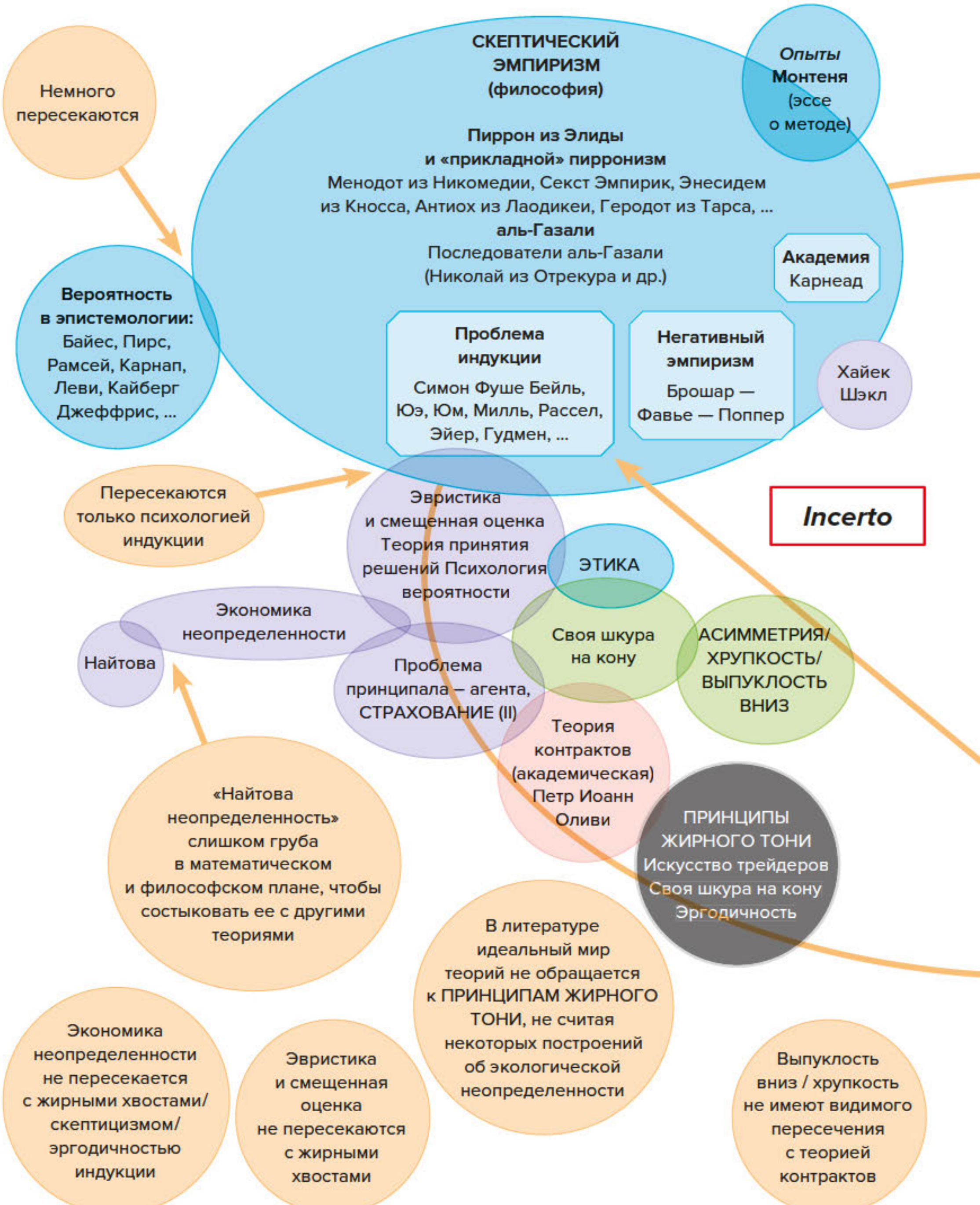
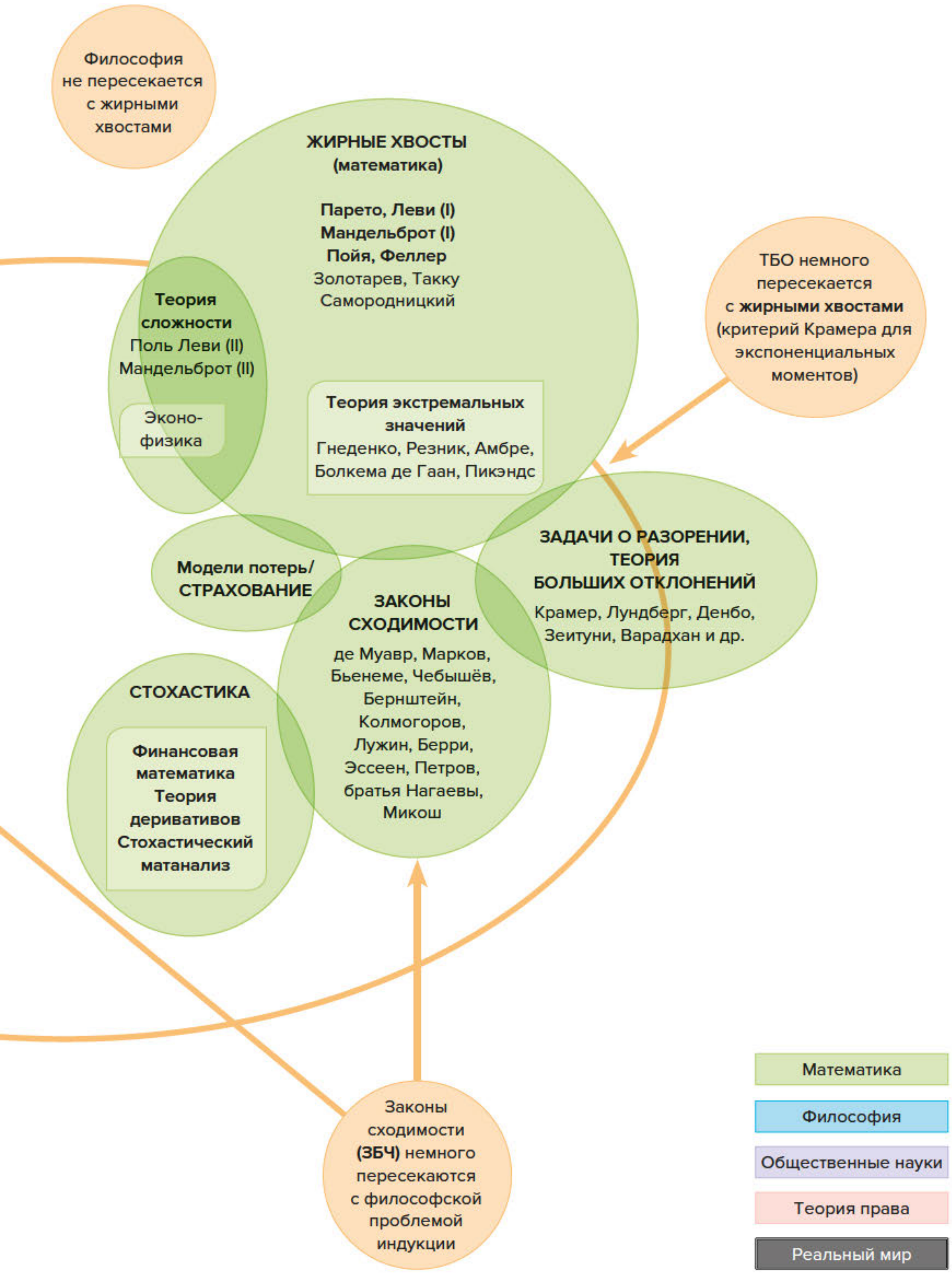
Генеалогическая карта цикла *Incerto*




