Э. НЕЛЬСОН

РАДИКАЛЬНО ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Перевод с английского
А. А. Рубана и С. С. Кутателадзе

Под редакцией
С. С. Кутателадзе

НОВОСИБИРСК
Издательство Института математики СО РАН
им. С. Л. Соболева
1995

RADICALLY ELEMENTARY PROBABILITY THEORY

BY
EDWARD NELSON

PRINCETON UNIVERSITY PRESS

PRINCETON, NEW JERSEY

1987

УДК 519.2

Нельсон Э. **Радикально элементарная теория вероятностей:** Пер. с англ. — Новосибирск: Изд-во ин-та математики СО РАН. 1995. — 124 с.

В этой книге выдающегося американского математика предлагается принципиально новый подход как к изложению основ, так и продвинутых тем теории вероятностей. Автору удалось предложить очень простую и в то же время весьма мощную «частотную» версию теории, использующую идеи современного инфинитезимального (нестандартного) анализа. Элементарность изложения делает книгу доступной широкому кругу студентов, преподавателей и научных работников всех специальностей, интересующихся теорией вероятностей или применяющих ее. Ил. 12. Библиогр. 3.

Оглавление

Предисловие к русскому изданию

Уже из названия лежащей перед читателем брошюры следует необычность ее содержания. Переводчик и редактор могли бы ограничить свое предисловие ссылкой на отмеченное обстоятельство, указав на естественные желание и необходимость ознакомить пользователей русского языка с новым подходом к основаниям теории вероятностей. При этом остались бы за скобкой приятный личный опыт и удовольствие от знакомства с книгой Э. Нельсона, которые не в малой мере служили источником решения подготовить перевод этой книги.

Специалистам по вероятностным процессам Э. Нельсон хорошо известен, в частности, своей книгой Dynamical Theories of Brownian Motion, 1967. Его результаты вошли в такие известные руководства по теории вероятностей, как, например, книги В. Феллера и М. Лоэва. Не менее известен Э. Нельсон и своими работами по нестандартному анализу. Его знаменитая статья "Internal Set Theory. A New Approach to Nonstandard Analysis," Bull. Amer. Math. Soc., 1977, **83**:6, pp. 1165–1198 — это общепризнанное кредо неоклассической установки нестандартного анализа, объявляющей универсум внутренних множеств главным объектом изучения математики.

Для читателя, впервые встретившегося с некоторыми словами из предыдущего предложения, поясним, что внутренние множества не что иное, как новое название для обычных множеств, повсеместно используемых в математике. Новое кредо постулирует возможность отличать стандартные и нестандарные множества среди привычных объектов.

Содержательный смысл понятия стандартности раскрывается тем, что из стандартных элементов могут быть составлены исключительно конечные стандартные множества; при этом в каждом бесконечном множестве есть хотя бы один нестандартный элемент *(принцип идеализации). Новые стандартные объекты возникают из уже заданных стандартных объектов с помощью теорем существования и единственности (принцип переноса). Таким образом, стандартные множества — это конкретные, явно определяемые (описанными процедурами) математические объекты. Все прочие объекты нестандартны, т. е. вводятся в рассмотрение косвенным образом как элементы ранее построенных множеств.*

Тем самым классический мир математических объектов расцвечивается и владеющий «цветным телевизором» (по выражению самого Э. Нельсона) наблюдатель видит новые ранее скрытые свойства объектов изучения. Например, в множестве вещественных чисел немедленно обнаруживаются бесконечно большие и бесконечно малые элементы. По-новому видится единство дискретного и непрерывного в устройстве числовой прямой. Доступная наблюдению часть последней оказывается составленной из «монад» стандартных чисел (такая «монада» образована представлениями данного числа с точностью до всех бесконечно малых). При этом монада «неделима» — бесконечно малые выдерживают умножение на любое стандартное вещественное число.

Принцип переноса в эквивалентной переформулировке показывает, что стандартные объекты хорошо представляют весь математический мир: обычное математическое утверждение, проверенное только для стандартных элементов, оказывается верным для всех элементов (разумеется, если все свободные параметры — условия утверждения — были стандартными). Учитывая, что стандартных элементов существенно меньше, чем всех вообще (число первых можно считать даже конечным, хотя и бесконечно большим), неудивительно, что принцип переноса позволяет существенно упростить доказательство обычных утверждений.

Однако пафос новых методов не в упрощении прежних, а в появлении принципиально новых возможностей математического моделирования, проявляющихся прежде всего в выработке новых понятий.

Книга Э. Нельсона не предполагает никаких предварительных знаний из теории внутренних множеств и олицетворяет главную возможность нового видения математического мира. Она демонстрирует возможность элементарного сугубо дискретного подхода к сложным задачам теории вероятностей. Стоит подчеркнуть, что теория стохастических процессов, которой в основном посвящена эта книга, никогда не принадлежала к элементарным главам теории вероятностей. В обстоятельных и толстых книгах, посвященных строгому изложению теории стохастических процессов, всегда отводится много места изложению предварительных сведений, без которых нельзя даже и подступиться к изучению самой теории. Можно сказать, что теперь положение радикально изменилось.

Новый подход, предложенный Э. Нельсоном, очень прост, нагляден и отвечает существу дела, что позволяет автору элементарно изложить, а читателю легко освоить весьма нетривиальные темы, которые при стандартном изложении требуют многих специальных сведений, тонкой и даже изощренной техники и большого объема сведений из продвинутых разделов теории меры и функционального анализа.

Книга написана с большим уважением к читателю. Если в ней кому-нибудь попадется трудное место, то можно смело гарантировать, что эта трудность вызвана существом предмета, а не избранным способом изложения, и что в любой другой книге трудности будут не просто большими, но большими значительно.

При переводе была поставлена задача возможно меньше отклоняться от оригинала. Естественно, некоторые потери при этом неизбежны. Русская (как впрочем, и английская) терминология в нестандартном анализе не устоялась. Достаточно сказать, что и сам этот несколько эпатажный термин «нестандартный» сейчас уступает место новому «инфинитезимальный». В этой связи мы использовали терминологию, представляющуюся нам как отвечающей существу дела, так и нормам математического русского (надеемся, и просто русского) языка. Для удобства читателя в указателе представлены термины на обоих языках.

Книга написана незаурядным автором. Ясны возникающие в этой связи трудности перевода. Наверняка нам не удалось в сколь-либо полной мере отразить ненавязчивый и тонкий юмор автора (ведь хорошая книга по математике не обязана быть нудной). Автор обращается к интеллекту другой культуры, что не так уж часто в интернациональной математической литературе. Например, в главе 9 речь идет о неком мифическом стихотворном сочинении “Owed to a Martingale” — «Должное мартингалу», и вряд ли у каждого потенциального читателя представляемого перевода возникнет ассоциация с поэзией Дж. Китса (ср. “Ode to a Nightingale” — «Ода соловью»). В главе 11 автор, говоря о биржевом игроке, играющем на повышение («быке»), использует местоимение “he,” а говоря о биржевом игроке, играющем на понижение («медведе»), использует местоимение “she.” Читатель, желающий полнее понять, что же именно хочет этим выразить автор, может поразмышлять над знаменитым “Personal Pronoun Pronouncement” из настольной книги многих математиков нашего времени M. D. Spivak “The Joy of *TEX*.” Можно найти и аналоги у больших мастеров беллетристики (в книге М. Колпакчи «Дружеские встречи с английским языком». СПб, 1993, стр. 66–68). Мы отказались при переводе подобных пассажей от всех попыток далеко отойти от оригинала в погоне за призраком художественности. Надеемся, что читатель учтет наши мотивы.

Некоторые опечатки оригинала связаны с компьютерным набором. Естественно, они были исправлены без всяких примечаний. Столь же естественно, что в перевод заведомо вкрались свои собственные опечатки. Хочется надеется, что их немного.

Мы искренне признательны профессору Э. Нельсону и издательству “Princeton University Press” как за разрешение подготовить и издать русский перевод этой книги, так и за постоянное и терпеливое содействие на всех этапах подготовки книги к изданию.

С. С. Кутателадзе
А. А. Рубан

Предисловие к английскому изданию

Теория вероятностей более чем любой другой раздел математики развивалась в тесной связи со своими приложениями. Так было в начале, когда Паскаль и Ферма рассматривали задачу о справедливом разделе ставок в азартной игре, также происходит и сейчас когда наиболее привлекательные исследования в теории вероятностей проводятся физиками, работающими в статистической механике.

Основания теории вероятностей были заложены уже более пятидесяти лет назад Колмогоровым. Я уверен, что помимо меня и многие другие специалисты по теории вероятностей, читающие студентам старших курсов свой предмет, чувствуют, что указанные основания с помощью теории меры служат в большей мере тому, чтобы успокоить нашу математическую совесть, а не тому, чтобы предоставить точное и острое оружие ученому, желающему применять теорию вероятностей.

Настоящая работа представляет собою попытку заложить новые основания теории вероятностей с помощью крошечного кусочка нестандартного анализа. Требуемая математическая подготовка чуть превышает то, чему учат в старших классах, и я надеюсь, что эта книга сделает глубокие результаты современной теории стохастических процессов сразу же доступными каждому, кто умеет складывать, умножать и рассуждать.

Реализация замысла стала возможной благодаря решению сохранить нестандартные формулировки упомянутых результатов. Представители нестандартного анализа обладают некоторым новым способом того, как думать о математике, и если этот способ не переводить назад в привычные термины, он представляется весьма элементарным.

Математики с полным основанием являются консерваторами и относятся с подозрением к новым идеям. Они пожелают спросить являются получаемые в книге результаты столь же сильными как и привычные результаты, и стоит ли им тратить время на изучение нестандартных методов. Эти вопросы затрагиваются в приложении, которое предполагает существенно более высокий уровень математических знаний по сравнению с требуемым в основном тексте. Однако я хочу подчеркнуть, что основной текст опирается лишь на самого себя.

Предисловие к русскому переводу

Мне приятно видеть этот перевод на русский язык, и я глубоко признателен профессору Кутателадзе и д-ру Рубану за их великолепную работу.

Первый русский перевод чего-либо, написанного мною, был осуществлен в 1943 г. Моему классу в школе было поручено описать какой-либо аспект жизни страны нашего великого союзника, и я выбрал «Российское образование». Мой отец, который в 1917 г. работал в Петрограде, помогая военнопленным, перевел заголовок, так что обложка моего доклада оказалась украшена кириллицей.

Позвольте мне воспользоваться случаем для того, чтобы выразить свою большую надежду. Эта книга — всего лишь начало, которое не приведет к успеху, если другие не продолжат начатое дело. В каждой профессии имеется тенденция скрывать свои тайны с помощью недоступного для посторонних языка. Давайте препятствовать этому. Теория вероятностей используется многими из тех, кто не является математиками, и поэтому важно развивать ее самым простым и доступным из возможных способом.

Надежда моя в том, что Вы, читатели этой книги, построите дворец теории вероятностей, открытый каждому, ищущему понимания.

Принстон *Эдвард Нельсон*
16 января 1995 г.

Благодарности

Я признателен Эрику Карлену, Мэю Геркхе, Клаусу Кайзеру и Брайену Уайту за полезные замечания и Пьеру Картье за критическое прочтение рукописи, исправление многих ошибок и внесение ценных предложений. Частично эта работа была выполнена при поддержке Национального научного фонда. Я благодарю Моргана Филлипса за шесть иллюстраций, с которыми не справился LATEX.

Глава 1

Случайные величины

Здесь приведены некоторые основные определения и неравенства теории вероятностей в случае конечного вероятностного пространства.

Конечное вероятностное пространство — это конечное множество Ω и строго положительная функция pr на Ω такая, что $\sum \mathrm{pr}(\omega) = 1$. При этом *случайная величина* на Ω — это функция $x : \Omega \to \mathbf{R}$, где $\mathbf{R}$ — множество вещественных чисел. *Математическое ожидание,* или *среднее,* случайной величины x — это число

$$\mathbf{E}x = \sum x(\omega)\mathrm{pr}(\omega).$$

Событие — это подмножество $A \subseteq \Omega$, а *вероятность* события A — это число

$$\Pr A = \sum_{\omega \in A} \mathrm{pr}(\omega).$$

Для события A определим случайную величину χ_A, называемую его *индикаторной* или *характеристической функцией,* следующим образом: $\chi_A(\omega) = 1$, если $\omega \in A$, и $\chi_A(\omega) = 0$, если $\omega \notin A$. При этом $\Pr A = \mathbf{E}\chi_A$. Дополнительное событие A^c определяется равенством $A^c = \Omega \setminus A$, иными словами, $\omega \in A^c$, если $\omega \in \Omega$ и $\omega \notin A$.

