы только сдвигаем распределение в хвосты. Мы щедро допускаем, что по закону больших чисел было установлено, что «загадка премии по акциям» на самом деле имела место и что акции действительно превосходили бонды по эффективности.

Итак, будем переключаться между двумя состояниями,  $(1+a)\sigma^2$  с вероятностью  $p$  и  $(1-a)\sigma^2$  с вероятностью  $1-p$ .

Перепишем (H.1) как

$$H_{a,p}(t) = \int_{-\infty}^{\infty} w_{\lambda, \alpha}(x) \left( p \varphi_{\mu, \sqrt{1+a}\sigma\sqrt{t}}(x) + (1-p) \varphi_{\mu, \sqrt{1-a}\sigma\sqrt{t}}(x) \right) dx. \quad (\text{H.3})$$

**Результат.** В итоге, как видим на Рисунках H.1 и H.2, эффекты второго порядка отменяют утверждения на основе «близорукости» боязни потерь. Это не значит, что близорукость не проявляется; она просто не может объяснить «премию по акциям» как внешний эффект (когда распределение дает другую доходность); эффект оказывается внутренним, происходящим из структуры функции стоимости Канемана — Тверски  $v(x)$ .

**Комментарий.** Мы использовали эвристику  $1 \pm a$  как иллюстрацию; можно было бы использовать полноценное распределение  $\sigma^2$  и получить аналогичные результаты.

Например, для гамма-распределения с плотностью  $f(v) = \frac{v^{y-1}}{\Gamma(y)} e^{-\frac{\alpha v}{V}} \left( \frac{\alpha}{V} \right)^y$  с математическим ожиданием  $V$ , отвечающим дисперсии в теории «премии по акциям».

Переписывая (H.3) с учетом этого выражения, получаем

$$\int_{-\infty}^{\infty} \int_0^{\infty} w_{\lambda, \alpha}(x) \varphi_{\mu, \sqrt{v}}(x) f(v) dv dx,$$

и для этой формулы есть явное выражение, только оно длинновато для нашего изложения.

**Настоящая проблема Бенарци и Талера.** Конечно, проблема была связана с толстыми хвостами и сходимостью по ЗБЧ, о которой мы рассказываем отдельно.

### Ошибка временных предпочтений для модели

Другой пример эффекта из-за рандомизации параметра можно описать как создание лишнего уровня неопределенности.

Автор этой книги однажды, к своему ужасу, видел, как некий Лейбсон [150] на конференции в Колумбийском университете развивал тезис в том духе, что люди ведут себя нерационально, если, например, сначала выберут вместо покупки одного сеанса массажа сегодня покупку сразу двух сеансов — завтра, но через год вернуться к прежнему поведению; что требуется исправить положение дел некоторой политикой. (Обзор по проблеме дисконтирования по времени и временных предпочтений смотрите в [96], при обсуждении экономистов, которые приписывают участникам переменную «ставку дисконтирования», выводимую по упрощенной модели)<sup>1</sup>.

На интуитивном уровне: пусть я узнал, что предложение о массаже исходит, вероятно, от мошенника. Понятно, что я переверну предпочтения, предпочту немедленную услугу долгосрочному абонементу с большой условной выгодой, условие которой состоит в том, что продавец абонементов не исчезнет до срока. Смоделируем это явление.

Во-первых, дисконтирование по времени должно иметь вид геометрической прогрессии, чтобы льготная цена не становилась отрицательной: линейное дисконтирование по времени вида  $Ct$ , где  $C$  — константа и  $t$  — время в будущем, исключено. Нужна формула вида  $C^t$  или, чтобы была видна ставка,  $(1+k)^t$ . Прогрессию можно математически упростить до показательной функции путем предельного перехода к непрерывному времени. Запишем экспоненциальный дисконт в виде  $e^{-kt}$ . Дисконтирование с таким мелководным профилем предотвращает «временную несостоятельность», гарантируя, что при  $\delta < t$

$$\lim_{t \rightarrow \infty} \frac{e^{-kt}}{e^{-k(t-\sigma)}} = e^{-k\sigma}.$$

Теперь добавим уровень стохастичности: стохастизируем коэффициент дисконтирования, который обозначим  $\lambda$ .

Теперь мы можем выразить  $H(t)$  только как

$$H(t) = \int e^{-\lambda t} \varphi(\lambda) d\lambda.$$

Нетрудно вывести для общего случая, теми же методами, что в Разделе 4.1, что при симметрической стохастизации с интенсивностью  $\Delta\lambda$  (то есть с вероятностью по  $\frac{1}{2}$  отклониться в любую сторону от центра распределения):

$$H'(t, \Delta\lambda) = \frac{1}{2} (e^{-(\lambda-\Delta\lambda)t} + e^{-(\lambda+\Delta\lambda)t})$$

$$\frac{H'(t, \Delta\lambda)}{H'(t, 0)} = \frac{1}{2} (e^{-(\lambda-\Delta\lambda)t} + e^{-(\lambda+\Delta\lambda)t}) = e^{-\lambda t} \operatorname{ch} \Delta\lambda t,$$

где  $\operatorname{ch}$  — гиперболический косинус; данное выражение сойдется к некоторому значению в будущем, когда временные льготы выйдут на горизонталь.

<sup>1</sup> Фармер и Геанокопелос [89] применяли аналогичный подход к гиперболическому дисконтированию (*Hyperbolic discounting*). — Прим. автора.

**Пример: гамма-распределение.** При гамма-распределении на носителе  $\mathbb{R}^+$  с параметрами  $\alpha$  и  $\beta$ ,  $\varphi(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1}}{\Gamma(\alpha)} e^{-\frac{\lambda}{\beta}}$  получаем

$$H(t, \alpha, \beta) = \int_0^\infty e^{-\lambda t} \frac{\beta^\alpha \lambda^{\alpha-1}}{\Gamma(\alpha)} e^{-\frac{\lambda}{\beta}} d\lambda = \beta^{-\alpha} \left( \frac{1}{\beta} + t \right)^\alpha,$$

откуда

$$\lim_{t \rightarrow \infty} \frac{H(t, \alpha, \beta)}{H(t - \delta, \alpha, \beta)} = 1.$$

Это значит, что предпочтения становятся горизонтальными в будущем независимо от крутизны наклона в настоящем, и это объясняет спад ставки дисконтирования в экономической литературе.

Более того, если подправить и нормировать распределение, когда

$$\varphi(\lambda) = \frac{1}{k} e^{-\frac{\lambda}{k}},$$

получаем *нормативно построенное* так называемое гиперболическое дисконтирование:

$$H(t) = \frac{1}{1 + kt},$$

которое оказывается отнюдь не эмпирически выявленной «патологией», как заявляют наивные исследователи, у которых на самом деле в моделях не хватало одного уровня неопределенности.

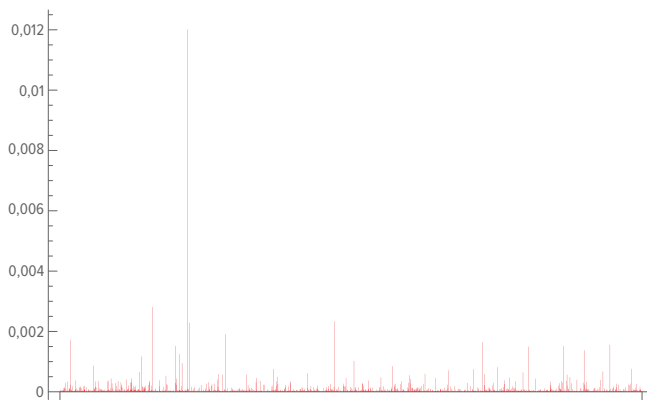
## **Часть VII**

# **ТОРГОВЛЯ ОПЦИОНАМИ И ЦЕНЫ ПРИ ЖИРНЫХ ХВОСТАХ**



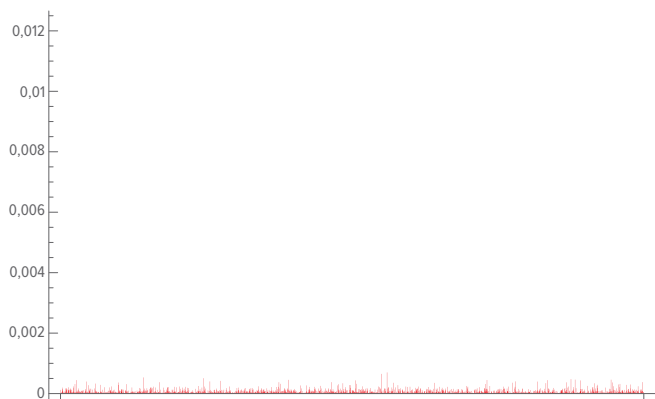
## НЕУДАЧИ ФИНАНСОВОЙ ТЕОРИИ КАСАТЕЛЬНО ЦЕН ОПЦИОНОВ<sup>†</sup>

Обсудим, почему теория опционов в так называемой неоклассической экономике терпит неудачу в реальном мире. Каким образом финансовая теория назначает цену финансовому продукту? Принципиальная разница в подходе, который Башелье предложил в 1900 году, [6], и современной финансовой модели Блэка — Шоулза — Мертона, [24] и [171], лежит в ряде центральных допущений, причем Башелье был ближе к реальности и практике трейдеров, традиция которой насчитывает не одно столетие.

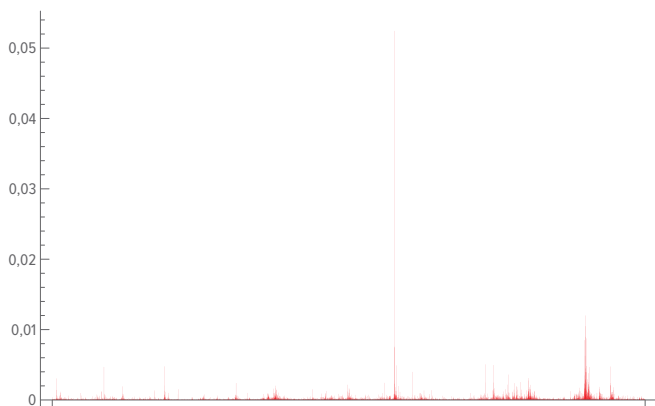


**Рисунок 20.1:** Ошибки хеджирования для портфеля опционов (в режиме ежедневного пересмотра) на протяжении 3000 дней, смоделированные при  $t$ -распределении Стьюдента с постоянной волатильностью и показателем хвоста  $\alpha = 3$ . Технически ошибки не должны сходиться за конечное время, так как их распределение имеет бесконечную дисперсию





**Рисунок 20.2:** Ошибки хеджирования для портфеля опционов (в режиме ежедневного пересмотра), смоделированные при тех же параметрах, но в мире Блэка — Шоулза (фантастическом)



**Рисунок 20.3:** Ошибки портфельного хеджирования (в режиме ежедневного пересмотра), включая крах фондового рынка в 1987

### 20.1. БАШЕЛЬЕ, А НЕ БЛЭК — ШОУЛЗ

Модель Башелье основана на ожидаемых итоговых страховых выплатах, а не на динамическом хеджировании. Это значит, что можно использовать любое распределение! Более формальное доказательство с использованием теории меры приведено в Главе 21, а пока поговорим на интуитивном уровне с минимумом математики.

Тот же метод в дальнейшем предлагался рядом исследователей, например Спренклом [216] в 1964, Бонессом [26] в 1964, Кассуфом и Торпом [253] в 1967 и Торпом [249] (опубликовано только в 1973).

Все они наткнулись на следующую проблему: как создать параметр риска — ставку дисконтирования рискованного актива, — чтобы он был совместим с портфельной теорией? Для модели САРМ (от *Capital Asset Pricing Model*, модель ценообразования капитальных активов) требуется, чтобы ценные бумаги проявляли ожидаемую доходность пропорционально рискованности. В подходе Блэка — Шоулза — Мертон цена опциона выводится из непрерывного по времени динамического хеджирования и только как свойство такого хеджирования; ниже мы расскажем о динамическом хеджировании подробнее. Благодаря этому методу опцион коллапсирует в детерминистскую платежную функцию и обеспечивает доходность независимо от рынка; следовательно, рискованная премия для опциона не требуется.

### 20.1.1. Искажения из-за идеализации

Суть наших претензий к подходу Блэка — Шоулза — Мертона в том, что их требования к динамическому хеджированию грубо идеализированы; они содержат следующие строгие условия. Оператор предполагается способным покупать и продавать на рынке без потерь на трение, то есть без транзакционных затрат. Процедура не допускает, чтобы цена зависела от последовательности потока, — если оператор продаст некоторое количество акций, это не повлияет на цену. Оператор знает распределение вероятностей, и оно гауссово с фиксированными неизменными коэффициентами в течение всего времени. Наконец, самое существенное ограничение: не допускаются масштабируемые скачки. В поздней редакции [Мертон, 1976] допускаются скачки, но только пуассоновы по времени ожидания и одинаковые по величине, в крайнем случае гауссовы.

Система не разрешает использовать степенные законы, ни практически, ни математически. Рассмотрим математику для потока динамических хеджей в уравнении Блэка — Шоулза — Мертона.

Без потери общности допустим, что процентная ставка  $r = 0$ , без риска. Каноническая модель Блэка — Шоулза — Мертона предусматривает продажу опциона «колл» и приобретение акций, которые обеспечивают хедж против мгновенных изменений в ценных бумагах. Следовательно, портфель  $\pi$ , локально «хеджированный» против рисков первого момента распределения, будет такой:

$$\pi = -C + \frac{\partial C}{\partial S} S, \quad (20.1)$$

где  $C$  — цена опциона «колл» и  $S$  — исходные ценные бумаги. Рассмотрим изменение в стоимости портфеля

$$\Delta\pi = -\Delta C + \frac{\partial C}{\partial S} \Delta S. \quad (20.2)$$

Разложением вблизи начальных значений  $S$  получаем изменения в портфеле в дискретном времени. Обычная теория опционов применяет гауссово распределение, где все порядки выше  $(\Delta S)^2$  и  $\Delta t$  быстро исчезают:

$$\Delta\pi = -\frac{\partial C}{\partial S} \Delta t - \frac{1}{2} \frac{\partial^2 C}{\partial S^2} \Delta S^2 + O(\Delta S^3). \quad (20.3)$$

Берем математическое ожидание от обеих частей равенства и видим, из (20.3), весьма жесткие ограничения на конечность моментов: все моменты должны сходиться. Если мы добавим слагаемое  $-\frac{1}{6} \frac{\partial^3 C}{\partial S^3} \Delta S^3$ , оно может оказаться значимым, если распределение вероятностей имеет значительные компоненты третьего или четвертого порядка. В самом деле, даже если  $n$ -я производная по  $S$  резко падает, это не устраняет последствия для опционов со страйк-ценой  $K$  далеко от центра распределения, потому что моменты растут непропорционально быстро.

Итак, мы настаиваем, что здесь требуется конечность всех моментов и спад их влияния; никакая аппроксимация недопустима. Признаем, что модель диффузионных скачков (Мертон, 1976) так легко не разваливается, поскольку в ней все моменты существуют. Беспокойство вызывает то, что при степенном законе бесконечны все моменты выше  $\alpha$ , и для такого закона уравнение портфеля Блэка — Шоулза — Мертона не сработает.

Как мы уже говорили, логика так называемого решения Блэка — Шоулза — Мертона, благодаря формуле Ито, состоит в том, что портфель коллапсирует в детерминистскую платежную функцию. Но посмотрим, насколько быстро и эффективно это работает на практике.

### 20.1.2. Фактическая процедура репликации

Платежную функцию опциона «колл» нужно реплицировать следующим потоком динамических хеджей, предел которого между  $t$  и  $T$

$$\lim_{\Delta t \rightarrow 0} \left( \sum_{i=1}^{n=T/\Delta t} \frac{\partial C}{\partial S} \bigg|_{S=S_{t+(i-1)\Delta t}, t=t+(i-1)\Delta t} (S_{t+i\Delta t} - S_{t+(i-1)\Delta t}) \right). \quad (20.4)$$

Мы разбили временной период на  $n$  приращений по  $\Delta t$ . Здесь коэффициент хеджирования  $\frac{\partial C}{\partial S}$  вычисляется на момент  $t+(i-1)\Delta t$ , но мы получаем неожиданную разность между ценой на момент инициации хеджа и результирующей ценой в момент  $t+i\Delta t$ .

Предполагается, что это сделает платежную функцию детерминистской *в пределе по времени*  $\Delta t \rightarrow 0$ . В гауссовом мире она будет одним из интегралов Ито — Маккина.

### 20.1.3. Провал: погрешность хеджирования может сделать модель непригодной

Вследствие показанных выше математических особенностей погрешность хеджирования при кубической  $\alpha$  оказывается такой же, как при процессе с бесконечной дисперсией. Более того, погрешность непропорционально сильно влияет на страйк-цены далеко от денег.

Коротко: динамическое хеджирование в мире степенного закона нисколько не устраняет риск.

## ДАЛЕЕ

В следующей главе, применяя теорию меры, покажем, почему опционы все-таки могут быть нейтральными к риску.

## ЕДИНСТВЕННАЯ МЕРА ДЛЯ ЦЕН ОПЦИОНОВ (БЕЗ ДИНАМИЧЕСКОГО ХЕДЖИРОВАНИЯ ИЛИ ПОЛНОГО РЫНКА)<sup>†</sup>

Мы представляем доказательство того, что при простых допущениях, таких как ограничения паритета «пут» — «колл», теоретико-вероятностная мера для оценки европейских опционов имеет среднее, выводимое из форвардной цены<sup>1</sup>, которое может быть или не быть нейтральным к риску при любом обычном распределении вероятностей, обходясь без доводов динамического хеджирования Блэка — Шоулза — Мертона и не требуя полных рынков и прочих сильных допущений. Мы подтверждаем, что эвристика, веками используемая трейдерами, более робастна, состоятельна и строга, чем считается в литературе по экономике. Мы также показываем, что цены опционов можно моделировать при распределениях с бесконечной дисперсией (и конечным средним).

### 21.1. ИСТОРИЯ ВОПРОСА

Методы оценки опциона использовались трейдерами веками и с высокой эффективностью (Хог и Талеб, [126]). Кроме того, оценки по математическому ожиданию суммарного выигрыша заставляют среднее по распределению вероятностей, используемому для цен опционов, становиться ценой форварда благодаря пут-колл-паритету, и, если цена форварда нейтральна к риску, опцион тоже будет нейтрален к риску. Считается, что доводы Блэка — Шоулза (Блэка и Шоулза, 1973, Мертон, 1973) состоят в допущении, что цены опционов нейтральны к риску благодаря динамическому хеджированию, поскольку опционы становятся излишни (ведь платежную функцию опциона можно построить как линейную комбинацию кэша и исходного актива, динамически корректируемую с течением времени). Это загадка, поскольку (1) динамическое хеджирование операционно неисполнимо на финансовом рынке из-за преобладания портфельных изменений в результате скачков; (2) доводы динамического хеджирования математически рушатся при жирных хвостах; требуется особый «мир

<sup>1</sup> Англ. *forward price* (иногда переводится также как «курс по сделкам на срок») — цена, указанная в «форварде», взаимном обязательстве покупателя и продавца совершить куплю-продажу в оговоренную дату по оговоренной цене. Форварды отличаются от фьючерсов тем, что не стандартизированы и не перепродаются на бирже.

Блэка — Шоулза» с рядом невозможных допущений, в том числе конечных биквадратных отклонений; (3) трейдеры используют тот же «довод нейтральности к риску» Блэка — Шоулза для оценки опционов по активам, не допускающим динамической репликации; (4) трейдеры систематически торгуют опционами в областях, где доводы нейтральности к риску неприменимы; (5) есть фундаментальные информационные ограничения, препятствующие сходимости стохастического интеграла<sup>1</sup>.

Было несколько предшественников у данного тезиса, объявляющего пут-колл-паритет достаточным ограничением, чтобы навязать некоторую структуру на уровне среднего по исходному распределению, например Дерман и Талеб (2005), Хог и Талеб (2010). Их подходы были эвристичны и робастны, но сочтены запальчивыми (Руффино и Трюссар [207]). Они также показали, что участникам нужно использовать среднее, нейтральное к риску. Эта глава решает следующие задачи:

- Переходит от «запальчивости» к формальным доказательствам.
- Использует полностью свободный от распределения, основанный на математическом ожидании подход и обосновывает довод нейтральности к риску без динамического хеджирования и без допущений о распределении.
- Устанавливает, даже в отсутствие нейтральности к риску, единственное распределение цен опционов. Цена форварда (или фьючерса) может включать в себя ожидания и отклоняться от цены арбитража (благодаря, скажем, регуляторным и иным ограничениям), и все же цены опционов могут следовать распределению, соответствующему среднему такого форварда.
- Показывает, как можно на практике иметь рынок опционов без «полноты» и выполнения теорем финансовой экономики.

Это достигается при всего лишь двух ограничениях: «горизонтальном» — о пут-колл-паритете и «вертикальном» — о том, что разные оценки разных страйк-цен приводят к теоретико-вероятностной мере, для которой доказывается единственность. Из экономических допущений требуется лишь существование и оправданность рынка форвардов; если единственная цена форварда не существует, бессмысленно обсуждать цены стандартных опционов. Мы также требуем, чтобы теоретико-вероятностные меры соответствовали распределениям с конечным первым моментом.

Предыдущие работы в этом направлении были следующие. Бриден и Лиценбергер [31] и Дюпир [72] показывают, как опционный спред обеспечивает единственную вероятностную меру; есть статьи, где строится более широкое множество отношений арбитража между опционами, например Карра и Мадана [37]<sup>2</sup>.

Однако (1) ни в одной из этих статей не строился мост между опционами «колл» и «пут» через форвард, переводящий взаимосвязь с языка арбитража между опцио-

1 Кроме того, упомянем научно-историческую загадку: точная формула, которая теперь называется «формулой Блэка — Шоулза — Мертона», была эвристически выведена (и использована) Эдвардом Торпом на основе математического ожидания без помощи динамического хеджирования; см. Торп [251]. — *Прим. автора.*

2 Смотрите также публикации Грина и Джерроу [114] и Нахмана [175]. Мы знали о возможности цен, нейтральных к риску, и без динамического хеджирования, со времен Харрисона и Крепса [123], но та теория нуждается в сильных, нереалистичных допущениях, таких как абсолютно полные рынки и многопериодное ядро цен. — *Прим. автора.*

нами, обеспечивающего вероятностное распределение, на язык необходимого выравнивания по среднему распределения форварда, и поэтому нейтрального к риску (в случае, где есть арбитраж форварда); (2) ни в одной из этих статей не доказывалось, что даже отсутствие второго момента (скажем, из-за бесконечной дисперсии) ничуть не мешает определять цены опционов. Наши методы и доказательства не пользуются дисперсией; (3) наш метод много проще, более прямой и робастный к изменениям в допущениях.

Мы не делаем допущений о полноте общего рынка. Опционы не являются избыточными ценными бумагами и не превращаются в таковые. В Таблице 21.1 сведены основные результаты статьи<sup>1,2</sup>.

**Таблица 21.1:** Основные практические различия между аргументацией динамического хеджирования и статическим пут-колл-паритетом со спредами по страйк-ценам

	Блэк — Шоулз — Мертон	Пут-колл-паритет со спредами
<b>Тип</b>	Непрерывная ребалансировка	Интерполяционный статистический хедж
<b>Предел</b>	Закон больших чисел по времени (горизонтальный)	Закон больших чисел по страйк-ценам (вертикальный)
<b>Допущения о рынке</b>	1) Непрерывные рынки без разрывов, без скачков	1) Допускаются разрывы и скачки. Можно рассмотреть непрерывные страйк-цены или допустимое число страйк-цен
	2) Возможность занимать и ссужать исходный актив во все даты	2) Возможность занимать и ссужать исходный актив для одной-единственной даты форварда
	3) Никаких транзакционных издержек при торговле активом	3) Низкие транзакционные издержки при торговле опционами
<b>Вероятностное распределение</b>	Требуется конечность всех моментов. Исключается класс распределений по медленно меняющейся функции	Требуется конечность 1-го момента (допустима бесконечная дисперсия)

1 Знаменитый парадокс Хаккансона звучит так: если рынки полные и опционы излишни, зачем кто-то станет ими пользоваться? Если рынки неполные, то опционы могут принести пользу, только как назначить им цену? Настоящее обсуждение, возможно, решает этот парадокс: рынки неполные, но мы можем назначить цену опционам. — *Прим. автора.*

2 Цены опционов не уникальны в абсолютном смысле: премия сверх фундаментальной стоимости может пробегать целый спектр значений; просто ограничения пут-колл-паритета заставляют меры, используемые для опционов «пут» и «колл», склеиться друг с другом и сравняться по математическому ожиданию с форвардом. Опционы — полноправная разновидность ценных бумаг, просто у них сильная связь с форвардом. — *Прим. автора.*

Продолжение таблицы 21.1

<b>Полнота рынка</b>	Достигается через динамическую полноту	Не требуется (в традиционном смысле)
<b>Реалистичность допущений</b>	Низкая	Высокая
<b>Сходимость</b>	Неопределенная; один большой скачок изменяет математическое ожидание	Робастная
<b>Настраивается по реальности</b>	Только после «подправки» среднеквадратических отклонений для каждой страйк-цены	Гибридно, с использованием специального распределения и адаптации к реальности

## 21.2. ДОКАЗАТЕЛЬСТВО

Определим  $C(S_{t_0}, K, t)$  и  $P(S_{t_0}, K, t)$  как европейские опционы соответственно «колл» и «пут» со страйк-ценой  $K$ , сроком  $t$  и стоимостью исходных ценных бумаг  $S_{t_0}$  в момент  $t_0$ ,  $t \geq t_0$  и возможной стоимостью  $S_t$  исходных ценных бумаг в момент  $t$ .

### 21.2.1. Случай 1: форвард как мера, нейтральная к риску

Определим  $r = \frac{1}{t-t_0} \int_{t_0}^t r_s ds$ , доход безрискового валютного фонда, и  $\delta = \frac{1}{t-t_0} \int_{t_0}^t \delta_s ds$ , выплата по активу (непрерывные дивиденды по акции или внешний процент по валютному займу).

Арбитражную цену на форвард  $F_t^Q$  можно представить как

$$F_t^Q = S_{t_0} \frac{(1+r)^{t-t_0}}{(1+\delta)^{t-t_0}} \approx S_{t_0} e^{(r-\delta)(t-t_0)} \quad (21.1)$$

по арбитражу, см. у Кейнса, 1924<sup>1</sup>. Назовем поэтому  $F_t^Q$  ценой фьючерса или форварда, полученной арбитражем при ставке, нейтральной к риску. Пусть  $F_t^P$  — фьючерс, для которого требуется «ожидаемый доход»  $m$ , связанный с риском, и ожидаемая цена форварда

$$F_t^P = S_{t_0} (1+m)^{t-t_0} \approx S_{t_0} e^{m(t-t_0)}. \quad (21.2)$$

**Замечание:** по арбитражу все торгуемые ценности с заданной ценой форварда  $S_{t_0}$  должны быть равны  $F_t^Q$ .

«Торгуемые» здесь означает не «фактически проданные», а лишь подлежащие репликации арбитража типа «кэш-энд-керри», то есть спекуляции путем заимствования кэша и хранения ценных бумаг в расчете на доход  $d$ , если доход охваченного форварда отклонится от  $r$ .

1 Возможно, отсылка к переписке в период работы над двухтомной книгой *A Treatise on Money*, опубликованной в 1930.

### 21.2.2. Вывод формул

Примем на дальнейшее, что  $F$  обладает собственной динамикой — независимо от попадания под случай 1 или 2 — и, следовательно, единственной вероятностной мерой  $Q$ .

Определим  $\Omega = [0, \infty) = A_K \cup A_K^c$ , где  $A_K = [0, K]$  и  $A_K^c = (K, \infty)$ .

Рассмотрим класс стандартных (упрощенных) вероятностных пространств  $(\Omega, \mu_i)$ , проиндексированных номером  $i$ , где  $\mu_i$  — вероятностная мера, т. е. мера, удовлетворяющая требованию  $\int_{\Omega} d\mu_i = 1$ .

#### Теорема 6

Для данного срока платежа по займу  $T$  существует единственная мера  $\mu_Q$ , задающая цены европейских опционов «пут» и «колл» по ожидаемому суммарному выигрышу.

Эта мера может быть нейтральной к риску в том смысле, что задает цену форварда  $F_t^Q$ , но может и не быть такой и задавать ставку дохода по акциям, охваченным форвардом.

#### Лемма 21.1

Для заданного срока платежа  $T$  существует две меры  $\mu_1$  и  $\mu_2$  для европейских опционов «колл» и «пут», с общим сроком исполнения и общими исходными ценными бумагами, связанными с оценкой согласно ожидаемому суммарному выигрышу, — которые единственны и таковы, что для любых опционов «колл» и «пут» со страйк-ценой  $K$  имеет место соответственно

$$C = \int_{\Omega} f_c d\mu_1, \quad (21.3)$$

и

$$P = \int_{\Omega} f_p d\mu_2, \quad (21.4)$$

где  $f_c$  и  $f_p$  — это соответственно  $(S_t - K)^+$  и  $(K - S_t)^+$ .

*Доказательство.* Для ясности, не теряя общности, зададим для  $r$  и  $\delta$  значение 0. В силу арбитража при пут-колл-паритете позитивное владение опционом «колл» («длинным») и негативное владение опционом «пут» («коротким») реплицирует торгуемый форвард; из-за изменчивости P/L возникает разница, которую, приписывая длинным опционам положительную стоимость и коротким — отрицательную, можно записать в виде:

$$C(S_{t_0}, K, t) - P(S_{t_0}, K, t) + K = F_t^P, \quad (21.5)$$

поскольку форвард  $F_t^P$  торгуемый.

Пут-колл-паритет остается в силе для всех страйк-цен, поэтому:

$$C(S_{t_0}, K + \Delta K, t) - P(S_{t_0}, K + \Delta K, t) + K + \Delta K = F_t^P \quad (21.6)$$

при всех  $K \in \Omega$ .



Заметим, что спред по опционам колл в единицах  $\frac{1}{\Delta K}$ :

$$C(S_{t_0}, K + \Delta K, t) - P(S_{t_0}, K + \Delta K, t)$$

приносит 1 доллар, если  $S_t > K + \Delta K$  (то есть отвечает индикаторной функции  $\mathbb{1}_{S > K + \Delta K}$ ), 0 долларов, если  $S_t \leq K$  (т. е.  $\mathbb{1}_{S \leq K}$ ), и столько долларов, сколько содержится в нашей единице  $\frac{1}{\Delta K}$ , умноженной на  $S_t - K$ , если  $K < S_t \leq K + \Delta K$ , — значит, между 0 и 1 долларом (смотрите у Бридена и Лиценбергера, 1978 [31]). Аналогичным образом рассмотрим обратную аргументацию для опциона «пут», при  $\Delta K < S_t$ .