Философия не пересекается с жирными хвостами
ЖИРНЫЕ ХВОСТЫ (математика)
Парето, Леви (I)
Мандельброт (I)
Пойя, Феллер
Золотарев, Такку
Самородницкий
Теория сложности
Поль Леви (II)
Мандельброт (II)
Эконо-физика
Теория экстремальных значений
Гнеденко, Резник, Амбре, Болкема де Гаан, Пикэндс
ТБО немного пересекается с **жирными хвостами** (критерий Крамера для экспоненциальных моментов)
ЗАДАЧИ О РАЗОРЕНИИ, ТЕОРИЯ БОЛЬШИХ ОТКЛОНЕНИЙ
Крамер, Лундберг, Денбо, Зеитуни, Варадхан и др.
Модели потерь/ СТРАХОВАНИЕ
ЗАКОНЫ СХОДИМОСТИ
де Муавр, Марков, Бьенеме, Чебышёв, Бернштейн, Колмогоров, Лужин, Берри, Эссеен, Петров, братья Нагаевы, Микош
СТОХАСТИКА
Финансовая математика
Теория деривативов
Стохастический матанализ
Законы сходимости **(ЗБЧ)** немного пересекаются с философской проблемой индукции
Математика
Философия
Общественные науки
Теория права
Реальный мир