Множество $\mathbf{R}^\Omega$ всех случайных величин на Ω является n-мерным векторным пространством, где n — число точек в Ω. Рассмотрим выражение $\mathbf{E}xy$, где x и y — две произвольные случайные величины. Тогда $\mathbf{E}xy = \mathbf{E}yx$, кроме того, $\mathbf{E}xy$ линейно по x и y и $\mathbf{E}xx > 0$ при $x \neq 0$. Таким образом, $\mathbf{E}xy$ обладает всеми свойствами скалярного произведения на n-мерном евклидовом пространстве. Евклидова норма $\sqrt{\mathbf{E}x^2}$ случайной величины x обозначена через $\|x\|_2$.

Математическое ожидание является линейной функцией на $\mathbf{R}^\Omega$, поэтому случайные величины со средним 0 образуют гиперплоскость. Ортогональное дополнение этой гиперплоскости представляет собой одномерное подпространство постоянных случайных величин. Мы отождествляем постоянную случайную величину, принимающую значение λ, с этим числом λ. При таком отождествлении отображение $x \mapsto \mathbf{E}x$ становится ортогональным проектором на подпространство постоянных случайных величин, а $x \mapsto x - \mathbf{E}x$ — ортогональным проектором на подпространство случайных величин с нулевыми математическими ожиданиями. Мы называем величину $\operatorname{Var} x = \mathbf{E}(x - \mathbf{E}x)^2$ *дисперсией* x, величину $\sqrt{\operatorname{Var} x}$ — *стандартным отклонением* x, величину $\mathbf{E}(x - \mathbf{E}x)(y - \mathbf{E}y)$ — *ковариацией* x и y, а величину

$$\frac{\mathbf{E}(x - \mathbf{E}x)(y - \mathbf{E}y)}{\sqrt{\operatorname{Var} x}\sqrt{\operatorname{Var} y}}$$

— *коэффициентом корреляции* x и y. Таким образом, если x и y — случайные величины, имеющие нулевые математические ожидания, то дисперсия x есть квадрат $\|x\|_2^2$ ее евклидовой нормы, стандартное отклонение x — это евклидова норма $\|x\|_2$, ковариация x и y — их скалярное произведение, а коэффициент корреляции x и y — косинус угла между ними.

Для случайных величин часто используются и другие нормы. Для $1 \le p < \infty$ положим $\|x\|_p = (\mathbf{E}|x|^p)^{1/p}$, и пусть $\|x\|_\infty = \max|x(\omega)|$. Ясно, что $\|x\|_p \le \|x\|_\infty$, а если ω_0 — точка, в которой $|x|$ достигает своего максимума, то

$$\|x\|_p \ge (|x(\omega_0)|^p \operatorname{pr}(\omega_0))^{1/p} \to \|x\|_\infty$$

при $p \to \infty$. Значит, $\|x\|_p \to \|x\|_\infty$ при $p \to \infty$.

Для $1 < p < \infty$ определим число p', называемое *сопряженным показателем* к p, с помощью равенства $p' = p/(p-1)$, так что

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

Неравенство Гёльдера утверждает, что

$$|\mathbf{E}xy| \le \|x\|_p \|y\|_{p'}. \tag{1.1}$$

Если x или y — нулевая случайная величина, то неравенство (1.1) тривиально. В остальных случаях после замены x на $x/\|x\|_p$ и y на $y/\|y\|_{p'}$ можем считать $\|x\|_p = \|y\|_{p'} = 1$. Тогда $\mathbf{E}|x|^p = \mathbf{E}|y|^{p'} = 1$, и нам нужно проверить, что $|\mathbf{E}xy| \le 1$. В силу неравенства $|\mathbf{E}xy| \le \mathbf{E}|xy|$, мы получим требуемое, если докажем, что $|xy|$ меньше, чем некоторая выпуклая комбинация $|x|^p$ и $|y|^{p'}$. Взяв очевидную выпуклую комбинацию, нам остается только показать, что

$$|xy| \le \frac{1}{p}|x|^p + \frac{1}{p'}|y|^{p'}.$$

Для доказательства прологарифмируем обе части неравенства. В силу вогнутости логарифмической функции логарифм правой части больше, чем

$$\frac{1}{p}\log|x|^p + \frac{1}{p'}\log|y|^{p'} = \log|xy|,$$

что и завершает доказательство неравенства Гёльдера.

Используя нормировки $\|x\|_p = \|y\|_{p'} = 1$, видим, что в (1.1) имеет место строгое неравенство за исключением случая $|\mathbf{E}xy| = \mathbf{E}|xy|$, т. е. ситуации, в которой x и y одного знака и $|x|^p = |y|^{p'}$. Однако если $x = \operatorname{sgn} y|y|^{p'/p}$, то $\mathbf{E}xy = 1$. Поэтому для всех случайных величин x будет

$$\|x\|_p = \max_{\|y\|_{p'}=1} |\mathbf{E}xy|. \tag{1.2}$$

Непосредственным следствием (1.2) является *неравенство Минковского*

$$\|x+z\|_p \le \|x\|_p + \|z\|_p. \tag{1.3}$$

Если считать, что $1' = \infty$ и $\infty' = 1$, то (1.1), (1.2) и (1.3) будут выполнены для всех p, $1 \le p \le \infty$.

Пусть f — выпуклая функция. По определению это означает, что

$$f\left(\sum x(\omega)\mathrm{pr}(\omega)\right) \le \sum f(x(\omega))\mathrm{pr}(\omega),$$

т. е.

$$f(\mathbf{E}x) \le \mathbf{E}f(x). \tag{1.4}$$

Последнее соотношение — это *неравенство Йенсена.* Если мы применим его к выпуклой функции $f(x) = |x|^p$, где $1 \le p < \infty$, то получим неравенство $|\mathbf{E}x|^p \le \mathbf{E}|x|^p$. Неравенство Йенсена, примененное к функции $|x|$, дает $\|x\|_1 \le \|x\|_p$, а к $|x|^r$, где $1 \le r < \infty$, — $\|x\|_r \le \|x\|_{rp}$. Таким образом, $\|x\|_p$ — возрастающая функция от p для $1 \le p \le \infty$.

Пусть f — положительная функция. Тогда для $\lambda > 0$

$$\mathbf{E}f(x) = \sum f(x(\omega))\mathrm{pr}(\omega) \ge \sum_{\omega \in \{f(x) \ge \lambda\}} f(x(\omega))\mathrm{pr}(\omega) \ge \lambda \Pr\{f(x) \ge \lambda\},$$

откуда

$$\Pr\{f(x) \ge \lambda\} \le \frac{\mathbf{E}f(x)}{\lambda}. \tag{1.5}$$

(Здесь $\{f(x) \ge \lambda\}$ — сокращение для $\{\omega \in \Omega : f(x(\omega)) \ge \lambda\}$; подобные сокращения обычны в теории вероятностей.) В частности, для $\lambda > 0$ и $p > 0$ имеем $\{|x| \ge \lambda\} = \{|x|^p \ge \lambda^p\}$ и потому согласно (1.5) будет

$$\Pr\{|x| \ge \lambda\} \le \frac{\mathbf{E}|x|^p}{\lambda^p}. \tag{1.6}$$

Это — *неравенство Чебышёва.*

Глава 2
Алгебры случайных величин

Множество $\mathbf{R}^\Omega$ всех случайных величин на Ω — это не только векторное пространство, но и алгебра. Под *алгеброй* $\mathcal{A}$ случайных величин мы всегда будем понимать подалгебру в $\mathbf{R}^\Omega$, содержащую константы. Таким образом, $\mathcal{A}$ есть множество случайных величин, содержащее константы и такое, что если $x \in \mathcal{A}$ и $y \in \mathcal{A}$, то $x + y \in \mathcal{A}$ и $xy \in \mathcal{A}$.

Структура алгебры $\mathcal{A}$ очень проста. Под *атомом* алгебры $\mathcal{A}$ мы понимаем какое-либо максимальное событие A такое, что каждая случайная величина из $\mathcal{A}$ постоянна на A. Таким образом, Ω разлагается на атомы. Это значит, что Ω есть объединение атомов и различные атомы не пересекаются. Если A — атом, а $\omega \notin A$, то по определению в $\mathcal{A}$ найдется элемент x, для которого $x(A) \neq x(\omega)$. Положим $x_\omega = (x - x(\omega))/(x(A) - x(\omega))$. Тогда $x_\omega \in \mathcal{A}$ и $x_\omega(A) = 1$, $x_\omega(\omega) = 0$. Следовательно, индикаторная функция χ_A входит в $\mathcal{A}$, так как

$$\chi_A = \prod_{\omega \notin A} x_\omega.$$

Таким образом, алгебра $\mathcal{A}$ состоит из всех случайных величин, постоянных на атомах $\mathcal{A}$. В свою очередь, для произвольного разбиения Ω множество всех случайных величин, постоянных на каждом событии разбиения, составляют некоторую алгебру случайных величин.

Отметим, что алгебра $\mathcal{A}$ случайных величин содержит произвольные функции своих элементов: если $f : \mathbf{R}^n \to \mathbf{R}$ и $x_1, \ldots, x_n$ — элементы алгебры $\mathcal{A}$, то суперпозицию $f(x_1, \ldots, x_n)$ также принадлежит $\mathcal{A}$.

В качестве примера рассмотрим множество Ω всех пар $\langle i, j \rangle$, $1 \leq i, j \leq 6$, и $\mathrm{pr}(\langle i, j \rangle) = 1/36$ для любой пары $\langle i, j \rangle \in \Omega$. Это модель для бросания пары игральных костей. Пусть $x(\langle i,j\rangle)=i$, $y(\langle i,j\rangle)=j$, а $z(\langle i,j\rangle)=i+j$. Пусть $\mathcal{A}$ — наименьшая алгебра, содержащая z. Атомы алгебры $\mathcal{A}$ указаны на рис. 2.1. Имеется 11 атомов, которые обозначены через $A_2,\ldots,A_{12}$, здесь индекс указывает значение величины z на соответствующем атоме.

Пусть $\mathcal{A}$ — некоторая алгебра случайных величин. Если A — атом из $\mathcal{A}$, то множество A само является конечным вероятностным пространством относительно вероятности pr_A, определенной для каждого $\omega\in A$ равенством

$$\mathrm{pr}_A(\omega)=\frac{1}{\Pr A}\mathrm{pr}(\omega).$$

Это значит, что *любая конструкция или теорема теории вероятностей может быть релятивизирована* (*снесена*) *на любую алгебру* $\mathcal{A}$ *случайных величин.* При такой релятивизации элементы из алгебры $\mathcal{A}$ играют роль констант.