В пределе при  $\Delta K \rightarrow 0$

$$\frac{\partial C(S_{t_0}, K, t)}{\partial K} = -\mathbb{P}(S_t > K) = -\int_{A_K^c} d\mu_1. \quad (21.7)$$

В силу той же аргументации:

$$\frac{\partial P(S_{t_0}, K, t)}{\partial K} = -\int_{A_K} d\mu_2 = 1 - \int_{A_K^c} d\mu_2. \quad (21.8)$$

Поскольку полузакрытые интервалы образуют всю борелеву  $\sigma$ -алгебру на  $\Omega$ , это доказывает, что меры  $\mu_1$  и  $\mu_2$  — единственные. ■

### Лемма 21.2

*Вероятностные меры опционов «пут» и «колл» совпадают, а именно для всякого борелевского множества  $A$  в  $\Omega$  эти две меры равны:  $\mu_1(A) = \mu_2(A)$ .*

*Доказательство.* Комбинируя формулы (21.5) и (21.6), деля на  $\frac{1}{\Delta K}$  и переходя к пределу при  $\Delta K \rightarrow 0$ , получаем:

$$-\frac{\partial C(S_{t_0}, K, t)}{\partial K} + \frac{\partial P(S_{t_0}, K, t)}{\partial K} = 1 \quad (21.9)$$

при всех значениях  $K$ , поэтому

$$\int_{A_K^c} d\mu_1 = \int_{A_K^c} d\mu_2, \quad (21.10)$$

поэтому  $\mu_1(A_K) = \mu_2(A_K)$  при всех  $K \in [0, \infty)$ . В силу справедливости этого равенства для всякого полузакрытого интервала оно распространяется на всякое борелевское множество. ■

### Лемма 21.3

*В силу статического арбитража опционы «пут» и «колл» должны оцениваться той же нейтральной к риску мерой  $\mu_Q$ , что и торгуемый форвард.*

*Доказательство.*

$$F_t^P = \int_{\Omega} F_t d\mu_Q; \quad (21.11)$$

из формулы (21.5)

$$\int_{\Omega} f_C(K) d\mu_1 - \int_{\Omega} f_P(K) d\mu_1 = \int_{\Omega} F_t d\mu_Q - K. \quad (21.12)$$

Дифференцируя обе части и пользуясь тем, что  $f_C - f_P = S_{t_0} + K$ , получаем, что производная<sup>1</sup> Радона — Никодима:

$$\frac{d\mu_Q}{d\mu_1} = 1 \quad (21.13)$$

для всех значений  $K$ . ■

### 21.3. СЛУЧАЙ ФОРВАРДА БЕЗ НЕЙТРАЛЬНОСТИ К РИСКУ

Рассмотрим случай, где форвард  $F_t$  наблюдаемый и торгуемый, и используем его только как исходную ценную бумагу с собственной динамикой. В этом случае мы можем полностью игнорировать динамику его номинального актива  $S$  или использовать «предполагаемую» ставку, без нейтральности к риску, привязывающую кэш к форварду,

$$m^* = \frac{\ln \frac{F}{S_{t_0}}}{t - t_0}.$$

Ставка  $m$  может охватывать премию за риск, трудности с финансированием, структурные или регуляторные препятствия для займов; для итогового результата это все равно.

В этой ситуации можно показать, что применимы в точности те же результаты, что выше, просто заменив меру  $\mu_Q$  на другую меру  $\mu_{Q^*}$ . Цены опционов останутся единственными<sup>2</sup>.

### 21.4. КОММЕНТАРИЙ

Мы заменили громоздкое и неудобоотслеживаемое динамическое хеджирование на простую, легкую задачу интерполяции и объяснили эффективность операторов опционов до Блэка — Шоулза, использовавших простую эвристику и правила, обходясь без теорем финансовой экономики.

Опционы могут оставаться не излишними и рынки неполными: наша аргументация исходит лишь из существования некоторых арбитражных цен (включая нейтральные к риску цены на уровне математического ожидания вероятностной меры). Этого достаточно, чтобы использовать любое из вероятностных распределений с конечным первым моментом, включая логнормальное, при котором воспроизводится модель Блэка — Шоулза.

Еще одно сравнение напоследок. При динамическом хеджировании пропуск одного-единственного хеджа или возникновение одного-единственного разрыва (хвостового события) может все разрушить — как мы упоминали, требуется ряд допущений помимо математических, сверх ряда строгих и крайне нереалистичных математиче-

<sup>1</sup> Производная меры относительно меры; обозначается  $\frac{d\nu}{d\mu}$  и определяется через интеграл Лебега критерием  $\nu(A) = \int_A \frac{d\mu}{d\mu} d(\mu(A))$  для всех измеримых множеств  $A \subset \Omega$ . Существование гарантируется теоремой

Радона — Никодима, если мера  $\nu$  в борелевом пространстве  $\Omega$  абсолютно непрерывна относительно меры  $\mu$  в том же борелевом пространстве. Теорема доказана в 1913 Радонам для  $\Omega = \mathbb{R}^n$  и в 1930 Никодимом для общего случая.

<sup>2</sup> Мы предполагали, для простоты доказательств, нулевую ставку дисконтирования; если она отлична от нуля, премии дисконтируются по ставке оператора арбитража. — *Прим. автора.*

ских ограничений. В классе жирнохвостых распределений повышение частоты хеджей не гарантирует снижение риска. Кроме того, аргументация для стандартного динамического хеджирования требует точного задания *нейтрального к риску* стохастического процесса между  $t_0$  и  $t$ , с неуклюжей эконометрикой, из-за чего обычно представлением процесса вообще не пользуются, а прибегают к реинжинирингу по ценам опционов, то есть к интерполяции, ориентированной на арбитраж.

У нас же, по методам на основе пут-колл-паритета, возможность отслеживать нейтральное к риску распределение гарантировано сложением страйк-цен, и, поскольку сумма вероятностей дает 1, число степеней свободы в реконструированной мере  $\mu_Q$  в области разрыва между страйк-ценой  $K$  и следующей по величине страйк-ценой  $K + \Delta K$ , существенно сокращается: мера в этом интервале ограничена разностью  $\int_{A_K^c}^c d\mu - \int_{A_{K+\Delta K}}^c d\mu$ . Другими словами, никакой отдельно взятый разрыв между страйк-ценами не может существенно повлиять на вероятностную меру, не говоря уже о первом моменте, от чего не застраховано динамическое хеджирование. На самом деле наш метод не отличается от стандартного ядерного сглаживания<sup>1</sup> статистических выборок, только примененного к распределению по страйк-ценам<sup>2</sup>.

Допущение о наличии страйк-цен естественное: раз мы обсуждаем практику опционов, значит, страйк-цены опционов существуют. Более того, маркет-мейкеры<sup>3</sup> способны, как автору довелось испытать на себе, добавлять внебиржевые страйк-цены по своему усмотрению, как им заблагорассудится.

## БЛАГОДАРНОСТИ

Питеру Карру, Марко Авельянеде, Эльетт Жиман, Рафаэлю Дуади, Гуру Губерману, Эспену Хогу и Хоссейну Каземи.

1 Англ. *kernel smoother*, или *kernel smoothing*, — сглаживание пикселей изображения или иной статистической информации, при котором в поправке для данной точки велик вес близких точек, составляющих «ядро» сглаживания.

2 О методах интерполяции подразумеваемого вероятностного распределения между страйк-ценами смотрите у Авельянеды и др. [4]. — Прим. автора.

3 Дословно «делатели рынка» (*market makers*) — участники рынка, действия которых оказывают главное влияние на курс акций.

## ТОРГОВЦЫ ОПЦИОНАМИ НЕ ПОЛЬЗУЮТСЯ ФОРМУЛОЙ БЛЭКА — ШОУЛЗА — МЕРТОНА\*\*

Торговцы опционами пользуются эвристически нащупанными формулами для цен и адаптируют их, изменяя хвосты и асимметрию, путем варьирования одного параметра: среднеквадратического отклонения гауссова распределения. У широкой публики формулы эти называются «формулами Блэка — Шоулза — Мертона» и приписываются одноименному открытию, хотя регулировка среднеквадратического отклонения ему противоречит. Мы же собрали исторические свидетельства, что (1) упомянутые Блэк, Шоулз и Мертон никаких формул не изобретали, а просто построили аргументацию в пользу известной и употребительной формулы и сделали ее приемлемой для истеблишмента экономистов, удалив параметр риска путем динамического хеджирования, (2) трейдеры, торгующие опционами, пользуются (есть свидетельства, что с 1902) сложной эвристикой и приемами, скорее совместимыми с прошлыми версиями формулы Луи Башелье и Эдварда Торпа (где допустимо выбирать из широкого спектра вероятностных распределений), и устраняют параметр риска при помощи пут-колл-паритета, (3) после 1973 торговцы опционами не начали пользоваться формулами Блэка — Шоулза — Мертона или подобными и продолжали придерживаться эвристики «снизу вверх», более робастной к редким влиятельным событиям. Глава опирается на исторически применявшиеся методы торговли и ссылается на источники XIX и XX веков, игнорируемые в финансовой литературе. Не надо повторять ложную атрибуцию нашей практики назначать цены опционов.

### 22.1. ПРЕРЫВАНИЕ ЦЕПИ ПЕРЕДАЧИ

Для нас, практиков, теории должны рождаться из практики<sup>1</sup>. Это объясняет нашу настороженность по поводу идеи «научного подхода», по которому практика должна настраиваться по теории. Хеджирование опционов, назначение цен и торговля — это

<sup>1</sup> Здесь под *практиками* понимаются те, кто ежедневно принимает решения о хеджировании опционов, взвешивает отношение P/L риска и ставит свою шкуру на кон, а не просто ассистент — финансовый математик, который пишет программы для вычисления цен, или ученый, оказывающий консалтинговые услуги. — *Прим. автора.*

ни философия, ни математика. Это развитое ремесло, которому трейдеры учатся у трейдеров (или в котором трейдеры копируют трейдеров), и приемы развиваются под давлением эволюционного отбора, снизу вверх. Это техника, а не наука.

Подчинись трейдеры научному подходу, их отрасль погибла бы, потому что, как мы увидим, предлагаемая эмпирическая и научная настройка цен и теории хеджирования в лучшем случае недостаточны и ненаучны (а в худшем случае методы хеджирования не устраняют риск, а создают). Наш подход в этой главе состоит в том, чтобы откопать исторические свидетельства техник, показывающих, как трейдеры делали бизнес с опционами в прошлом.

Как мы покажем, опционы использовались весьма активно и до современных финансов. Приемы и эвристически нащупанные методы торговли опционами и управления рисками держателей деривативов были разработаны на протяжении прошлого столетия и с большим эффектом применены операторами. Параллельно математики произвели свои изыскания. Но мейнстримная литература по экономике не признала их вклада и заменила предложенные рядом экономистов повторные открытия и последующие переформулировки на свою линию. Есть доказательства проблемы с атрибуцией формулы опционов Блэка — Шоулза — Мертона: она разработана, использована и адаптирована робастным образом предшествующей традицией исследователей, она давно использовалась эвристически держателями опционных портфелей. Более того, в порядке научно-исторического парадокса точная формула Блэка — Шоулза — Мертона была записана и использована Эдвардом Торпом, причем, дополнительный парадокс, была сочтена нестрогой, хотя была робастной и реалистичной. Приходим к выводу, что (1) инновация Блэка — Шоулза — Мертона сводилась к аргументации в пользу неоклассической финансовой математики, попросту к мысленному эксперименту<sup>1</sup>, и что (2) не видно, чтобы трейдеры использовали эту аргументацию и соответствующую версию формулы.

Давно пора вернуть честь открытия тем, кому она принадлежит.

## 22.2. ВВЕДЕНИЕ И РЕЗЮМЕ

### 22.2.1. Теория Блэка — Шоулза была аргументом в дискуссии

Торговцы опционами называют формулой Блэка — Шоулза — Мертона ту формулу, которой пользуются, не подозревая, что она мало походит на то, что предложили Блэк и Шоулз, 1973, и Мертон, 1973, — во всяком случае, меньше, чем на другие формулы, предложенные за последний век. На самом деле из всех предложенных за долгую историю формул лишь одна незадачливая была уязвима к прыжкам и хвостовым событиям, — и это как раз была формула Блэка — Шоулза — Мертона.

Начнем с потерь при переводе: Блэк и Шоулз [25] и Мертон [172] вообще не выдвинули новой формулы опциона, они выдвинули теоретико-экономический

<sup>1</sup> Мы здесь ставим под вопрос идею путаного мысленного эксперимента в гипотетическом мире, без предсказательной силы, научной или практической. Тот факт, что аргументация Блэка — Шоулза — Мертона работает в платоническом мире и смотрится элегантно, ничего не значит; всегда можно сотворить платонический мир, в котором твоя формула будет работать или в котором получается строгое доказательство, — такая процедура суть не что иное, как обратный инжиниринг. — *Прим. автора.*

довод, построенный на новом способе вывода уже известной формулы. Этот довод, как мы увидим, крайне уязвим из-за связанных с ним допущений. К этому времени основы хеджирования и цен опционов уже были заложены, и куда более прочные. Аргументация Блэка — Шоулза — Мертона, в простых словах, заключалась в том, что опцион можно хеджировать определенным методом, так называемым динамическим хеджированием, и тем превратить его в свободный от риска инструмент; портфель не будет стохастическим, в отличие от его содержимого. На самом деле Блэк, Шоулз и Мертон занимались, по сути, продвижением хорошо известной формулы в современный им экономический истеблишмент, причем в ходе этого маркетинга исказили суть метода.



**Рисунок 22.1:** Луи Башелье, предложивший формулу опционов, базирующуюся на математическом ожидании. Она обоснована строже, чем аргументация Блэка — Шоулза в пользу динамического хеджирования, поскольку не требует тонкохвостого распределения. Мало кто знает, что так называемое открытие Блэка — Шоулза заключалось не в выводе новой формулы, а в призыве устранить математическое ожидание исходных ценных бумаг

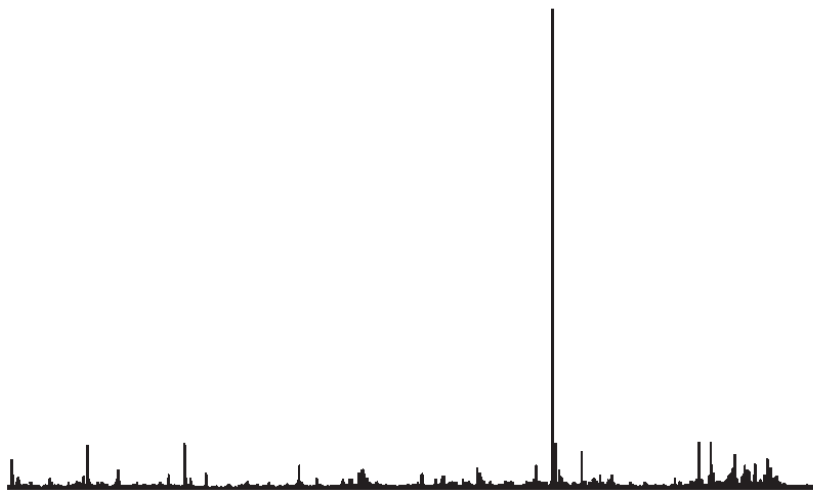
Предложенная аргументация потребовала странных, надуманных допущений: некоторой ликвидности на уровне транзакций, осведомленности о вероятности будущих событий (в неоклассическом стиле Эрроу — Дебрё<sup>1</sup>) и, что еще критичнее, некоторой математической структуры, требующей тонкохвостой, умеренной случайности — об этом ниже<sup>2</sup>. Это странная и неприемлемая аргументация, если смотреть с точки зрения клинических, базирующихся на наблюдениях исследований, а не того стиля, что сложился в неоклассической экономике. В простых словах аргументация динами-

1 В модели рынка Эрроу — Дебрё доказывалось существование равновесия. Допущения: индивиды-потребители могут свободно обмениваться экономическими благами, а фирмы — производители экономических благ распределяют доходы между индивидами (владеющими долями дохода фирм).

2 Из всех неуместных допущений Блэка — Шоулза, низводящих их построение до статуса мысленного эксперимента (элегантного с виду, как и вся современная портфельная теория), именно некоторая осведомленность о будущей дисперсии случайной величины (или о будущем вероятностном распределении) создает противоречие с практикой; рынок ужирняет хвосты, и мысленный эксперимент Блэка — Шоулза проваливается. — Прим. автора.

ческого хеджирования опасна для практики, потому что толкает вас к риску банкротиться; она не имеет смысла, если только ваша задача не состоит в защите неоклассической экономической теории. Аргументация и формулы Блэка — Шоулза — Мертона следуют в русле построения общей теории равновесия сверху вниз, отталкиваясь от тех допущений, что операторы работают в условиях полной оведомленности о вероятностном распределении будущих результатов, и еще ряда предположений, которые, как мы увидим, математически неверны; главное из них — о возможности устранить риски с помощью непрерывной торговли, что на самом деле работает только в узком специальном случае тонкохвостых распределений. И не только изъяны мешают применять эту теорию на практике: торговцы опционами вообще не клюют на теории и тем более не клюют на умозрительные общие теории равновесия, находя их слишком рискованными и не отвечающими стандартам надежности. Попросту говоря, теория о том, как должно быть, недостаточно хороша, чтобы принимать по ней решения в условиях неопределенности (тем более если хронически расходится с эмпирическими наблюдениями). Участники могут принимать решения с учетом умозрительных теорий, но не так, чтобы опираться на хрупкие стороны теорий при управлении рисками.

Тем не менее профессиональные трейдеры и исследователи, включая авторов этой книги (а также, увы, Шведскую королевскую академию наук) пребывали в плену иллюзии, будто все пользуются формулой Блэка — Шоулза — Мертона. Этот миф все более укреплялся в литературе и бизнес-школах параллельно с забвением источников или взглядом на источники как на досадные случайности (Мертон [174]).



**Рисунок 22.2:** Типичное «устранение рисков», достигаемое в аргументации Блэка — Шоулза — Мертона. Перед нами отклонения динамически хеджируемого портфеля (причем вполне типичного). БШМ действительно «сглаживает» отклонения, но подвергает оператора риску массивных хвостовых событий, напоминающих лопнувший пузырь LTCM\*. Другие формулы опционов робастны к редким событиям и не притязают на «устранение рисков»

\* *Long-Term Capital Management* — хедж-фонд, основанный в 1994 с участием Шоулза и Мертона. В кризис 1998 понес убытки, подвергся рекапитализации и в 2000 был ликвидирован.

Мы здесь обсудим, как наш реальный мир и его обитатели понимали цены опционов и хеджирование, и будем опираться на фактические действия торговцев опционами за последние сто с лишним лет, включая настоящее время.

Есть общая проблема. Как мы уже говорили, торговцы опционами, подобно многим профессионалам, выработали цепь передачи своего ремесла. Проблема же в том, что эта цепь часто прерывается, поскольку университеты не стали хранителями накопленного операторами искусства. За годы опыта приобретено множество робастных, эмпирически нащупанных наработок, но экономический истеблишмент отказался упоминать их и признавать. Из-за этого трейдерам периодически приходится учиться разным вещам заново. Например, провал динамического хеджирования в 1987, практиковавшегося такой фирмой, как «Лиланд, О'Брайен, Рубинштейн»<sup>1</sup>, не нашел заметного отражения в академической литературе после события (Мертон [174], Рубинштейн [205], Росс [203]); напротив, динамическое хеджирование осталось в мнении этих авторов стандартным образом действий<sup>2</sup>.

Есть центральные элементы реального мира, способные ускользнуть от академического исследования в отсутствие обратной связи с практикой (когда речь о прикладной науке) и приводить к расхождению, которое мы наблюдаем между лабораторией и жизнью. Этим объясняется, почему так много ученых так часто достигают гладких доходов и затем терпят банкротство, используя свои собственные теории<sup>3</sup>. Мы начнем с другого конца, начнем с многолетнего опыта торговли опционами, где были совершены миллионы хеджей и тысячи продаж опционов. Этот опыт, в сочетании с изучением забытого и игнорируемого древнего знания о ценах опционов и торговле ими, поможет нам разоблачить мифы о ценах опционов и хеджировании. По существу, этих мифов два:

- будто мы ждали формулы опционов Блэка — Шоулза — Мертона, чтобы начать торговлю опционами, задавать их цены и управлять опционными портфелями. На самом деле после аргументации Блэка, Шоулза и Мертона риски только увеличились, отбросив тех, кто прислушался, назад в искусстве управления рисками. А вообще-то это миф, будто трейдеры полагаются на теории, тем более общие теории равновесия, когда задают цены опционов;
- будто мы пользуемся формулой цен опционов Блэка — Шоулза — Мертона. Мы ею не пользуемся.

1 *Leland O'Brien Rubinstein Associates, Inc.: Portfolio Insurance*, LOR — американская финансово-консультационная фирма. Основана в 1980. До краха 1987 рекомендовала страхование портфеля: задать нижний предел портфеля ценных бумаг, создав по существу опцион «пут» по широкому рыночному индексу.

2 Пример того, как ошибки так и не достигают уровня сознания: Марк Рубинштейн был в 1995 назван финансовым инженером года по версии Международной ассоциации финансовых инженеров (*International Association of Financial Engineers*). Страхование портфеля не упоминалось, провал динамического хеджирования не упоминался. — *Прим. автора*.

3 О характерной реакции на редкое событие смотрите здесь: «Нынешнюю среду мы в Квантландии запомним надолго, — заявил м-р Ротман из Чикагского университета, возглавлявший инвестиционный фонд квантативно-математических финансовых стратегий, а затем поступивший в *Lehman Brothers*. — События, которые согласно модели могли происходить только раз в 10 000 лет, повторялись три дня подряд» (*Wall Street Journal*, 11.08.2007, р. В3). — *Прим. автора*.



Обсуждая эти мифы, мы сосредоточимся на литературе о теории опционов, построенной снизу вверх; на книгах, сосланных в потайные закоулки библиотек. Речь будет только о дошедших до нас письменах; еще больше могла содержать фактическая практика торговцев опционами, ныне утерянная.

### 22.3. МИФ 1: ТРЕЙДЕРЫ НЕ ЗАДАВАЛИ ЦЕНЫ ОПЦИОНАМ ДО БШМ

Принято считать, что теория Блэка — Шоулза — Мертона дала трейдерам возможность вычислять дельта-хедж (по исходному активу) и задавать цены опционам. Эта точка зрения весьма спорная, как исторически, так и аналитически.

Опционы активно торговались как минимум с 1600-х годов, как описывает Йозеф де ла Вега; отсюда следует, что существовало в той или иной форме искусство опционной торговли, с эвристическим методом устанавливать цены опционов и работать с их рисками. Де ла Вега описывает торговлю опционами в Нидерландах, указывая, что операторы обладают некоторым опытом в ценах опционов и хеджировании. Он в общих чертах описывает нечто вроде пут-колл-паритета, хотя в его книге никоим образом не ставилась задача обучить читателя технике торговли опционами. Мы настаиваем на давнем использовании пут-колл-паритета по следующей причине: предметом гордости Блэка — Шоулза — Мертона является устранение необходимости дрейфа с учетом риска по исходным ценным бумагам, чтобы сделать торговлю нейтральной к риску. Но для этого не требуется динамическое хеджирование; достаточно простого пут-колл-паритета (Дерман и Талеб, 2005), о чем мы поговорим ниже. По-видимому, именно это удаление рискованной премии подтолкнуло Нобелевский комитет присудить Мертону и Шоулзу премию, которая в те годы называлась Премия Шведского национального банка по экономическим наукам памяти Альфреда Нобеля: выдающийся вклад Блэка, Шоулза и Мертона состоял в том, чтобы доказать избыточность учета рискованной премии при оценке стоимости опциона. Это не значит, что рискованная премия исчезает; просто она уже включена в биржевую цену. Поначалу именно за это удаление влияния дрейфа на стоимость опциона, на основании мысленного эксперимента, их работа стала цитироваться, хотя то же самое механическим образом присутствовало в любой форме торговли и конвертации с помощью намного более простых технических приемов.

История опционов богаче, чем показано в обычной литературе. Форвардные контракты, судя по всему, прослеживаются аж до глиняных табличек Междуречья 1750 года до н. э. Гелдерблом и Йост [104] показывают, что амстердамские торговцы зерном пользовались опционами и форвардами уже в 1550.

В конце 1800-х и в начале 1900-х активно действуют рынки опционов в Лондоне и Нью-Йорке, а также в Париже и ряде других европейских бирж. К 1870-м рынки опционов развились и усложнились; Каирис и Валерио (1997) обсуждают рынок для опционов по ценным бумагам в США 1870-х, косвенно показывая, что трейдеры той эпохи уже научились задавать цену хвостовым событиям<sup>1</sup>.

<sup>1</sup> Историческое описание рынка остается информативным, пока Каирис и Валерио [140] не начинают попытки судить, имела ли в 1870-е тенденция недооценивать или переоценивать опционы (методами в стиле Блэка — Шоулза — Мертона). Период охватывает одно хвостовое событие, панику в сентябре 1873. Каирис и Валерио находят, что придерживание опционов «пут» было выгодным, но считают панику на рынке всего лишь единичным событием: «Пут-контракты, однако, выиграли от финансовой паники,

Между некоторыми из тех рынков имела даже активная арбитражная торговля опционами. Есть длинный список утраченных работ о торговле опционами: мы обнаружили следы минимум десяти германских монографий и диссертаций, написанных между концом 1800-х и эпохой гиперинфляции<sup>1</sup>.

## 22.4. МЕТОДЫ И ВЫВОД ФОРМУЛ

Один из дошедших до нас информативных источников, Нельсон [176], красноречивее всяких монографий: торговец опционами и арбитражер, он опубликовал «Азбуку опционов и арбитража», где поделился своим опытом в начале XX века. Согласно Нельсону (1904), телеграфные компании передавали до 500 сообщений в час и порядка 2000–3000 сообщений в день между рынками Лондона и Нью-Йорка, и на передачу сообщения тратилось менее минуты. Эвристически, не теоретизируя — стиль, воспроизведенный в *Динамическом хеджировании* [225], — Нельсон описывает множество клинически строгих аспектов своего арбитражного бизнеса: расходы на пересылку акций, расходы на страхование акций, расходы по выплате процентов, возможности прямого обмена акциями между долгим держателем ценных бумаг в Нью-Йорке



Рисунок 22.3: Эспен Хог (соавтор данной главы) вместе с Мандельбротом и автором этой книги в 2007 году

поразившей рынок в сентябре 1873. Считая данное событие одноразовым, мы повторили анализ пут-опционов, исключив контракты, подписанные до и продолжавшие действовать на момент биржевой паники». И заключают, ссылаясь на экономическую литературу, в которой делается вывод об общей переоценке опционов в 1950, 1960 и 1970-е: «Как показывает наш анализ, в целом опционные контракты переоценивались и не были привлекательны для приобретения мелкими инвесторами. — И добавляют: — Эмпирически мы находим, что и пут-, и колл-опционы регулярно переоценивались относительно теоретической модели оценки». С этими результатами спорит практик Нельсон (1904): «Большинство великих торговцев опционами на опыте убедились, что в конечном итоге в выигрыше остаются те, кто вкладывает деньги через опционы, а не забирает». — *Прим. автора.*

<sup>1</sup> Часть этого списка: Bielschowsky, R. (1892): *Ueber die rechtliche Natur der Prämiengeschäfte*, Bresl. Genoss. — Buchdr; Granichstaedten-Czerva, R. (1917): *Die Prämiengeschäfte an der Wiener Börse*, Frankfurt am Main; Holz, L. (1905): *Die Prämiengeschäfte*, Thesis (doctoral) — Universität Rostock; Kitzing, C. (1925): *Prämiengeschäfte: Vorprämien-, Rückprämien-, Stelligen- u. Nochgeschäfte; Die solidesten Spekulationsgeschäfte mit Versicherung auf Kursverlust*, Berlin; Leser, E. (1875): *Zur Geschichte der Prämiengeschäfte*; Szkolny, I. (1883): *Theorie und Praxis der Prämiengeschäfte nach einer originalen Methode dargestellt*, Frankfurt am Main; Author Unknown (1925): *Das Wesen der Prämiengeschäfte*, Berlin: Eugen Bab & Co., Bankgeschäft. — *Прим. автора.*

и коротким в Лондоне, чтобы сэкономить на пересылке и страховании, и ряд других технических приемов.

Официальный канон финансовой экономики не содержит источников вне финансовой экономики; этот механизм обсуждает Талеб (2007) [227]. Согласно официальной литературе по опционам, пут-колл-паритет впервые был полноценно описан Столлом [219], но ни он, ни его коллеги не упоминают Нельсона. Притом что аргументация пут-колл-паритета полностью осознана и подробно описывается Нельсоном, который, в свою очередь, часто ссылается на Хиггинса (1902) [129]. В качестве примера того, как Нельсон (1904) ссылается на Хиггинса (1902), процитируем:

Стоит отметить, что сделки типа опцион «колл» заключаются чаще, чем «пут», что, возможно, объясняется склонностью большинства биржевых игроков смотреть на вещи оптимистически и, соответственно, ожидать подъема цен, а не падения.

Однако эта склонность покупать опционы «колл» и упорно избегать опционов «пут» не приводит к тенденции дорогих «коллов» и дешевых «пут», потому что, как можно доказать, искусный опционный дилер способен конвертировать «пут» в «колл», «колл» в «пут» и несколько «коллов» в «пут-энд-колл», то есть, по сути, любой опцион в любой другой, играя на бирже против первого опциона. Мы можем, следовательно, заключить с приемлемой точностью, что в каждый момент «колл» по некоторым акциям стоит столько же, сколько «пут» по тем же акциям, и половину того, сколько стоит «пут-энд-колл».

«Пут-энд-колл» — это просто пара опционов «пут» и «колл» с одинаковой страйк-ценой и сроком; то, что мы бы назвали двойным опционом<sup>1</sup>. Нельсон описывает пут-колл-паритет во всех подробностях, на нескольких страницах. Статическое нейтральное к рынку дельта-хеджирование также было известно в то время; так, Нельсон пишет в своей книге:

Наученные долгим опытом продавцы опционов в Лондоне, продав опцион «колл», немедленно приобретают половину акций, против которых играли, продавая «колл»; а продав опцион «пут», немедленно продают половину этих акций.