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 A, B и т. д.

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.

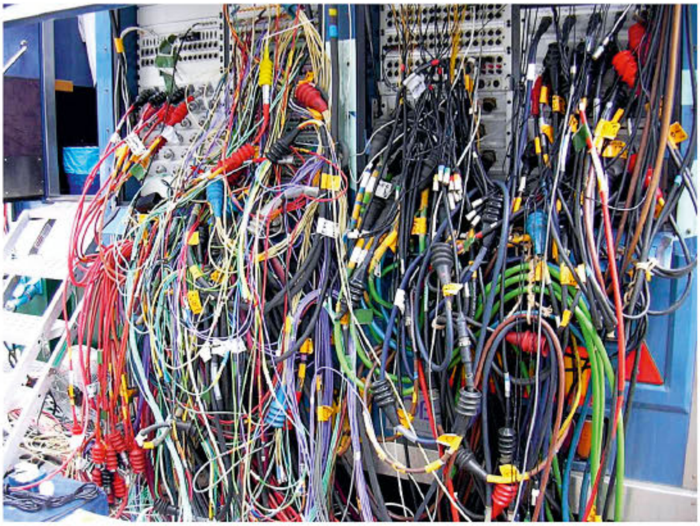


Рисунок 1.2: *Усложнение из-за непонимания. Что творится в головах профессионалов, когда они применяют статистику и анализ данных, не имея ясного представления об основных понятиях. © Wikimedia*

«Стандартная» статистика по большей части работает на основе теорем, выведенных для тонких хвостов. Чтобы работать с предасимптотикой[1] 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»).

1 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 степенно́му 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 Яркину, 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[1] 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).

2
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 прочий французский встречаются в Главе 11 и других местах, где этот формализм имеет смысл.

$\mathbb{E}$ — оператор математическое ожидание[1].

$\mathbb{V}$ — оператор дисперсия[2].

$\mathbb{M}$ — среднее абсолютное отклонение;[3] 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 $\mathbb{E}X=\int_{\mathbb{R}} x f_X(x)\,\mathrm{d}x$, 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 $\mathbb{V}X=\int_{\mathbb{R}}(x-\mathbb{E}X)^2 f_X(x)\,\mathrm{d}x$, 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 $\mathbb{M}X=\int_{\mathbb{R}}|x-\mathbb{E}X|\,f_X(x)\,\mathrm{d}x$, где $f_X(x)$ — плотность вероятности.

$\varphi(.)$ и $f(.)$ 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[1].

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

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

$\chi(t) = \mathbb{E}\ \mathrm{e}^{\mathrm{i}tX_s}$ — характеристическая функция случайной величины X_S. Иногда для аргумента $t \in \mathbb{R}$ используется другая буква — ω. Сама характеристическая функция иногда обозначается заглавной Ψ[3].

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

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

при всяком действительном x, при котором F непрерывна.

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

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

$\xrightarrow{a.s.}$ означает сходимость почти наверное[4], то есть более сильное требование:

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

S_n обычно обозначает сумму n слагаемых.

α, а также α_P и α_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 α.