Например, исходя из понятия математического ожидания $\mathbf{E}x$ случайной величины x, мы определяем *условное математическое ожидание,* или *условное среднее,* $\mathbf{E}_{\mathcal{A}}x$ как такой элемент алгебры $\mathcal{A}$, который на каждом атоме A совпадает с математическим ожиданием случайной величины x относительно pr_A. Таким образом, если A_ω — атом, содержащий ω, то

$$\mathbf{E}_{\mathcal{A}}x(\omega)=\frac{1}{\Pr A_\omega}\sum_{\eta\in A_\omega}x(\eta)\mathrm{pr}(\eta).$$

Математическое ожидание линейно:

$$\mathbf{E}(\lambda_1x_1+\lambda_2x_2)=\lambda_1\mathbf{E}x_1+\lambda_2\mathbf{E}x_2,\quad \lambda_1,\lambda_2\in\mathbf{R}.$$

Условное математическое ожидание $\mathcal{A}$-линейно:

$$\mathbf{E}_{\mathcal{A}}(y_1x_1+y_2x_2)=y_1\mathbf{E}_{\mathcal{A}}x_1+y_2\mathbf{E}_{\mathcal{A}}x_2,\quad y_1,y_2\in\mathcal{A}.$$

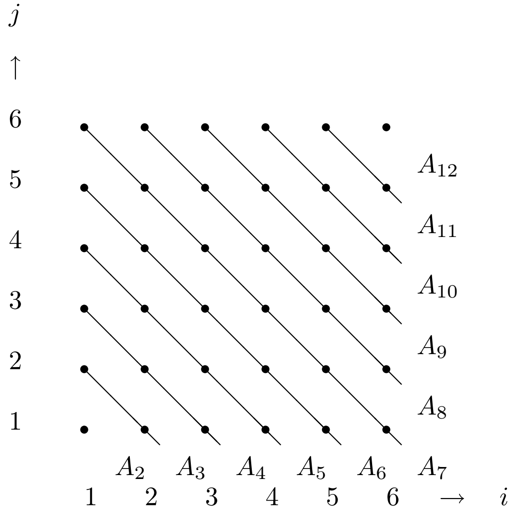


Рис. 2.1. Игра в кости

Математическое ожидание сохраняет константы, т. е. $\mathbf{E}\lambda = \lambda$, если λ — константа. Условное математическое ожидание сохраняет элементы из $\mathcal{A}$, т. е. $\mathbf{E}_{\mathcal{A}}(y) = y$, если $y \in \mathcal{A}$. Математическое ожидание — это ортогональный проектор на пространство констант, и, значит, $\mathbf{E}(x - \mathbf{E}x)^2 \le \mathbf{E}(x - \lambda)^2$ для любой константы λ. С учетом релятивизации $\mathbf{E}_{\mathcal{A}}(x - \mathbf{E}_{\mathcal{A}}x)^2 \le \mathbf{E}_{\mathcal{A}}(x - y)^2$ для всех $y \in \mathcal{A}$. Отметим, что если $\mathcal{B} \subseteq \mathcal{A}$, то $\mathbf{E}_{\mathcal{B}}\mathbf{E}_{\mathcal{A}} = \mathbf{E}_{\mathcal{B}}$. В частности, $\mathbf{E}\mathbf{E}_{\mathcal{A}} = \mathbf{E}$, так как $\mathbf{E}$ есть условное математическое ожидание относительно тривиальной алгебры, состоящей из констант. Поэтому $\mathbf{E}(x - \mathbf{E}_{\mathcal{A}}x)^2 \le \mathbf{E}(x - y)^2$ для любого $y \in \mathcal{A}$, так что $\mathbf{E}_{\mathcal{A}}$ есть ортогональный проектор на $\mathcal{A}$.

Величина $\mathbf{E}_{\mathcal{A}}(x)$ иначе обозначается через $\mathbf{E}\{x \mid \mathcal{A}\}$, а для условного математического ожидания случайной величины x относительно алгебры, порожденной $x_1, \ldots, x_n$, будем использовать обозначение $\mathbf{E}\{x \mid x_1, \ldots, x_n\}$. В примере на рис. 2.1 в силу симметрии $\mathbf{E}\{x \mid z\} = \mathbf{E}\{y \mid z\}$, так что $\mathbf{E}\{x \mid z\} = \frac{1}{2}\mathbf{E}\{z \mid z\} = \frac{1}{2}z$.

Определим *условную вероятность* $\Pr_{\mathcal{A}} B$ события B относительно алгебры $\mathcal{A}$ с помощью релятивизации как $\Pr_{\mathcal{A}} B = \mathbf{E}_{\mathcal{A}}\chi_B$. Таким образом,

$$\mathrm{Pr}_{\mathcal{A}}B(\omega) = \frac{\Pr(B \bigcap A_\omega)}{\Pr A_\omega},$$

где A_ω — атом из $\mathcal{A}$, содержащий ω. Отметим, что это (вообще говоря, непостоянная) случайная величина.

Релятивизация неравенства Гёльдера дает:

$$|\mathbf{E}_{\mathcal{A}}xy| \le (\mathbf{E}_{\mathcal{A}}|x|^p)^{1/p}(\mathbf{E}_{\mathcal{A}}|y|^{p'})^{1/p'}.$$

Релятивизация неравенства Йенсена состоит в следующем: для выпуклой функции f будет

$$f(\mathbf{E}_{\mathcal{A}}x) \le \mathbf{E}_{\mathcal{A}}f(x).$$

Релятивизация неравенства Чебышёва гласит:

$$\mathrm{Pr}_{\mathcal{A}}\{f(x) \ge y\} \le \frac{\mathbf{E}_{\mathcal{A}}f(x)}{y}$$

для положительной функции f и $y > 0$ из $\mathcal{A}$. Из неравенства Йенсена следует, что

$$|\mathbf{E}_{\mathcal{A}}x|^p \le \mathbf{E}_{\mathcal{A}}|x|^p,$$

а так как $\mathbf{E}\mathbf{E}_{\mathcal{A}} = \mathbf{E}$, то $\|\mathbf{E}_{\mathcal{A}}x\|_p \le \|x\|_p$, $1 \le p \le \infty$.

Пусть $\mathcal{A}$ — некоторая алгебра случайных величин. Через $\mathrm{at}(\mathcal{A})$ обозначим множество всех атомов алгебры $\mathcal{A}$. Не только каждый элемент $A \in \mathrm{at}(\mathcal{A})$ является конечным вероятностным пространством относительно pr_A, но и множество $\mathrm{at}(\mathcal{A})$ само есть конечное вероятностное пространство относительно вероятности

$$\mathrm{pr}'_{\mathcal{A}}(A) = \mathrm{Pr}(A)$$

для $A \in \mathrm{at}(\mathcal{A})$. Будем говорить, что исходное вероятностное пространство $\langle \Omega, \mathrm{pr}\rangle$ *расслоено* над $\langle \mathrm{at}(\mathcal{A}), \mathrm{pr}'_{\mathcal{A}}\rangle$ *слоями* $\langle A, \mathrm{pr}_A\rangle$. Наглядно это можно представить на рис. 2.1, поворачивая его на 45° по часовой стрелке. Условные математические ожидания относительно $\mathrm{pr}'_{\mathcal{A}}$ обозначаются через $\mathbf{E}'_{\mathcal{A}}$, а для вероятности на множестве атомов используется обозначение $\mathrm{Pr}'_{\mathcal{A}}$. Отметим, что $\mathbf{E}'_{\mathcal{A}}\mathbf{E}_{\mathcal{A}}x = \mathbf{E}x$.

Частный случай расслоения — это произведение. Предположим, что $\langle \Omega_1, \mathrm{pr}_1\rangle$ и $\langle \Omega_2, \mathrm{pr}_2\rangle$ — конечные вероятностные пространства. Тогда $\Omega_1 \times \Omega_2$ — конечное вероятностное пространство относительно $\mathrm{pr}_1 \times \mathrm{pr}_2$, где

$$\mathrm{pr}_1 \times \mathrm{pr}_2(\langle \omega_1, \omega_2\rangle) = \mathrm{pr}_1(\omega_1)\mathrm{pr}_2(\omega_2).$$

Пусть $\mathcal{A}_1$ — алгебра всех случайных величин, которые являются функциями, зависящими только от аргумента ω_1. Тогда $\mathrm{at}(\mathcal{A}_1)$ состоит из всех множеств вида $\{\langle \omega_1, \omega_2\rangle : \omega_1 = \eta_1\}$, где η_1 — произвольный элемент из Ω_1.

Глава 3
Стохастические процессы

Слово «стохастический» означает случайный, а слово «процесс» в этом контексте означает функцию, так что стохастический процесс — это функция, значения которой суть некоторые случайные величины. Пусть T — конечное множество, а $\langle \Omega, \mathrm{pr} \rangle$ — конечное вероятностное пространство. *Стохастический процесс, параметризованный элементами множества* T *и определенный на* $\langle \Omega, \mathrm{pr} \rangle$, — это функция $\xi : T \to \mathbf{R}^{\Omega}$. Говоря о «стохастическом процессе», мы всегда будем подразумевать, что он параметризован элементами конечного множества и определен на конечном вероятностном пространстве. Введем обозначения $\xi(t, \omega)$ для значения процесса $\xi(t)$ в точке ω и $\xi(\cdot, \omega)$ для функции $t \mapsto \xi(t, \omega)$. Таким образом, каждая функция $\xi(t)$ есть случайная величина, $\xi(t, \omega)$ — вещественное число, а $\xi(\cdot, \omega)$ — это вещественнозначная функция на T, называемая *траекторией* или *выборочным путем* процесса.

Пусть Λ_ξ — это множество всех траекторий стохастического процесса ξ. Тогда Λ_ξ — конечное подмножество конечномерного векторного пространства $\mathbf{R}^T$ всех функций из T в $\mathbf{R}$. Определим *вероятностное распределение процесса* pr_ξ равенством

$$\mathrm{pr}_\xi(\lambda) = \mathrm{Pr}\{\xi(t) = \lambda(t) \text{ для каждого } t \in T\}$$

для всех λ из Λ_ξ. Тогда $\langle \Lambda_\xi, \mathrm{pr}_\xi \rangle$ — конечное вероятностное пространство. Два стохастических процесса ξ и $\xi^{'}$, параметризованные элементами одного и того же конечного множества T, но определенные, возможно, на различных конечных вероятностных пространствах, называют *эквивалентными,* если $\Lambda_\xi = \Lambda_{\xi'}$ и $\mathrm{pr}_\xi = \mathrm{pr}_{\xi'}$, т. е. если они имеют одни и те же траектории с теми же самыми вероятностями. *Теория* *вероятностей имеет дело только с теми свойствами стохастических процессов, которые являются общими для всех эквивалентных стохастических процессов.* Отметим, что функция, переводящая t в отображение вычисления $\lambda \mapsto \lambda(t)$, есть стохастический процесс, определенный на $\langle \Lambda_\xi, \mathrm{pr}_\xi \rangle$ и эквивалентный процессу ξ. Таким образом, при изучении стохастического процесса можно, не теряя общности, предполагать, что он определен на пространстве своих траекторий.

Когда множество T состоит из единственного элемента, мы можем отождествить стохастический процесс, параметризованный T, с соответствующей случайной величиной. Иначе говоря, случайная величина — это просто частный случай стохастического процесса. Если x — случайная величина, то $\mathrm{pr}_x(\lambda) = \Pr\{x = \lambda\}$, где $\lambda \in \Lambda_x$. Здесь

$$\Lambda_x = \{\lambda \in \mathbf{R} : \Pr\{x = \lambda\} \neq 0\}.$$

Легко видеть, что

$$\mathbf{E}x = \sum \lambda \mathrm{pr}_x(\lambda). \tag{3.1}$$

Правая часть равенства (3.1) есть математическое ожидание тождественной функции λ на $\langle \Lambda_x, \mathrm{pr}_x \rangle$. (Если бы (3.1) было неверно, то согласно констатации из предыдущего абзаца математическое ожидание не имело бы отношения к теории вероятностей.)

Случайные величины стохастического процесса ξ называются *независимыми,* если

$$\mathrm{pr}_\xi(\lambda) = \prod_{t \in T} \mathrm{pr}_{\xi(t)}(\lambda(t))$$

для всех $\lambda \in \Lambda_\xi$. Предположим для примера, что $T = \{1, \ldots, \nu\}$, а x_0 — случайная величина на $\langle \Omega, \mathrm{pr} \rangle$. Тогда $\langle \Omega^\nu, \mathrm{pr}^\nu \rangle$, где

$$\mathrm{pr}^\nu(\omega_1, \ldots, \omega_\nu) = \mathrm{pr}(\omega_1) \cdots \mathrm{pr}(\omega_\nu)$$

— конечное вероятностное пространство, а случайные величины x_n, определенные равенствами

$$x_n(\omega_1, \ldots, \omega_\nu) = x_0(\omega_n), \quad 1 \leq n \leq \nu,$$

независимы. Такой стохастический процесс описывает ν повторных независимых наблюдений заданной случайной величины x_0.

События $A_1, \dots, A_\nu$ называются *независимыми,* если независимы их индикаторные функции. Пусть в примере с бросанием игральной кости (рис. 2.1) событие A означает, что i нечетно, событие B означает, что j нечетно, а событие C означает, что $i + j$ нечетно. Тогда A и B независимы; A и C независимы; B и C независимы; однако A, B и C не независимы. Событие A ничего не говорит нам о событии C, событие B ничего не говорит нам о событии C, в то же время совокупность событий A и B говорит нам о событии C все. Именно на этом самом принципе строится хороший детективный сюжет.