Чтобы правильно понять это наблюдение, учтем, что стандартные опционы в Лондоне тех лет заключались «на деньгах» (о чем ясно сообщает Нельсон); кроме того, все стандартные опционы в Лондоне были в европейском стиле. Опционы «в деньгах» или «вне денег» торговались в Лондоне лишь эпизодически и считались «вычурными опционами» (*fancy options*). Отсюда, как и из книги Нельсона в целом, ясно, что опционные дилеры отлично понимали, что дельта по опционам «на деньгах» примерно 50 %. Несомненно, торговля опционами «на деньгах» в Лондоне тех лет настраивалась под страйк-цену, равную форварду «на деньгах», чтобы цены опционов «пут» и «колл» сравнились. Сегодня мы знаем, что опционы, равные форварду «на деньгах», при небольшом сроке имеют дельту, близкую к 50 % (или

<sup>1</sup> Более жаргонный вариант этого термина по-русски — дословный: «стеллаж» или «стрэддл», от *straddle*. Жаргонные названия есть и у модификаций двойного опциона с неравным количеством путов и коллов, хотя по-прежнему с одинаковым страйком и сроком: два колла и один пут называются «стрэп», от *strap*, а два пута и один колл называются «стрип», от *strip*.

минус 50 % для опционов «пут»). Типичные опционы в Лондоне тех лет заключались на срок один месяц.

Нельсон также указывает, в общих чертах, на динамическое дельта-хеджирование, и что оно лучше работает в теории, чем на практике (смотрите Хог [125]). Из всех подробностей у Нельсона очевидно, что в начале 1900-х опционы активно торговались и что опционные трейдеры тех лет отнюдь не чувствовали себя беспомощными при определении цен и хеджировании.

Еще один опционный трейдер Герберт Файлер торговал опционами с 1919 по 1960-е. Файлер (1959) описывает вполне активный рынок опционов в Нью-Йорке и Европе в начале 1920-х и в 1930-е. Файлер упоминает, что из-за Второй мировой войны европейские биржи закрылись. Далее он упоминает, что в Лондоне торговля опционами возобновилась только в 1958. В начале 1900-х опционные трейдеры Лондона считались самыми умудренными, согласно [177]. Не исключено, что Вторая мировая война и прекращение торговли опционами на долгие годы были причиной забвения робастных принципов опционного арбитража, которые пришлось заново открывать профессорам финансовой математики, таким как Столл.

Задолго до него, в 1908, Винценс Бронзин опубликовал книгу, где вывел несколько формул для цен опционов, включая близкую к той, что ныне называется формулой Блэка — Шоулза — Мертона, см. также у Хафнера и Циммермана (2007, 2009) [117]. Бронзин обосновал свою оценку опциона, нейтрального к риску, на робастных принципах арбитража, таких как пут-колл-паритет и привязка опционов «пут» и «колл» к цене форварда способом, который был переоткрыт Дерманом и Талебом (2005)<sup>1</sup>. На самом деле ограничения пут-колл-паритета достаточно, чтобы устранить необходимость инкорпорировать будущие доходы в исходных ценных бумагах; оно выравнивает опционы по цене форварда<sup>2</sup>.

А затем в 1910 и Анри Дойч описывает пут-колл-паритет, хотя не так подробно, как Хиггинс и Нельсон. А затем в 1961 и Рейнах описывает пут-колл-паритет, с подробностями (еще один текст, обычно игнорируемый академической публикой). Трейдеры на Нью-Йоркской фондовой бирже, специализирующиеся на применении пут-колл-паритета для конвертации опционов «пут» в опционы «колл» или коллов в путы, назывались в те годы Конвертерами (*Converters*). Рейнах (1961) [198]:

Не могу подтвердить числами, но, по моей оценке, более 60 процентов всех опционов «колл» обязаны своим появлением существованию Конвертеров.

1 Аргументация Дермана и Талеба (2005) [63] имела в [225], но осталась незамеченной. — *Прим. автора.*

2 Руффино и Трюссар (2006) [204] признают, что этот автор, возможно, решил задачу рискованной премии случайно (не понимая, что пут-колл-паритет интенсивно использовался на протяжении истории), но якобы неудовлетворительно. На самом деле аргументация Бронзина кажется неудовлетворительной для того, кто далее усложняет представление о мире рядом инструментов финансовой математики, таких как «стохастические ставки дисконтирования», и в то же время неправомерно упрощает картину, ограничиваясь гауссовыми распределениями, допускающими динамическое хеджирование. Руффино и Трюссар пишут, что использование нестохастической ставки дисконтирования, обычное для опционов «колл» и «пут», несовместимо с современной теорией ценообразования капитальных активов. Учитывая, что мы не видели ни одного практика, который бы пользовался стохастической ставкой дисконтирования, мы, подобно нашим предшественникам — опционным трейдерам, находим, что пут-колл-паритет достаточен и прекрасно справляется сам. Ситуация напоминает ученых, которые учат птиц летать и потом гордятся успехами подопечных, хотя настаивали их неправильно. — *Прим. автора.*

Другими словами, конвертеры (дилеры), по сути маркет-мейкеры, были способны обрабатывать и хеджировать большую часть риска путем статического хеджирования опционов опционами. Рейнах пишет, что был опционным трейдером (конвертером), и приводит примеры, как он и его коллеги были склонны хеджировать и включать в арбитражные сделки одни опционы против других опционов, используя те опционы, которые охватывались конвертируемыми облигациями:

Авторы опционов и трейдеры выработали и другие процедуры, чтобы извлекать прибыли, выписывая Пути и Коллы. По большей части они слишком специальные для всех, кроме матерых профи. Согласно одной такой процедуре, нужно, приобретя конвертируемый бонд, выписать Коллы на акции, в которые можно конвертировать этот бонд. Если акции по Коллам будут затребованы, можно будет сконвертировать бонд и предоставить акции.

Хиггинс, Нельсон, Рейнах — все они описывают, как важен пут-колл-паритет и как важно хеджировать опционы опционами. Опционные трейдеры никоим образом не были беспомощны в хеджировании или задании цен до формулы Блэка — Шоулза — Мертона. Опираясь на простые принципы арбитража, им удавалось хеджировать опционы робастнее, чем по Блэку — Шоулзу — Мертону. Мы уже упомянули, что статическое нейтральное к рынку дельта-хеджирование было описано Хиггинсом и Нельсоном в 1902 и 1904. Уильям Ганн (1937) тоже обсуждает нейтральное к рынку дельта-хеджирование для опционов «на деньгах», хотя не столь подробно, как Нельсон (1904). Кроме того, Ганн указывает некоторые виды дополнительного динамического хеджирования.

Миллс (1927) приводит примеры описания скачков и жирных хвостов в литературе до эпохи современной портфельной теории. Он пишет: «(...) распределение может оказаться далеким от гауссова из-за влияния одного-двух экстремальных изменений цен».

#### 22.4.1. Формулы опционов и дельта-хеджирование

Что возвращает нас к формулам ценообразования опционов. Первая узнаваемая формула принадлежит Башелье (1900) [5]. Спренкл в 1961 [215] расширил работу Башелье, предположив, что цена актива следует логнормальному распределению, а не гауссову. Кроме того, модель Башелье избегала дисконтирования (что не суть важно, потому что на многих рынках, в частности на рынках США, опционные премии выплачиваются в момент истечения срока).

Джеймс Бонесс (1964) [26] тоже предположил логнормальную цену актива. Он выводит формулу для цены опциона «колл», практически совпадающую с формулой Блэка — Шоулза — Мертона 1973 года, только Блэк, Шоулз и Мертон при выводе опирались кто на непрерывное динамическое дельта-хеджирование, кто на CAPM и смогли освободиться от ожидаемой доходности. Другими словами, открытием Блэка — Шоулза — Мертона была признана не столько формула, сколько подход к ее выводу. На это указывал, например, Рубинштейн (2006) [206]:

По-настоящему важным для финансово-математической теории инвестирования стала не сама формула, а способ, которым она была выведена. Формулу десятью годами раньше уже вывели Кейс Спренкл [215] и Джеймс Бонесс [26].

Самуэльсон (1969) и Торп (1969) опубликовали формулы ценообразования опционов, отчасти похожие на формулы Бонесса и Спренкла. Торп (2007) заявлял, что запрограммировал в своем компьютере формулу, тождественную формуле Блэка — Шоулза — Мертона, за несколько лет до того, как Блэк, Шоулз и Мертон опубликовали свою теорию.

Перейдем теперь к дельта-хеджированию. Как уже упоминалось, статическое нейтральное к рынку дельта-хеджирование несомненно описывалось Хиггинсом и Нельсоном в 1902 и 1904. Торп и Кассуф (1967) представили нейтральное к рынку статическое дельта-хеджирование подробнее, и не только для опционов «на деньгах», но и для опционов с любой дельтой. В своей статье 1969 Торп кратко описал нейтральное к рынку статическое дельта-хеджирование, а также кратко указал в направлении некоторого динамического дельта-хеджирования не как на центральный инструмент ценообразования, а как на средство управления рисками. Файлер также указывает на динамическое хеджирование опционов, но не показывает больших знаний о том, как вычислять дельту. Еще один игнорируемый забытый текст — брошюра, опубликованная в 1970 финансово-консультационной фирмой Arnold Bernhard & Co. Авторы несомненно знакомы с нейтральным к рынку статическим дельта-хеджированием, или, как они его называют, сбалансированным хеджем при любом уровне страйк-цены и цены актива. В этой брошюре приводится множество примеров, как приобрести варранты или конвертируемые бонды и построить нейтральный к рынку дельта-хедж, создав нужное количество коротких позиций по тем же акциям. Арнольд Бернхард и Компания также опубликовали дельты для большого числа варрантов и конвертируемых бондов, которые распространили среди инвесторов на Уолл-стрит.

Если сравнивать с Торпом и Кассуфом (1967), Блэк, Шоулз и Мертон продвинули идею динамического хеджирования на шаг вперед. Блэк и Шоулз писали (1973):

Если хедж обрабатывается непрерывно, то упомянутые выше аппроксимации становятся точными равенствами, а доходы по хеджированной позиции — полностью независимыми от изменений в стоимости акций. Доходность по хеджированной позиции становится просто гарантированной. На это нам указал Роберт Мертон.

Возможно, математически это блестящая идея, но торговля опционами не математическая теория. Оторванная от реальности теоретическая идея не будет робастной на практике. Удивительно скорее то, что лишь один принцип опционные трейдеры должны были отвергнуть и действительно не стали использовать: подход, названный в честь той самой формулы; эту тему мы обсудим далее.

## **22.5. МИФ 2: СОВРЕМЕННЫЕ ТРЕЙДЕРЫ ПОЛЬЗУЮТСЯ БЛЭКОМ — ШОУЛЗОМ**

Трейдеры не выполняют Оценку.

Во-первых, технически цена не то же самое, что оценка. Для оценки требуется сильная система теоретических предположений с соответствующей хрупкостью по предположениям и по структуре модели. Для трейдеров цена, построенная для покупки опциона, когда неизвестно будущее вероятностное распределение, никакая не оценка, а риторический прием. Такая цена может изменяться. Такая цена не отра-



жает воззрений трейдера. Она может быть установлена и по имеющемуся у него портфелю акций.

Различие между ценой и оценкой важно: трейдеры — инженеры, они мыслят в рамках ограниченной рациональности (и могут вовсе не интересоваться никакой разновидностью вероятностной рациональности), они не допущены в информационную прозрачность о будущих состояниях мира и их вероятностях. Поэтому им не нужна общая теория образования цены; им достаточно знать, что нельзя набрать «голландскую книгу» ставок (подставляться под арбитраж) и поэтому надо соблюдать стандартные ограничения: пут-колл-паритет, а еще запреты: торговать опционом «колл» со страйк-ценой  $K$  по цене ниже  $K + \Delta K$  (отрицательные спреды «колл» и «пут»), делать опцион «колл» со страйк-ценой  $K$  и опцион «колл» со страйк-ценой  $K + 2\Delta K$  дороже, чем два опциона со страйк-ценой  $K + \Delta K$  (негативные бабочки), делать отрицательные горизонтальные календарные спреды (при низких процентных ставках) и так далее. Число степеней свободы для трейдеров сокращается, потому что им приходится соблюдать пут-колл-паритет и совместимость с остальными опционами на рынке.

В этом смысле трейдерам не нужно выполнять оценку с помощью некоторого ценообразующего ядра до срока погашения ценной бумаги; вместо этого им нужно произвести цену опциона, совместимую с остальными инструментами на рынках, без учета времени владения, которое стохастично. Им не нужна наука, построенная сверху вниз.

### 22.5.1. Когда мы оцениваем стоимость?

Если бы вы нашли трейдеров, выброшенных на необитаемый остров, и одному из них потребовалось бы произвести цену опциона и держать этот опцион до окончания срока, и на местном рынке не было бы форварда, то тогда, пожалуй, потребовалась бы оценка, только трейдерская книга получилась бы маленькая. И в этом мысленном эксперименте есть искажение против реальности: люди не станут торговать опционами, если только речь не о профессионалах, а если они опционные трейдеры, у них должна быть книга с величинами торговой маржи. Если торговой маржи не существует, вряд ли трейдеры смогут создать позицию, кроме самой малой (пренебрежимой), недостаточной для динамического хеджирования. (Мы и не слышали, чтобы нашлось много неразорившихся трейдеров или учреждений, которым удалось работать в вакууме аргументации Блэка — Шоулза — Мертона.) К невозможности такого хеджирования мы сейчас и обратимся.

## 22.6. О МАТЕМАТИЧЕСКОЙ НЕВОЗМОЖНОСТИ ДИНАМИЧЕСКОГО ХЕДЖИРОВАНИЯ

Обсудим напоследок серьезные изъяны в концепции динамического хеджирования. В нем допускается, в смысле требуется, чтобы у вероятностного распределения существовали все моменты<sup>1</sup>.

<sup>1</sup> По-видимому, Мертон (1992) соглашался с неприменимостью динамического хеджирования, но, возможно, полагал, что недостатки исчезнут, поскольку предсказал, что финансовый мир входит «по спирали в динамическую полноту». Спустя пятнадцать лет если мы и продвинулись по спирали, то уж точно не к «динамической полноте», а от нее. — *Прим. автора.*

Допустим, что распределение доходов имеет безмасштабное, или фрактальное свойство, которое, в упрощенной формулировке, звучит так: при большом  $x$  (т. е. в хвостах) отношение вероятностей  $\frac{\mathbb{P}(X > nx)}{\mathbb{P}(X > x)}$  зависит от  $n$ , но не от  $x$ . Скажем,  $X$  — ежедневный доход по ценным бумагам, и нет причин, по которым  $\mathbb{P}(X > 20\%) / \mathbb{P}(X > 10\%)$  отличается от  $\mathbb{P}(X > 15\%) / \mathbb{P}(X > 7,5\%)$ . Такое самоподобие при всех масштабах порождает распределение с хвостом степенного закона, он же хвост Парето; в таком хвосте выше точки перехода вероятность  $\mathbb{P}(X > x) = Kx^{-\alpha}$ . Такое случается; глядя на миллионы порций данных, такое свойство характерно чуть ли не для всех рынков, если пренебречь ошибкой выборки. Несметное количество эмпирических примеров этого закона смотрите у Мандельброта (1963), задолго до Блэка — Шоулза — Мертона (1973) и диффузионных скачков Мертона (1976); смотрите их также у Стэнли и др. (2000) и Габе и др. (2003). Доводы предположить безмасштабность такие: даже если в конце концов распределение становится тонкохвостым, выше некоторого значения  $X$ , мы все равно не знаем этой точки; пребывая в эпистемологической темноте относительно границы безмасштабности, мы вынуждены отодвинуть эту границу в бесконечность.

Как возражение на эти «истинно жирные хвосты» встречается тот довод, что, хотя ежедневная доходность ведет себя именно так, все же действие центральной предельной теоремы сделает сумму большого числа наблюдений гауссовой, если  $\alpha$  превышает 2. Эта аргументация рассыпается из-за доасимптотического поведения масштабируемых распределений: Бушо и Поттерс (2003) и Мандельброт и Талеб (2007) настаивают, что предасимптотика фрактальных распределений замедляет и делает ничтожным действие центральной предельной теоремы в хвостах. Более того, растет погрешность выборки, так как для долгих периодов мы собираем мало данных и, следовательно, мало хвостовых эпизодов; тем самым мы склонны создавать иллюзию более тонких хвостов. К тому же точка, в которой накопление приводит к истончению хвостов, неприменима при динамическом хеджировании, где оператор вынужден опираться на часто обновляемые данные и статистические показатели. Безмасштабность в периоде динамического хеджа сохраняется, так что высокие моменты подвержены бурному росту до бесконечности и не дают сформировать портфель динамических хеджей. Разложение в ряд Тейлора просто невозможно, потому что моменты порядка выше 2 критически важны, и один из моментов уходит в бесконечность.

Механика динамического хеджирования следующая. Допустим, без потери общности, нейтральную к риску процентную ставку 0. Канонический пакет Блэка — Шоулза — Мертона заключается в том, чтобы продать опцион «колл» и приобрести акции, обеспечивающие хедж против мгновенных изменений в этих ценных бумагах. Таким образом, портфель  $\pi$ , локально «хеджированный» против риска первого момента распределения, такой:

$$\pi = -C + \frac{\partial C}{\partial S} S,$$

где  $C$  — цена опциона «колл», а  $S$  — цена исходных ценных бумаг. Рассмотрим изменение стоимости портфеля за шаг дискретного времени

$$\Delta\pi = -\Delta C + \frac{\partial C}{\partial S} \Delta S.$$



Разлагая его в ряд вблизи начальных значений  $S$ , находим изменения портфеля в дискретном времени. Общепринятая теория опционов применяет гауссиану, у которой порядки выше  $\Delta S^2$  быстро исчезают.

Беря математическое ожидание от обеих частей равенства, видим строгие требования к конечности моментов: все моменты должны сходиться. Добавление слагаемого порядка  $\Delta S^3$  может оказаться существенно при вероятностном распределении с кубическими и биквадратными компонентами. Даже если  $n$ -я производная по  $S$  резко падает, это все равно не стабилизирует хеджи для опционов со страйк-ценой  $K$  далеко от центра распределения, где высшие моменты  $S$  растут непропорционально быстро. Итак, нужны все моменты, и никакая аппроксимация недопустима. Логика так называемого решения Блэка — Шоулза — Мертона заключалась в том, что по формуле Ито портфель коллапсирует в детерминистскую платежную функцию. Но посмотрим, насколько быстро или эффективно это сработает на практике.

На самом деле репликация происходит так: платеж к опциону «колл» нужно реплицировать потоком динамических хеджей, предел которого, как можно видеть, находится между  $t$  и  $T$ :

$$\lim_{\Delta t \rightarrow 0} \sum_{i=1}^{n=T/\Delta t} \frac{\partial C}{\partial S} \bigg|_{t=t+(i-1)\Delta t} S = S_{t+(i-1)\Delta t} (S_{t+i\Delta t} - S_{t+(i-1)\Delta t}). \quad (22.1)$$

Эта политика не соответствует стоимости опциона «колл»: разность остается стохастической (хотя по Блэку — Шоулзу должна съежиться), если только дело не происходит в фантастическом мире, где такое устранение риска возможно.

В дальнейших работах Мертона видим непоследовательность, из-за которой становится непонятно, что, собственно, утверждает теория: у Мертона (1976) видим согласие с использованием опционных деривативов в стиле Башелье, если есть скачки и разрывы, обходясь без динамического хеджирования, но только если цена исходных акций не коррелирует с рынком. Похоже на признание, что аргументация за динамическое хеджирование применима только к некоторым ценным бумагам: тем, которые не скачут и коррелируют с рынком.

### 22.6.1. Сбивающая с толку робастность гауссианы

Своим успехом формула, выведенная в окончательном виде Торпом и названная формулой Блэка — Шоулза — Мертона, обязана простому свойству гауссианы: как гауссово представимо любое вероятностное распределение, даже с жирными хвостами, если варьировать среднеквадратическое отклонение  $\sigma$  на уровне плотности случайной величины. Это не значит, что вы пользуетесь гауссианой; это не значит, что гауссиана особенно экономное средство описания (ведь приходится назначать отдельную  $\sigma$  каждому уровню цены). Это значит только то, что гауссианой можно представить любое распределение, если вместо коэффициента  $\sigma$  подставить функцию, в том числе зависящую от страйк-цены и оставшегося срока.

Такая улыбка волатильности, то есть варьирование одного параметра, чтобы получить  $\sigma(K)$ , или поверхность волатильности, если варьировать два параме-

тра,  $\sigma(S, t)$ , — это то, чем, по сути, занимались, каждый по-своему, Дюпир (1994, 2005) [72, 73] и Дерман [61, 64], смотрите об этом у Гезерала (2006 [103]). Они допускали волатильность процесса — не потому, что процесс обязательно волатильный, а только лишь как метод подогнать цены опционов под гауссиану. Более того, хотя у гауссова распределения второй момент конечный (и все высшие моменты конечны), масштабируемое распределение с бесконечной дисперсией можно представить поверхностью волатильности гауссова распределения. Остается одно строгое ограничение на параметр  $\sigma$  — он должен быть одинаков для опционов «пут» и «колл» с одинаковой страйк-ценой (если оба европейского стиля), и дрейф должен быть дрейфом форварда.

Ирония в том, что на самом деле улыбка волатильности противоречит теории Блэка — Шоулза — Мертона. Из-за этого появились сотни, если не тысячи статей с попытками расширить модель Блэка — Шоулза — Мертона (насколько ее понимали авторы статей) и охватить стохастическую волатильность и диффузионные скачки. Ряд исследователей при этом выражали удивление, как мало трейдеров на деле используют модели стохастической волатильности. Модель ничего не позволяет сказать о форме *улыбки волатильности* и как она должна меняться со временем; поверхность волатильности и ее эволюция со временем определяются методом хеджирования, который робастен и согласован с безарбитражной поверхностью волатильности.

Другими словами, поверхность волатильности можно использовать как карту, но не как территорию. И ею пользуются, только было бы глупо этим оправдывать модель Блэка — Шоулза — Мертона; повторим: гауссиана не дает использовать вероятностные распределения, отличные от гауссова, тогда как вывод законов нединамического хеджирования (Башелье, Торп) не основан на гауссиане.

## 22.6.2. Поток заказов и опционы

Очевидно, что опционные трейдеры не обязаны интересоваться вероятностными распределениями на момент окончания срока; для них это такая же абстракция, как метафизика. Помимо ограничений пут-колл-паритета, которые уже были полностью разработаны в 1904, можно хеджировать инвестиционный риск опционов другими опционами. У этого метода есть то важное следствие, что из-за хеджирования опционов опционами цены опционов начинают опираться на спрос и предложение. Это, в отличие от теории Блэка — Шоулза — Мертона (1973), основанной на идеализированном мире геометрического броуновского движения<sup>1</sup> с непрерывным по времени дельта-хеджированием, а не на спросе и предложении опционов, просто не должно влиять на цену опционов. Если кто-то желает купить больше опционов, маркет-мейкеры могут просто создать их динамическим дельта-хеджированием, которое будет прекрасным заменителем самого опциона.

Здесь приходится сделать важное заявление: опционные трейдеры не оценивают шансы редких событий, ориентируясь на цены опционов «без денег». Они реагируют на спрос и предложение. Идея подразумеваемого вероятностного распределения — это всего лишь утверждение типа оглядки на «голландскую книгу».

<sup>1</sup> Случайный процесс, в котором логарифм величины ведет себя как координата броуновской частицы.

### 22.6.3. Башелье — Торп

Часто и недальновидно выдвигаемый довод приписывает успех объема опционов высокому качеству формулы Блэка — Шоулза, и это слабый довод. Особенно с учетом того факта, что опционы имели успех в разные периоды и в разных местах.

Более того, есть свидетельства, что, хотя Чикагская биржа опционов и формула Блэка — Шоулза — Мертона появились около 1973, модель «редко использовалась трейдерами» до 1980-х (О'Конелл, 2001). Когда один из авторов этой книги, Талеб, стал брокером товарной биржи в 1992, почти через два десятка лет после Блэка — Шоулза — Мертона, то, к своему удивлению, обнаружил, что многие трейдеры продолжали задавать цены опционов без таблиц, по бабочке и по конверсии, не пользуясь никакими формулами.

Еще в 1975 книга, написанная финансовыми математиками, ссылалась на Торпа и Кассуфа (1967), а не на Блэка и Шоулза (1973), хотя и включила последних в список литературы. Остер (1975):

Сидни Фрайд писал о хеджах на варранты до 1950 года, но только в 1967 книга<sup>1</sup> *Обыграв рынок* Эдварда Торпа и Шина Кассуфа строго, но просто объяснила широкой читательской аудитории хедж шорта по варранту лонгом по обычной акции.

В заключение такое замечание. К сожалению, все формулы до Блэка — Шоулза — Мертона, от первой (Башелье) до последней (Торп), используют безмасштабное распределение. Идея явным образом удалить математическое ожидание из форварда впервые была представлена Кейнсом (1924) и позже Блау (1944) — потому что лонг по коллу и шорт по путу той же страйк-цены равны форварду. Эта взаимосвязь через арбитраж, очевидно, понималась еще в 1904.

Нетрудно связать взрывной рост объема опционов с компьютерной эпохой и упрощением обработки транзакций, а также с долгой полосой мирного экономического роста и отсутствия гиперинфляции. Удалив пропаганду, видим, что развитие финансовой науки скорее было сопутствующим явлением, а не причиной торговли опционами. Повторим: если вы учили птиц летать, это не дает вам права поставить себе в заслугу, что они научились.

Вот почему мы приписываем формулу Башелье — Торпу. Мы пользовались ею всю дорогу и давали ей ложные названия, в честь неправильного метода и в честь не тех авторов. Отсюда не следует, что о динамическом хеджировании нужно забыть; просто это не центральная часть парадигмы ценообразования. Оно привело к описанию некоторого стохастического процесса, которое еще может пригодиться, если рынки когда-нибудь «придут по спирали в динамическую полноту». Но не в настоящем.

1 *Beat the Market* by Edward O. Thorp and Sheen T. Kassouf, Random House, 1967.

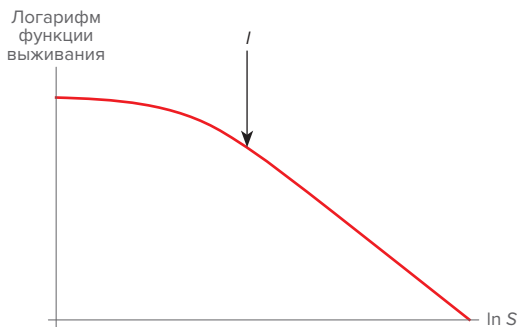
## ЦЕНООБРАЗОВАНИЕ ОПЦИОНОВ ПРИ СТЕПЕННЫХ ЗАКОНАХ: РОБАСТНАЯ ЭВРИСТИКА<sup>\*\*</sup>

В этой (исследовательской) главе<sup>1</sup> мы построим эвристику, которая по данной цене опциона в хвостах со страйк-ценой  $K$  расширяет цены на все опционы «колл» и страйк-цены  $> K$  и на все опционы «пут» и страйк-цены  $< K$ , в предположении, что продолжение хвостов вырождается в то, что мы назвали «константой Караматы», поведение за «точкой Караматы», в области, где воцаряется строгий закон Парето.

Эвристика дает относительные цены опционов, зависящие от единственного параметра — показателя хвоста  $\alpha$ , при некоторых мягких арбитражных ограничениях.

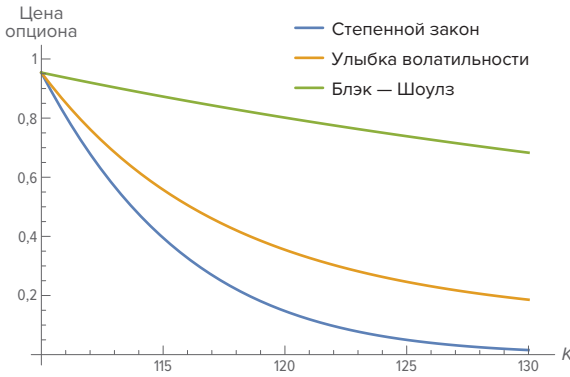
Обычно встречающиеся ограничения, такие как конечность дисперсии, отпадают.

Эта эвристика позволяет критически изучить поверхность волатильности и протестировать теории ошибочной оценки и переоценки относительной стоимости хвостового опциона, которые обычно основываются на тонкохвостых моделях и модификациях формулы Блэка — Шоулза — Мертона.



**Рисунок 23.1:** Точка Караматы, в которой медленно меняющуюся функцию можно смело заменить константой  $L(S) = l$ . Константа разная в зависимости от использования цены  $S$  или геометрической доходности — но асимптотический наклон один и тот же, он отвечает показателю хвоста  $\alpha$ .

<sup>1</sup> С участием сотрудников хедж-фонда *Universa Investments*: Брэндон Яркин, Читпьюнит Манн, Дамир Делич и Марк Шпицнагель.



**Рисунок 23.2:** Мы показываем прямую цену опциона Блэка — Шоулза (при постоянной волатильности), цену при улыбке волатильности (когда коэффициент масштаба растет в хвостах) и цену опционов при степенном законе. В упрощенном случае, когда исходный актив распределен по степенному закону, цена опционов линейно зависит от страйк-цены  $K$

### 23.1. ВВЕДЕНИЕ

Класс степенных законов обычно определяют через свойство функции выживания следующим образом. Пусть случайная величина  $X$  принадлежит классу распределений с правым хвостом степенного закона, то есть:

$$\mathbb{P}(X > x) = L(x) x^{-\alpha}, \quad (23.1)$$

где  $L: [x_{\min}, +\infty) \rightarrow (0, +\infty)$  — медленно меняющаяся функция, определенная по критерию  $\lim_{x \rightarrow +\infty} \frac{L(kx)}{L(x)} = 1$  при всех  $k > 0$  [22].

Говорят, что функция выживания случайной величины  $X$  принадлежит классу правильно меняющихся функций RV. Функция  $f: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  меняется на бесконечности с показателем  $\alpha$  (пишут  $f \in \text{RV}_\alpha$ ), когда

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^{-\alpha}.$$

В более практичном варианте существует точка, где  $L(x)$  приближается к своему пределу,  $l$ , и становится константой, как на Рисунке 23.1; мы называем этот предел «константой Караматы». За этой точкой хвост степенного закона калибруется стандартными методами, по формуле Хилла. Распределение в этой зоне Б. Мандельброт называл сильным законом Парето [162], [75].