$\mathcal{N}(\mu_1, \sigma_1^2)$ — нормальное (гауссово) распределение со средним μ_1 и дисперсией σ_1^2[5].

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, r. 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 $\bar{F}(.)$, $S(.)$, $R(.)$ часто просто выписывают разность $1-F(.)$.

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 Φ, без параметров. Тогда для произвольного нормального распределения $\mathcal{N}(\mu, \sigma^2)$

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

$\mathcal{S}(\alpha_S, \beta, \mu, \sigma)$ — устойчивое распределение с показателем хвоста $\alpha_S \in (0, 2]$, коэффициентом симметрии β 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).

δ(.) — дельта-функция Дирака.

ϑ(.) — тета-функция Хевисайда[2].

erf(.) — функция ошибок, представляющая собой интеграл плотности гауссова распределения[3]

$$\operatorname{erf} z = \frac{2}{\sqrt{\pi}} \int_0^z \mathrm{d}t\, \mathrm{e}^{-t^2}.$$

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 μ и масштабом σ плотность вероятности $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$, дисперсия σ^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 $\lim_{p\to\infty} \kappa_n^p = 0$.

1 Логнормальное распределение — это распределение случайной величины e^Y, где Y — гауссова случайная величина. Если гауссова случайная величина Y имеет среднее μ и дисперсию σ^2, то логнормальная случайная величина $X = \mathrm{e}^Y$ имеет среднее $X_0 = \mathrm{e}^{\mu+\sigma^2/2}$ и дисперсию $(\mathrm{e}^{\sigma^2}-1)X_0^2$. Величина $X = \mathrm{e}^Y$ 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 $\mathrm{e}^{\mu} = X_0\, \mathrm{e}^{-\sigma^2/2}$, мода $\mathrm{e}^{\mu-\sigma^2} = X_0\, \mathrm{e}^{-3\sigma^2/2}$, коэффициент асимметрии $(\mathrm{e}^{\sigma^2}+2)\sqrt{\mathrm{e}^{\sigma^2}-1}$, коэффициент эксцесса $\mathrm{e}^{4\sigma^2} + 2\,\mathrm{e}^{3\sigma^2} + 3\,\mathrm{e}^{2\sigma^2} - 6$.

2 (*Heaviside step function*) Функция ступенька в нуле, $\vartheta(x) \triangleq \begin{cases} 1, & x \ge 0 \\ 0, & x < 0 \end{cases}$. Если нужна ступенька в произвольной точке K, используют $\vartheta_K(x) = \begin{cases} 1, & x \ge K \\ 0, & \text{иначе} \end{cases} = \vartheta(x - K)$; можно представить и через индексную функцию $\mathbb{1}_{x \ge K}$.

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

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

3 Связь с нормальным распределением такая: интегрируется функция плотности $f(t) = \frac{1}{\sqrt{\pi}}\mathrm{e}^{-t^2}$ 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: $\Phi(z) = \frac{1}{2}\left(1 + \operatorname{erf}\frac{z}{\sqrt{2}}\right)$.

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

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

$\|.\|_p$ — норма; в этой книге[1] применяется к действительному вектору $\mathbf{X} = (X_1, \ldots, X_n)^T$ и определяется как

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

Обратите внимание, что компоненты вектора берутся по абсолютной величине.

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

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

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

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

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

$${}_pF_q(a;b;z) \triangleq \sum_{k=1}^{\infty} \frac{(a_1)_k \ldots (a_p)_k}{(b_1)_k \ldots (b_q)_k} \frac{z^k}{k!}$$

с использованием символа Похгаммера[2] $(a)_n = \prod_{i=0}^{n-1} (a+i)$.

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:

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

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

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

для всех $k > 0$ [22].

---

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

2 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_α[1].

Давайте уточним: функция $f\colon \mathbb{R}^+ \to \mathbb{R}^+$ меняется на бесконечности с показателем ρ, то есть $f \in \mathrm{RV}_\rho$, когда

$$\lim_{x\to\infty} \frac{f(tx)}{f(t)} = x^{\rho}.$$ [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 $X_1, X_2, \ldots, X_n$ — 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 $\overline{X}_n = \frac{1}{n}\,(X_1 + X_2 + \ldots + X_n)$ сходится к математическому ожиданию, $\overline{X}_n \to \mu$ при $n \to \infty$.