Глава 4
Внешние понятия

Пусть x_0 — случайная величина с нулевым средним и единичной дисперсией. Рассмотрим ν независимых наблюдений $x_1, \ldots, x_\nu$ величины x_0.

> Если ν — большое число, то почти наверное для любого большого $n \leq \nu$ среднее арифметическое $(x_1 + \cdots + x_n)/n$ примерно равно 0.

Это утверждение является интуитивным выражением усиленного закона больших чисел. Оно неточно, потому что мы не объяснили, что значит «большой», «почти наверное» и «примерно равно».

Теперь коротко поясним, как формулируется усиленный закон больших чисел в традиционной математике. Конечная последовательность $x_1, \ldots, x_\nu$ заменяется актуально бесконечной последовательностью $x_1, x_2, \ldots$. Чтобы сделать это, необходимо осуществить бесконечное число декартовых произведений исходного вероятностного пространства Ω на себя. Даже если Ω — конечное вероятностное пространство, такая бесконечная декартова степень Ω будет содержать несчетное множество точек. Каждая индивидуальная точка имеет нулевую вероятность. Только некоторые (измеримые) множества представляют собой события, а вероятность события уже больше не является суммой вероятностей своих точек. Только некоторые (измеримые) функции служат случайными величинами. Математическое ожидание становится интегралом, и только некоторые (интегрируемые) случайные величины имеют математические ожидания. При этом усиленный закон больших чисел превращается в утверждение о том, что, за исключением некоторого события с нулевой вероятностью, для любого $\varepsilon > 0$ существует такое натуральное число m, что для любого $n \geq m$ будет $|(x_1 + \cdots + x_n)/n| \leq \varepsilon$.

Мы избираем другой подход, который имеет то достоинство, что он остается в элементарных рамках конечных вероятностных пространств. Мы сохраняем конечную последовательность $x_1, \ldots, x_\nu$, однако считаем число ν нестандартным. Под бесконечно малым числом, или инфинитезималью, мы понимаем вещественное число, чья абсолютная величина меньше, чем обратная величина некоторого нестандартного натурального числа, и говорим, что два вещественных числа примерно равны, если их разность бесконечно мала. Какое-то свойство считается выполненным почти наверное, если для любого не инфинитезимального положительного ε существует событие N, для которого $\Pr N \leq \varepsilon$, и рассматриваемое свойство выполняется на N^c. Если теперь слово «большой» использовать как синоним слова «нестандартный», формулировка усиленного закона больших совпадет с ранее приведенным интуитивным утверждением.

Традиционный подход связан с идеализацией, потому что никто реально не может провести бесконечное число наблюдений. Второй подход также связан с идеализацией, потому что никто реально не может провести нестандартное число наблюдений. На самом деле, иметь дело с идеализациями в природе математики. Выбор формализма должен быть основан на эстетических соображениях, подобных ясности выражения, простоте и силе. Различные формализмы ни в коей мере не исключают друг друга, и если посмотреть на хорошо известный материал со свежей точки зрения, то можно прояснить суть дела.

Посмотрим, как возникает понятие нестандартного числа. Пусть $\mathbf{N}$ — множество всех натуральных чисел $0, 1, 2, \ldots$. Основным свойством множества $\mathbf{N}$ является *теорема индукции,* которая гласит, что если S — подмножество множества $\mathbf{N}$, содержащее 0 и обладающее тем свойством, что из $n \in S$ следует $n + 1 \in S$, то $S = \mathbf{N}$. Если теперь $\mathbf{A}(n)$ — некоторая формула традиционной математики типа «n простое и $n + 2$ простое» или «$n \geq m$», то мы можем образовать подмножество $S = \{n \in \mathbf{N} : \mathbf{A}(n)\}$ всех натуральных чисел n, для которых справедлива формула $\mathbf{A}(n)$. Однако формула должна быть формулой принятого в математике языка. Множества не являются объектами реального мира, они суть формальные математические объекты и существуют только тогда, когда формальные правила математики утверждают их существование. Например, нет смысла рассматривать множество $S = \{n \in \mathbf{N} : \mathbf{A}(n)\}$, если $\mathbf{A}(n)$ — это утверждение, что «по моему мнению, n не чрезмерно велико».

После работы Гёделя начала тридцатых годов выяснилось, что основные интуитивные системы математики, такие как $\mathbf{N}$, не могут быть полностью описаны никакой системой аксиом. Чтобы пояснить, что это значит, давайте добавим к языку традиционной математики новый неопределяемый предикат «стандартный». При этом утверждение «n стандартно» лишено смысла в традиционной математике. Назовем формулу *внутренней,* если она не включает в себя предикат «стандартный» — в этом случае она является формулой традиционной математики; в противном случае мы называем ее *внешней.* Простейший пример внешней формулы — «натуральное число n стандартно». Другой пример внешней формулы — «x бесконечно мало», поскольку по определению это означает, что существует нестандартное натуральное число ν такое, что $|x| \leq 1/\nu$. *Только внутренние формулы можно использовать для образования подмножеств.* (Например, бессмысленно говорить «множество всех стандартных натуральных чисел» или «множество всех бесконечно малых вещественных чисел».) Нарушение этого правила мы называем *незаконным введением множества.*

Сделаем следующие предположения:

(1) *число* 0 *стандартно*,

(2) *если $n \in \mathbf{N}$ стандартно, то и $n+1$ стандартно.*

При этом невозможно доказать, что каждое $n \in \mathbf{N}$ стандартно. (Это утверждение не противоречит теореме индукции — оно просто показывает, что нельзя доказать существование такого подмножества $S \subseteq \mathbf{N}$, что натуральное число n принадлежит S тогда и только тогда, когда n стандартно.) Иначе говоря, предыдущему не противоречит такое предположение:

(3) *существует нестандартное натуральное число $n \in \mathbf{N}$.*

Мы также принимаем следующее предположение:

(4) *если утверждение $\mathbf{A}(0)$ верно и для каждого стандартного числа n из выполнения $\mathbf{A}(n)$ следует справедливость $\mathbf{A}(n+1)$, то $\mathbf{A}(n)$ имеет место для всех стандартных чисел n.*

В (4) $\mathbf{A}(n)$ — это любая формула, внутренняя или внешняя. Сформулированное предположение называют *внешней индукцией.* Оно является дополнением к обычной индукции, которая, как мы видели, может нарушаться для внешних формул. (Конечно, обычная индукция продолжает выполняться для обычных, т. е. внутренних, формул. Ничто не меняется в традиционной математике; просто мы строим более богатый язык, чтобы обсуждать те же самые математические объекты, что и прежде.)

Используя внешнюю индукцию, мы можем легко доказать, что *любое нестандартное натуральное число больше каждого стандартного натурального числа* (для нестандартного натурального числа ν в (4) надо взять $\mathbf{A}(n)$ — «$\nu \geq n$»), что *сумма двух стандартных натуральных чисел стандартна* (для стандартного натурального числа m надо взять $\mathbf{A}(n)$ — «$n+m$ стандартно») и что *произведение двух стандартных натуральных чисел стандартно* (возьмем стандартное натуральное число m, считая, что $\mathbf{A}(n)$ — это «nm стандартно», и используем только что отмеченный факт о сумме).

Другое предположение, которое мы иногда будем использовать, называется *принципом последовательности.* Пусть $\mathbf{A}(n, x)$ — формула, внутренняя или внешняя. Если для любого стандартного номера n существует такой элемент x, что выполняется формула $\mathbf{A}(n, x)$, то, конечно, существует такой элемент x_0, что выполняется формула $\mathbf{A}(0, x_0)$, и существует такой элемент x_1, что выполняется формула $\mathbf{A}(1, x_1)$, и существует такой элемент x_2, что выполняется формула $\mathbf{A}(2, x_2)$, и так далее. Мы предполагаем:

> (*5) *Если для любого стандартного натурального числа n существует элемент x такой, что формула $\mathbf{A}(n, x)$ верна, то существует последовательность $n \mapsto x_n$ такая, что для любого стандартного n будет выполнено $\mathbf{A}(n, x_n)$.*

В качестве примера использования принципа последовательности см. доказательство теоремы 6.1. Результаты, полученные с использованием принципа последовательности, будут отмечены звездочкой (*).

Отметим, что согласно (2) не существует наименьшего нестандартного натурального числа. Мы можем представлять натуральные числа расположенными на ленте (рис. 4.1). Стандартные натуральные числа, пока дело касается внутренних свойств, ведут себя точно так же, как вся система $\mathbf{N}$. Но $\mathbf{N}$ содержит как стандартные, так и нестандартные натуральные числа. Заметим, что мы не начали с левой части ленты с тем чтобы добавить к ней изобретенную правую часть.

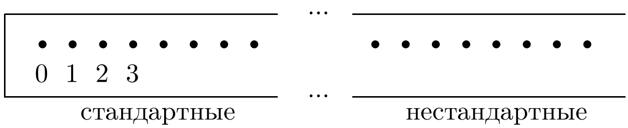


Рис. 4.1. Натуральные числа

Скорее, мы с самого начала взяли целую ленту, а затем добавили к нашему языку предикат, который позволил нам различать две части ленты. Использование этого нового предиката «стандартный» похоже на применение цвета в телевизоре: картинка та же самая, но мы можем заметить различия, которые нельзя было увидеть прежде.

В течение долгого времени неполнота аксиоматических систем воспринималась математиками как недостаток. Гений А. Робинсона в начале шестидесятых обратил неполноту в продуктивное средство, показав, что благодаря ей может быть достигнуто огромное упрощение математических рассуждений.

Глава 5

Инфинитезимали, или бесконечно малые числа

Введем некоторые полезные внешние понятия, относящиеся к полю вещественных чисел $\mathbf{R}$.

Вещественное число x называется *бесконечно малым,* или *инфинитезимальным,* или просто *инфинитезималью,* если $|x| \leq 1/\nu$ для некоторого нестандартного натурального числа ν. Так как нестандартное число ν больше любого стандартного числа n, отсюда вытекает, что если x — инфинитезималь, то $|x| \leq 1/n$ для каждого стандартного натурального n. Верно и обратное: если $|x| \leq 1/n$ для любого $n \in \mathbf{N}$, то $x = 0$, т. е. x — инфинитезималь; в противном случае, пусть μ — наименьшее натуральное число такое, что $|x| \geq 1/\mu$. Тогда μ нестандартно, значит, и $\nu = \mu - 1$ также нестандартно. Но $|x| \leq 1/\nu$, таким образом, x — инфинитезималь.

Вещественное число x называется *доступным*, если $|x| \leq n$ для некоторого стандартного числа $n \in \mathbf{N}$, в противном случае x называется *недоступным.* Слова «конечный» и «бесконечный» используются как синонимы для слов «доступный» и «недоступный» соответственно, но так как они уже имеют внутренний смысл, их использование может привести к недоразумению, как, например, в высказывании «этот интеграл конечен».

Если x и y — вещественные числа, то будем писать:

$x \simeq y$, если число $x - y$ бесконечно мало;

$x \lesssim y$, если $x \leq y + \alpha$ для некоторой инфинитезимали α;

$x \gtrsim y$, если $y \lesssim x$;

$x \ll y$, если $x < y$ и $x \not\simeq y$;

$x \gg y$, если $y \ll x$.

Мы можем читать формулы следующим образом:

$x \simeq y$ — «*x примерно равно y*» или «*x бесконечно близко* к *y*»;
$x \lesssim y$ — «*x ослабленно меньше y*»;
$x \gtrsim y$ — «*x ослабленно больше y*»;
$x \ll y$ — «*x усиленно меньше y*»;
$x \gg y$ — «*x усиленно больше y*»;
$x \gg 0$ — «*x усиленно положителен*».

Расширенная числовая прямая $\overline{\mathbf{R}}$ состоит из $\mathbf{R}$ и двух точек $-\infty$ и ∞. Мы пишем $-\infty < x < \infty$ для любого $x \in \mathbf{R}$. По определению $x \simeq \infty$ означает, что x положительно и недоступно, $x \ll \infty$ (или $\infty \gg x$) значит, что $x \not\simeq \infty$, и $-\infty \ll x$ (или $x \gg -\infty$), что $x \not\simeq -\infty$. Таким образом, $|x| \ll \infty$ тогда и только тогда, когда число x доступно, и $|x| \simeq \infty$ тогда и только тогда, когда число x недоступно.