### 23.2. ЦЕНООБРАЗОВАНИЕ «КОЛЛОВ» ЗА КОНСТАНТОЙ КАРАМАТЫ

Теперь определим цену европейского опциона «колл»  $C(K)$  со страйк-ценой  $K$  и ценой исходного актива  $S$ , где  $K, S \in (0, +\infty)$ , как  $(S - K)^+$ , и оценка выполняется при вероятностной мере  $P$ ; тогда можно задать цену опциона как  $\mathbb{E}_P(S - K)^+ = \int_K^\infty (S - K) dP$ . Это позволяет нам немедленно доказать следующее.

#### 23.2.1. Первый подход, $S$ в классе правильно меняющихся функций

Начнем с упрощенного примера, чтобы построить интуитивное понимание. Пусть  $S$  имеет функцию выживания в классе правильно меняющихся функций  $\text{RV}_\alpha$  согласно (23.1). При всех  $K > l$  и  $\alpha > 1$

$$C(K) = \frac{K^{1-\alpha} l^\alpha}{\alpha - 1}. \quad (23.2)$$

### Замечание 21

Обратим внимание, что параметр  $l$ , если выведен из существующей цены опциона, содержит всю необходимую информацию о вероятностном распределении ниже  $S=l$ , и тогда при данном параметре  $\alpha$  не приходится оценивать среднее, «волатильность» (то есть коэффициент масштаба) и прочие атрибуты.

Допустим, что  $\alpha$  задана внешним образом (выведена при подгонке распределений или просто взята из прошлого опыта; так или иначе, предполагается, что флуктуации  $\alpha$  минимальны [239]). Заметим, что  $C(K)$  инвариантна к калибровкам распределений, и единственный нужный параметр,  $l$ , будучи константой, исчезает в отношениях. Теперь допустим, что дана рыночная цена «якорного» хвостового опциона  $C_m$  со страйк-ценой  $K_1$  и для остальных опционов рассматривается отношение их цены к цене этого опциона. Мы можем просто сгенерировать все дальнейшие страйк-цены, начиная с  $l = ((\alpha - 1)C_m K_1^{\alpha-1})^{1/\alpha}$  и применяя формулу (23.2).

### Результат 1: Относительное ценообразование при распределении для $S$

При  $K_1, K_2 \leq l$

$$C(K_2) = \left( \frac{K_2}{K_1} \right)^{1-\alpha} C(K_1). \quad (23.3)$$

Удобная формула, в которой устранены все параметры распределений: чтобы построить механизм однозначного ценообразования, нужны только цена хвостового опциона и  $\alpha$ .

### Замечание 22: Как не перепутать разные функции $L$ и показатели хвоста $\alpha$

Показатель хвоста  $\alpha$  и константа Караматы  $l$  должны соответствовать заданному распределению рассматриваемого исходного актива. Показатель хвоста  $\alpha$  для стоимости  $S$  в классе правильно меняющихся функций согласно формуле (23.1), из которой выводится формула (23.2), отличен от показателя хвоста для доходности  $R = \frac{S - S_0}{S_0} \in \text{RV}_\alpha$ . У величин  $S$  и  $R$  разные графики Ципфа и прочие представления.

1. Даже если  $\mathbb{P}(X > x) = L_a(x)x^{-\alpha}$  и  $\mathbb{P}\left(\frac{X - X_0}{X_0} > \frac{x - X_0}{X_0}\right) = L_b(x)x^{-\alpha}$  и константа  $\alpha$  будет одна и та же, все равно разные  $L(\cdot)$  будут выходить на уровень константы с разной скоростью.

2. Если  $R_l = \ln \frac{S}{S_0}$ , эта величина вообще не принадлежит классу правильно меняющихся функций; смотрите Теорему 7.

Причина, по которой  $\alpha$  остается одной и той же, в том, что это безмасштабный атрибут хвоста.

### Теорема 7: Логарифм доходности

Пусть  $S$  — случайная величина доходности с функцией выживания  $\varphi_S(s) = L(s)s^{-\alpha} \in \text{RV}_\alpha$ , где  $L(\cdot)$  — медленно меняющаяся функция. Пусть  $R_l$  — логарифм доходности,  $R_l = \ln \frac{S}{S_0}$ . Функция выживания  $\varphi_{R_l}(r_l)$  не принадлежит классу  $\text{RV}_\alpha$ .

*Доказательство* непосредственное. Преобразование<sup>1</sup>  $\varphi_{R_l}(r_l) = L(s)s^{-\frac{\ln \ln s}{\ln s}}$ . ■

Заметим, что, даже если на практике для динамичной ситуации требуется непрерывное начисление процентов [229], в нашем подходе предполагается, что эта динамика охвачена ценой якорного опциона, выбранного для анализа, или  $l$ . Более того, вне далекого хвоста нет осязаемых различий между  $\ln \frac{S}{S_0}$  и  $\frac{S - S_0}{S_0}$ .

### 23.2.2. Второй подход, геометрическая доходность от $S$ в классе правильно меняющихся функций

Теперь применим аналогичный подход к реалистичному случаю, где доходность  $\frac{S - S_0}{S_0}$  распределена по Парето. Рассмотрим, при  $R > l$ , стоимость  $S = (1 + R)S_0$ , где  $S_0$  — начальная стоимость исходного актива и  $R \sim \mathcal{P}(l, \alpha)$ , распределение Парето I с функцией выживания

$$\left( \frac{K - S_0}{lS_0} \right)^{-\alpha}, \quad K > S_0(1 + l), \quad (23.4)$$

и подгоним ее под  $C_m$ , задав

$$l = \frac{(\alpha - 1)^{1/\alpha} C_m^{1/\alpha} (K - S_0)^{1 - \frac{1}{\alpha}}}{S_0};$$

как и в предыдущем случае, это покажет, что практически вся информация о распределении содержится в  $l$ .

Пусть  $\frac{S - S_0}{S_0}$  в классе правильно меняющихся функций. При  $S \geq S_0(1 + l)$

$$C(K, S_0) = \frac{(lS_0)^\alpha (K - S_0)^{1-\alpha}}{\alpha - 1}. \quad (23.5)$$

1 Можно представить функцию выживания логарифма доходности  $R_l$  и как функцию  $r_l$ . Обратное преобразование  $S = S_0 e^{R_l}$ , поэтому  $\varphi_{R_l}(r_l) = \mathbb{P}(R_l > r_l) = \mathbb{P}\left(\ln \frac{S}{S_0} > \ln \frac{S}{S_0}\right) = \mathbb{P}(\ln S > \ln s) = \varphi_S(s) = L(s)s^{-\alpha} = L(S_0 e^{r_l})(S_0 e^{r_l})^{-\alpha} = S_0^{-\alpha} L(S_0 e^{r_l}) e^{-\alpha r_l}$ , и это не только не степенной закон, но даже, в случае совсем медленной функции  $L$ , экспоненциальный закон, как хвост гауссианы.

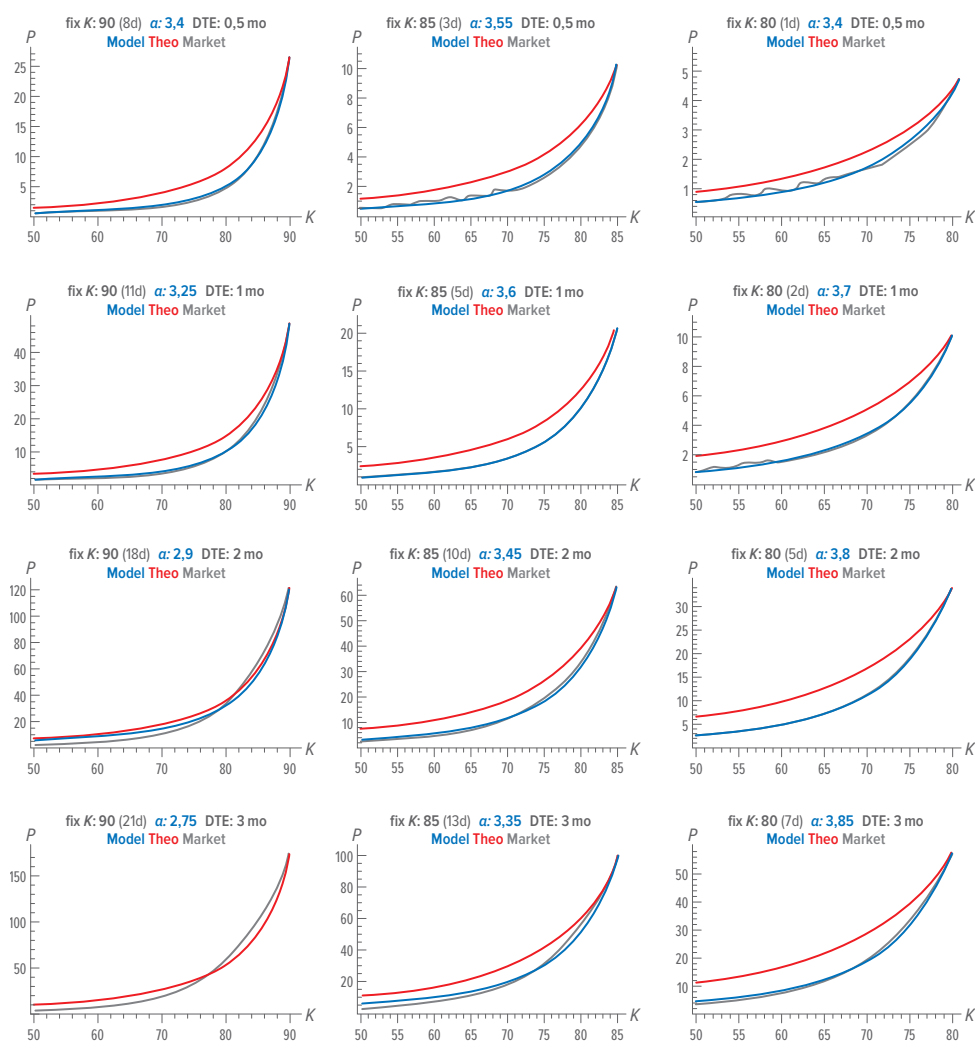
2 В самом деле, вне далекого хвоста  $\frac{S}{S_0} \approx 1$ , так что  $\frac{S - S_0}{S_0} = \frac{S}{S_0} - 1 \approx \ln \frac{S}{S_0}$ .

Таким образом, мы можем переписать формулу (23.3), устранив  $I$ :

**Результат 2: Относительные цены под распределением для  $\frac{S - S_0}{S_0}$**

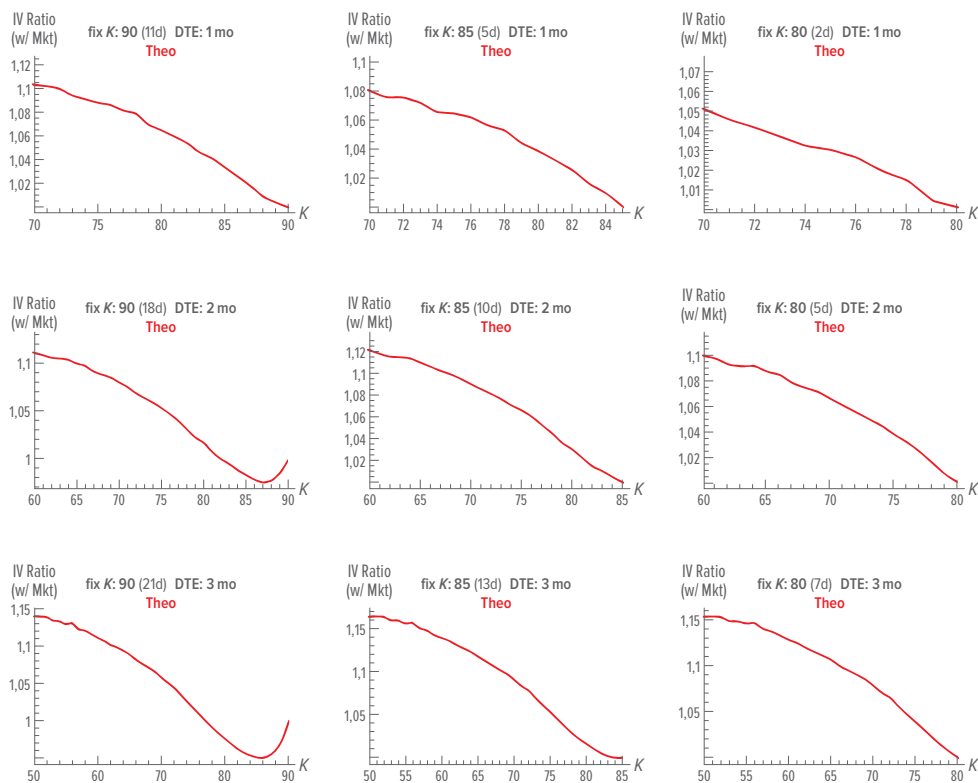
При  $K_1, K_2 \geq (1 + I)S_0$

$$C(K_2) = \left( \frac{K_2 - S_0}{K_1 - S_0} \right)^{1-\alpha} C(K_1). \quad (23.6)$$

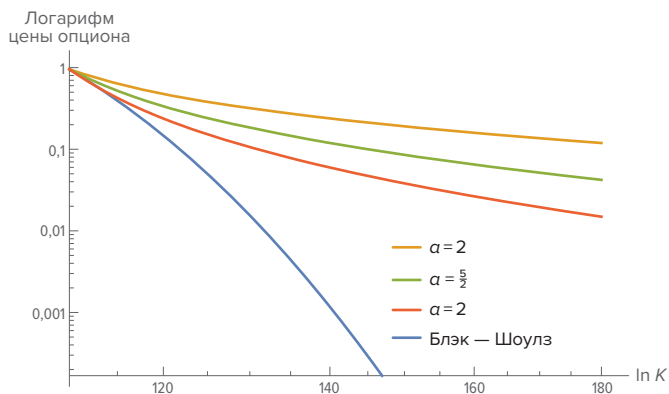


**Рисунок 23.3:** Цены опционов «пут» в индексе S&P 500, если использовать как якорь режим «фиксировать  $K$ » (позаимствовав 31 декабря 2018) и сгенерировать цены опционов по показателю хвоста  $\alpha$ , соответствующему ценам рынка (синие кривые, Model), и по показателю  $\alpha = 2,75$  (красные кривые). Видим, что цены рынка имеют тенденцию (1) ложиться на степенной закон (соответствует стохастической волатильности с подправленными параметрами), но (2) со значением  $\alpha$ , приуменьшающим жирность хвостов. Это показывает, как грубо ошибаются модели, обосновывающие переоценку хвостов





**Рисунок 23.4:** Те же результаты, что на Рисунке 23.3, но выраженные через предполагаемую волатильность. Мы подогадали цены по предполагаемой волатильности для страйк-цен нижней половины (якоря 90, 85 и 80), используя отношения нашей модели к рынку



**Рисунок 23.5:** Интуитивное понимание по дважды логарифмическому графику при второй калибровке

**Замечание 23**

В отличие от методов ценообразования в классе модификаций Блэка — Шоулза (модели стохастической и локальной волатильности смотрите в обзорах Дюпира, Дермана и Гезерала [74], [102], [60]) в нашей модели ценообразования опционов конечность дисперсии, вообще говоря, не требуется, как показано в [229]. Требуется только  $\alpha > 1$ , то есть конечный первый момент.

**23.3. ЦЕНЫ ОПЦИОНОВ «ПУТ»**

Рассмотрим теперь страйк-цены опционов «пут» (или соответствующих опционов «колл» в отрицательном хвосте, где цены можно определить арбитражом пут-колл-паритета). В отличие от опционов «колл» будет достаточно рассмотреть отклонения  $\frac{S - S_0}{S_0}$ , не переходя к логарифмической доходности (или к рассмотрению стоимости  $S$ , взятой отдельно).

Мы строим отрицательную сторону по отрицательной доходности исходного актива. Обозначим  $R$  доходность,  $S = (1 - R)S_0$ , и пусть  $R > l > 0$  распределена по Парето в положительной области, с плотностью  $f_R(r) = \alpha l^\alpha r^{-\alpha-1}$ . Теоретико-вероятностным преобразованием и масштабированием функции плотности исходного актива получаем:

$$f_S(s) = -\frac{\alpha \left( -\frac{s - S_0}{lS_0} \right)^{-\alpha-1}}{lS_0} \lambda, \quad s \in [0, (1-l)S_0],$$

где константа масштабирования

$$\lambda = \frac{1}{(-1)^{\alpha+1}(l^\alpha - 1)}$$

задана так, чтобы интеграл от  $f_S(s)$  давал 1. Впрочем, параметр  $\lambda$  близок к 1, и масштабирующая поправка незначительна в приложениях, где  $\sigma\sqrt{t} \leq \frac{1}{2}$  ( $\sigma$  — это эквивалентная предполагаемая волатильность Блэка — Шоулза и  $t$  — время до окончания опциона).

Замечательным образом устранились оба параметра, и  $l$ , и константа масштабирования  $\lambda$ .

**Результат 3: Цены опционов «пут»**

При  $K_1, K_2 \leq (1-l)S_0$ ,

$$P(K_2) = P(K_1) \frac{(-1)^{1-\alpha} S_0^{-\alpha} ((\alpha-1)K_2 + S_0) - (K_2 - S_0)^{1-\alpha}}{(-1)^{1-\alpha} S_0^{-\alpha} ((\alpha-1)K_1 + S_0) - (K_1 - S_0)^{1-\alpha}}. \quad (23.7)$$

### 23.4. ГРАНИЦЫ АРБИТРАЖА

Очевидно, что не существует арбитража для страйк-цен выше базовой  $K_1$  в приведенных выше формулах. Потому что мы можем подтвердить результат Бридена — Лиценбергера [32], где плотность реконструируется по второй производной опциона относительно страйк-цены  $\frac{\partial^2 C(K)}{\partial K^2} \Big|_{K \geq K_1} = \alpha K^{-\alpha-1} L^{-\alpha} \geq 0$ .

Однако остается возможность арбитража между страйк-ценами  $K_1 + \Delta K$ ,  $K_1$  и  $K_1 - \Delta K$ , если нарушить следующую границу. Пусть  $BSC(K, \sigma(K))$  — стоимость опциона «колл» по Блэку — Шоулзу при страйк-цене  $K$  и волатильности  $\sigma(K)$ , зависящей от страйк-цены и времени  $t$  до окончания срока. Тогда

$$C(K_1 + \Delta K) + BSC(K_1 - \Delta K) \geq 2C(K_1), \quad (23.8)$$

где  $BSC(K_1, \sigma(K_1)) = C(K_1)$ . Чтобы удовлетворить неравенству (23.8), требуется неравенство спредов «колл», взятое в пределе:

$$\frac{\partial BSC(K, \sigma(K))}{\partial K} \Big|_{K=K_1} \geq \frac{\partial C(K)}{\partial K} \Big|_{K=K_1}. \quad (23.9)$$

Такой арбитраж задает нижнюю границу для показателя хвоста  $\alpha$ . Предполагая для простоты нулевую доходность, получаем:

$$\alpha \geq \frac{1}{-\ln(K - S_0) + \ln t + \ln S_0} \ln \left( \frac{1}{2} \operatorname{erfc} \frac{t\sigma(K)^2 + 2\ln K - 2\ln S_0}{2\sqrt{2}\sqrt{t}\sigma(K)} - \frac{\sqrt{S_0}\sqrt{t}\sigma'(K)}{\sqrt{2\pi}} K^{\frac{\ln S_0}{t\sigma(K)^2} + \frac{1}{2}} \exp \left( -\frac{\ln^2 K + \ln^2 S_0}{2t\sigma(K)^2} - \frac{1}{8} t\sigma(K)^2 \right) \right). \quad (23.10)$$

### 23.5. КОММЕНТАРИИ

Как видно на Рисунке 23.5, модели стохастической волатильности и аналогичные адаптации (скажем, диффузионные скачки или варианты стандартного распределения Пуассона) обычно проваливаются «в хвостах», вне зоны начальной калибровки. Делались неудачные попытки экстраполировать цены опционов по подправленному тонкохвостому распределению, а не Парето; сделанные на этой основе многочисленные заявления в финансовой литературе о «переоценке» хвостовых опционов и псевдопсихологические рассуждения о «боязни риска» некорректны. Предложенные методы позволяют нам подойти к подобным заявлениям с большим реализмом.

И последнее. Заметим, что наш подход показывает не абсолютную погрешность цен для хвостовых опционов, а лишь погрешность относительно данной страйк-цены, которая ближе «к деньгам».

### БЛАГОДАРНОСТИ

Бруно Дюпиру, Питеру Карру, студентам Политехнического института Нью-Йоркского университета.

## ЧЕТЫРЕ ОШИБКИ В ФИНАНСОВОЙ МАТЕМАТИКЕ\*\*

Обсудим комментарии Джеффа Хоулмена (сделанные, между прочим, в то время, что он был старшим сотрудником по управлению рисками в крупном хедж-фонде), опубликованные в *Quantitative Finance*, чтобы проиллюстрировать четыре поучительные ошибки:

- 1) путать хвосты (4-й момент и выше) с волатильностью (2-й момент);
- 2) упускать из виду неравенство Йенсена при вычислении потенциала доходности;
- 3) анализировать результаты хеджирования в отрыве от эффективности исходного актива;
- 4) забывать о роли денег как масштаба цен.

Отзыв мистера Хоулмена на книгу «Антихрупкость» (4 декабря 2013) изобилует фактическими, логическими и аналитическими ошибками. Мы упомянем здесь только важные, затрагивающие сообщества специалистов по управлению рисками и финансовой математике; они поучительны как коренные ошибки, которые грозят начинающим финматематикам и риск-менеджерам.

### 24.1. ПУТАНИЦА МЕЖДУ ВТОРЫМ И ЧЕТВЕРТЫМ МОМЕНТАМИ

Начинающим важно избегать следующей элементарной ошибки. Мистер Хоулмен неправильно описывает связь между VIX (контрактами по индексу волатильности) и пари на хвостовые события. Вспомним, что наличие хвостовых событий свидетельствует о непропорционально сильном влиянии хвостов на свойства распределения, что математически равносильно слабому влиянию «туловища»<sup>1</sup>.

М-р Хоулмен, видимо, понял вторую часть насчет атрибутов жирнохвостости с точностью до наоборот. Неверно принимать VIX за индикатор хвостовых собы-

<sup>1</sup> Суть бросается в глаза в любой справочной таблице: эксцесс, который служит стандартной мерой жирнохвостости, вычисляется нормированием четвертого момента по квадрату дисперсии. — Прим. автора.

тий. Главное влияние на VIX оказывают опционы «на деньгах», соответствующие центру распределения и скорее близкие ко второму моменту, чем к четвертому (опционы «на деньгах» на самом деле имеют линейную платежную функцию и соответствуют условному первому моменту). Как объяснялось примерно 17 лет назад в *Динамическом хеджировании* (Талеб, 1997) (смотрите Приложение) в разговоре о таких хвостовых пари, или «ставках на четвертый момент», если вы ставите на жирность хвоста и большую роль экстремальных событий, продавайте опционы «вокруг денег» (такие, как контракты по VIX) и приобретайте опционы в хвостах с целью извлечь второй момент и достичь по нему нейтральности (в некотором роде «нейтральности к рынку»). Для нейтральности требуется «короткая волатильность» в туловище, потому что, будь эксцесс выше, событий в центре распределения стало бы меньше.

Математический механизм объясняется в технической серии *Incerto*: жирные хвосты означают «высокие пики» в распределении, потому что чем жирнее хвосты, тем больше времени рынок проводит между  $\mu - \sqrt{\frac{5-\sqrt{17}}{2}}\sigma$  и  $\mu + \sqrt{\frac{5-\sqrt{17}}{2}}\sigma$ , где  $\sigma$  — среднее квадратическое отклонение и  $\mu$  — среднее распределения (мы здесь использовали гауссиану для простоты, но аргументация применима ко всем однокорбым, колоколообразным распределениям, так называемым полувывпуклым вверх). И «высокие пики» означают меньше отклонений, отличных от хвостовых событий: больше спокойных времен, а не меньше. Чтобы прочувствовать последствия для цен опционов, рекомендую читателю контрольный вопрос, который предлагаю студентам после первой лекции по дериватам: «Что случится с опционами на деньгах, если ужирнить хвосты?»; правильный ответ: их стоимость упадет<sup>1</sup>.

Практически, как показано в более сложной аргументации в статье в «Квонтитив файненс» (Талеб и Дуади, 2013), наш коэффициент хрупкости имеет обратную чувствительность к событиям вокруг центра распределения, поскольку — как получается из рассмотрения вероятности выживания — что хрупко, то чувствительно к хвостовым шокам и не должно критически меняться в туловище (в противном случае оно бы разбилось).

## 24.2. НЕУЧЕТ НЕРАВЕНСТВА ЙЕНСЕНА ПРИ АНАЛИЗЕ ДОХОДНОСТИ ОПЦИОНОВ

Вот ошибка, которую любой ценой следует избегать при обсуждении стратегий волатильности, да и любых финансовых вопросов. М-р Хоулмен как будто упустил из виду неравенство Йенсена, без которого владеть опционом бессмысленно, как расписано во

<sup>1</sup> Технический вопрос: где начинается хвост? Как мы видели в Разделе 4.3, для широкого класса симметричных распределений с хвостами степенного закона хвосты начинаются в точках:

$$\pm \sqrt{\frac{5\alpha + \sqrt{(\alpha+1)(17\alpha+1)} + 1}{2(\alpha-1)}} \sigma, \text{ где } \alpha \text{ бесконечна в случае гауссианы со стохастической волатильностью и } \sigma —$$

среднеквадратическое отклонение. «Хвосты» располагаются между 2 и 3 среднеквадратическими отклонениями. Это вытекает из эвристического определения хрупкости как эффекта второго порядка: хвост — это часть распределения, выпуклая вниз по погрешности оценки коэффициента масштаба. На практике исторически сложившийся коэффициент, среднеквадратическая ошибка смещена вниз из-за эффектов малой выборки (как мы любим повторять, жирные хвосты подчеркивают эффекты малой выборки), и отклонения будут > 2–3 среднеквадратических ошибок. — Прим. автора.

всех подробностях в книге «Антихрупкость». Первый признак, что человек упускает из виду эффект выпуклости вниз, — критический просчет при наивном допущении, как опционы реагируют на VIX:

1 доллар инвестиций 1 января 2007 по стратегии покупать и откатывать краткосрочные фьючерсы VIX поднялся бы до 4,84 доллара 20 ноября 2008, а затем потерял 99 % стоимости за следующие четыре с половиной года, придя к 31 мая 2013 на уровне 0,05 доллара<sup>1</sup>.

В данном рассуждении доходность опциона недооценивается на порядки величины. М-р Хоулмен анализирует хвостовую стратегию, используя VIX (или фьючерсы VIX) как пример инвестиции в финансовые опционы, но такой пример математически хромает из-за эффектов второго порядка; с тем же успехом можно пытаться оценить инвестиции в горнолыжный курорт, анализируя температурные фьючерсы. Допустим, что опционная стратегия — периодический откат: опцион в пяти среднеквадратических отклонениях от денег<sup>2</sup> растет по стоимости в 16 раз, если его предположительная волатильность поднимается на 4, а теряет, если волатильность уходит в 0, всего лишь одну свою стоимость. При 10 среднеквадратических отклонениях разница уже будет в 144 раза. И, чтобы показать величину ускорения, заметим, что, если бы торговались опционы в двадцати среднеквадратических отклонениях, разница была бы в 210 000 раз<sup>3</sup>. Вторая критическая ошибка в данной дискуссии: м-р Хоулмен в своих вычислениях исключает платежи по опционам «в деньгах».

Нужно помнить, что VIX — не цена, а обратная функция; индекс, производный от цены. «Волатильность» не продается на рынке, как помидоры; операторы покупают опционы, соответствующие такого рода обратной функции, и зависимость далека от линейной. Хотя по сравнению с хвостовыми опционами VIX ближе к линейной зависимости, этот индекс тоже выпуклый вниз по фактической волатильности рынка, где-то между дисперсией и среднеквадратическим отклонением, поскольку «стрип опционов», охватывающий все страйк-цены, должен сравняться с дисперсией (Гезерал, 2006). Предлагаем желающим задачку. Пусть VIX «покупается», скажем, при 10 % — то есть входящие в него опционы покупаются при сочетании волатильностей, которое соответствует VIX на этом уровне. Пусть доходы в квадратах. Из-за нелинейности пакет может выиграть от эпизода 4 %-й волатильности, за которым следует эпизод 15 %, дав средний уровень 9,5 %; между тем м-р Хоулмен полагает или пытается убедить читателя, что случится убыток 0,5 процентного пункта — вопреки реальности, где неровности второго порядка в колебаниях волатильности важнее, чем этот эффект первого порядка.

1 М-р Хоулмен в той же дискуссии демонстрирует доказательства жалкой доходности индексного опциона «пут», который, как мы сказали выше, реагирует на волатильность, а не на хвостовые события. На биржевом жаргоне он называется «пут для молокососов» (*sucker put*). — *Прим. автора.*

2 Мы используем предположительную волатильность как точку отсчета для ее среднеквадратической ошибки. — *Прим. автора.*

3 Автор был свидетелем такого события: при ликвидации хедж-фонда Виктора Нидерхоффера опционы продавали по 0,05 доллара, а затем покупали обратно уже по 38 долларов; на этом обанкротилась брокерская фирма *Refco*, причем опционы так и не приблизились к деньгам: это был просто панический взлет предполагаемой волатильности. — *Прим. автора.*

### 24.3. НЕРАЗРЫВНАЯ СВЯЗЬ МЕЖДУ СТРАХОВКОЙ И ПРЕДМЕТОМ СТРАХОВАНИЯ

Нельзя вычислять стоимость страховки, не сдвигая ее на доходность, генерируемую пакетами, которые иначе не были бы куплены. Даже если бы м-р Хоулмен в приведенном выше примере правильно определил знак волатильности, все равно при анализе эффективности стратегии покупать опционы для защиты от хвостовых событий нельзя не добавить доходность самого портфеля; это все равно что подсчитывать затратную часть страховки без учета эффективности актива, составляющего предмет страхования, — без этого страховку не купили бы. За тот период, который он обсуждает, рынок вырос больше чем на 100 %; будет разумно сравнить доход с заработком инвестора (и, конечно, избавиться от возни с контрактами по VIX и сосредоточиться на мелких инвестициях в хвостовые опционы, где возможна напористая политика). Многие инвесторы (включая автора) держались бы подальше от рынка или перестали бы вкладывать в него дополнительные фонды, не будь такой страховки.