Конечность дисперсии не обязательна (однако весьма желательна: если дисперсия и прочие высшие моменты распределения конечны, то $\overline{X}_n$ 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 $\overline{X}_n$ — это среднее по выборке первых n величин. Тогда по мере приближения n к бесконечности центрированное и нормированное среднее $\sqrt{n}(\overline{X}_n - \mu)$ сходится по распределению к гауссову [20] [21]

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

---

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 $\lim_{t\to\infty}\left|\frac{f(tx)}{f(t)}\right|$. Согласно теореме Караматы, такая функция представима в виде степенного закона t^ρ 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
$\lim_{t\to\infty}\left|\frac{f(tx)}{f(t)}\right| = \lim_{t\to\infty}\left|\frac{L(tx)(tx)^\rho}{L(t)t^\rho}\right| = \lim_{t\to\infty}\left|\frac{L(tx)}{L(t)}\right| \cdot \lim_{t\to\infty}\left|\frac{(tx)^\rho}{t^\rho}\right| = 1 \cdot \lim_{t\to\infty}|x^\rho| = x^\rho$.

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

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

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[1]. 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[2].

Пусть $X_1, \ldots, X_n, \ldots$ — случайные величины н. о. р. с конечным математическим ожиданием, то есть $\mathbb{E}\,X<+\infty$. Пусть $S_n=X_1+X_2+\ldots+X_n$ — 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:

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

где $n_0, n=1, 2, \ldots$ и $n>n_0\ge 1$; соответственно

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

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 κ_{n_0}.

2.2.6. Эллиптическое распределение

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

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 распределениям 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 и независимы, и некоррелированы. Иначе обстоит дело с многомерными формами t-распределения Стьюдента или распределения Коши.

2.2.8. Устойчивое распределение (устойчивое по Леви)

Это обобщение ЦПТ.

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

$$\frac{S_n - a_n}{b_n} \xrightarrow{D} X_S, \tag{2.4}$$

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

Свойства $\mathcal{S}$ будут должным образом определены и рассмотрены в следующей главе. Пока заметим, что про случайную величину X_S говорят, что она следует устойчивому (или α-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:

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

Ограничения: $-1 \leq \beta \leq 1$ и $0 < \alpha_S \leq 2$[1].

2.2.9. Многомерное устойчивое распределение

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

2.2.10. Точка Караматы

См. **Класс степенного закона**.

2.2.11. Субэкспоненциальность

Естественной границей между Медиокристаном[2] и Экстремистаном служит субэкспоненциальный класс, обладающий следующим свойством.

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

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

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

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 α-устойчивое по Леви, следует в пределе сумма n независимых случайных величин при $n \to \infty$; 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 μ 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. Эксцесс 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 — большему из двух, — тогда как вклад другого пренебрежимо мал[1].

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

a) $\lim\limits_{x\to\infty} \dfrac{\mathbb{P}(S_n > x)}{\mathbb{P}(X > x)} = n,$

b) $\lim\limits_{x\to\infty} \dfrac{\mathbb{P}(S_n > x)}{\mathbb{P}(M_n > x)} = 1.$

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:

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

для всех ε 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 α степенями свободы как удобное распределение степенного закона с двумя хвостами. При $\alpha = 1$ оно превращается в распределение Коши, а при $\alpha \to \infty$, 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

---

1 Имеется в виду, что в хвосте события редкие и при некотором x в хвосте вероятность превзойти x увеличится вдвое, если разрешить две попытки: $\mathbb{P}(X_1 > x$ или $X_2 > x) \approx 2\mathbb{P}(X_1 > x)$. Вообще говоря, если разрешить суммировать результаты двух попыток, шансы превзойти x еще увеличатся: $\mathbb{P}(X_1 + X_2 > x) > 2\mathbb{P}(X_1 > x$ 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.