Постарайтесь зрительно представить себе отношения $\simeq$, $\lesssim$ и $\ll$ на вещественной числовой прямой. Для невооруженного глаза $\simeq$ выглядит как $=$, $\lesssim$ как $\leq$, а $\ll$ как $<$.

Ниже приведен ряд предложений, которые легко вытекают из определений.

1. $x \simeq 0$ тогда и только тогда, когда x бесконечно мало.

2. $x \simeq 0$ тогда и только тогда, когда $|x| \leq \varepsilon$ для любого $\varepsilon \gg 0$.

3. Бесконечно малые доступны.

4. Пусть $x \neq 0$. Тогда $x \simeq 0$ в том и только в том случае, если число $1/x$ недоступно.

5. $|x| \simeq \infty$ тогда и только тогда, когда $1/x \simeq 0$.

6. Если числа x и y доступны, то доступны также и числа $x + y$ и xy.

7. Если x и y бесконечно малы, то $x + y$ и xy также бесконечно малы.

8. Если $x \simeq 0$ и $|y| \ll \infty$, то $xy \simeq 0$.

9. $x \lesssim y$ и $y \lesssim x$ тогда и только тогда, когда $x \simeq y$.

10. Если $x \simeq y$ и $y \simeq z$, то $x \simeq z$.

11. Для любого $n \in \mathbf{N}$ число n стандартно тогда и только тогда, когда оно доступно.

12. Для любого $n \in \mathbf{N}$ число n нестандартно тогда и только тогда, когда оно недоступно.

Теорема 5.1. *Если натуральное число n стандартно и $x_i \simeq y_i$, $i = 1, \ldots, n$, то*

$$\sum_{i=1}^{n} x_i \simeq \sum_{i=1}^{n} y_i.$$

Доказательство. Нужно показать, что $\sum_{i=1}^{n}(x_i - y_i) \simeq 0$. Привлекаем (7) и внешнюю индукцию. $\square$

Утверждение теоремы, вообще говоря, неверно для недоступного числа n: возьмем, например, $x_i = 1/n$ и $y_i = 0$.

Если $x \neq 0$ и $y \neq 0$, то будем говорить, что $x \sim y$ (словами: x *асимптотически равно* y или x и y *одного порядка*) при $x/y \simeq 1$.

Теорема 5.2. *Если* $0 \ll |x|, |y| \ll \infty$ *(т. е. если числа* x *и* y *не инфинитезимальны и доступны), то* $x \sim y$ *тогда и только тогда, когда* $x \simeq y$.

Доказательство. Пусть $0 \ll |x|, |y| \ll \infty$. Предположим, что $x \sim y$. Тогда $x/y = 1 + \alpha$, $\alpha \simeq 0$, т. е. $x - y = \alpha y$. Но согласно (8) $\alpha y \simeq 0$, поэтому $x \simeq y$. Обратно, предположим, что $x \simeq y$, тогда $x - y = \alpha$, $\alpha \simeq 0$, так что $x/y = 1 + \alpha/y$. Но согласно (5) и (8) $\alpha/y \simeq 0$, и, следовательно, $x \sim y$. $\square$

Теорема 5.3. *Если* $x_i > 0$, $y_i > 0$ *и* $x_i \sim y_i$, $i = 1, \ldots, n$, *то*

$$\sum_{i=1}^{n} x_i \simeq \sum_{i=1}^{n} y_i.$$

Доказательство. Пусть $x > 0$ и $y > 0$. Тогда $x \sim y$ в том и только в том случае, если $(1-\varepsilon)y \leq x \leq (1+\varepsilon)y$ для любого $\varepsilon \gg 0$. Пусть $\varepsilon \gg 0$. Тогда

$$(1-\varepsilon)y_i \leq x_i \leq (1+\varepsilon)y_i.$$

Поэтому

$$(1-\varepsilon)\sum_{i=1}^{n} y_i \leq \sum_{i=1}^{n} x_i \leq (1+\varepsilon)\sum_{i=1}^{n} y_i. \quad \square$$

В теореме 5.3 не требуется, чтобы число n было доступным. Вместе теоремы 5.2 и 5.3 объясняют, почему срабатывает интегрирование: при сложении недоступного числа инфинитезималей мы совершаем лишь бесконечно малую ошибку, всего лишь при условии, что сумма доступна и относительная ошибка, сделанная в каждом слагаемом, составляет его бесконечно малую *часть*.

Я уже подчеркивал, что правила нашей теории не позволяют нам образовывать подмножества, соответствующие внешним свойствам. Во многих случаях мы можем даже доказать, что таких множеств просто не существует.

Теорема 5.4. *Не существует множеств* A_1, A_2, A_3, A_4, A_5 *таких, что (для любых* n *и* x*) выполняется*

$n \in A_1$ *тогда и только тогда, когда* $n \in \mathbf{N}$ *и* n *стандартно*;

$n \in A_2$ *тогда и только тогда, когда* $n \in \mathbf{N}$ *и* n *нестандартно*;

$x \in A_3$ *тогда и только тогда, когда* $x \in \mathbf{R}$ *и* x *доступно*;

$x \in A_4$ *тогда и только тогда, когда* $x \in \mathbf{R}$ *и* x *недоступно*;

$x \in A_5$ *тогда и только тогда, когда* $x \in \mathbf{R}$ *и* x *бесконечно мало.*

Доказательство. Существование множества A_1 нарушило бы теорему индукции; если бы A_2 существовало, мы могли бы положить $A_1 = \mathbf{N} \setminus A_2$; если бы A_3 существовало, мы могли бы положить $A_2 = \mathbf{N} \setminus A_3$; если бы A_4 существовало, мы могли бы положить $A_3 = \mathbf{R} \setminus A_4$; а если бы A_5 существовало, мы могли бы положить $A_4 = \{x \in \mathbf{R} : \ 1/x \in A_5\}$. $\square$

Эта теорема выглядит как отрицательный результат, но она очень полезна. Например, если $\mathbf{A}(x)$ — *внешняя* формула, и мы доказали, что для любого бесконечно малого числа x выполнена формула $\mathbf{A}(x)$, то можно заключить о существовании отличного от инфинитезимали числа x, для которого выполняется $\mathbf{A}(x)$, потому что иначе мы могли бы положить $A_5 = \{x \in \mathbf{R} : \mathbf{A}(x)\}$. Этот эффект называют *переполнением*, или *принципом переполнения.*

Пусть $x_1, x_2, \ldots$ — такая последовательность вещественных чисел, что $x_n \simeq 0$ для любого доступного номера n. Так как «$x_n \simeq 0$» — внешняя формула, мы не можем непосредственно использовать принцип переполнения, чтобы доказать, что $x_n \simeq 0$ для некоторого недоступного номера n. Тем не менее, сам этот результат, известный под названием *лемма Робинсона,* верен, а его доказательство получается заменой формулы «$x_n \simeq 0$» более слабой внешней формулой:

Теорема 5.5. *Пусть* $x_1, x_2, \ldots$ *— такая последовательность вещественных чисел, что* $x_n \simeq 0$ *для любого доступного номера* n*. Тогда найдется недоступное число* ν *такое, что* $x_n \simeq 0$ *для любого номера* $n \leq \nu$.

Доказательство. Рассмотрим множество S таких чисел m, что $|x_n| \leq 1/n$ для всех $n \leq m$. Это множество содержит все стандартные числа и, по принципу переполнения (теорема 5.4), оно содержит нестандартное (недоступное) натуральное число ν. Пусть $n \leq \nu$. Если номер n доступен, то, по предположению, $x_n \simeq 0$. Если же номер n недоступен, то $\mid x_n \mid \leq 1/n \simeq 0$, так как $\nu \in S$. Таким образом, $\nu \simeq \infty$ и $x_n \simeq 0$ для любого номера $n \leq \nu$. $\square$

Если бы мы попытались построить контрпример, рассмотрев такую последовательность: $x_n = 0$, когда номер n доступен, и $x_n = 1$, когда номер n недоступен, то мы совершили бы незаконное введение множества. Такой последовательности не существует. Ведь последовательность — это некоторая функция, функция — это некоторое множество, а мы не можем пользоваться внешними свойствами для определения множеств.

Глава 6
Внешние аналоги внутренних понятий

Пусть T — конечное подмножество вещественной прямой $\mathbf{R}$. На протяжении всей книги мы будем пользоваться следующими обозначениями: a — это первый элемент множества параметров T, последний элемент в T — это b; мы полагаем $T' = T \setminus \{b\}$, для $t \in T'$ его преемник — это $t + dt$, и для любой функции $\xi : T \to \mathbf{R}$ будем считать $d\xi(t) = \xi(t + dt) - \xi(t)$. Если для отображения $\xi : T \to \mathbf{R}$ написано $\xi(t)$, то при этом подразумевается, что $t \in T$. Если $0 \ll b - a \ll \infty$ и каждое dt бесконечно мало, то назовем T *около-интервалом.*

Бо́льшая часть понятий анализа лишается смысла, когда их применяют к функции, областью определения которой является конечное множество точек. Поэтому нет опасности недоразумений, если использовать знакомую терминологию для внешних аналогов подобных понятий. Однако иногда удобно вставлять модификатор типа «около» или «примерно» перед знакомым термином, в случае, если он используется как элементарный внешний аналог.

Пусть $T = \{1, \ldots, \nu\}$, где ν — недоступное натуральное число. Это элементарный аналог для $\mathbf{N}^+ = \mathbf{N} \setminus \{0\}$. По аналогии с понятием сходимости бесконечной последовательности мы говорим, что последовательность $x_1, \ldots, x_\nu$ является (около-)*сходящейся*, если существует такое число x, что $x_n \simeq x$ для всех недоступных номеров $n \leq \nu$. В этом случае мы также говорим, что последовательность $x_1, \ldots, x_\nu$ (около-)*сходится к* x. Отметим, что если последовательность сходится к x, то она будет также сходиться и к y тогда и только тогда, когда $x \simeq y$.

На рис. 6.1 значения x_n нельзя отличить от x невооруженным глазом при условии, что номер $n \leq \nu$ недоступен. Возможно, из этого рисунка и не видно, что около-сходимость схватывает интуитивное понятие все более и более тесного приближения к x. Что мы можем сказать о значениях x_n для $n \ll \infty$? Пусть $\varepsilon \gg 0$, и пусть n_ε — наименьшее из чисел таких, что $|x_n - x| \le \varepsilon$ для любого n, $n_\varepsilon \le n \le \nu$. Тогда $n_\varepsilon \ll \infty$, ибо если $n_\varepsilon \simeq \infty$, то $n_\varepsilon - 1 \simeq \infty$ и $|x_{n_\varepsilon - 1} - x| \le \varepsilon$ (рис. 6.2).

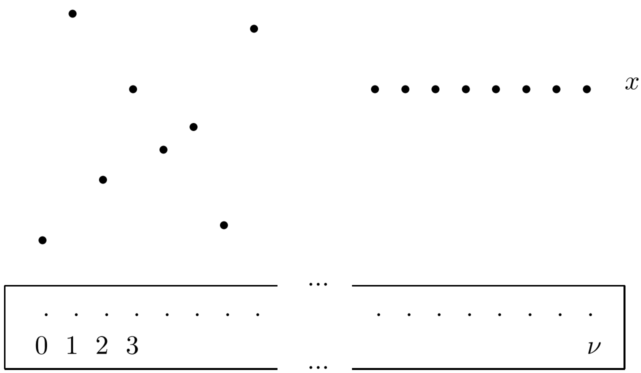


Рис. 6.1. Около-сходимость

Для каждого данного внутреннего понятия нет однозначного способа построения элементарного внешнего аналога. Пусть T — подмножество $\mathbf{R}$, и пусть $\xi : T \to \mathbf{R}$. Будем говорить, что ξ *допускает k колебаний порядка ε или ε-колебаний,* если в T существуют такие элементы $t_0 < \cdots < t_k$, что

$$|\xi(t_0) - \xi(t_1)| \ge \varepsilon,\ |\xi(t_1) - \xi(t_2)| \ge \varepsilon, \ldots, |\xi(t_{k-1}) - \xi(t_k)| \ge \varepsilon$$

(при этом $t_0, \ldots, t_k$ называем *параметрами k ε-колебаний*). Эти понятия внутренние. Бесконечная последовательность сходится тогда и только тогда, когда для любого $\varepsilon > 0$ найдется число k такое, что наша последовательность не допускает k ε-колебаний. Это наводит на мысль ввести следующее внешнее определение. Будем говорить, что бесконечная или конечная последовательность (или любая функция $\xi : T \to \mathbf{R}$) имеет *доступное колебание*, если ни для каких $\varepsilon \gg 0$ и $k \simeq \infty$ она не допускает k ε-колебаний. Таким образом, свойство последовательности иметь доступное колебание, как и свойство около-сходимости, есть внешний аналог внутреннего понятия сходимости. Но это более слабое свойство: пусть число $i \le \nu$ недоступно, и пусть $x_n = 0$, если $n \le i$, и $x_n = 1$, если $i < n \le \nu$.