### 24.4. НЕОБХОДИМОСТЬ МАСШТАБА ЦЕН В ФИНАНСОВОЙ МАТЕМАТИКЕ

Есть и более глубокая ошибка в анализе.

«Штангу»<sup>1</sup> определяют в «Антихрупкости»<sup>2</sup> как инвестиционную стратегию с двугорбым профилем, по которому часть портфеля — вложения в «объект, сохраняющий стоимость» (защищенный от инфляции и пр.), а остальное — вложение в рискованные ценные бумаги. М-р Хоулмен пускается в нигилистический дискурс о несуществовании защищенного от риска инвестиционного объекта (в духе софистики: кто-то утверждает, что на суше обычно безопаснее, чем в море, но что, если вдруг землетрясение?).

Известные выкладки Блэка и Шоулза опираются на безрисковый актив как точку отсчета; литература примерно с 1977 года перешла с термина «кэш» (наличные деньги) на *numéraire* (объект, сохраняющий стоимость), а также приняла идею работы с разными валютами, что технически означает работу с несколькими мерами вероятности. *Numéraire*<sup>3</sup> определяется как условная денежная единица, в которую можно переводить все остальные денежные единицы (то есть практически это такая валютная корзина, что происходящие внутри ее отклонения не сказываются на благосостоянии инвестора). К сожалению, без объекта, сохраняющего стоимость, невозможно определить вероятностную меру и исследовать количественные показатели в финансовой математике; общая валютная единица, в которую переводится все остальное, необходима. В эмоциональной тираде м-р Хоулмен не только отвергает штангу, но и запрещает применять оператор математического ожидания к любой экономической

1 Жаргонный термин для стратегии портфеля с кратко- и долгосрочными бумагами без среднесрочных; от *barbell*, буквально гантеля или штанга в форме гантели.

2 «Если вы размещаете 90 процентов средств в скучной наличке (предположим, вы защищены от инфляции) или в чем-то вроде «объекта, сохраняющего стоимость», а 10 процентов — в очень рискованных, максимально рискованных ценных бумагах, вы не сможете потерять больше 10 процентов средств, в то время как ваши доходы могут быть велики» (перевод Н. Караева).

3 Читается «нумерёр»; от *fr.* счетное, числовое, номерное.

величине, тем самым нападая на все публикации по финансовой математике и сам журнал «Куонтитатив файненс», в котором публикует свою рецензию.

Очевидно, что гневная рецензия м-ра Хоулмена насыщена ошибочными и бестолковыми заявлениями, но я не сомневаюсь, что читатель «Куонтитатив файненс» их заметил, и, как мы уже говорили, цель этой дискуссии — научиться избегать аналитических ошибок в финансовой математике.

И последнее: автор приветствует критику со стороны финансового сообщества, если она содержит доводы по существу, а не беспорядочные выпады против ложных целей или, как в случае м-ра Хоулмена, против основ данной науки.

## 24.5. ПРИЛОЖЕНИЕ (СТАВКИ НА ХВОСТЫ РАСПРЕДЕЛЕНИЯ)

Из книги «Динамическое хеджирование», страницы 264–265:

Ставка на четвертый момент означает «лонг» или «шорт»<sup>1</sup> по «волатильности» «волатильности». Она делается с помощью опционов «вне денег» или с помощью календарных спредов<sup>2</sup>. Пример: «пропорциональный обратный спред»<sup>3</sup> — метод, при котором покупаются опционы «вне денег», в больших объемах, и продаются опционы «на деньгах», в меньших объемах, но с таким расчетом, чтобы обмен удовлетворял «кредитному правилу» (т. е. приносил кэш на начальном этапе). Кредитное правило требует большей возни, когда используются опционы «в деньгах». В этом случае нужно вычислить текущее значение основной части каждого опциона, пользуясь правилом пут-колл-паритета, чтобы пересчитать эти опционы по опционам «вне денег».

Обмен на Рисунке 24.1 был достигнут приобретением путов «вне денег» и коллов «вне денег» и продажей меньшего объема «стеллажей» «на деньгах» с тем же остающимся сроком.

На Рисунке 24.2 показан второй метод, в котором покупаются 60-дневные опционы в одном объеме и продаются, в объеме 80 % первого объема, 20-дневные опционы. Оба обмена создают позицию, выигрывающую от жирных хвостов и высоких пиков. Однако у этих обменов разная вега-чувствительность;<sup>4</sup> при этом модифицированная вега у них близка к плоской.

1 На жаргоне трейдеров сделка «лонг», или «длинная», — это покупка актива в надежде потом продать по выросшей цене, сделка «шорт», или «короткая», — это продажа актива в надежде потом купить его же по упавшей цене; от *long* и *short*.

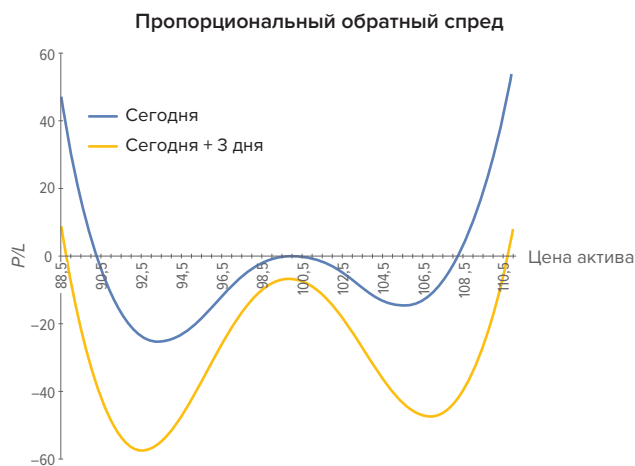
2 От *calendar*, или *calendar spread*. Продажа опциона с коротким сроком и покупка опциона с той же страйк-ценой, но с долгим сроком.

3 От *ratio backspread* или *ratio reverse spread*.

4  $V = \frac{\partial V}{\partial \sigma}$ , мера чувствительности стоимости  $V$  опциона к предполагаемой волатильности  $\sigma$  цены исходного актива. Обозначается нестандартной заглавной буквой «ню», представляющей собой увеличенных размеров строчную букву «ню»; чтение «вега» (*vega*) предположительно появилось из-за путаницы с латинской  $V$  и в подражание двусложным названиям греческих букв.

Есть несколько показателей риска опционов, обозначаемых греческими буквами и более экзотическими символами, под общим названием «греки» (*Greeks*); «вега»  $V$  принадлежит к «грекам» первого порядка.





**Рисунок 24.1:** Первый метод извлечь прибыль из четвертого момента (из книги «Динамическое хеджирование», 1997)



**Рисунок 24.2:** Второй метод извлечь прибыль из четвертого момента (из книги «Динамическое хеджирование», 1997)

Смотрите Раздел 4.3. Туловище, плечи и хвосты, где мы предполагаем, что хвосты начинаются при том уровне выпуклости вниз отрезка вероятностного распределения, который отвечает масштабу распределения.

## ОГРАНИЧЕНИЯ ХВОСТОВОГО РИСКА И МАКСИМАЛЬНАЯ ЭНТРОПИЯ<sup>†</sup>

Выбор портфеля анализировался в финансовой литературе, по сути, при двух центральных предположениях: о полном знании совместного вероятностного распределения доходов по ценным бумагам, составляющим искомый портфель, и о том, что предпочтения инвесторов выражаются функцией полезности. В реальном мире операторы строят портфели при ограничениях риска, выражаемых клиентами и регуляторами и сказывающихся на максимальных потерях, которые могут образоваться в заданный период при заданном доверительном уровне (так называемая стоимость под риском). Интересно, что в финансовой литературе серьезная дискуссия о том, как мало известно с теоретико-вероятностной точки зрения о многомерной плотности для доходов от активов, не считается важной.

Мы же, напротив, хотим высветить эти проблемы и затем уложить их в систему максимизации энтропии, чтобы представить реальную неосведомленность об «истинных» вероятностных распределениях, одномерных и многомерных, для доходов от торгуемых ценных бумаг. В этих рамках мы выясняем, каков оптимальный портфель при ряде ограничений риска понижающей тенденции. На свет выходят два интересных результата: (i) ограничения левого хвоста достаточно сильны, чтобы переопределить остальные соображения, рассматриваемые в обычной теории; (ii) «портфель штанги» (максимальная определенность / низкий риск в одном пакете и максимальная неопределенность в другом), хорошо знакомый трейдерам, в нашем построении появляется естественным образом.

### 25.1. РИСК ЛЕВОГО ХВОСТА КАК ЦЕНТРАЛЬНОЕ ОГРАНИЧЕНИЕ ПОРТФЕЛЯ

С точки зрения клиента — оператора или активного инвестора, работающего в институциональной системе параметров, принципиально использовать нормативные обязательные лимиты на хвостовые потери, чтобы задать уровни риска в портфелях; для

банков это обязательное требование регулятора со времен Базеля II<sup>1</sup>. Расчеты опираются на стресс-тесты, стоп-лоссы<sup>2</sup>, стоимость под риском (VaR), ожидаемые потери (ожидаемые при том условии, что превысят VaR; ожидаемые потери называют также условной стоимостью под риском, CVaR) и подобные методы ограничения потерь, не заботясь о полезности. В частности, запасы на финансовые транзакции калибруются клиринговыми фирмами и биржами по хвостовым потерям, которые оцениваются по вероятностному распределению и по стресс-тестам. (В терминологии активных инвестиций стоп-лосс — это ордер с целью полностью или частично выйти из рискованного вложения по триггеру, в случае заранее определенных номинальных потерь. Базелем II обычно называют рекомендации по банковским законам и нормативам, выпущенные Базельским комитетом по банковскому надзору. Стоимость под риском, VaR, определяется как величина  $K$  пороговых потерь, для которых вероятность того, что данный портфель в данном горизонте времени понесет потери тяжелее данного значения, составляет  $\varepsilon$ . Стресс-тест — это проверка эффективности в случае произвольно заданного отклонения параметров исходных активов.) Информация, заложенная в выборе константы, является, мягко говоря, важной статистической характеристикой аппетита к риску и формы желательного распределения.

Операторы озабочены не столько колебаниями портфеля, сколько риском истощения запасов в некотором временном окне. Кроме того, они не знают совместного вероятностного распределения компонентов в своем портфеле (не считая туманного представления об ассоциации и хеджах), но могут управлять потерями естественными методами размещения средств, учитывающими максимальный риск. (Идея заменить дисперсию на риск может показаться очень странной практикующим активным инвесторам. Цель современной портфельной теории, понизить дисперсию, несовместима с предпочтениями рационального инвестора, вопреки боязни риска, поскольку минимизирует также вариативность в области прибылей — кроме узкого круга ситуаций, где есть уверенность в средних будущих доходах, и натянутого примера с инвестором, которому разрешено вкладывать только в случайные величины с симметричным вероятностным распределением и/или симметричной платежной функцией. Стоп-лоссы и процедуры управления хвостовым риском нарушают такую симметрию.) Возможно, используются, хотя и не напрямую, и обычные идеи функции полезности и дисперсии, поскольку эта информация заложена в ограничении хвостовых потерь.

Поскольку методы управления рисками, подобные стоп-лоссу или VaR (и ожидаемых потерь), учитывают только один отрезок распределения: отрицательную часть области определения, — мы можем получить двойственный подход, подражая разделению портфеля на конструкцию в стиле «штанги», чтобы инвестор мог вкладываться в позиции в разных частях распределения доходов. Здесь наше определение «штанги» означает портфель со смесью двух экстремальных свойств, подобно линейной комбинации максимального консерватизма для доли  $w$  портфеля, где  $w \in (0, 1)$ , и максимального (или высокого) риска для оставшейся доли  $(1 - w)$ .

Исторически финансовая теория предпочитала параметрические, менее робастные методы. Идея, будто принимающий решения имеет ясное и безошибочное зна-

1 Принятые в 2004 рекомендации Базельского комитета по банковскому надзору в области банковского регулирования.

2 Англ. *stop order* или *stop-loss order* — приказ продавать или покупать некоторые акции, если они достигнут заданной стоп-цены; цель — ограничить убытки.

ние о распределении будущих платежей, оказалась живуча вопреки практической и теоретической несостоятельности — например, требуемые корреляции слишком неустойчивы, чтобы получать точные оценки. Такой подход опирается на уверенность в распределении и его параметрах, и это может быть интересно для исследователя, но не приспособлено к решениям об активных инвестициях. (Корреляции неустойчивы неустойчивым образом, поскольку совместные доходы по активам не эллиптические; смотрите у Бушо и Шишпортика (2012) [42].)

Есть, грубо говоря, две традиции: в одной, уважаемой экономическим истеблишментом (представленным по большей части Марковицем [166]), принятие решений сильно опирается на параметры, а в другой — на менее сильные допущения, известные как критерий Кэлли (Кэлли, 1956 [142], смотрите у Белла и Коувера, 1980 [15].) (В отличие от подхода минимальной дисперсии метод Кэлли — разработанный примерно в те же годы, что метод Марковица, — не требует знать совместное распределение или функцию полезности. На практике требуется динамически корректируемое отношение ожидаемой прибыли к доходам худшего случая, позволяющее избежать разорения. Очевидно, что погрешность моделирования имеет меньшие последствия при использовании критерия Кэлли; смотрите у Торпа (1969) [250], Хэя (2000) [119], Маклейна, Зембы и Блазенко [157]. Обсуждение различий между двумя подходами смотрите в возражении Самуэльсона против критерия Кэлли и логарифмических размеров у Торпа (2010) [252].) Кроме того, метод Кэлли связан и с управлением левым хвостом благодаря пропорциональным инвестициям, которые автоматически сокращают портфель в случае потерь; но для первоначального метода требуется жесткий, непараметрический сценарий наихудшего случая, то есть для ценных бумаг требуется нижняя граница отклонений, аналогично игре в казино, и достичь такого в финансовой области можно только с бинарными опционами. Вдобавок критерий Кэлли требует точного знания некоторых параметров будущих доходов, таких как среднее. Наш подход выходит за этот метод, приспособляясь к большей неопределенности относительно доходов, и оператор может управлять левым хвостом только через деривативы и прочие средства страхования или построив динамический портфель на основе стоп-лоссов. (Су, У, Цзян и Сон (2014) [266] противопоставляют среднюю дисперсию максимальной энтропии и используют энтропию для построения робастных портфелей.) Вкратце, мы прикручиваем к системе механизм прерывания потерь, но в остальном допускаем максимальную неопределенность относительно доходов. Точнее, мы приравняем распределение доходов к максимально энтропийному расширению ограничений, выраженных в статистических ожиданиях, на поведение в левом хвосте, а также на ожидание доходов или логарифма доходности в неопасной зоне. (Заметим, что все время используем шенноновскую энтропию<sup>1</sup>. Есть и другие меры информации, такие как энтропия Цаллиса<sup>2</sup> [256] и энтропия Реньи<sup>3</sup> [137],

1 Информационная энтропия по Шеннону; прирост энтропии приравнивается к утраченной неопределенности.

2 Энтропия Цаллиса — обобщение стандартной энтропии Гиббса на неэкстенсивные (неаддитивные) системы, где сильное взаимодействие создает новые степени свободы. Энтропия Гиббса обобщает энтропию Больцмана на случай, где микросостояния системы не равновероятны. Энтропия Больцмана, с точностью до постоянного множителя и выбора единиц измерения, — это логарифм числа микросостояний, отвечающих данному макросостоянию.

3 Энтропия Реньи — обобщение шенноновской энтропии, энтропии Хартли, энтропии столкновений и min-энтропии; допускает дробное число измерений фрактальной системы. Служит мерой разнообразия, неопределенности и случайности для широкого класса систем.

обобщающая шенноновскую энтропию; возможно, какие-то из этих мер более удобны для вычислений в отдельных случаях. Тем не менее шенноновская энтропия остается наиболее изученной мерой для систем максимизации.)

Здесь под «поведением в левом хвосте» подразумеваются жесткие, институциональные ограничения, которые мы обсудили выше. Мы описываем форму и исследуем прочие свойства результирующего распределения, так называемого *maxent*-распределения<sup>1</sup>. Помимо математического результата, выявляющего связь между допустимыми хвостовыми потерями (VaR) и ожидаемыми доходами в системе среднего и дисперсии гауссианы, наш вклад содержит две новации: (1) исследование формы распределения доходов от построения портфеля при более естественных ограничениях, чем те, что налагаются методом среднего и дисперсии, и (2) использование стохастической энтропии для представления остаточной неопределенности.

Методы VaR и CVaR не свободны от погрешности — параметрический VaR славится неэффективностью как отдельный метод контроля рисков. Однако эти методы можно сделать робастными, используя построения, которые, после уплаты цены за страховку, больше не зависят от параметрических предположений. Достичь этого можно с помощью деривативов или органического построения (очевидно, что, если иметь 80 % портфеля в ценных бумагах, сохраняющих стоимость, риск потерять более 20 % нулевой независимо от возможных моделей доходности; колебания в объекте, сохраняющем стоимость, не считаются рискованными).

Мы используем «чистую робастность» VaR и нулевые потери с помощью «жесткого упора» или страховки, представляющие в нашей статье особый случай построения, которые мы выше назвали «штангой».

Нелишне заметить, что, согласно давно известной в экономике идее, инвестор может построить портфель на основе двух разных категорий риска, смотрите у Хикса (1939) [128]. Современная портфельная теория выдвигает теорему о «разделении» взаимных фондов, гласящую, что все инвесторы способны получить желаемые портфели, смешивая два взаимных фонда, один с безрисковым активом и один с оптимальным портфелем по среднему и дисперсии, касательных к заданным ограничениям; смотрите у Тобина (1958) [254], Марковица (1959) [167] и, в других версиях, у Мертон (1972) [170] и Росса (1978) [202]. В нашем случае безрисковый актив — часть хвоста, в которой риск задан в точности равным нулю. Заметим, что в традиционной финансовой экономике у рискованной части портфеля должна быть минимальная дисперсия; по нашему методу для рискованной части представление прямо противоположное.

### 25.1.1. Штанга с точки зрения Э. Т. Джейнса

Наш подход — ограничивать только то, что поддается ограничению (робастным образом) и максимизировать энтропию во всем остальном, — вторит примечательному прозрению Э. Т. Джейнса в статье «Как нам следует использовать энтропию в экономике?» [134]:

Может оказаться, что макроэкономическая система не подчиняется (или не вполне подчиняется) стимулам, которые должны диктовать ее поведение согласно существующим теориям; возможно, она просто движется в направлении роста энтропии, ограниченного законами сохранения, которые наложены Природой и Правительством.

<sup>1</sup> Вероятностное распределение с максимальной энтропией (от *maximum entropy probability distribution*).

## 25.2. ВЕРНЕМСЯ К ЗАДАНИЮ СРЕДНЕГО И ДИСПЕРСИИ

Пусть  $\vec{X} = (X_1, \dots, X_m)$  обозначает доходы от актива за данный период с совместной плотностью  $g(\vec{x})$ , средними доходами  $\vec{\mu} = (\mu_1, \dots, \mu_m)$  и ковариационной матрицей  $\Sigma$  размера  $m \times m$ ,  $\Sigma_{ij} = \mathbb{E}(X_i X_j) - \mu_i \mu_j$ ,  $1 \leq i, j \leq m$ . Допустим, что  $\mu$  и  $\Sigma$  можно надежно оценить по данным.

Тогда доходы от портфеля с весами  $\vec{w} = (w_1, \dots, w_m)$  будут

$$X = \sum_{i=1}^m w_i X_i,$$

и среднее и дисперсия этой величины будут

$$\mathbb{E}(X) = \vec{w} \vec{\mu}^T, \mathbb{V}(X) = \vec{w} \Sigma \vec{w}^T.$$

В стандартной портфельной теории минимизируют  $\mathbb{V}X$  по всем  $\vec{w}$ , удовлетворяющим  $\mathbb{E}X = \mu$  при некотором фиксированном желательном среднем доходе  $\mu$ . Эквивалентным образом максимизируется ожидаемый доход  $\mathbb{E}X$ , удовлетворяющий некоторой фиксированной дисперсии  $\mathbb{V}X$ . В этой системе дисперсия служит суррогатом риска.

Чтобы проследить связь с нашим подходом на основе энтропии, рассмотрим следующие два стандартных случая.

(1) Мир нормального распределения: совместное распределение  $g(\vec{x})$  доходов от актива представляет собой многомерную гауссиану  $\mathcal{N}(\vec{\mu}, \Sigma)$ . Предположение о нормальности равносильно предположению о том, что у  $g(\vec{x})$  максимальная (шенноновская) энтропия среди всех многомерных распределений при заданных статистических параметрах первого и второго порядка,  $\vec{\mu}$  и  $\Sigma$ . Более того, при фиксированном среднем  $\mathbb{E}X$  минимизировать дисперсию  $\mathbb{V}X$  равносильно тому, чтобы минимизировать энтропию (неопределенность) величины  $X$ . (Это верно, поскольку совместная нормальность делает  $X$  одномерной нормальной величиной при любом выборе весов, а энтропия случайной величины с распределением  $\mathcal{N}(\mu, \sigma^2)$

$$H = \frac{1}{2} (1 + \ln 2\pi\sigma^2).$$

Это естественно в мире с полной информацией. (Идея энтропии как средней неопределенности выдвинута Филиппатосом и Уилсоном (1972) [188]; смотрите у Чжоу и др. (2013) [270] обзор по энтропии в финансовой экономике и у Джорджеску-Регена (1971) [107] по экономике в целом.)

(2) Неизвестное многомерное распределение. Поскольку мы предполагаем, что можем судить о структуре второго порядка, мы все еще можем выполнить программу Марковица, т. е. выбирать веса портфеля в поисках эффективности, оптимальной по среднему и дисперсии, задавая  $\mathbb{E}X = \mu$  and  $\mathbb{V}X = \sigma^2$ . Однако мы не знаем распределения величины дохода  $X$ . Заметим, что предположение о нормальном распределении  $X \sim \mathcal{N}(\mu, \sigma^2)$  равносильно предположению о максимальной энтропии  $X$ , поскольку, повторим, нормальное распределение имеет максимальную энтропию при данном среднем и дисперсии, смотрите [188].

Наша стратегия в том, чтобы обобщить второй сценарий, заменив дисперсию  $\sigma^2$  на два ограничения стоимости под риском в левом хвосте, и смоделировать доходы портфеля как максимально энтропийное расширение этих ограничений совместно с ограничением на общую эффективность или на рост портфеля в неопасной зоне.

### 25.2.1. Анализ ограничений

Пусть  $X$  имеет плотность вероятности  $f(x)$ . Далее везде будем считать, что  $K < 0$  — нормирующая константа, выбранная для согласования с капиталом активного инвестора. При любых  $\varepsilon > 0$  и  $v_- < K$  ограничения на стоимость под риском следующие.

(1) Хвостовая вероятность:

$$\mathbb{P}(X \leq K) = \int_{-\infty}^K f(x) dx = \varepsilon.$$

(2) Ожидаемые потери (CVaR):

$$\mathbb{E}(X \mid X \leq K) = v_-.$$

В предположении (1) ограничение (2) равносильно тому, что

$$\mathbb{E}(X \mid X \leq K) = \int_{-\infty}^K xf(x) dx = \varepsilon v_-.$$

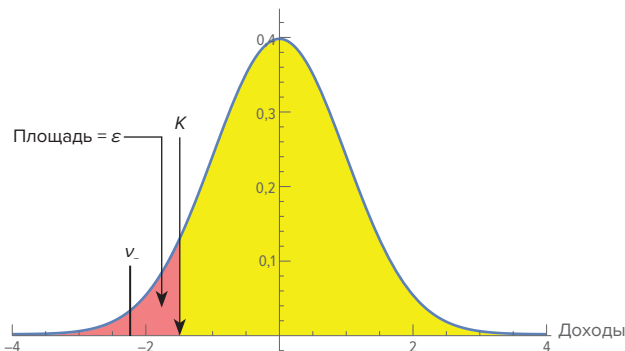
При заданных параметрах стоимости под риском  $\mathcal{G} = (K, \varepsilon, v_-)$ , пусть  $\Omega_{\text{var}}(\mathcal{G})$  обозначает множество функций плотности вероятности  $f$ , удовлетворяющих двум ограничениям. Заметим, что  $\Omega_{\text{var}}(\mathcal{G})$  выпукла вниз:

$$\text{если } f_1, f_2 \in \Omega_{\text{var}}(\mathcal{G}), \text{ то } \alpha f_1 + (1 - \alpha) f_2 \in \Omega_{\text{var}}(\mathcal{G}).$$

Позже мы добавим еще одно ограничение, касающееся общего среднего.

### 25.3. ВЕРНЕМСЯ К ГАУССОВУ СЛУЧАЮ

Пусть величина  $X$  — гауссова со средним  $\mu$  и дисперсией  $\sigma^2$ . В принципе будет возможно удовлетворить ограничениям VaR, поскольку есть два свободных параметра. В самом деле, как показано ниже, ограничения левого хвоста задают среднее и дисперсию; смотрите Рисунок 25.1. Однако выполнение ограничений VaR накладывает любопытные ограничения на  $\mu$  и  $\sigma$  и ведет к естественному неравенству в духе «бесплатных обедов не бывает».



**Рисунок 25.1:** Если задать  $K$  (стоимость под риском), вероятность  $\varepsilon$  превысить  $K$  и ожидаемые при этом потери, при гауссовом распределении не остается свободы:  $\sigma$  и  $\mu$  заданы, и построения, согласно портфельной теории, теряют значение

Пусть  $\eta(\varepsilon)$  — это  $\varepsilon$ -квантиль стандартного нормального распределения, т. е.  $\eta(\varepsilon) = \Phi^{-1}(\varepsilon)$ , где  $\Phi$  — кумулятивная функция стандартного нормального распределения, имеющего плотность  $\phi(x)$ . Кроме того, зададим

$$B(\varepsilon) = \frac{1}{\varepsilon \eta(\varepsilon)} \phi(\eta(\varepsilon)) = \frac{1}{\sqrt{2\pi} \varepsilon \eta(\varepsilon)} \exp\left(-\frac{\eta(\varepsilon)^2}{2}\right).$$

### Предложение 25.1

Если величина  $X \sim \mathcal{N}(\mu, \sigma^2)$  и удовлетворяет двум ограничениям VaR, то среднее и дисперсия имеют вид

$$\mu = \frac{v_- + KB(\varepsilon)}{1 + B(\varepsilon)}, \quad \sigma = \frac{K - v_-}{\eta(\varepsilon)(1 + B(\varepsilon))}.$$

Более того,  $B(\varepsilon) < -1$  и  $\lim_{\varepsilon \downarrow 0} B(\varepsilon) = -1$ .

Доказательство приведено в Приложении. Ограничения VaR непосредственно ведут к двум линейным уравнениям относительно  $\mu$  и  $\sigma$ :

$$\mu + \eta(\varepsilon)\sigma = K, \quad \mu - \eta(\varepsilon)B(\varepsilon)\sigma = v_-.$$

Рассмотрим условия, при которых ограничения VaR допускают *положительный* средний доход  $\mu = \mathbb{E}X > 0$ . Во-первых, из приведенного выше линейного уравнения относительно  $\mu$  и  $\sigma$ , содержащего  $\eta(\varepsilon)$  и  $K$ , мы видим, что  $\sigma$  растет<sup>1</sup> с ростом  $\varepsilon$  при любом фиксированном среднем  $\mu$  и что  $\mu > 0$  тогда и только тогда, когда  $\sigma > \frac{K}{\eta(\varepsilon)}$ , т. е. для дисперсии нужно принять нижнюю границу, растущую с  $\varepsilon$ , — это поведение выглядит логично. Во-вторых, из выражения для  $\mu$  в Предложении 25.1 видим, что

$$\mu > 0 \Leftrightarrow |v_-| > KB(\varepsilon)^2.$$

Следовательно, единственная возможность иметь положительный ожидаемый доход — принять достаточно большой риск, выраженный рядом компромиссов между параметрами риска  $\mathcal{R}$ , удовлетворяющими приведенному выше неравенству.

Этот тип ограничения применим также к более общему случаю симметричных распределений, поскольку ограничения на левые хвосты влияют через симметрию структуры на положение и масштаб. Например, в случае  $t$ -распределения Стюдента с коэффициентом масштаба  $s$ , коэффициентом положения  $m$  и показателем хвоста  $\alpha$  применимо такое же линейное соотношение между  $s$  и  $m$ :  $s = (K - m)\kappa(\alpha)$ , где

$$\kappa(\alpha) = -\frac{i \sqrt{\Gamma_{2\alpha}^{-1}\left(\frac{\alpha}{2}, \frac{1}{2}\right)}}{\sqrt{\alpha} \sqrt{\Gamma_{2\alpha}^{-1}\left(\frac{\alpha}{2}, \frac{1}{2}\right)} - 1}, \quad \Gamma_{(\cdot)}^{-1}(\cdot, \cdot) \text{ — обратная регуляризованная неполная бета-функция}$$

$$\text{ция и } s \text{ — решение уравнения } \varepsilon = \frac{1}{2} I_{\frac{\alpha s^2}{(k-m)^2 + \alpha s^2}}\left(\frac{\alpha}{2}, \frac{1}{2}\right).$$

1 В самом деле, в посылке Предложения 25.1 стоимость под риском,  $K$ , зафиксирована, а  $\varepsilon$ -квантиль,  $\eta(\varepsilon)$ , при повышении вероятности  $\varepsilon$  уменьшается.

2 В самом деле, величины  $v_-$ ,  $K$ ,  $B(\varepsilon)$  и  $1 + B(\varepsilon)$  все отрицательные. Чтобы дробь  $\mu = \frac{v_- + KB(\varepsilon)}{1 + B(\varepsilon)}$  была положительна, числитель должен быть отрицателен,  $v_- < -KB(\varepsilon)$ .