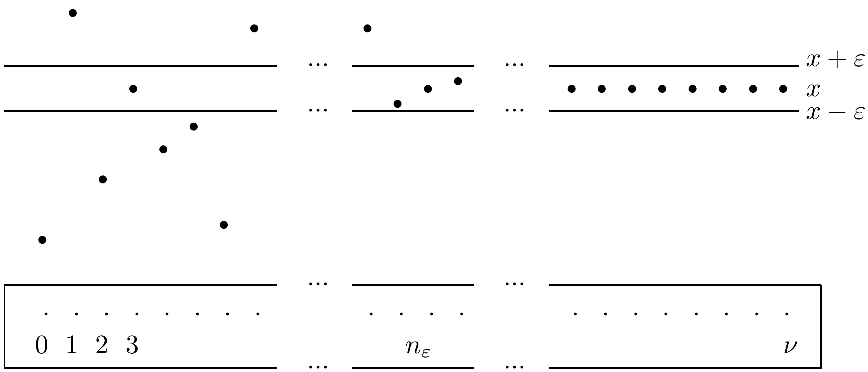


Рис. 6.2. Снова около-сходимость

Тогда последовательность $x_1, \ldots, x_\nu$ имеет доступное колебание, но не сходится. С другой стороны, если последовательность $x_1, \ldots, x_\nu$ сходится, то, как легко видеть (см. рис. 6.2), она имеет доступное колебание.

Как понятие около-сходимости, так и понятие доступного колебания будут важны для изучения флуктуаций стохастических процессов. Около-сходимость выражает порядковое — ординальное — свойство сходимости (начиная с некоторой точки нет ε-колебаний), тогда как доступность колебания выражает мощностное — кардинальное — свойство сходимости (существует только доступное число ε-колебаний).

Пусть $\mathbf{A}(n)$ — любая формула, внутренняя или внешняя. *Внешний принцип наименьшего числа* утверждает, что если существует такое стандартное число n, что $\mathbf{A}(n)$ верна, то существует такое наименьшее число m (с необходимостью стандартное), что для него $\mathbf{A}(m)$ тоже верна. Это может быть доказано внешней индукцией точно так же, как по индукции доказывается обычный принцип наименьшего числа.

***Теорема 6.1.** *Пусть $x_1, \ldots, x_\nu$ — последовательность доступного колебания, $\nu \simeq \infty$. Тогда существует такое недоступное число $\mu \le \nu$, что последовательность $x_1, \ldots, x_\mu$ сходится.*

Доказательство. Пусть $j \ll \infty$. Множество всех k таких, что $x_1, \ldots, x_\nu$ не допускает k колебаний порядка $1/j$, содержит все натуральные числа $k \simeq \infty$ и потому согласно принципу переполнения оно содержит некоторое число $k \ll \infty$. По внешнему принципу наименьшего числа существует наименьшее число l, $l \leq k$, такое, что не существует *недоступных* номеров $l+1$ колебаний порядка $1/j$. Если $l = 0$, положим $\mu_j = \nu$. Если $l > 0$, то существуют недоступные числа $n_0, \ldots, n_l$, которые являются параметрами l колебаний порядка $1/j$. В этом случае полагаем $\mu_j = n_0$.

Теперь $\infty \simeq \mu_j \leq \nu$ и для любого недоступного номера $n \leq \mu_j$ имеем $|x_n - x_{\mu_j}| < 1/j$. Согласно принципу последовательности существует последовательность натуральных чисел $j \mapsto \mu_j$ такая, что для любого номера $j \ll \infty$ эти свойства выполнены. Сделаем возникшую последовательность убывающей, полагая $\widetilde{\mu}_j = \inf_{l \leq \mu_j} \mu_l$. По лемме Робинсона, примененной к величинам, обратным $\widetilde{\mu}_j$, существует такое недоступное натуральное число j, что $\widetilde{\mu}_j$ недоступно. Пусть $\mu = \widetilde{\mu}_j$ для такого j. Таким образом, число μ недоступно, $\mu \leq \nu$ и $\mu \leq \mu_j$ для любого номера $j \ll \infty$. Для любого недоступного номера $n \leq \mu$ выполнено

$$|x_n - x_\mu| \leq |x_n - x_{\mu_j}| + |x_{\mu_j} - x_\mu| \leq \frac{2}{j}$$

для каждого номера $j \ll \infty$, т. е. $x_n \simeq x_\mu$. Таким образом, последовательность $x_1, \ldots, x_\mu$ сходится. □

Будем говорить, что сумма $\sum_{i=1}^{\nu} x_i$ (около-)*сходится*, если последовательность частичных сумм $y_n = \sum_1^n x_i$, $n = 1, \ldots, \nu$, сходится; сумма *имеет доступное колебание,* если последовательность частичных сумм имеет доступное колебание. (Так как символ «$\sum_{i=1}^{\nu} x_i$» обозначает число, а не последовательность, здесь допущена некоторая вольность речи.) Заметим, что если сумма $\sum_{i=1}^{\nu} x_i$ сходится, то она сходится к своей сумме $x = \sum_{i=1}^{\nu} x_i$ (и к любому числу y, которое бесконечно близко к x), и что $\sum_{i=1}^{\nu} x_i$ сходится тогда и только тогда, когда «хвосты» $\sum_{i=n}^{\nu} x_i$ бесконечно малы для любого недоступного номера $n \leq \nu$.

Следующие импликации верны всегда, но ни одна из обратных к ним, вообще говоря, не выполняется:

$$\begin{array}{ccccc} & & \sum_{i=1}^{\nu} |x_i| \text{ сходится} & \Rightarrow & \sum_{i=1}^{\nu} x_i \text{ сходится} \\ & & \Downarrow & & \Downarrow \\ \sum_{i=1}^{\nu} |x_i| \ll \infty & \Rightarrow & \sum_{i=1}^{\nu} |x_i| \text{ имеет доступное колебание} & \Rightarrow & \sum_{i=1}^{\nu} x_i \text{ имеет доступное колебание} \end{array}$$

Если $x_i \ll \infty$ для любого $i \ll \infty$, то справедлива также импликация

$$\sum_{i=1}^{\nu} |x_i| \text{ сходится} \Rightarrow \sum_{i=1}^{\nu} |x_i| \ll \infty.$$

Это верно потому, что наименьшее n, для которого $\sum_{i=n}^{\nu} |x_i| \le 1$, должно быть доступным и

$$\sum_{i=1}^{n-1} |x_i| \le (n-1) \max_{1 \le i \le n-1} |x_i| \ll \infty.$$

Наконец, если $|x_i| \ll \infty$ для любого $i = 1, \ldots, \nu$, то, как легко заметить,

$$\sum_{i=1}^{\nu} |x_i| \ll \infty \Longleftrightarrow \sum_{i=1}^{\nu} |x_i| \text{ имеет доступное колебание.}$$

Пусть $\xi : T \to \mathbf{R}$. Будем говорить, что функция ξ (около-)*непрерывна в точке* t, если $\xi(s) \simeq \xi(t)$ для любого $s \simeq t$. Будем говорить, что функция ξ (около-)*непрерывна* (на T), если она (около-)непрерывна в каждой точке $t \in T$. Например, пусть T — около-интервал, $a > 0$, и пусть $\xi(t) = 1/t$. Тогда функция ξ непрерывна в точке t тогда и только тогда, когда $t \gg 0$, так что она непрерывна тогда и только тогда, когда $a \gg 0$.

Пусть функция ξ непрерывна в точке t и $\varepsilon \gg 0$. Пусть Δ_t — множество всех δ таких, что $|\xi(s) - \xi(t)| \le \varepsilon$ для любого числа s, если $|s - t| \le \delta$. Тогда Δ_t содержит все $\delta \simeq 0$. По принципу переполнения Δ_t содержит некоторое число $\delta \gg 0$. Таким образом, функция ξ непрерывна в точке t тогда и только тогда, когда для любого $\varepsilon \gg 0$ найдется

$\delta \gg 0$ такое, что $|\xi(s) - \xi(t)| \leq \varepsilon$ для любого числа s, $|s - t| \leq \delta$. Пусть теперь функция ξ непрерывна на T, и пусть $\varepsilon \gg 0$. Для произвольного числа $t \in T$ пусть δ_t — наибольшее обратное целого числа из Δ_t. Тогда $\delta_t \gg 0$. Пусть $\delta = \min_t \delta_t$. Тогда число δ также строго положительно (просто потому, что оно равно δ_t для некоторого t из конечного множества T). Таким образом, функция ξ непрерывна на T в том и только в том случае, если для любого числа $\varepsilon \gg 0$ найдется такое число $\delta \gg 0$, что для любых чисел s и t неравенство $|s - t| \leq \delta$ влечет неравенство $|\xi(s) - \xi(t)| \leq \varepsilon$. Следовательно, около-непрерывность в точке t есть элементарный аналог непрерывности, а около-непрерывность на T — это элементарный аналог равномерной непрерывности.

Приведем несколько простых и полезных иллюстраций этих понятий. По внешней индукции $e^n \ll \infty$ для любого стандартного числа $n \in \mathbf{N}$. Поэтому $e^t \ll \infty$ для любого $t \ll \infty$. По теореме о среднем $e^{t+h} = e^t + e^{\overline{t}}h$ с некоторым $\overline{t}$ между t и $t + h$. Итак, если $t \ll \infty$ и $h \simeq 0$, то $e^{t+h} \simeq e^t$, следовательно, функция $t \mapsto e^t$ непрерывна на T, если $b \ll \infty$. Аналогично

$$\log(t + h) = \log t + h/\overline{t} \simeq \log t$$

для $h \simeq 0$ и $0 \ll t$, т. е. функция $t \mapsto \log t$ непрерывна на T, если $a \gg 0$. Пусть $|x| \ll \infty$ и $n \simeq \infty$. Тогда

$$\log\left(1 + \frac{x}{n}\right)^n = n \log\left(1 + \frac{x}{n}\right) = n\left(\frac{x}{n} - \frac{1}{2}\left(\frac{\overline{x}}{n}\right)^2\right) \simeq x,$$

где $\overline{x}$ лежит между 0 и x, откуда

$$\left(1 + \frac{x}{n}\right)^n \simeq e^x.$$

Таким образом, последовательность $n \mapsto (1 + x/n)^n$, $n = 1, \ldots, \nu$, где $\nu \simeq \infty$, а x доступно, сходится к e^x. По формуле Тейлора с остаточным членом в форме Лагранжа

$$e^x = \sum_{n=0}^{\nu} \frac{x^n}{n!} + e^{\overline{x}} \frac{x^{\nu+1}}{(\nu + 1)!},$$

где $\overline{x}$ лежит между 0 и x. Если x доступно, то и $e^{\overline{x}}$ доступно , и если $\nu \simeq \infty$, то легко видеть, что остаток бесконечно мал. Поэтому если $|x| \ll \infty$, а $\nu \simeq \infty$, то $\sum_{n=0}^{\nu} x^n/n!$ сходится к e^x.

Вариацией, или *полной вариацией*, функции $\xi : T \to \mathbf{R}$ назовем сумму $\sum_{t\in T'} |d\xi(t)|$. Будем говорить, что ξ имеет *доступную вариацию*, если ее вариация доступна. Это элементарный аналог понятия функции ограниченной вариации. Будем говорить, что ξ (около-)*абсолютно непрерывна,* если для любого подмножества $S \subseteq T'$ из $\sum_{t\in S} dt \simeq 0$ следует $\sum_{t\in S} |d\xi(t)| \simeq 0$. Ясно, что абсолютно непрерывные функции непрерывны.

Теорема 6.2. *Пусть T — около-интервал. Если функция ξ абсолютно непрерывна, то она имеет доступную вариацию.*

Доказательство. Рассмотрим множество всех δ таких, что для любого $S \subseteq T'$ из $\sum_{t\in S} dt \leq \delta$ следует $\sum_{t\in S} |d\xi(t)| \leq 1$. Так как введенное множество содержит все $\delta \simeq 0$, оно содержит и некоторое $\delta \gg 0$. Значит, множество T' может быть разбито на $n = [(b-a)/\delta] + 1$ подмножеств S таких, что последний элемент каждого из них равен первому элементу следующего, причем $\sum_{t\in S} |d\xi(t)| \leq 1$. Поэтому $\sum_{t\in T'} |d\xi(t)| \leq n \ll \infty$. $\square$

Глава 7
Свойства, выполняющиеся почти всюду

Конечные вероятностные пространства обычно используются только при исследовании комбинаторных задач, но мы хотим сделать их основой для обсуждения классических предельных теорем теории вероятностей и центральных тем современной теории стохастических процессов. Фундаментальным внешним понятием, которое предоставляет нам эту возможность, является следующее. Пусть $\langle \Omega, \mathrm{pr} \rangle$ — конечное вероятностное пространство, и пусть $\mathbf{A}(\omega)$ — внутренняя или внешняя формула. Мы говорим, что формула $\mathbf{A}(\omega)$ выполнена *почти всюду* (п. в.), или *почти наверное* (п. н.), на пространстве $\langle \Omega, \mathrm{pr} \rangle$, если для любого числа $\varepsilon \gg 0$ существует такое событие N, $\Pr N \leq \varepsilon$, что формула $\mathbf{A}(\omega)$ выполнена для каждой точки $\omega \in N^c$.

Если $\mathbf{A}(\omega)$ — внутренняя формула, то мы можем образовать множество

$$\{\mathbf{A}\} = \{\omega \in \Omega : \mathbf{A}(\omega)\}.$$

Формула $\mathbf{A}(\omega)$ имеет место п. в. тогда и только тогда, когда $\Pr\{\mathbf{A}\} \simeq 1$. Но некоторые из наиболее интересных рассматриваемых нами свойств внешние, и нам нужна такая формулировка предыдущего абзаца, чтобы избежать запрещенного образования множества. Однако независимо от того, является ли формула $\mathbf{A}(\omega)$ внутренней или внешней, интуитивное содержание, доставляемое утверждением о том, что $\mathbf{A}(\omega)$ выполняется п. в., это почти уверенность: для заданного $\varepsilon \gg 0$, к примеру для $\varepsilon = 10^{-100}$, существует такое событие N, для которого $\Pr N \leq \varepsilon$, что, за возможным исключением точек из N, формула $\mathbf{A}(\omega)$ всегда верна.

Теорема 7.1. *Пусть x — случайная величина. Тогда следующие утверждения эквивалентны*:

(i) $x \simeq 0$ *п. в.*;

(ii) $\Pr\{|x| \geq \lambda\} \simeq 0$ *для любого* $\lambda \gg 0$;

(iii) *существует* $\lambda \simeq 0$ *такое, что* $\Pr\{|x| \geq \lambda\} \simeq 0$.

Доказательство. Предположим, что выполнено условие (i), и пусть $\lambda \gg 0$, $\varepsilon \gg 0$. Тогда существует событие N, для которого $\Pr N \leq \varepsilon$ и $x(\omega) \simeq 0$ для любой точки $\omega \in N^c$. При этом $\{|x| \geq \lambda\} \subseteq N$ и, следовательно, $\Pr\{|x| \geq \lambda\} \leq \varepsilon$. Так как число $\varepsilon \gg 0$ произвольно, $\Pr\{|x| \geq \lambda\} \simeq 0$. Таким образом, (i) $\Rightarrow$ (ii). Предположим выполненным условие (ii). Тогда множество всех λ таких, что $\Pr\{|x| \geq \lambda\} \leq \lambda$, содержит все $\lambda \gg 0$, а также по принципу переполнения оно содержит некоторое $\lambda \simeq 0$. Следовательно, (ii) $\Rightarrow$ (iii). Наконец, импликация (iii) $\Rightarrow$ (i) очевидна. $\square$

Пока мы рассматриваем одну случайную величину x, если $x \simeq 0$ п. в., мы можем спокойно считать, что $x = 0$ практически для всех целей — вероятность обнаружить невооруженным глазом какое-либо отклонение ее от 0 меньше, чем 10^{-100}. Положение радикально меняется, когда мы рассматриваем недоступное число случайных величин $x_1, \ldots, x_\nu$, каждая из которых бесконечно мала п. в. Предположим, что сутки разделены на ν равных частей бесконечно малой продолжительности $1/\nu$, что у нас есть устройство, чей отказ приводит к несчастному случаю, что вероятность отказа в любой период равна c/ν, где $0 \ll c \ll \infty$, и что различные периоды независимы. Если мы обозначим через x_n индикаторную функцию отказа в n-й период, то $x_n \simeq 0$ п. в. для каждого номера n (на самом деле $x_n = 0$ п. в.). Однако фактически нас интересует $\max x_n$, индикаторная функция несчастного случая в произвольный момент суток. Учитывая условие независимости, вероятность того, что в течение суток не произойдет несчастного случая, составляет

$$\left(1 - \frac{c}{\nu}\right)^\nu \simeq e^{-c} \ll 1.$$

Пусть $x_1, \ldots, x_\nu$ — конечная последовательность случайных величин, причем ν недоступно. Будем говорить, что $x_1, \ldots, x_\nu$ (около-)*сходится к x по вероятности,* если $x_n \simeq x$ п. в. для любого недоступного номера $n \leq \nu$. Как показывает приведенный выше пример, это условие не слишком обременительно. Более интересен вопрос, сходится ли $x_1, \ldots, x_\nu$ к x п. в. При сходимости по вероятности исключительному множеству N можно разрешить зависеть от n, но при сходимости п. в. этого делать нельзя.

Теорема 7.2. *Пусть $x_1, \ldots, x_\nu$ — случайные величины. Последовательность $x_1, \ldots, x_\nu$ сходится к 0 п. в. тогда и только тогда, когда для любого $\lambda \gg 0$ и любого недоступного $n \le \nu$*

$$\Pr\{\max_{n \le i \le \nu} |x_i| \ge \lambda\} \simeq 0. \tag{7.1}$$

Доказательство. Пусть

$$M(n, \lambda) = \{\max_{n \le i \le \nu} |x_i| \ge \lambda\}.$$

Предположим, что последовательность $x_1, \ldots, x_\nu$ сходится к 0 п. в., и пусть $\lambda \gg 0$, $\varepsilon \gg 0$. Тогда существует такое событие N, $\Pr N \le \varepsilon$, что последовательность $x_1, \ldots, x_\nu$ сходится к 0 на N^c. Если n недоступно, то $M(n, \lambda) \subseteq N$, и, значит, $\Pr M(n, \lambda) \le \varepsilon$. Так как $\varepsilon \gg 0$ произвольно, будет $\Pr M(n, \lambda) \simeq 0$.

Наоборот, предположим, что $\Pr M(n, \lambda) \simeq 0$, если $n \simeq \infty$ и $\lambda \gg 0$. Пусть $\varepsilon \gg 0$, и для $j \in \mathbf{N}$, $j \ne 0$, пусть n_j — наименьшее натуральное число такое, что

$$\Pr M\left(n_j, \frac{1}{j}\right) \le \frac{\varepsilon}{2^j}.$$

Положим

$$N = \bigcup_{j=1}^{\infty} M\left(n_j, \frac{1}{j}\right).$$

(Хотя это и не существенно, но множества $M(n_j, 1/j)$ для достаточно больших j пусты, поскольку пространство Ω конечно.) Имеем $\Pr N \le \varepsilon$. Заметим, что если j доступно, то доступно и n_j. В противном случае мы бы получили, что $n_j - 1 \simeq \infty$ и $1/j \gg 0$, так что вероятность

$$\Pr M\left(n_j - 1, \frac{1}{j}\right)$$

по предположению стала бы бесконечно малой и, значит, $\le \varepsilon/2^j$, что противоречит определению n_j. Следовательно, если $\omega \in N^c$ и $n \simeq \infty$, то $|x_n(\omega)| \le 1/j$ для любого $j \ll \infty$, и поэтому $x_n(\omega) \simeq 0$. Так как $\varepsilon \gg 0$ произвольно, последовательность $x_1, \ldots, x_\nu$ сходится к 0 п. в. $\square$

Отметим, что по теореме 7.1 соотношение (7.1) эквивалентно утверждению: для любого недоступного $n \le \nu$ выполняется

$$\max_{n \le i \le \nu} |x_i| \simeq 0 \text{ п. в.}$$

Теорема 7.2 имеет

Следствие. *Пусть ξ — стохастический процесс, параметризованный элементами конечного подмножества $T \subseteq \mathbf{R}$, и пусть $t \in T$. Процесс ξ непрерывен в точке t п. н. тогда и только тогда, когда для любых чисел $\lambda \gg 0$ и $h \simeq 0$*

$$\Pr\{\max_{|s-t| \le h} |\xi(s) - \xi(t)| \ge \lambda\} \simeq 0. \tag{7.2}$$

Доказательство. Пусть $\nu \simeq \infty$ и для $n \le \nu$

$$x_n = \max_{|s-t| \le 1/n} |\xi(s) - \xi(t)|.$$

Тогда процесс ξ непрерывен в точке t в том и только в том случае, если последовательность x_n сходится к 0 и условие (7.1) эквивалентно условию (7.2). □

Теорема 7.3 (Бореля — Кантелли, ординальная версия). *Пусть $A_1, \ldots, A_\nu$ — события, $k(\omega)$ — наибольшее число k такое, что $\omega \in A_k$, и $k(\omega) = 0$, если ω не принадлежит никакому A_n.*

(i) *Если $\sum_{n=1}^{\nu} \Pr A_n$ сходится, то число k доступно п. н.*

(ii) *Если события A_n независимы, то $\sum_{n=1}^{\nu} \Pr A_n$ сходится тогда и только тогда, когда число k доступно п. н.*

Доказательство. Чтобы доказать (i), положим $\varepsilon \gg 0$ и возьмем наименьшее число j такое, что $\Pr \bigcup_{n=j}^{\nu} A_n \le \varepsilon$. Тогда j доступно, так как в противном случае выполнялось бы соотношение

$$\Pr \bigcup_{n=j-1}^{\nu} A_n \le \sum_{n=j-1}^{\nu} \Pr A_n \simeq 0.$$

Но $k(\omega) \le j-1 \simeq \infty$ для любой точки $\omega \in \left(\bigcup_{n=j}^{\nu} A_n\right)^c$. Таким образом, $k \ll \infty$ п. н.

Так как $1 - \lambda \leq e^{-\lambda}$ (выпуклая функция лежит над своей касательной), в случае (ii) имеем

$$\Pr \bigcap_{n=i}^{\nu} A_n^c = \prod_{n=i}^{\nu}(1 - \Pr A_n) \leq \exp\Bigl(-\sum_{n=i}^{\nu} \Pr A_n\Bigr).$$

Если $k \ll \infty$ п. н., то для $i \simeq \infty$ левая часть равенства бесконечно близка к 1, так что $\sum_{n=i}^{\nu} \Pr A_n \simeq 0$. Вместе с (i) это доказывает (ii).