### 25.3.1. Смесь двух нормальных распределений

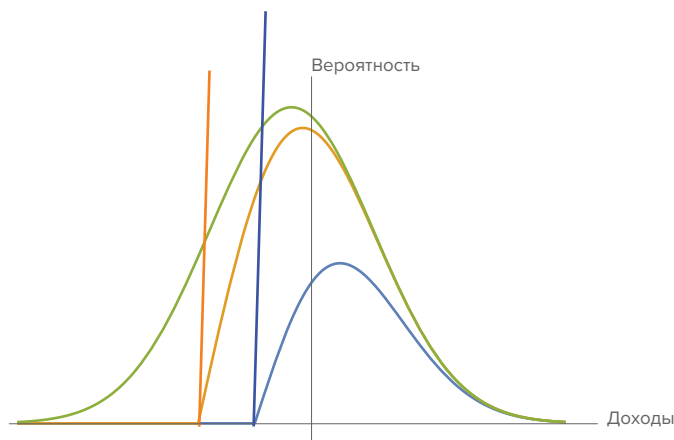
Во многих прикладных науках смесь двух нормальных распределений дает полезное и естественное обобщение гауссианы; в финансовой математике соответствующая модель получила название «гипотеза смешанного распределения» (*Mixture Distribution Hypothesis*, MDH) — имеется в виду смесь двух нормальных распределений, — и была широко исследована (смотрите, например, у Ричардсона и Смита (1995) [200]). Э. Жиман и Т. Ан (1996) [2] показывают, как бесконечная смесь нормальных распределений акций возникает при введении «стохастических часов», отвечающих за неравномерную интенсивность информационного потока на финансовых рынках. Кроме того, опционные трейдеры давно используют смешивание для представления жирных хвостов и оценки чувствительности портфеля к росту эксцесса (модель *DvegaDvol*); смотрите у Талеба (1997) [225]. Наконец, Бриго и Меркурио (2002) [34] с помощью смеси двух нормальных распределений откалибровали асимметрию в фоновых опционах.

Рассмотрим смесь

$$f(x) = \lambda \mathcal{N}(\mu_1, \sigma_1^2) + (1 - \lambda)(\mu_2, \sigma_2^2).$$

Интуитивно простым и привлекательным подходом кажется зафиксировать общее среднее  $\mu$  и принять  $\lambda = \varepsilon$  и  $\mu_1 = v_-$ ; в таком случае  $\mu_2$  придется взять  $\frac{\mu - \varepsilon v_-}{1 - \varepsilon}$ . Тогда ограничения левого хвоста приблизительно выполняются при достаточно малых  $\sigma_1, \sigma_2$ . В самом деле, когда  $\sigma_1 = \sigma_2 \approx 0$ , плотность по существу состоит из двух пиков (нормальных распределений низкой дисперсии), причем левый сцентрирован по  $v_-$  и правый сцентрирован по  $\frac{\mu - \varepsilon v_-}{1 - \varepsilon}$ . В крайнем случае слева имеем дельта-функцию Дирака, как мы увидим далее.

**Динамический стоп-лосс, краткий комментарий.** Можно задать такой уровень  $K$ , что ниже его не будет массы, и результаты будут зависеть от точности выполнения такого стоп-ордера. Распределение справа от стоп-лосса больше не похоже на стандартную гауссиану, поскольку скашивается в положительную сторону согласно расстоянию стоп-уровня от среднего. Завершим это обсуждение иллюстрациями на Рисунке 25.2.



**Рисунок 25.2:** Динамический стоп-лосс играет роль поглощающего барьера, создавая дельта-функцию Дирака в точке исполнения стоп-ордера

## 25.4. МАКСИМАЛЬНАЯ ЭНТРОПИЯ

Из комментариев и анализа, приведенных выше, очевидно, что на практике плотность  $f$  доходов  $X$  неизвестна; в частности, ни одна теория ее не находит. Допустим, что можем уточнять параметры портфеля, чтобы удовлетворить ограничениям VaR и, возможно, еще одному ограничению на ожидаемую величину некоторой функции  $X$  (например, общее среднее). Тогда мы бы хотели вычислить интересующие нас вероятности и математические ожидания, например,  $P(X > 0)$  или вероятность потерять больше  $2K$ , или ожидаемую прибыль при условии, что  $X > 0$ . Одна из возможных стратегий состоит в том, чтобы получить такие оценки и предсказания при самых непредсказуемых обстоятельствах, совместимых с ограничениями. То есть использовать максимально энтропийное расширение (*maximum entropy extension*, МЕЕ) ограничений как модель для  $f(x)$ .

«Дифференциальная энтропия» для  $f$  — это  $h(f) = -\int f(x) \ln f(x) dx$ . (В общем случае этот интеграл, возможно, не существует.) Энтропия выпукла вверх на том пространстве плотностей, для которого определена. В общем случае МЕЕ определяется как

$$f_{\text{МЕЕ}} = \operatorname{argmax}_{f \in \Omega} h(f),$$

где  $\Omega$  — пространство плотностей, удовлетворяющее набору ограничений вида  $\mathbb{E}\varphi_j(X) = c_j, j = 1, \dots, M$ . Как известно, в предположении непустого  $\Omega$  максимально-энтропийная  $f_{\text{МЕЕ}}$  единственная и (вдали от границ допустимой области) представима как экспоненциальное распределение функций ограничений, — то есть в виде

$$f_{\text{МЕЕ}}(x) = C^{-1} \exp \sum_j \lambda_j \varphi_j(x),$$

где  $C = C(\lambda_1, \dots, \lambda_M)$  — нормирующая константа. (Эта формула возникает, если продифференцировать соответствующий функционал  $J(f)$  относительно энтропии, приравняв интеграл к единице и наложить ограничения методом множителей Лагранжа.) В специальных случаях, приведенных ниже, мы используем это представление, чтобы найти МЕЕ при наших ограничениях.

В нашем случае мы хотим максимизировать энтропию при ограничениях VaR совместно с любыми другими, какие понадобятся наложить. В самом деле, ограничения VaR сами по себе не диктуют МЕЕ, поскольку не ограничивают плотность  $f(x)$  при  $x > K$ . Энтропию можно сделать неограниченно большой, допуская  $f$  быть тождественно равной  $C = \frac{1-\varepsilon}{N-K}$  по всем  $K < x < N$  и допуская  $N \rightarrow \infty$ . Но допустим, что добавили одно или несколько ограничений на поведение  $f$ , которые совместимы с ограничениями VaR в том смысле, что множество плотностей  $\Omega$ , удовлетворяющее всем ограничениям, не пусто. Здесь  $\Omega$  будет зависеть от параметров VaR,  $\vartheta = (K, \varepsilon, \nu_-)$ , и от параметров дополнительных ограничений.

### 25.4.1. Случай А: ограничиваем глобальное среднее

Простейший случай — когда мы добавляем ограничение на средний доход, т. е. фиксируем  $\mathbb{E}X = \mu$ .

Поскольку  $\mathbb{E}X = \mathbb{P}(X \leq K)\mathbb{E}(X | X \leq K) + \mathbb{P}(X > K)\mathbb{E}(X | X > K)$ , добавление ограничения на среднее равносильно добавлению ограничения

$$\mathbb{E}(X | X > K) = v_+,$$

где  $v_+$  удовлетворяет требованию  $\varepsilon v_- + (1 - \varepsilon)v_+ = \mu$ .

Определим

$$f_-(x) \begin{cases} \frac{1}{K - v_-} \exp\left(-\frac{K - x}{K - v_-}\right), & \text{если } x < K \\ 0, & \text{если } x \geq K \end{cases}$$

и

$$f_+(x) \begin{cases} \frac{1}{v_+ - K} \exp\left(-\frac{x - K}{v_+ - K}\right), & \text{если } x > K \\ 0, & \text{если } x \leq K. \end{cases}$$

Нетрудно проверить, что обе функции при интегрировании дают 1. Тогда

$$f_{\text{MEE}}(x) = \varepsilon f_-(x) + (1 - \varepsilon) f_+(x)$$

будет максимально энтропийным расширением при трех ограничениях. Во-первых, очевидно, что

$$1. \int_{-\infty}^K f_{\text{MEE}}(x) dx = \varepsilon;^1$$

$$2. \int_{-\infty}^K x f_{\text{MEE}}(x) dx = \varepsilon v_-;^2$$

$$3. \int_K^{\infty} x f_{\text{MEE}}(x) dx = (1 - \varepsilon) v_+;^3$$

то есть все ограничения соблюдены. Во-вторых,  $f_{\text{MEE}}(x)$  представима экспонентой от наших функций ограничений:

$$f_{\text{MEE}}(x) = C^{-1} \exp(-(\lambda_1 x + \lambda_2 \mathbb{1}_{x \leq K} + \lambda_3 x \mathbb{1}_{x \leq K})).$$

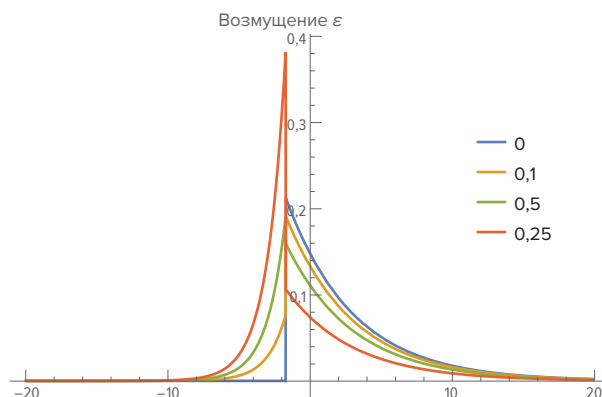
Форма  $f_-$  зависит от соотношения между  $K$  и ожидаемыми потерями  $v_-$ . Чем ближе  $v_-$  к  $K$ , тем быстрее спадает хвост. При  $v_- \rightarrow K$  функция  $f_-$  сходится к единичному пику<sup>4</sup> при  $x = K$  (Рисунки 25.3 и 25.4).

1 Это ограничение вида  $\mathbb{E}\varphi_1(X) = c_1$ , где  $\varphi_1(x) = f_-(x)$  и  $c_1 = \varepsilon$ .

2 Это ограничение вида  $\mathbb{E}\varphi_2(X) = c_2$ , где  $\varphi_2(x) = x f_-(x)$  и  $c_2 = \varepsilon v_-$ .

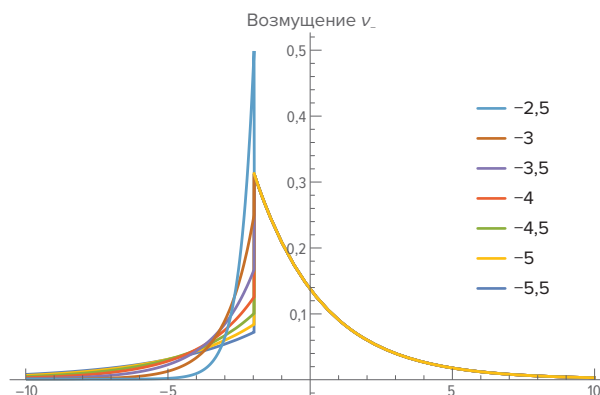
3 Это ограничение вида  $\mathbb{E}\varphi_3(X) = c_3$ , где  $\varphi_3(x) = x f_+(x)$  и  $c_3 = (1 - \varepsilon) v_+$ .

4 По-видимому, имеется в виду пик плотности, площадь под которым 1 и ширина которого  $\downarrow 0$ .



**Рисунок 25.3:** Случай А: Влияние различных значений  $\varepsilon$  на форму распределения\*

\* Использованы  $K = -1,75$ ,  $v_- = -3,05$ ,  $v_+ = 2,75$ .



**Рисунок 25.4:** Случай А: Влияние различных значений  $v_-$  на форму распределения\*

\* Примерно такое семейство графиков получается при  $K = -2$ ,  $\varepsilon = 0,25$ ,  $v_+ = 0,3$ .

### 25.4.2. Случай В: ограничиваем абсолютное среднее

Если вместо этого ограничить абсолютное среднее:

$$\mathbb{E}|X| = \int |x| f(x) dx = \mu,$$

то максимально энтропийное расширение будет не столь очевидно, но все-таки найти его можно. Определим  $f_-(x)$  как выше и примем

$$f_+(x) = \begin{cases} \frac{\lambda_1}{2 - \exp \lambda_1 K} \exp(-\lambda_1 |x|), & \text{если } x > K \\ 0, & \text{если } x \leq K \end{cases}.$$

Тогда можно выбрать  $\lambda_1$  так, что

$$\varepsilon v_- + (1 - \varepsilon) \int_K^\infty |x| f_+(x) dx = \mu.$$

### 25.4.3. Случай С: степенные законы для правого хвоста

Если мы полагаем, что фактические доходы имеют «жирные хвосты», в частности, что правый хвост убывает не экспоненциально (как плотность при нормальном и экспоненциальном распределениях), а по степенному закону, то можем добавить такое

ограничение к ограничениям VaR, вместо работы со средним или абсолютным средним. Глядя на экспоненциальное представление МЕЕ, понятно, что плотность  $f_+(x)$  будет иметь степенной закон, а именно

$$f_+(x) = \frac{1}{C(\alpha)} (1 + |x|)^{-(1+\alpha)}, \quad x \geq K$$

при  $\alpha > 0$ , если ограничение имеет вид

$$\mathbb{E}(\ln(1 + |X|) | X > K) = A.$$

Более того, еще раз обратимся к теории МЕЕ, пользуясь тем, что параметр получен минимизацией логарифма нормирующей функции. В этом случае легко показать, что

$$C(\alpha) = \int_K^\infty (1 + |x|)^{-(1+\alpha)} dx = \frac{1}{\alpha} (2 - (1-K)^{-\alpha}).$$

Отсюда следует, что  $A$  и  $\alpha$  удовлетворяют уравнению

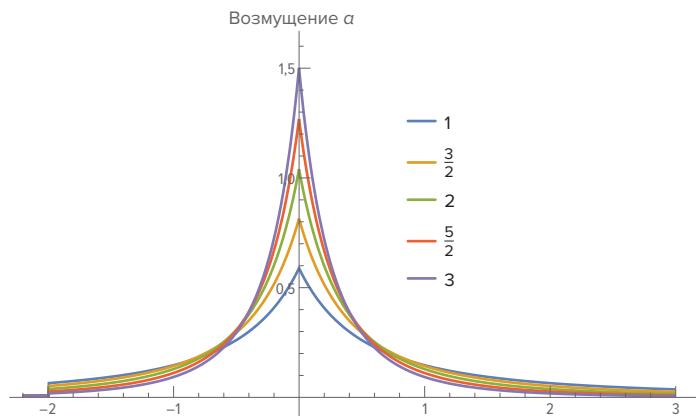
$$A = \frac{1}{\alpha} - \frac{\ln(1-K)}{2(1-K)^\alpha - 1}.$$

Мы можем относиться к этому уравнению как определяющему скорость спада  $\alpha$  при заданном  $A$  или же как определяющему значение ограничения  $A$ , необходимое для получения конкретного  $\alpha$  для степенного закона.

Тогда итоговое максимально энтропийное расширение ограничений VaR совместно с ограничением на логарифм доходности будет

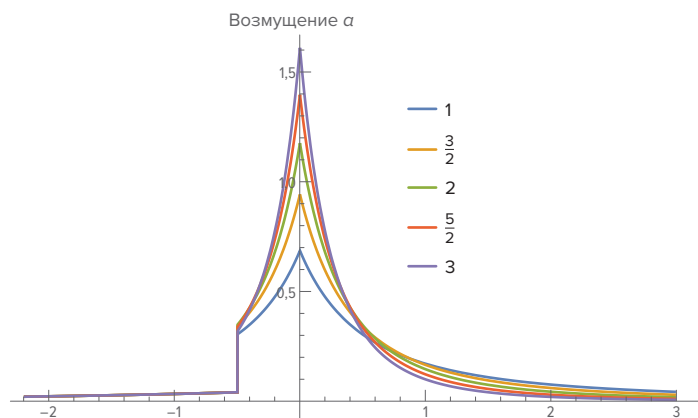
$$f_{\text{МЕЕ}}(x) = \varepsilon \mathbb{1}_{x \leq K} \frac{1}{K - v_-} \exp\left(-\frac{K-x}{K-v_-}\right) + (1-\varepsilon) \mathbb{1}_{x > K} \frac{(1+|x|)^{-(1+\alpha)}}{C(\alpha)}$$

(смотрите Рисунки 25.5 и 25.6).



**Рисунок 25.5:** Случай C: Влияние различных значений  $\alpha$  на форму жирного хвоста максимума энтропийного распределения\*

\* Примерно такое семейство графиков получается при  $K = -2$ ;  $\varepsilon = 0,01$ ;  $v_- = -3$ ;  $v_+ = 0$ . — Прим. перев.

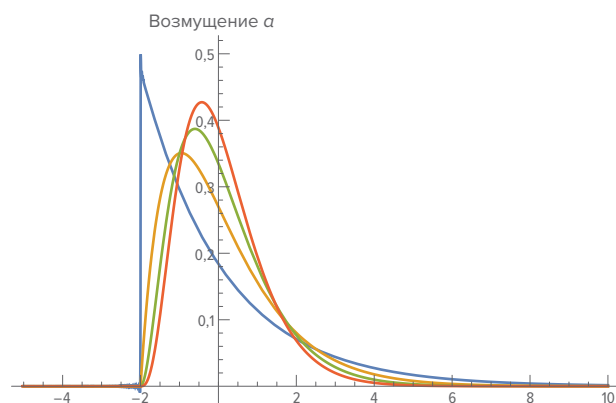


**Рисунок 25.6:** Случай С: Влияние различных значений  $\alpha$  на форму жирнотхвостого максимално энтропийного распределения (при  $K$ , более близком к нулю)\*

\* Примерно такое семейство графиков получается при  $K = -0,5$ ;  $\varepsilon = 0,1$ ;  $v_- = -3$ ;  $v_+ = 0$ . — Прим. перев.

#### 25.4.4. Расширение на несколько периодов: комментарий

Рассмотрим поведение в течение нескольких периодов. Используя наивный подход, мы суммируем эффективность так, как будто никакой реакции на прошлые доходы не происходит. Можно видеть, что Случай А приближается к обычной гауссиане, а Случай С этого не делает (Рисунок 25.7).



**Рисунок 25.7:** Средний доход за много периодов при наивной стратегии Случая А, где размер позиции не зависит от эффективности в прошлом и допустима независимая оценка размеров. Периоды четко складываются в стандартную гауссиану и, как показывает формула (25.1), сжимаются в дельта-функцию Дирака в точке среднего значения

Для Случая А характеристическую функцию можно записать как

$$\Psi^A(t) = \frac{e^{iKt} (t(K - \varepsilon v_- - (1 - \varepsilon)v_+) - i)}{(t(K - v_-) - i)(-1 + i t(v_+ - K))}.$$

С помощью свертки можно вывести, что  $\Psi^A(t)^n$  сходится к сумме  $n$  гауссиан. Характеристическая функция в пределе среднего стратегий, то есть

$$\lim_{n \rightarrow \infty} \Psi^A\left(\frac{t}{n}\right)^n = e^{it(v_+ - \varepsilon(v_+ - v_-))}, \quad (25.1)$$

представляет собой характеристическую функцию дельта-функции Дирака, демонстрируя заметный эффект закона больших чисел и обеспечивая тот же результат, что гауссово распределение со средним  $v_+ - \varepsilon(v_+ - v_-)$ .

В Случае С, при степенном законе, сходимость к гауссиане медленная и только при  $\alpha \geq 2$ .

## 25.5. КОММЕНТАРИИ И ЗАКЛЮЧЕНИЕ

Заметим, что стоп-лосс играет большую роль в определении стохастических свойств, чем в составлении портфеля. Упрощая, жесткий упор срабатывает не в ответ на отклонения отдельных компонент, а при отклонениях портфеля в целом. Это освобождает анализ от фокуса на отдельных компонентах портфеля, когда мы знаем только хвост и управляем только хвостом — через деривативы или органическое построение.

В заключение отметим, что большинство финансово-математических статей, в которых фигурирует энтропия, использовали как критерий оптимизации минимизацию энтропии. Например, Фрителли (2000) [98] выявляет однозначность «меры минимально энтропийного мартингала» при некоторых условиях и доказывает, что минимизация энтропии эквивалентна максимизации экспоненциальной полезности конечного капитала. Мы же, не обращаясь ни к какому критерию полезности, предложили максимизацию энтропии как признак неопределенности в распределении стоимости активов. При ограничениях VaR и ожидаемых потерь мы получили полную общность «портфеля штанги» как оптимального решения, расширив до весьма общей ситуации подход, известный по теореме о разделении на два фонда.

## 25.6. ПРИЛОЖЕНИЕ / ДОКАЗАТЕЛЬСТВА

*Доказательство* Предложения 25.1: Поскольку  $X \sim \mathcal{N}(\mu, \sigma^2)$ , ограничение хвостовой вероятности

$$\varepsilon = \mathbb{P}(X < K) = \mathbb{P}\left(Z < \frac{K - \mu}{\sigma}\right) = \Phi\left(\frac{K - \mu}{\sigma}\right).$$

По определению,  $\Phi(\eta(\varepsilon)) = \varepsilon$ . Значит,

$$K = \mu + \eta(\varepsilon)\sigma. \quad (25.2)$$

Для ограничения ожидаемых потерь

$$\begin{aligned} \mathbb{E}(X | X < k) &= \int_{-\infty}^K \frac{x}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx = \\ &= \mu\varepsilon + \sigma \int_{-\infty}^{(K - \mu)/\sigma} x\varphi(x) dx = \mu\varepsilon - \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{(K - \mu)^2}{2\sigma^2}\right). \end{aligned}$$

Поскольку  $\mathbb{E}(X | X < k) = \varepsilon v_-$ , и в силу определения  $B(\varepsilon)$ , получаем, что

$$v_- = \mu - \eta(\varepsilon)B(\varepsilon)\sigma. \quad (25.3)$$

Решение (25.2) и (25.3) относительно  $\mu$  и  $\sigma^2$  дает выражения в Предложении 25.1.

Наконец, в силу симметрии стандартного нормального распределения к «неравенству верхнего хвоста», получаем при  $x < 0$ , что  $\Phi(x) \leq \frac{\varphi(x)}{-x}$ . Выбором  $x = \eta(\varepsilon) = \Phi^{-1}(\varepsilon)$  достигаем  $\varepsilon = \mathbb{P}(X < \eta(\varepsilon)) \leq \varepsilon B(\varepsilon)$  или  $1 + B(\varepsilon) \leq 0$ . Поскольку неравенство верхнего хвоста является асимптотически точным, при  $x \rightarrow -\infty$  имеем  $B(0) = -1$ , на чем доказательство завершено.

## ПЕРСОНАЛИИ

Абрамовиц, Милтон (Milton Abramowitz, 1915–1958) и Айрин Энн Стеган (Irene Ann Stegun, 1919–2008) — американские математики.

Авельянеда, Марко (Marco Avellaneda, 1955–2022) — американский математик.

Амбре, Поль (Paul Embrechts, р. 1953) — бельгийский математик.

Баз, Джамиль (Jamil Baz, р. 1959) — ливанский и американский экономист; сотрудник Математического института Оксфордского университета.

Бак, Пер (Per Bak, 1948–2002) — датский физик-теоретик.

Барберис, Николас (Nicholas C. Barberis, р. 1971) — профессор финансов в Йельской школе менеджмента.

Бар-Ям, Янир (Yaneer Bar-Yam, р. 1959) — американский физик, специалист по сложным системам.

Баффетт, Уоррен (Warren Buffett; р. 1930) — американский инвестор; его состояние на ноябрь 2021 оценивалось в 105,2 млрд долларов (десятый самый богатый человек в мире).

Башелье, Луи Жан-Батист Альфонс (Louis Jean-Baptiste Alphonse Bachelier, 1870–1946) — французский математик.

Бейль, Пьер (Pierre Bayle, 1647–1706) — философ и богослов.

Белл, Роберт (Robert M. Bell) — американский физик, работает в AT&T Inc.

Бенарци, Шломо (Shlomo Benartzi, р. 1968) — израильский и американский поведенческий экономист.

Бернулли, Якоб (Jakob Bernoulli, 1655–1705) — швейцарский математик.

Бернштейн, Сергей Натанович (1880–1968) — советский математик. Предложил первую аксиоматику в теории вероятностей; продолжил исследования петербургской школы Чебышёва — Маркова по предельным теоремам; разработал теорию слабозависимых случайных величин; исследовал стохастические дифференциальные уравнения.

Биллингсли, Патрик (Patrick Paul Billingsley, 1925–2011) — американский математик и актер.

Блазенко, Джордж (George W. Blazenko) — канадский статистик, работает в Университете Британской Колумбии.

Блэк, Фишер (Fischer Sheffey Black, 1938–1995) — американский экономист.

Бокль, Генри Томас (Henry Thomas Buckle, 1821–1862) — английский историк, автор «Истории цивилизации в Англии».

Болкема, А. (A. Balkema) — нидерландский математик.

Больцман, Людвиг (Ludwig Eduard Boltzmann, 1844–1906) — австрийский физик.



Большев, Логин Николаевич (1922–1978) — советский математик, работал в области математической статистики и теории вероятностей.

Бонесс, Артур Джеймс (Arthur James Boness, 1900–1978) — американский экономист.

Борель, Эмиль (Émile Borel, 1871–1956) — французский математик.

Брайер, Гленн (Glenn Wilson Brier, 1913–1998) — американский метеоролог.

Брайтон, Генри (Henry Brighton) — сотрудник Института гуманитарных и цифровых исследований при международном университете Тилбурга (Нидерланды).

Брейман, Лео (Leo Breiman, 1928–2005) — американский математик.

Бриго, Дамиано (Damiano Brigo, р. 1966) — итальянский и британский экономист. Работал в миланском Институте Св. Павла, теперь в Имперском колледже Лондона.

Бриден, Дуглас (Douglas T. Breeden) — американский финансовый математик, профессор Дюкского университета, штат Северная Каролина.

Бронзин, Винченс (Vinzenc Bronzin, 1872–1970) — итальянский математик.

Вапник, Владимир Наумович (р. 1936) — советский и американский математик.

Варадхан, Шриниваса (Sathamangalam Ranga Iyengar Srinivasa Varadhan, р. 1940) — американский математик индийского происхождения.

Вега, Йозеф (Хосе) де ла (Joseph De La Vega, 1650–1692) — амстердамский торговец алмазами сефардского происхождения, известный как знаток финансов, моралист и поэт.

Вейбулл, Валодди (Waloddi Weibull, 1887–1979) — шведский инженер и математик.

Вейк, Й. Ван дер (J. van der Wijk) — нидерландский экономист, на его статью *Inkomensen Vermogensverdeling* в журнале Publication of the Nederlandsch Economisch Instituut (том 26, 1939) принято ссылаться в статистике.

Винсор, Чарльз (Charles Paine Winsor, 1895–1951) — американский инженер и биостатистик.

Витгенштейн, Людвиг (Ludwig Josef Johann Wittgenstein, 1889–1951) — австрийско-британский философ. Ему принадлежит такое замечание: то, что выглядит как проверка длины стола мерной планкой (Maßstab), иногда является проверкой мерной планки столом. В наше время Линейкой Витгенштейна называют принцип: если в модели много неизвестных параметров, применение модели будет работать не на оценку моделируемого явления, а на оценку модели.

Гаан, Лоренс де (Laurens de Haan, р. 1937) — нидерландский математик.

Габе, Ксавье (Xavier Gabaix, р. 1971) — французский и американский экономист. В 2003 в соавторстве опубликовал статью *A Theory of Power-Law Distributions in Financial Market Fluctuations*.

Ганн, Уильям Делберт (William Delbert Gann, 1878–1955) — американский трейдер, автор книг по биржевой торговле.

Гега, Доминик (Dominique Guéguan) — французский экономист.

Гезерал, Джим (Jim Gatheral) — американский ученый, специалист по финансовой математике.

Гелл-Манн, Мюррей (Murray Gell-Mann, 1929–2019) — американский физик-теоретик, автор основополагающей работы по странности и очарованию элементарных частиц.

Гиббс, Джозайя (Josiah Willard Gibbs, 1839–1903) — американский физик.

Гигеренцер, Герд (Gerd Gigerenzer; р. 1947) — немецкий психолог, член Леопольдины. Директор отдела «Адаптивное поведение и познание» и директор Хардинг-центра по оценке рисков в Институте человеческого развития Общества Макса Планка в Берлине. Работает над ограниченной рациональностью, эвристикой и эффективными деревьями принятия решений.

Гливенко, Валерий Иванович (1896–1940) — советский математик и логик.

Гнеденко, Борис Владимирович (1912–1995) — советский математик.

Голди, Чарльз (Charles M. Goldie, p. ~1940) — английский математик.

Голдстейн, Дэниэл (Daniel G. Goldstein, p. 1969) — американский ученый, специалист в когнитивной психологии. Внес вклад в теорию и эксперименты эвристики распознавания и эвристики выбора оптимального варианта.

Грам, Йёрген Педерсен (Jørgen Pedersen Gram, 1850–1916) — датский математик.

Губерман, Гур (Gur Huberman) — американский исследователь в области финансовой математики.

Гумбель, Эмиль Юлиус (Emil Julius Gumbel, 1891–1966) — немецкий и американский математик и политический публицист.

Дагум, Камило (Camilo Dagum, 1925–2005) — аргентинский математик, специалист по эконометрике.

Дебрё, Жерар (Gerard Debreu, 1921–2004) — американский экономист французского происхождения. Лауреат Нобелевской премии по экономике 1983 года.

Дембо, Амир (Amir Dembo, p. 1958) — израильский и американский математик.

Дерман, Эммануэль (Emanuel Derman, p. 1945) — южноафриканский и американский финансовый аналитик, бизнесмен и писатель.

Джейнс, Эдвин Томпсон (Edwin Thompson Jaynes, 1922–1998) — американский физик и специалист по математической статистике.

Джорджеску-Реген, Николас (Николае) (Nicholas Georgescu-Roegen, Nicolae Georgescu, 1906–1994) — румынский и американский экономист, математик и статистик.

Диаконис, Перси Уоррен (Persi Warren Diaconis, p. 1945) — американский математик и иллюзионист.

Донскер, Монро (Monroe David Donsker, 1924–1991) — американский математик.

Дуади, Рафаэль (Raphael Douady) — французский экономист.

Дюпир, Бруно (Bruno Dupire, p. 1958) — французский и американский финансовый математик.

Жибра, Робер (Robert Gibrat, 1904–1980) — французский инженер, сформулировал Правило пропорционального роста фирмы: скорость роста фирмы пропорциональна текущему размеру фирмы. В результате распределение фирм по размеру должно быть логнормальным.

Жиман, Дональд (Donald Jay Geman) — американский математик.

Жиман, Эльетт (Hélyette Geman) — французская исследовательница в области финансовой математики.

Звет, Виллем ван (Willem Rutger van Zwet, 1934–2020) — датский математик.

Зеитунни, Офер (Ofer Zeitouni, p. 1960) — израильский математик.

Земба, Уильям (William T. Ziemba) — канадский статистик, работает в Университете Британской Колумбии.

Золотарёв, Владимир Михайлович (1931–2019) — советский и российский математик.

Йор, Марк (Marc Yor; p. 1949–2014) — французский математик.

Ито, Киёси (Kiyosi Itô, 1915–2008) — японский математик.

Каземи, Хоссейн (Hossein B. Kazemi) — американский исследователь в области финансовой математики.

Канеман, Даниел (Daniel Kahneman, p. 1934) — израильско-американский психолог, один из основоположников поведенческой экономики.

Кантелли, Франческо Паоло (Francesco Paolo Cantelli, 1875–1966) — итальянский математик.

Карамата, Йован (Јован Карамата, 1902–1967) — сербский математик.

Карацас, Иоан (Ioannis Karatzas, р. ~1954) — американский математик греческого происхождения.

Карр, Питер (Peter Carr, 1958–2022) — американский математик, один из авторов модели цен деривативов CGMY и соответствующего распределения, «умеренно устойчивого».

Кассуф, Шин (Sheen T. Kassouf, 1928–2005) — американский экономист.

Кейнс, Джон Мейнард (John Maynard Keynes, 1883–1946) — английский экономист, основатель кейнсианского направления в экономической науке.

Келли, Джон (John Larry Kelly Jr., 1923–1965) — американский инженер и математик, автор финансовой стратегии ставок, известной как критерий Келли.

Кириан, Бенедикт Ф. (Benedict F. Kiernan, р. 1953) — австралийский историк.

Козлов, Валерий Васильевич (р. 1950) — советский и российский математик и механик.

Колмогоров, Андрей Николаевич (1903–1987) — советский математик, один из основоположников современной теории вероятностей.

Коувер, Томас (Thomas M. Cover, 1938–2012) — американский физик, работал в Стэнфордском университете.

Краммер, Харальд (Harald Cramér, 1893–1985) — шведский математик и статистик.

Крейн, Гарри (Harry Crane) — американский математик, сотрудник Ратгерского университета, Нью-Джерси.

Куммер, Эрнст (Ernst Eduard Kummer, 1810–1893) — немецкий математик.

Кэлли, Джон (John Larry Kelly Jr., 1923–1965) — американский физик, работал в Лабораториях Белла.

Лебег, Анри Леон (Henri Léon Lebesgue, 1875–1941) — французский математик. Интеграл Лебега обобщает традиционный интеграл, чтобы интегрировать быстро меняющиеся функции; при этом, по Лебегу, разрешается мельче разбивать область там, где интегрируемая функция изменяется быстрее.

Леви, Поль Пьер (Paul Pierre Lévy) — французский математик. В 1937 году в книге *Théorie de l'addition des variables aléatoires* дал обобщенную версию ЦПТ.

Лежандр, Адриен Мари (Adrien-Marie Legendre, 1752–1833) — французский математик.

Лейбсон, Дэвид (David I. Laibson, р. 1966) — американский экономист, один из крупнейших специалистов в области нейроэкономики и поведенческой экономики.

Линник, Юрий Владимирович (1914–1972) — советский математик в области теории вероятностей, математической статистики и теории чисел.

Лиценбергер, Роберт (Robert Litzenberger) — американский финансовый математик, профессор Уортонской школы бизнеса при Пенсильванском университете.

Ломакс К. (K. S. Lomax) — британский и американский статистик. На его статью *Business Failures: Another Example of the Analysis of Failure Data* (1954) принято ссылаться среди исследователей распределений Парето.

Лоу, Эндрю (Andrew Wen-Chuan Lo, р. 1960) — профессор финансов и экономики в Школе менеджмента Слоуна при Массачусетском технологическом институте.

Лоэв, Мишель (Michel Loève, 1907–1979) — французский и американский математик.

Лулакис, Михалис (Michalis Loulakis) — греческий и американский математик.

Лундберг, Филип (Ernst Filip Oskar Lundberg, 1876–1965) — шведский математик и статистик.

Ляпунов, Александр Михайлович (1857–1918) — русский математик и механик, основоположник теории устойчивости равновесия и движения механических систем с конечным числом параметров. Также работал в области дифференциальных уравнений, гидродинамики, теории вероятностей.

Мадан, Дилип (Dilip Madan) — американский финансовый математик, профессор Мэрилендского университета в Колледж-Парке, один из основателей *Bachelier Finance Society* и руководитель семинаров *Quant*.

Мадала, Гангадхаррао (Gangadharrao Soundalayarao Maddala, 1933–1999) — индийский и американский экономист.

Маккин, Генри (Henry P. McKean, Jr., p. 1930) — американский математик.

Маклейн, Л. (L. C. Maclean) — канадский статистик, работает в Университете Далхаузи.

Мандельброт, Бенуа (Benoît B. Mandelbrot, 1924–2010) — французский и американский математик, создатель фрактальной геометрии.

Манн, Читпунит (Chitpuneet Mann) — американский финансовый аналитик, сотрудник *Universa Investments*.

Марков, Андрей Андреевич (1856–1922) — русский математик, внесший вклад в теорию вероятностей, математический анализ и теорию чисел. Отец Андрея Андреевича Маркова-младшего (1903–1979), основоположника советской школы конструктивной математики.

Марковиц, Гарри Макс (Harry Max Markowitz, p. 1927) — американский экономист, предложивший в 1952 теорию, которая высоко оценивает эффективность инвестиционного портфеля.

Марченко, Владимир Александрович (p. 1922) — советский и украинский математик.

Маскерони, Лоренцо (Lorenzo Mascheroni, 1750–1800) — итальянский математик.

Мейер, Корнелис (Cornelis Simon Meijer, 1904–1974) — голландский математик.

Меллин, Ялмар (Robert Hjalmar Mellin, 1854–1933) — финский математик.

Меркурио, Фабио (Fabio Mercurio, p. 1966) — итальянский и американский экономист. Работал в миланском Институте Св. Павла, теперь в нью-йоркском Центре количественной аналитики при корпорации Блумберга.

Мёрт, Анри Дойч де ла (Henri Deutsch de la Meurthe, 1846–1919) — французский бизнесмен («нефтяной король Европы») и заядлый сторонник ранней авиации.

Мертон, Роберт (Robert Carhart Merton, p. 1944) — американский экономист, лауреат Нобелевской премии по экономике 1997 года.

Мехра, Раджниш (Rajnish Mehra, p. 1950) — индийский и американский специалист по финансовой экономике.

Миланович, Бранко (Branko Milanović, p. 1953) — сербско-американский экономист.

Миллс, Фредерик (Frederick Cecil Mills, 1892–1964) — американский экономист и статистик.

Муавр, Абрахам де (Abraham de Moivre, 1667–1754) — английский математик французского происхождения, ученик Ньютона.

Нагаев, Сергей Викторович (p. 1932) — советский и российский математик, автор трудов в области цепей Маркова и больших отклонений.

Нидерхоффер, Виктор (Victor Niederhoffer, p. 1943) — американский экономист и менеджер.

Никодим, Отто Мартин (Otton Marcin Nikodym, 1887–1974) — польский математик.

О’Конелл, Мартин (Martin P. O’Connell) — автор книги *The Business of Options: Time-Tested Principles and Practice*, Wiley, 2001.

Оксендал, Бернт (Bernt Karsten Øksendal, p. 1945) — норвежский математик.

Орнштейн (Орнстин), Леонард (Leonard Salomon Ornstein, 1880–1941) — нидерландский физик.

Пастур, Леонид Андреевич (p. 1937) — советский и украинский математик и физик.

Петерс, Оле (Ole Peters) — специалист по неэргодическим стохастическим процессам, сотрудник Лондонской математической лаборатории.

Петров, Валентин Владимирович (р. 1931) — советский и российский математик, специалист в области предельных теорем теории вероятностей и вероятностных неравенств.

Пикетти, Том (Thomas Piketty, р. 1971) — французский экономист.

Пикэндс, Джеймс (James Pickands III, ок. 1932–2022) — американский математик.

Питерс, Оле (Ole Peters) — современный англо-американский математик и физик, специалист в теории игр.

Питман, Эдвин (Edwin James George Pitman, 1897–1993) — австралийский математик.

Поппер, Карл (Karl Raimund Popper, 1902–1994) — австрийский и британский философ и социолог.

Поттерс, Марк (Marc Potters) — физик и один из руководителей *Capital fund Management*.

Похгаммер, Лео Август (Leo August Pochhammer, 1841–1920) — немецкий математик.

Прескотт, Эдвард (Edward C. Prescott, р. 1940) — американский экономист, лауреат Нобелевской премии 2004 года.

Радон, Иоганн (Johann Karl August Radon, 1887–1956) — австрийский математик.

Робертс, Расс (Russell David «Russ» Roberts, р. 1954) — сотрудник Гуверовского института при Стэнфордском университете и ведущий подкаста EconTalk.

Рейнах, Энтони (Anthony M. Reinach) — американский экономист.

Реньи, Альфред (Rényi Alfréd, 1921–1970) — венгерский математик.

Ричардсон, Мэтью (Matthew Richardson) — американский экономист, работал в Уортонской школе бизнеса при Пенсильванском университете, теперь в Стерновской школе бизнеса при Нью-Йоркском университете.

Романовский, Всеволод Иванович (1879–1954) — русский и советский математик, основатель ташкентской математической школы, автор трудов по математической статистике.

Росс, Стивен Алан (Stephen Alan Ross, 1944–2017) — американский экономист.

Рубинштейн, Марк (Mark Edward Rubinstein, 1944–2019) — финансовый экономист и финансовый инженер.

Руффино, Дориана (Doriana Ruffino) — американский экономист, занимающая высокие посты в Федеральной резервной системе США.

Саймон, Герберт Александер (Herbert Alexander Simon, 1916–2001) — американский ученый в области социальных, политических и экономических наук.

Салмон, Феликс (Felix Salmon, р. 1972) — американский финансовый журналист.

Самородницкий, Геннадий Пенхосович (р. ~1954) — советский и американский шахматист, математик и физик.

Самуэльсон, Пол (Paul Anthony Samuelson, 1915–2009) — американский экономист, лауреат Нобелевской премии по экономике 1970 года.

Санстейн, Касс (Cass Robert Sunstein, р. 1954) — американский ученый, специалист в области права и поведенческой экономики.

Слуцкий, Евгений Евгеньевич (1880–1948) — российский и советский математик, статистик и экономист. Один из создателей современной теории случайных функций.

Смирнов, Николай Васильевич (1900–1966) — советский математик, один из создателей непараметрических методов математической статистики и теории предельных распределений порядковых статистик.

Смит, Том (Thomas (Tom) Smith, р. 1958) — австралийский экономист.

Сон, Сюэфэн (Xuefeng Song, р. 1980) — китайский и финский специалист по математической статистике.

Сорнетт, Дидье (Didier Sornette, p. 1957) — французский и швейцарский физик и экономист.

Спренкл, Кейс (Case Middleton Sprenkle, p. 1935) — американский экономист, автор первой модели ценообразования опционов.

Столл, Ганс (Hans Reiner Stoll, 1939–2020) — американский экономист германского происхождения.

Стэнли, Юджин (Harry Eugene Stanley, p. 1941) — американский физик, специалист по статистике.

Су, Инйин (Yingying Xu) — китайский и американский специалист по эконометрике.

Такку, Мюрад (Murad Taqqu, p. ~1924) — американский математик иракского происхождения.

Талер, Ричард (Richard H. Thaler, p. 1945) — американский экономист. Лауреат премии по экономике памяти Альфреда Нобеля 2017 года за вклад в область поведенческой экономики.

Тверски (Тверский), Амос (Amos Tversky, 1937–1996) — израильский психолог, пионер когнитивной науки.

Типпетт, Леонард (Leonard Tippett, 1902–1985) — английский статистик.

Тобин, Джеймс (James Tobin; 1918–2002) — американский экономист, лауреат премии по экономике памяти Альфреда Нобеля 1981 года.

Тойгелс, Йозеф (Jozef L. Teugels) — бельгийский математик.

Торп, Эдвард (Edward Oakley Thorp, p. 1932) — американский математик и экономист, специалист в теории игр.

Трюссар, Джонатан (Jonathan Treussard) — американский экономист, занимающий высокие посты в финансовых корпорациях.

У, Чжуву (Zhuwu Wu) — китайский и американский специалист по математической статистике.

Уишарт, Джон (John Wishart, 1898–1956) — шотландский математик и статистик.

Уленбек, Джордж (Георг) (George Eugene Uhlenbeck, 1900–1988) — нидерландский и американский физик.

Учайкин, Владимир Васильевич (p. 1941) — советский и российский математик и физик.

Феллер, Уильям (William Feller, 1906–1970) — американский математик.

Феррейра, Ана (Ana Ferreira, p. ~1970) — португальский специалист по стохастическим процессам.

Филиппатос, Джордж (George C. Philippatos, 1938–2016) — американский экономист, работал в Уортонской школе бизнеса при Пенсильванском университете.

Финетти, Бруно де (Bruno de Finetti, 1906–1985) — итальянский математик, специалист по теории вероятностей и математической статистике.

Фишер, Роналд Эйлмер (Sir Ronald Aylmer Fisher, 1890–1962) — английский статистик и биолог.

Фишхофф, Барух (Baruch Fischhoff, p. 1946) — американский математик и психолог, один из основоположников теории принятия решений и специалист в области управления рисками.

Фонтанари, Андреа (Andrea Fontanari) — итальянский экономист.

Фрайд, Сидни (Sidney Fried) — автор биржевых учебников.

Фреше, Морис Рене (Maurice René Fréchet, 1878–1973) — французский математик.

Фрителли, Марко (Marco Frittelli) — современный итальянский экономист, сотрудник Миланского университета.



Хайек, Фридрих Август фон (Friedrich August von Hayek, 1899–1992) — австро-британский экономист и политический философ. Представитель новой австрийской школы экономики. Обосновал преимущество свободных цен перед плановой экономикой.

Хёффдинг, Василий (Wassily Hoeffding, 1914–1991) — финский математик.

Хикс, Джон (Sir John Richard Hicks, 1904–1989) — британский экономист, лауреат премии по экономике памяти Альфреда Нобеля 1972 года.

Хинчин, Александр Яковлевич (1894–1959) — советский математик.

Хог, Эспен (Espen Gaarder Haug) — норвежский математик.

Хэй, Джон (John Haigh) — британский статистик, работает в Университете Сассекса.

Цаллис, Константино (Constantino Tsallis, p. 1943) — бразильский физик греческого происхождения.

Цварц, Берт (Bert Zwart) — сотрудник Центра математики и информатики, Амстердам.

Цвивович, Диего (Diego Zviovich) — американский прикладной математик, специалист по риску.

Цзян, Лон (Long Jiang) — китайский специалист по математической статистике.

Ципф (Зиф), Джордж Кингсли (George Kingsley Zipf, 1902–1950) — американский лингвист и экономист.

Чебышёв, Пафнутий Львович (1821–1894) — основоположник петербургской математической школы.

Червоненкис, Алексей Яковлевич (1938–2014) — советский и российский ученый в области информатики.

Чернов, Герман (Herman Chernoff, p. 1923) — американский математик.

Чжоу, Ронси (Rongxi Zhou) — китайский математик, работает в Пекинском научно-техническом университете.

Чирилло, Паскуале (Pasquale Cirillo) — итальянский экономист.

Чистяков, Владимир Павлович (p. 1934) — советский и российский математик.

Шарлье, Карл (Carl Charlier, 1862–1934) — шведский астроном.

Шеннон, Клод (Claude Elwood Shannon, 1916–2001) — американский инженер, криптоаналитик и математик.

Ширяев, Альберт Николаевич (p. 1934) — советский и российский математик, специалист в области стохастических процессов и финансовой математики.

Шишпортик, Реми (Rémy Chicheportiche) — французский экономист.

Шоулз, Майрон (Myron Samuel Scholes, p. 1941) — американский экономист канадского происхождения, лауреат Нобелевской премии 1997 года.

Шрив, Стивен (Steven Eugene Shreve, p. ~1950) — американский германист и математик.

Эйлер, Леонард (Leonhard Euler, 1707–1783) — швейцарский, прусский и российский математик и механик.

Элиазар, Иддо Исаак (Iddo Isaac Eliazar, p. ~1970) — израильский математик.

Эрмит, Шарль (Charles Hermite, 1822–1901) — французский математик.

Эрроу, Кеннет (Kenneth Joseph Arrow, 1921–2017) — американский экономист, лауреат премии по экономике памяти Альфреда Нобеля 1972 года.

Юл, Джордж Удни (George Udny Yule, 1871–1951) — шотландский статистик.

Юм, Дэвид (David Hume, 1711–1776) — шотландский философ.

Яркин, Брэндон (Brandon Yarkin) — американский экономист и инвестор, главный оперативный директор *Universa Investments*, Майами.

## БИБЛИОГРАФИЯ

- [1] Inmaculada B Aban, Mark M Meerschaert, and Anna K Panorska. Parameter estimation for the truncated pareto distribution. *Journal of the American Statistical Association*, 101(473):270–277, 2006.
- [2] Thierry Ané and Hélyette Geman. Order flow, transaction clock, and normality of asset returns. *The Journal of Finance*, 55(5):2259–2284, 2000.
- [3] Kenneth J Arrow, Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O Ledyard, Saul Levmore, Robert Litan, Paul Milgrom, Forrest D Nelson, et al. The promise of prediction markets. *Science*, 320(5878):877, 2008.
- [4] Marco Avellaneda, Craig Friedman, Richard Holmes, and Dominick Samperi. Calibrating volatility surfaces via relative-entropy minimization. *Applied Mathematical Finance*, 4(1):37–64, 1997.
- [5] L. Bachelier. Theory of speculation in: P. Cootner, ed., 1964, *The random character of stock market prices*,. MIT Press, Cambridge, Mass, 1900.
- [6] Louis Bachelier. *Théorie de la spéculation*. Gauthier-Villars, 1900.
- [7] Kevin P Baland and HL MacGillivray. Kurtosis: a critical review. *The American Statistician*, 42(2):111–119, 1988.
- [8] August A Balkema and Laurens De Haan. Residual life time at great age. *The Annals of probability*, pages 792–804, 1974.
- [9] August A Balkema and Laurens De Haan. Limit distributions for order statistics. i. *Theory of Probability & Its Applications*, 23(1):77–92, 1978.
- [10] August A Balkema and Laurens de Haan. Limit distributions for order statistics. ii. *Theory of Probability & Its Applications*, 23(2):341–358, 1979.
- [11] Shaul K Bar-Lev, Idit Lavi, and Benjamin Reiser. Bayesian inference for the power law process. *Annals of the Institute of Statistical Mathematics*, 44(4):623–639, 1992.
- [12] Nicholas Barberis. The psychology of tail events: Progress and challenges. *American Economic Review*, 103(3):611–616, 2013.
- [13] Jonathan Baron. *Thinking and deciding*, 4th Ed. Cambridge University Press, 2008.
- [14] Norman C Beaulieu, Adnan A Abu-Dayya, and Peter J McLane. Estimating the distribution of a sum of independent lognormal random variables. *Communications, IEEE Transactions on*, 43(12):2869, 1995.
- [15] Robert M Bell and Thomas M Cover. Competitive optimality of logarithmic investment. *Mathematics of Operations Research*, 5(2):161–166, 1980.
- [16] Shlomo Benartzi and Richard Thaler. Heuristics and biases in retirement savings behavior. *Journal of Economic perspectives*, 21(3):81–104, 2007.



- [17] Shlomo Benartzi and Richard H Thaler. Myopic loss aversion and the equity premium puzzle. *The quarterly journal of Economics*, 110(1):73–92, 1995.
- [18] Shlomo Benartzi and Richard H Thaler. Naive diversification strategies in defined contribution saving plans. *American economic review*, 91(1):79–98, 2001.
- [19] Sergei Natanovich Bernshtein. Sur la loi des grands nombres. *Communications de la Société mathématique de Kharkow*, 16(1):82–87, 1918.
- [20] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.
- [21] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- [22] Nicholas H Bingham, Charles M Goldie, and Jef L Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.
- [23] Giulio Biroli, J-P Bouchaud, and Marc Potters. On the top eigenvalue of heavy-tailed random matrices. *EPL (Europhysics Letters)*, 78(1):10001, 2007.
- [24] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. 81:637–654, May–June 1973.
- [25] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654, 1973.
- [26] A. J. Boness. Elements of a theory of stock-option value. 72:163–175, 1964.
- [27] Jean-Philippe Bouchaud, Marc Mézard, Marc Potters, et al. Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, 2002.
- [28] Jean-Philippe Bouchaud and Marc Potters. *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge University Press, 2003.
- [29] Olivier Bousquet, Stéphane Boucheron, and Gábor Lugosi. Introduction to statistical learning theory. In *Advanced lectures on machine learning*, pages 169–207. Springer, 2004.
- [30] George Bragues. Prediction markets: The practical and normative possibilities for the social production of knowledge. *Episteme*, 6(1):91–106, 2009.
- [31] D. T. Breeden and R. H. Litzenberger. Price of state-contingent claims implicit in option prices. 51:621–651, 1978.
- [32] Douglas T Breeden and Robert H Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651, 1978.
- [33] Henry Brighton and Gerd Gigerenzer. Homo heuristicus and the bias– variance dilemma. *In Action, Perception and the Brain*, pages 68–91. Springer, 2012.
- [34] Damiano Brigo and Fabio Mercurio. Lognormal-mixture dynamics and calibration to market volatility smiles. *International Journal of Theoretical and Applied Finance*, 5(04):427–446, 2002.
- [35] Peter Carr. *Bounded brownian motion*. NYU Tandon School of Engineering, 2017.
- [36] Peter Carr, Hélyette Geman, Dilip B Madan, and Marc Yor. Stochastic volatility for lévy processes. *Mathematical finance*, 13(3):345–382, 2003.
- [37] Peter Carr and Dilip Madan. *Optimal positioning in derivative securities*. 2001.
- [38] Lars-Erik Cederman. Modeling the size of wars: from billiard balls to sandpiles. *American Political Science Review*, 97(01):135–150, 2003.
- [39] Bikas K Chakrabarti, Anirban Chakraborti, Satya R Chakravarty, and Arnab Chatterjee. *Econophysics of income and wealth distributions*. Cambridge University Press, 2013.
- [40] David G Champenowne. A model of income distribution. *The Economic Journal*, 63(250):318–351, 1953.
- [41] Shaohua Chen, Hong Nie, and Benjamin Ayers-Glassey. Lognormal sum approximation with a variant of type iv pearson distribution. *IEEE Communications Letters*, 12(9), 2008.
- [42] Rémy Chicheportiche and Jean-Philippe Bouchaud. The joint distribution of stock returns is not elliptical. *International Journal of Theoretical and Applied Finance*, 15(03), 2012.

- [43] VP Chistyakov. A theorem on sums of independent positive random variables and its applications to branching random processes. *Theory of Probability & Its Applications*, 9(4):640–648, 1964.
- [44] Pasquale Cirillo. Are your data really pareto distributed? *Physica A: Statistical Mechanics and its Applications*, 392(23):5947–5962, 2013.
- [45] Pasquale Cirillo and Nassim Nicholas Taleb. Expected shortfall estimation for apparently infinite-mean models of operational risk. *Quantitative Finance*, pages 1–10, 2016.
- [46] Pasquale Cirillo and Nassim Nicholas Taleb. On the statistical properties and tail risk of violent conflicts. *Physica A: Statistical Mechanics and its Applications*, 452:29–45, 2016.
- [47] Pasquale Cirillo and Nassim Nicholas Taleb. What are the chances of war? *Significance*, 13(2):44–45, 2016.
- [48] Pasquale Cirillo and Nassim Nicholas Taleb. Tail risk of contagious diseases. *Nature Physics*, 2020.
- [49] Open Science Collaboration et al. Estimating the reproducibility of psychological science. *Science*, 349(6251): aac4716, 2015.
- [50] Rama Cont and Peter Tankov. *Financial modelling with jump processes*, volume 2. CRC press, 2003.
- [51] Harald Cramér. On the mathematical theory of risk. *Centraltryckeriet*, 1930.
- [52] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.
- [53] Camilo Dagum. Inequality measures between income distributions with applications. *Econometrica*, 48(7):1791–1803, 1980.
- [54] Camilo Dagum. *Income distribution models*. Wiley Online Library, 1983.
- [55] Anirban DasGupta. *Probability for statistics and machine learning: fundamentals and advanced topics*. Springer Science & Business Media, 2011.
- [56] Herbert A David and Haikady N Nagaraja. *Order statistics*, 2003.
- [57] Bruno De Finetti. *Probability, induction, and statistics*, 1972.
- [58] Bruno De Finetti. *Philosophical Lectures on Probability: collected, edited, and annotated by Alberto Mura*, volume 340. Springer Science & Business Media, 2008.
- [59] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38. Springer Science & Business Media, 2009.
- [60] Kresimir Demeterfi, Emanuel Derman, Michael Kamal, and Joseph Zou. A guide to volatility and variance swaps. *The Journal of Derivatives*, 6(4):9–32, 1999.
- [61] Kresimir Demeterifi, Emanuel Derman, Michael Kamal, and Joseph Zou. More than you ever wanted to know about volatility swaps. Working paper, Goldman Sachs, 1999.
- [62] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953, 2007.
- [63] E. Derman and N. Taleb. The illusion of dynamic delta replication. *Quantitative Finance*, 5(4):323–326, 2005.
- [64] Emanuel Derman. The perception of time, risk and return during periods of speculation. Working paper, Goldman Sachs, 2002.
- [65] Marco Di Renzo, Fabio Graziosi, and Fortunato Santucci. Further results on the approximation of log-normal power sum via pearson type iv distribution: a general formula for log-moments computation. *IEEE Transactions on Communications*, 57(4), 2009.
- [66] Persi Diaconis and David Freedman. On the consistency of bayes estimates. *The Annals of Statistics*, pages 1–26, 1986.

[67] Persi Diaconis and Sandy Zabell. Closed form summation for classical distributions: variations on a theme of de moivre. *Statistical Science*, pages 284–302, 1991.

[68] Cornelius Frank Dietrich. *Uncertainty, calibration and probability: the statistics of scientific and industrial measurement*. Routledge, 2017.

[69] NIST Digital Library of Mathematical Functions. <http://dlmf.nist.gov/>, Release 1.0.19 of 2018–06–22. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller and B. V. Saunders, eds.

[70] Daniel Dufresne. Sums of lognormals. In *Proceedings of the 43rd actuarial research conference*. University of Regina, 2008.

[71] Daniel Dufresne et al. The log-normal approximation in financial and other computations. *Advances in Applied Probability*, 36(3):747–773, 2004.

[72] Bruno Dupire. Pricing with a smile. 7(1), 1994.

[73] Bruno Dupire. Exotic option pricing by calibration on volatility smiles. In *Advanced Mathematics for Derivatives: Risk Magazine Conference*, 1995.

[74] Bruno Dupire et al. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[75] Danny Dyer. Structural probability bounds for the strong pareto law. *Canadian Journal of Statistics*, 9(1):71–77, 1981.

[76] Iddo Eliazar. Inequality spectra. *Physica A: Statistical Mechanics and its Applications*, 469:824–847, 2017.

[77] Iddo Eliazar. Lindy’s law. *Physica A: Statistical Mechanics and its Applications*, 486:797–805, 2017.

[78] Iddo Eliazar and Morrel H Cohen. On social inequality: Analyzing the rich–poor disparity. *Physica A: Statistical Mechanics and its Applications*, 401:148–158, 2014.

[79] Iddo Eliazar and Igor M Sokolov. Maximization of statistical heterogeneity: From shannon’s entropy to gini’s index. *Physica A: Statistical Mechanics and its Applications*, 389(16):3023–3038, 2010.

[80] Iddo I Eliazar and Igor M Sokolov. Gini characterization of extreme-value statistics. *Physica A: Statistical Mechanics and its Applications*, 389(21):4462–4472, 2010.

[81] Iddo I Eliazar and Igor M Sokolov. Measuring statistical evenness: A panoramic overview. *Physica A: Statistical Mechanics and its Applications*, 391(4):1323–1353, 2012.

[82] Paul Embrechts. *Modelling extremal events: for insurance and finance*, volume 33. Springer, 1997.

[83] Paul Embrechts and Charles M Goldie. On convolution tails. *Stochastic Processes and their Applications*, 13(3):263–278, 1982.

[84] Paul Embrechts, Charles M Goldie, and Noël Veraverbeke. Subexponentiality and infinite divisibility. *Probability Theory and Related Fields*, 49(3):335–347, 1979.

[85] M Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo* (1884–1940), 27(1):247–271, 1909.

[86] Michael Falk et al. On testing the extreme value index via the pot-method. *The Annals of Statistics*, 23(6):2013–2035, 1995.

[87] Michael Falk, Jürg Hüsler, and Rolf-Dieter Reiss. *Laws of small numbers: extremes and rare events*. Springer Science & Business Media, 2010.

[88] Kai-Tai Fang. Elliptically contoured distributions. *Encyclopedia of Statistical Sciences*, 2006.

[89] Doyne James Farmer and John Geanakoplos. Hyperbolic discounting is rational: Valuing the far future with uncertain discount rates, 2009.

[90] J Doyne Farmer and John Geanakoplos. Power laws in economics and elsewhere. In *Santa Fe Institute*, 2008.

- [91] William Feller. 1971an introduction to probability theory and its applications, vol. 2.
- [92] William Feller. An introduction to probability theory, 1968.
- [93] Baruch Fischhoff, John Kadvany, and John David Kadvany. Risk: A very short introduction. Oxford University Press, 2011.
- [94] Ronald Aylmer Fisher and Leonard Henry Caleb Tippet. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In Mathematical Proceedings of the Cambridge Philosophical Society, volume 24, pages 180–190. Cambridge University Press, 1928.
- [95] Andrea Fontanari, Pasquale Cirillo, and Cornelis W Oosterlee. From concentration profiles to concentration maps. new tools for the study of loss distributions. Insurance: Mathematics and Economics, 78:13–29, 2018.
- [96] Shane Frederick, George Loewenstein, and Ted O’donoghue. Time discounting and time preference: A critical review. Journal of economic literature, 40(2):351–401, 2002.
- [97] David A Freedman. Notes on the dutch book argument “. Lecture Notes, Department of Statistics, University of Berkley at Berkley, <http://www.stat.berkeley.edu/~census/dutchdef.pdf>, 2003.
- [98] Marco Frittelli. The minimal entropy martingale measure and the valuation problem in incomplete markets. Mathematical finance, 10(1):39–52, 2000.
- [99] Xavier Gabaix. Power laws in economics and finance. Technical report, National Bureau of Economic Research, 2008.
- [100] Xavier Gabaix. Power laws in economics: An introduction. Journal of Economic Perspectives, 30(1):185–206, 2016.
- [101] Armengol Gasull, Maria Jolis, and Frederic Utzet. On the norming constants for normal maxima. Journal of Mathematical Analysis and Applications, 422(1):376–396, 2015.
- [102] Jim Gatheral. The Volatility Surface: a Practitioner’s Guide. John Wiley & Sons, 2006.
- [103] Jim Gatheral. The Volatility Surface: A Practitioner’s Guide. New York: John Wiley & Sons, 2006.
- [104] Oscar Gelderblom and Joost Jonker. Amsterdam as the cradle of modern futures and options trading, 1550–1650. William Goetzmann and K. Geert Rouwenhorst, 2005.
- [105] Andrew Gelman and Hal Stern. The difference between “significant” and “not significant” is not itself statistically significant. The American Statistician, 60(4):328–331, 2006.
- [106] Donald Geman, Hélyette Geman, and Nassim Nicholas Taleb. Tail risk constraints and maximum entropy. Entropy, 17(6):3724, 2015.
- [107] Nicholas Georgescu-Roegen. The entropy law and the economic process, 1971. Cambridge, Mass, 1971.
- [108] Gerd Gigerenzer and Daniel G Goldstein. Reasoning the fast and frugal way: models of bounded rationality. Psychological review, 103(4):650, 1996.
- [109] Gerd Gigerenzer and Peter M Todd. Simple heuristics that make us smart. Oxford University Press, New York, 1999.
- [110] Corrado Gini. Variabilità e mutabilità. Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi, 1912.
- [111] BV Gnedenko and AN Kolmogorov. Limit Distributions for Sums of Independent Random Variables (1954).
- [112] Charles M Goldie. Subexponential distributions and dominated-variation tails. Journal of Applied Probability, pages 440–442, 1978.
- [113] Daniel Goldstein and Nassim Taleb. We don’t quite know what we are talking about when we talk about volatility. Journal of Portfolio Management, 33(4), 2007.
- [114] Richard C Green, Robert A Jarrow, et al. Spanning and completeness in markets with contingent claims. Journal of Economic Theory, 41(1):202–210, 1987.

- [115] Emil Julius Gumbel. *Statistics of extremes*, 1958.
- [116] Laurens Haan and Ana Ferreira. *Extreme value theory: An introduction*. Springer Series in Operations Research and Financial Engineering, 2006.
- [117] Wolfgang Hafner and Heinz Zimmermann. Amazing discovery: Vincenz bronzin's option pricing models. 31:531–546, 2007.
- [118] Torben Hagerup and Christine Rüb. A guided tour of chernoff bounds. *Information processing letters*, 33(6):305–308, 1990.
- [119] John Haigh. The kelly criterion and bet comparisons in spread betting. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(4):531–539, 2000.
- [120] Peter Hall. On the rate of convergence of normal extremes. *Journal of Applied Probability*, 16(2):433–439, 1979.
- [121] Mahmoud Hamada and Emiliano A Valdez. Capm and option pricing with elliptically contoured distributions. *Journal of Risk and Insurance*, 75(2):387–409, 2008.
- [122] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge university press, 1952.
- [123] J Michael Harrison and David M Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, 20(3):381–408, 1979.
- [124] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*, springer series in statistics, 2009.
- [125] Espen G. Haug. *Derivatives: Models on Models*. New York: John Wiley & Sons, 2007.
- [126] Espen Gaarder Haug and Nassim Nicholas Taleb. Option traders use (very) sophisticated heuristics, never the black–scholes–merton formula. *Journal of Economic Behavior & Organization*, 77(2):97–106, 2011.
- [127] Friedrich August Hayek. The use of knowledge in society. *The American economic review*, 35(4):519–530, 1945.
- [128] John R Hicks. *Value and capital*, volume 2. Clarendon press Oxford, 1939.
- [129] Leonard R. Higgins. *The Put-and-Call*. London: E. Wilson, 1902.
- [130] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.
- [131] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.
- [132] HM James Hung, Robert T O'Neill, Peter Bauer, and Karl Kohne. The behavior of the p-value when the alternative hypothesis is true. *Biometrics*, pages 11–22, 1997.
- [133] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.
- [134] E. T. Jaynes. *How should we use entropy in economics?* 1991.
- [135] Johan Ludwig William Valdemar Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906.
- [136] Hedegaard Anders Jessen and Thomas Mikosch. Regularly varying functions. *Publications de l'Institut Mathématique*, 80(94):171–192, 2006.
- [137] Petr Jizba, Hagen Kleinert, and Mohammad Shefaat. Rényi's information transfer between financial time series. *Physica A: Statistical Mechanics and its Applications*, 391(10):2971–2989, 2012.
- [138] Valen E Johnson. Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences*, 110(48):19313–19317, 2013.
- [139] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

- [140] Joseph P Kairys Jr and NICHOLAS VALERIO III. The market for equity options in the 1870s. *The Journal of Finance*, 52(4):1707–1723, 1997.
- [141] Ioannis Karatzas and Steven E Shreve. *Brownian motion and stochastic calculus* springer-verlag. New York, 1991.
- [142] John L Kelly. A new interpretation of information rate. *Information Theory, IRE Transactions on*, 2(3):185–189, 1956.
- [143] Gideon Keren. Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica*, 77(3):217–273, 1991.
- [144] Christian Kleiber and Samuel Kotz. *Statistical size distributions in economics and actuarial sciences*, volume 470. John Wiley & Sons, 2003.
- [145] Andrei Nikolaevich Kolmogorov. On logical foundations of probability theory. In *Probability theory and mathematical statistics*, pages 1–5. Springer, 1983.
- [146] Andrey Kolmogorov. Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.*, 4:83–91, 1933.
- [147] Samuel Kotz and Norman Johnson. *Encyclopedia of Statistical Sciences*. Wiley, 2004.
- [148] VV Kozlov, T Madsen, and AA Sorokin. Weighted means of weakly dependent random variables. *MOSCOW UNIVERSITY MATHEMATICS BULLETIN C/C OF VESTNIK-MOSKOVSKII UNIVERSITET MATHEMATIKA*, 59(5):36, 2004.
- [149] Jean Laherrere and Didier Sornette. Stretched exponential distributions in nature and economy: “fat tails” with characteristic scales. *The European Physical Journal B-Condensed Matter and Complex Systems*, 2(4):525–539, 1998.
- [150] David Laibson. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478, 1997.
- [151] Deli Li, M Bhaskara Rao, and RJ Tomkins. The law of the iterated logarithm and central limit theorem for l-statistics. Technical report, PENNSYLVANIA STATE UNIV UNIVERSITY PARK CENTER FOR MULTIVARIATE ANALYSIS, 1997.
- [152] Sarah Lichtenstein, Baruch Fischhoff, and Lawrence D Phillips. Calibration of probabilities: The state of the art. In *Decision making and change in human affairs*, pages 275–324. Springer, 1977.
- [153] Sarah Lichtenstein, Paul Slovic, Baruch Fischhoff, Mark Layman, and Barbara Combs. Judged frequency of lethal events. *Journal of experimental psychology: Human learning and memory*, 4(6):551, 1978.
- [154] Michel Loève. *Probability Theory. Foundations. Random Sequences*. New York: D. Van Nostrand Company, 1955.
- [155] Filip Lundberg. I. Approximerad framställning af sannolikhetsfunktionen. II. Återförsäkring af kollektivrisker. Akademisk afhandling... af Filip Lundberg,... Almqvist och Wiksells boktryckeri, 1903.
- [156] HL MacGillivray and Kevin P Balanda. Mixtures, myths and kurtosis. *Communications in Statistics-Simulation and Computation*, 17(3):789–802, 1988.
- [157] LC MacLean, William T Ziemba, and George Blazenko. Growth versus security in dynamic investment analysis. *Management Science*, 38(11):1562–1585, 1992.
- [158] Dhruv Madeka. Accurate prediction of electoral outcomes. arXiv preprint arXiv:1704.02664, 2017.
- [159] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. The m4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4):802–808, 2018.
- [160] Spyros Makridakis and Nassim Taleb. *Decision making and planning under low levels of predictability*, 2009.



- [161] Benoit Mandelbrot. A note on a class of skew distribution functions: Analysis and critique of a paper by ha simon. *Information and Control*, 2(1):90–99, 1959.
- [162] Benoit Mandelbrot. The pareto-levy law and the distribution of income. *International Economic Review*, 1(2):79–106, 1960.
- [163] Benoit Mandelbrot. The stable paretian income distribution when the apparent exponent is near two. *International Economic Review*, 4(1):111–115, 1963.
- [164] Benoit B Mandelbrot. New methods in statistical economics. In *Fractals and Scaling in Finance*, pages 79–104. Springer, 1997.
- [165] Benoît B Mandelbrot and Nassim Nicholas Taleb. *Random jump, not random walk*, 2010.
- [166] Harry Markowitz. Portfolio selection\*. *The journal of finance*, 7(1):77–91, 1952.
- [167] Harry M Markowitz. *Portfolio selection: efficient diversification of investments*, volume 16. Wiley, 1959.
- [168] RARD Maronna, Douglas Martin, and Victor Yohai. *Robust statistics*. John Wiley & Sons, Chichester. ISBN, 2006.
- [169] R. Mehera and E. C. Prescott. The equity premium: a puzzle. *Journal of Monetary Economics*, 15:145–161, 1985.
- [170] Robert C Merton. An analytic derivation of the efficient portfolio frontier. *Journal of financial and quantitative analysis*, 7(4):1851–1872, 1972.
- [171] Robert C. Merton. The relationship between put and call prices: Comment. 28(1):183–184, 1973.
- [172] Robert C. Merton. Theory of rational option pricing, 4:141–183, Spring 1973.
- [173] Robert C. Merton. Option pricing when underlying stock returns are discontinuous. 3:125–144, 1976.
- [174] Robert C Merton and Paul Anthony Samuelson. *Continuous-time finance*, 1992.
- [175] David C Nachman. Spanning and completeness with options. *The review of financial studies*, 1(3):311–328, 1988.
- [176] S. A. Nelson. *The A B C of Options and Arbitrage*. The Wall Street Library, New York, 1904.
- [177] S. A. Nelson. *The A B C of Options and Arbitrage*. New York: The Wall Street Library, 1904.
- [178] Hansjörg Neth and Gerd Gigerenzer. *Heuristics: Tools for an uncertain world. Emerging trends in the social and behavioral sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 2015.
- [179] Donald J Newman. *A problem seminar*. Springer Science & Business Media, 2012.
- [180] Hong Nie and Shaohua Chen. Lognormal sum approximation with type iv pearson distribution. *IEEE Communications Letters*, 11(10), 2007.
- [181] John P Nolan. Parameterizations and modes of stable distributions. *Statistics & probability letters*, 38(2):187–195, 1998.
- [182] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.
- [183] Joel Owen and Ramon Rabinovitch. On the class of elliptical distributions and their applications to the theory of portfolio choice. *The Journal of Finance*, 38(3):745–752, 1983.
- [184] T. Mikosch P. Embrechts, C. Kluppelberg. *Modelling Extremal Events*. Springer, 2003.
- [185] Vilfredo Pareto. La courbe des revenus. *Travaux de Sciences Sociales*, pages 299–345, 1896 (1964).
- [186] O. Peters and M. Gell-Mann. Evaluating gambles using dynamics. *Chaos*, 26(2), 2016.

- [187] T Pham-Gia and TL Hung. The mean and median absolute deviations. *Mathematical and Computer Modelling*, 34(7–8):921–936, 2001.
- [188] George C Philippatos and Charles J Wilson. Entropy, market risk, and the selection of efficient portfolios. *Applied Economics*, 4(3):209–220, 1972.
- [189] Charles Phillips and Alan Axelrod. *Encyclopedia of Wars:(3-Volume Set)*. Infobase Pub., 2004.
- [190] James Pickands III. Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131, 1975.
- [191] Thomas Piketty. *Capital in the 21st century*, 2014.
- [192] Thomas Piketty and Emmanuel Saez. The evolution of top incomes: a historical and international perspective. Technical report, National Bureau of Economic Research, 2006.
- [193] Iosif Pinelis. Characteristic function of the positive part of a random variable and related results, with applications. *Statistics & Probability Letters*, 106:281–286, 2015.
- [194] Steven Pinker. *The better angels of our nature: Why violence has declined*. Penguin, 2011.
- [195] Dan Pirjol. The logistic-normal integral and its generalizations. *Journal of Computational and Applied Mathematics*, 237(1):460–469, 2013.
- [196] EJG Pitman. Subexponential distribution functions. *J. Austral. Math. Soc. Ser. A*, 29(3):337–347, 1980.
- [197] Svetlozar T Rachev, Young Shin Kim, Michele L Bianchi, and Frank J Fabozzi. *Financial models with Lévy processes and volatility clustering*, volume 187. John Wiley & Sons, 2011.
- [198] Anthony M. Reinach. *The Nature of Puts & Calls*. New York: The Bookmailer, 1961.
- [199] Lewis F Richardson. Frequency of occurrence of wars and other fatal quarrels. *Nature*, 148(3759):598, 1941.
- [200] Matthew Richardson and Tom Smith. A direct test of the mixture of distributions hypothesis: Measuring the daily flow of information. *Journal of Financial and Quantitative Analysis*, 29(01):101–116, 1994.
- [201] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.
- [202] Stephen A Ross. Mutual fund separation in financial theory — the separating distributions. *Journal of Economic Theory*, 17(2):254–286, 1978.
- [203] Stephen A Ross. *Neoclassical finance*. Princeton University Press, 2009.
- [204] Francesco Rubino, Antonello Forgione, David E Cummings, Michel Vix, Donatella Gnuli, Geltrude Mingrone, Marco Castagneto, and Jacques Marescaux. The mechanism of diabetes control after gastrointestinal bypass surgery reveals a role of the proximal small intestine in the pathophysiology of type 2 diabetes. *Annals of surgery*, 244(5):741–749, 2006.
- [205] Mark Rubinstein. *Rubinstein on derivatives*. Risk Books, 1999.
- [206] Mark Rubinstein. *A History of The Theory of Investments*. New York: John Wiley & Sons, 2006.
- [207] Doriana Ruffino and Jonathan Treussard. Derman and taleb’s ‘the illusions of dynamic replication’: a comment. *Quantitative Finance*, 6(5):365–367, 2006.
- [208] Harold Sackrowitz and Ester Samuel-Cahn. P values as random variables— expected p values. *The American Statistician*, 53(4):326–331, 1999.
- [209] Gennady Samorodnitsky and Murad S Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance*, volume 1. CRC Press, 1994.
- [210] D Schleher. Generalized gram-charlier series with application to the sum of log-normal variates (corresp.). *IEEE Transactions on Information Theory*, 23(2):275–280, 1977.
- [211] Jun Shao. *Mathematical Statistics*. Springer, 2003.



- [212] Herbert A Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.
- [213] SK Singh and GS Maddala. A function for size distribution of incomes: reply. *Econometrica*, 46(2), 1978.
- [214] Didier Sornette. Critical phenomena in natural sciences: chaos, fractals, selforganization, and disorder: concepts and tools. Springer, 2004.
- [215] C. M. Sprenkle. Warrant prices as indicators of expectations and preferences. *Yale Economics Essays*, 1(2):178–231, 1961.
- [216] C. M. Sprenkle. Warrant Prices as Indicators of Expectations and Preferences: in P. Cootner, ed., 1964, *The Random Character of Stock Market Prices*,. MIT Press, Cambridge, Mass, 1964.
- [217] AJ Stam. Regular variation of the tail of a subordinated probability distribution. *Advances in Applied Probability*, pages 308–327, 1973.
- [218] Stephen M Stigler. Stigler’s law of eponymy. *Transactions of the New York academy of sciences*, 39(1 Series II):147–157, 1980.
- [219] Hans R Stoll. The relationship between put and call option prices. *The Journal of Finance*, 24(5):801–824, 1969.
- [220] Cass R Sunstein. Deliberating groups versus prediction markets (or hayek’s challenge to habermas). *Episteme*, 3(3):192–213, 2006.
- [221] Giitiro Suzuki. A consistent estimator for the mean deviation of the pearson type distribution. *Annals of the Institute of Statistical Mathematics*, 17(1):271–285, 1965.
- [222] E. Schechtman S. Yitzhaki. *The Gini Methodology: A primer on a statistical methodology*. Springer, 2012.
- [223] N N Taleb and R Douady. Mathematical definition, mapping, and detection of (anti) fragility. *Quantitative Finance*, 2013.
- [224] Nassim N Taleb and G Martin. The illusion of thin tails under aggregation (a reply to jack treynor). *Journal of Investment Management*, 2012.
- [225] Nassim Nicholas Taleb. *Dynamic Hedging: Managing Vanilla and Exotic Options*. John Wiley & Sons (Wiley Series in Financial Engineering), 1997.
- [226] Nassim Nicholas Taleb. *Incerto: Antifragile, The Black Swan, Fooled by Randomness, the Bed of Procrustes, Skin in the Game*. Random House and Penguin, 2001–2018.
- [227] Nassim Nicholas Taleb. Black swans and the domains of statistics. *The American Statistician*, 61(3):198–200, 2007.
- [228] Nassim Nicholas Taleb. Errors, robustness, and the fourth quadrant. *International Journal of Forecasting*, 25(4):744–759, 2009.
- [229] Nassim Nicholas Taleb. Finiteness of variance is irrelevant in the practice of quantitative finance. *Complexity*, 14(3):66–76, 2009.
- [230] Nassim Nicholas Taleb. *Antifragile: things that gain from disorder*. Random House and Penguin, 2012.
- [231] Nassim Nicholas Taleb. Four points beginner risk managers should learn from jeff holman’s mistakes in the discussion of antifragile. *arXiv preprint arXiv:1401.2524*, 2014.
- [232] Nassim Nicholas Taleb. The meta-distribution of standard p-values. *arXiv preprint arXiv:1603.07532*, 2016.
- [233] Nassim Nicholas Taleb. Stochastic tail exponent for asymmetric power laws. *arXiv preprint arXiv:1609.02369*, 2016.
- [234] Nassim Nicholas Taleb. Election predictions as martingales: an arbitrage approach. *Quantitative Finance*, 18(1):1–5, 2018.

- [235] Nassim Nicholas Taleb. How much data do you need? an operational, preasymptotic metric for fat-tailedness. *International Journal of Forecasting*, 2018.
- [236] Nassim Nicholas Taleb. *Skin in the Game: Hidden Asymmetries in Daily Life*. Penguin (London) and Random House (N.Y.), 2018.
- [237] Nassim Nicholas Taleb. *Technical Incerto, Vol 1: The Statistical Consequences of Fat Tails, Papers and Commentaries*. Monograph, 2019.
- [238] Nassim Nicholas Taleb. Common misapplications and misinterpretations of correlation in social science. Preprint, Tandon School of Engineering, New York University, 2020.
- [239] Nassim Nicholas Taleb. *The Statistical Consequences of Fat Tails*. STEM Academic Press, 2020.
- [240] Nassim Nicholas Taleb, Elie Canetti, Tidiane Kinda, Elena Loukoianova, and Christian Schmieder. A new heuristic measure of fragility and tail risks: application to stress testing. *International Monetary Fund*, 2018.
- [241] Nassim Nicholas Taleb and Pasquale Cirillo. Branching epistemic uncertainty and thickness of tails. arXiv preprint arXiv:1912.00277, 2019.
- [242] Nassim Nicholas Taleb and Raphael Douady. On the super-additivity and estimation biases of quantile contributions. *Physica A: Statistical Mechanics and its Applications*, 429:252–260, 2015.
- [243] Nassim Nicholas Taleb and Daniel G Goldstein. The problem is beyond psychology: The real world is more random than regression analyses. *International Journal of Forecasting*, 28(3):715–716, 2012.
- [244] Nassim Nicholas Taleb and George A Martin. How to prevent other financial crises. *SAIS Review of International Affairs*, 32(1):49–60, 2012.
- [245] Nassim Nicholas Taleb and Avital Pilpel. I problemi epistemologici del risk management. Daniele Pace (a cura di). *Economia del rischio. Antologia di scritti su rischio e decisione economica*, Giuffrè, Milano, 2004.
- [246] Nassim Nicholas Taleb and Constantine Sandis. The skin in the game heuristic for protection against tail events. *Review of Behavioral Economics*, 1:1–21, 2014.
- [247] NN Taleb and J Norman. *Ethics of precaution: Individual and systemic risk*, 2020.
- [248] Jozef L Teugels. The class of subexponential distributions. *The Annals of Probability*, 3(6):1000–1011, 1975.
- [249] Edward Thorp. A corrected derivation of the black-scholes option model. Based on private conversation with Edward Thorp and a copy of a 7 page paper Thorp wrote around 1973, with disclaimer that I understood Ed. Thorp correctly, 1973.
- [250] Edward O Thorp. Optimal gambling systems for favorable games. *Revue de l'Institut International de Statistique*, pages 273–293, 1969.
- [251] Edward O Thorp. Extensions of the black-scholes option model. *Proceedings of the 39th Session of the International Statistical Institute, Vienna, Austria*, pages 522–29, 1973.
- [252] Edward O Thorp. *Understanding the kelly criterion. The Kelly Capital Growth Investment Criterion: Theory and Practice*, World Scientific Press, Singapore, 2010.
- [253] Edward O. Thorp and S. T. Kassouf. *Beat the Market*. New York: Random House, 1967.
- [254] James Tobin. Liquidity preference as behavior towards risk. *The review of economic studies*, pages 65–86, 1958.
- [255] Jack L Treynor. Insights-what can taleb learn from markowitz? *Journal of Investment Management*, 9(4):5, 2011.
- [256] Constantino Tsallis, Celia Anteneodo, Lisa Borland, and Roberto Osorio. Nonextensive statistical mechanics and economics. *Physica A: Statistical Mechanics and its Applications*, 324(1):89–100, 2003.

- [257] Vladimir V Uchaikin and Vladimir M Zolotarev. *Chance and stability: stable distributions and their applications*. Walter de Gruyter, 1999.
- [258] Aad W Van Der Vaart and Jon A Wellner. Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer, 1996.
- [259] Willem Rutger van Zwet. *Convex transformations of random variables*, volume 7. Mathematisch centrum, 1964.
- [260] SR Srinivasa Varadhan. *Large deviations and applications*, volume 46. SIAM, 1984.
- [261] SR Srinivasa Varadhan. *Stochastic processes*, volume 16. American Mathematical Soc., 2007.
- [262] José A Villaseñor-Alva and Elizabeth González-Estrada. A bootstrap goodness of fit test for the generalized pareto distribution. *Computational Statistics & Data Analysis*, 53(11):3835–3841, 2009.
- [263] Eric Weisstein. Wolfram MathWorld. Wolfram Research [www.wolfram.com](http://www.wolfram.com), 2017.
- [264] Rafał Weron. Levy-stable distributions revisited: tail index  $> 2$  does not exclude the levy-stable regime. *International Journal of Modern Physics C*, 12(02):209–223, 2001.
- [265] Heath Windcliff and Phelim P Boyle. The 1/n pension investment puzzle. *North American Actuarial Journal*, 8(3):32–45, 2004.
- [266] Yingying Xu, Zhuwu Wu, Long Jiang, and Xuefeng Song. A maximum entropy method for a robust portfolio problem. *Entropy*, 16(6):3401–3415, 2014.
- [267] Yingying Yang, Shuhe Hu, and Tao Wu. The tail probability of the product of dependent random variables from max-domains of attraction. *Statistics & Probability Letters*, 81(12):1876–1882, 2011.
- [268] Jay L Zagorsky. Do you have to be smart to be rich? the impact of iq on wealth, income and financial distress. *Intelligence*, 35(5):489–501, 2007.
- [269] IV Zaliapin, Yan Y Kagan, and Federic P Schoenberg. Approximating the distribution of pareto sums. *Pure and Applied geophysics*, 162(6–7):1187–1228, 2005.
- [270] Rongxi Zhou, Ru Cai, and Guanqun Tong. Applications of entropy in finance: A review. *Entropy*, 15(11):4909–4931, 2013.
- [271] Zolotarev V. M. *One-dimensional stable distributions*, volume 65. American Mathematical Soc., 1986.
- [272] Zolotarev V. M. On a new viewpoint of limit theorems taking into account large deviations // *Selected Translations in Mathematical Statistics and Probability*. 9:153, 1971.

ТАЛЕБ НАССИМ НИКОЛАС

## СТАТИСТИЧЕСКИЕ ПОСЛЕДСТВИЯ ЖИРНЫХ ХВОСТОВ

О новых вычислительных подходах к принятию решений

Редакторы А. Захарова, М. Терехова  
Художественный редактор М. Левыкин  
Технический редактор Л. Синицына  
Корректоры С. Луконина, О. Левина  
Верстка Н. Козель, Т. Коровенкова

В оформлении обложки использована иллюстрация  
© tinkivinki/Shutterstock.com

Подписано в печать / Баспаға қол қойылды 20.10.2023.

Формат 72 × 100  $\frac{1}{16}$ . Гарнитура «Times».

Бумага офсетная. Печать офсетная. Усл. печ. л. 39,9.

Доп. тираж 4000 экз. В-СНМ-28300-03-Р. Заказ №

Дата изготовления / Өндірілген күні 16.11.2023.

Срок службы (годности): не ограничен. / Қызмет (жарамдылық) мерзімі: шектелмеген.

Условия хранения: в сухом помещении. / Сақтау шарттары: құрғақ үй-жайда.

Изготовитель: Өндіруші:

ООО «Издательская Группа «Азбука-Аттикус» –	«Издательская Группа «Азбука-Аттикус» ЖШҚ –
обладатель товарного знака КоЛибри	КоЛибри тауар белгісінің иесі,
115093, Москва, вн. тер. г. муниципальный округ	115093, Мәскеу, қ. іш. аум. Даниловский
Даниловский, пер. Партийный, д. 1, к. 25	муниципалдық округі, Партийный т.ш., 1-үй, к. 25
Тел. (495) 933-76-01, факс (495) 933-76-19	Тел. (495) 933-76-01, факс (495) 933-76-19
E-mail: sales@atticus-group.ru	E-mail: sales@atticus-group.ru

Филиал ООО «Издательская Группа	Санкт-Петербург қ.
«Азбука-Аттикус»	«Издательская Группа
в г. Санкт-Петербурге	«Азбука-Аттикус» ЖШҚ филиалы,
191123, Санкт-Петербург,	191123, Санкт-Петербург,
Воскресенская набережная, д. 12, лит. А	Воскресенская жағалауы, 12-үй, А лит.
Тел. (812) 327-04-55	Тел. (812) 327-04-55
E-mail: trade@azbooka.spb.ru	E-mail: trade@azbooka.spb.ru
www.azbooka.ru; www.atticus-group.ru	www.azbooka.ru; www.atticus-group.ru
Отпечатано в России.	Ресейде басып шығарылған.

Техникалық реттеу туралы РФ заңнамасына сай басылымның сәйкестігін  
растау туралы мәліметтерді мына адрес бойынша алуға болады:  
<http://atticus-group.ru/certification/>.

Знак информационной продукции  
(Федеральный закон № 436-ФЗ от 29.12.2010 г.)

Товар соответствует требованиям ТР ТС 007/2011 «О безопасности продукции,  
предназначенной для детей и подростков».



Ақпараттық өнім белгісі (29.12.2010 ж. № 436-ФЗ федералдық заң)  
Тауар КО ТР 007/2011 «Балалар мен жасөспірімдерге арналған  
өнімдердің қауіпсіздігі туралы» талаптарына сәйкес келеді.

В «Черном лебеде» Талеб показал, что повсюду нас поджидают невероятные и непредсказуемые события, в «Антихрупкости» открыл положительные стороны неопределенности и превратил ее в нечто желанное и необходимое для роста и развития, в «Одураченных случайностью» исследовал удачу, вероятность, человеческие ошибки и принятие решений в мире, который мы не понимаем, а в книге «Рискуя собственной шкурой» рассказал об искажениях и сложных системах, о равенстве и неравенстве, о рациональности и принятии риска. Ценным дополнением к перечисленным бестселлерам знаковой серии Incerto становится эта книга из серии The Technical Incerto Collection, посвященная тем классам статистических распределений, от которых можно ждать экстремальных событий. Талеб показывает, как использовать эти распределения для статистических выводов и принятия решений.

**ЕСЛИ ВЫ НЕ ДРУЖИТЕ С ГРАФИКАМИ И ФОРМУЛАМИ**, то из этой книги почерпнете информацию только про скандалы и разоблачение горе-ученых: создателя портфельной теории, который управлял своими инвестициями по старинке, но продал публике теорию, из которой произросли хедж-фонды и финансовые кризисы; философа, объявившего о конце войн, который жонглирует терминами статистики, но на самом деле не продвинулся дальше рассуждений; психологов, якобы открывших порок в человеческой интуиции: склонность преувеличивать риск; уютно устроившихся шарлатанов, насочинявших дурно обоснованные, неработающие модели и выступающих с «экспертными» предсказаниями результатов выборов, ущерба от пандемии нового коронавируса, силы будущих финансовых кризисов, землетрясений и цунами.

**ЕСЛИ ВЫ УЧИЛИ МАТЕМАТИЧЕСКУЮ СТАТИСТИКУ**, эта книга поможет вам переучиться. Вы узнаете, как вовремя заметить, что вставшие перед вами проблемы не похожи на задачи про гауссово распределение и вообще распределение с тонким хвостом. Вы узнаете, как оценивать данные и как принимать решения, когда ваши величины следуют распределению Парето 20 на 80 или иному распределению с жирным хвостом и, возможно, с бесконечно большим эксцессом, коэффициентом асимметрии, дисперсией или даже средним. Или когда хвост тонкий, но недостаточно тонкий, чтобы набрать столько данных, сколько требуется для оценки истинного значения параметров. Вы научитесь не полагаться на  $p$ -значение как способ отличить научное открытие от случайного шума в данных. Вы научитесь критически относиться к общедоступному программному обеспечению, рисующему линейный тренд по толпе разбросанных точек или доверительный интервал вокруг точки на графике, полученной усреднением множества измерений.

**ЕСЛИ ВЫ СТУДЕНТ ИЛИ УЧЕНЫЙ**, эта книга — бесценный мастер-класс. Впервые под одной обложкой собраны исследовательские статьи Талеба и его учеников, где в неповторимом талебовском стиле живо и ярко прослеживается ход мысли прикладного математика, сталкивающегося с жизненной задачей, не зная, можно ли ее решить аналитически, с чего начать, за что хвататься, — но настроенного пустить в ход, если понадобится, весь арсенал классической и современной математики и всю мощь компьютеров.