Теорема 7.4 (Бореля — Кантелли, кардинальная версия). *Пусть $A_1, \ldots, A_\nu$ — события и $K(\omega)$ — число таких номеров n, что $\omega \in A_n$.*

(i) *Если $\sum_{n=1}^{\nu} \Pr A_n \ll \infty$, то число K п. н. доступно.*

(ii) *Если события A_n независимы, то $\sum_{n=1}^{\nu} \Pr A_n \ll \infty$ тогда и только тогда, когда число K п. н. доступно. Иными словами, $\sum_{n=1}^{\nu} \Pr A_n \simeq \infty$ тогда и только тогда, когда число K п. н. недоступно.*

Доказательство. Для доказательства утверждения (i) заметим, что $\sum_{n=1}^{\nu} \Pr A_n = \mathbf{E}K$, и согласно неравенству Чебышёва для $p = 1$ имеем $\Pr\{K \geq l\} \leq \mathbf{E}K/l$. Но если $\mathbf{E}K \ll \infty$, то для любого $\varepsilon \gg 0$ найдется такое $l \ll \infty$, что $\mathbf{E}K/l \leq \varepsilon$, так что $K < l \ll \infty$ за исключением события вероятности, не большей ε, и, следовательно, $K \ll \infty$ п. н.

Чтобы доказать (ii), нам нужно показать только (когда события A_n независимы), что если $\mathbf{E}K \simeq \infty$, то п. н. $K \simeq \infty$. Из неравенства Чебышёва для $p = 2$ получаем $\Pr\{|K - \mathbf{E}K| \geq \lambda\} \leq \mathrm{Var}K/\lambda^2$. Но в силу независимости

$$\mathrm{Var}K = \sum \mathrm{Var}\chi_{A_n} \leq \sum \mathbf{E}\chi_{A_n}^2 = \mathbf{E}K.$$

Следовательно, $\Pr\{|K - \mathbf{E}K| \geq \lambda\} \leq \mathbf{E}K/\lambda^2$. Допустим, что $\mathbf{E}K \simeq \infty$, и возьмем $\lambda = \frac{1}{2}\mathbf{E}K$. Тогда $\Pr\{|K - \mathbf{E}K| \geq \frac{1}{2}\mathbf{E}K\} \simeq 0$. Таким образом, за исключением события бесконечно малой вероятности $|K - \mathbf{E}K| < \frac{1}{2}\mathbf{E}K$, так что $K \simeq \infty$ п. н. □

Если события A_n независимы, то либо $K \ll \infty$ п. н., либо $K \simeq \infty$ п. н. Вообще говоря, это неверно для k из теоремы 7.3: пусть $A_n = \varnothing$ для всех $n \leq \nu$, за исключением $n = i$, где $i \simeq \infty$, и пусть $\Pr A_i = \frac{1}{2}$. Тогда события A_n независимы, но $\Pr\{k = 0\} = \frac{1}{2}$ и $\Pr\{k = i\} = \frac{1}{2}$. В этом примере конечная сумма $\sum \Pr A_n$ доступна, но не сходится.

Сходящийся ряд вероятностей сходится к доступному значению, и если $k \ll \infty$, то $K \ll \infty$; таким образом, в ординальной версии как предположение, так и заключение в (i) сильнее. Язык традиционной математики не очень хорошо приспособлен для проведения различия между кардинальной и ординальной версиями: если мы имеем актуально бесконечное множество событий $A_1, A_2, \ldots$, то утверждение кардинальной версии, устанавливающей, что произошло только конечное число событий K, эквивалентно ординальному предложению, что после некоторого момента k ни одно из них не происходит. Все же разница между ними имеет известное практическое значение. Из проведенных выше рассуждений ясно, что для оценки K достаточно оценки $\sum \Pr A_n$, тогда как если мы хотим оценить k, то нам нужна оценка, показывающая малость остатков ряда $\sum \Pr A_n$. Для примера о несчастном случае, обсуждавшегося ранее в этой главе, неверно, что число k доступно п. н., но, по крайней мере, мы можем чувствовать себя достаточно комфортно от сознания того, что в течение суток может почти наверное случиться лишь доступное число несчастных случаев.

Глава 8

Суммируемые случайные величины

Если x — случайная величина, а a — константа, то определим *срезку* случайной величины $x^{(a)}$ равенством

$$x^{(a)} = x\chi_{\{|x \le a|\}}.$$

Таким образом, $x^{(a)}(\omega) = x(\omega)$, если $|x(\omega)| \le a$, иначе $x^{(a)}(\omega) = 0$.

Говорят, что случайная величина x *суммируема*, или L^1-*суммируема*, или, наконец, что x есть L^1-*случайная величина*, если $\mathbf{E}|x - x^{(a)}| \simeq 0$ для любого $a \simeq \infty$. Так как

$$\mathbf{E}|x - x^{(a)}| = \sum_{|\lambda| > a} |\lambda| \mathrm{pr}_x(\lambda),$$

отсюда вытекает, что случайная величина x суммируема тогда и только тогда, когда последовательность

$$n \mapsto \sum_{|\lambda| \le n} |\lambda| \mathrm{pr}_x(\lambda),$$

$n = 1, \ldots, \nu$, где $\nu \ge \|x\|_\infty$, сходится, или, как мы будем говорить для краткости, ряд $\sum |\lambda| \mathrm{pr}_x(\lambda)$ (около-)*сходится.* Таким образом, если случайная величина x суммируема, то $\mathbf{E}|x| \ll \infty$. Обратное, вообще говоря, неверно: предположим, что $\mathrm{pr}(\omega) \simeq 0$, и пусть $x = \mathrm{pr}(\omega)^{-1}\chi_{\{\omega\}}$. Тогда $\mathbf{E}|x| \ll \infty$, но для $a \simeq \infty$, $a < \mathrm{pr}(\omega)^{-1}$, имеем $\mathbf{E}|x - x^{(a)}| = 1 \not\simeq 0$.

Теорема 8.1 (Радона — Никодима и обратная к ней). *Случайная величина x суммируема тогда и только тогда, когда математическое ожидание $\mathbf{E}|x|$ доступно и $\mathbf{E}|x|\chi_M \simeq 0$ для любых событий M, у которых $\Pr M \simeq 0$.*

Доказательство. Предположим, что случайная величина x суммируема и $\Pr M \simeq 0$. Пусть $a \simeq \infty$ такое, что $a \Pr M \simeq 0$ (например, пусть $a = 1/\sqrt{\Pr M}$). Тогда

$$\mathbf{E}|x|\chi_M \leq \mathbf{E}|x^{(a)}|\chi_M + \mathbf{E}|x - x^{(a)}|\chi_M \leq a \Pr M + \mathbf{E}|x - x^{(a)}| \simeq 0.$$

Наоборот, предположим, что $\mathbf{E}|x| \ll \infty$ и для любого события M, вероятность которого бесконечно мала, выполнено $\mathbf{E}|x|\chi_M \simeq 0$. Пусть $a \simeq \infty$ и $M = \{|x| > a\}$. Тогда (по неравенству Чебышёва для $p = 1$) $\Pr M \leq \mathbf{E}|x|/a \simeq 0$, так что $\mathbf{E}|x|\chi_M \simeq 0$, т. е. $\mathbf{E}|x - x^{(a)}| \simeq 0$. $\square$

Из этого критерия следует, что если случайные величины x и y суммируемы, то суммируема и случайная величина $x + y$; если x суммируема и $|y| \leq |x|$, то y суммируема; если случайная величина x суммируема, а $\|y\|_\infty \leq \infty$, то случайная величина yx суммируема.

Теорема 8.2 (Лебега). *Если случайные величины x и y суммируемы и $x \simeq y$ п. в., то $\mathbf{E}x \simeq \mathbf{E}y$.*

Доказательство. Обозначим $z = x - y$. Тогда $z \simeq 0$ п. в., и (по теореме 7.1) существует такое число $\alpha \simeq 0$, что $\Pr\{|z| > \alpha\} \simeq 0$. Но $|z| \leq |z|\chi_{\{|z|>\alpha\}} + \alpha$, и так как z суммируема, из теоремы 8.1 вытекает, что $\mathbf{E}|z| \simeq 0$. Следовательно, $\mathbf{E}x \simeq \mathbf{E}y$. $\square$

При $1 < p < \infty$ говорят, что случайная величина x является *p-суммируемой* (или *L^p-суммируемой*, или что x есть *L^p-случайная величина*), если $|x|^p$ суммируема (при $p = 2$ — квадратично суммируемой); и что случайная величина x *существенно ограничена*, или что x есть *L^∞-случайная величина*, если $\|x\|_\infty \ll \infty$. Если случайная величина x p-суммируема, а случайная величина y p'-суммируема, где p' — показатель, сопряженный к p, то согласно неравенству

$$|xy| \leq \frac{1}{p}|x|^p + \frac{1}{p'}|y|^{p'},$$

доказанному в гл. 1, случайная величина xy суммируема. Аналогично, если $p \gg 1$ и $\mathbf{E}|x|^p \ll \infty$, то случайная величина x суммируема, так как для $\alpha \simeq \infty$ выполнено

$$\sum_{|\lambda|>a} |\lambda| \mathrm{pr}_x(\lambda) \leq \sum_{|\lambda|>a} \frac{|\lambda|^p}{a^{p-1}} \mathrm{pr}_x(\lambda) \leq \frac{1}{a^{p-1}} \mathbf{E}|x|^p \simeq 0. \square$$

Теорема 8.3. *Пусть случайная величина x p-суммируема, где $1 \leq p \leq \infty$, и пусть $\mathcal{A}$ — алгебра случайных величин. Тогда случайная величина $\mathbf{E}_\mathcal{A} x$ p-суммируема.*

Доказательство. Для $p = \infty$ это очевидно. Для $1 \le p < \infty$ из релятивизированного неравенства Йенсена следует, что $|\mathbf{E}_{\mathcal{A}} x|^p \le \mathbf{E}_{\mathcal{A}} |x|^p$. Таким образом, нам надо провести доказательство только для $p = 1$. Пусть случайная величина x суммируема. Тогда $\mathbf{E}|\mathbf{E}_{\mathcal{A}} x| \le \mathbf{E}|x| \ll \infty$. Пусть $\Pr M \simeq 0$ и $a = 1/\sqrt{\Pr M}$. Тогда

$$\mathbf{E}|\mathbf{E}_{\mathcal{A}} x|\chi_M \le \mathbf{E}|\mathbf{E}_{\mathcal{A}} x^{(a)}|\chi_M + \mathbf{E}|\mathbf{E}_{\mathcal{A}}(x - x^{(a)})|\chi_M$$

$$\le a \Pr M + \mathbf{E}|x - x^{(a)}| \simeq 0.$$

По теореме 8.1 случайная величина $\mathbf{E}_{\mathcal{A}} x$ суммируема.

Теорема 8.4. *Пусть случайная величина x суммируема и $\mathcal{A}$ — алгебра случайных величин. Тогда случайная величина x суммируема почти на каждом атоме из $\mathcal{A}$.*

Доказательство. Пусть $\varepsilon \gg 0$. Для каждого $n \in \mathbf{N}$ пусть a_n — наименьшее натуральное число такое, что

$$\Pr{}'_{\mathcal{A}} \left\{ \mathbf{E}_{\mathcal{A}} |x - x^{(a_n)}| \ge \frac{1}{n} \right\} \le \frac{\varepsilon}{2^n}$$

(см. определение $\Pr'_{\mathcal{A}}$ в гл. 2.) Я утверждаю, что если $n \ll \infty$, то $a_n \ll \infty$. Чтобы доказать это, заметим, что $\mathbf{E}_{\mathcal{A}}|x - x^{(a)}|$ — случайная величина на $\operatorname{at}(\mathcal{A})$, математическое ожидание которой

$$\mathbf{E}'_{\mathcal{A}} \mathbf{E}_{\mathcal{A}} |x - x^{(a)}| = \mathbf{E}|x - x^{(a)}|.$$

Из неравенства Чебышёва вытекает, что вероятность

$$\Pr{}'_{\mathcal{A}} \left\{ \mathbf{E}_{\mathcal{A}} |x - x^{(a_n)}| \ge \frac{1}{n} \right\} \le n \mathbf{E}|x - x^{(a)}|$$

бесконечно мала (и тем самым $\le \varepsilon/2^n$), если $a \simeq \infty$.

Пусть N — множество тех атомов, где

$$\mathbf{E}_{\mathcal{A}} |x - x^{(a_n)}| \ge \frac{1}{n}$$

для некоторого n. Тогда $\Pr'_{\mathcal{A}} N \le \varepsilon$. На атомах, не входящих в N, выполняется неравенство

$$\mathbf{E}_{\mathcal{A}} |x - x^{(a)}| < \frac{1}{n}$$

для каждого $n \ll \infty$, если $a \simeq \infty$, так как $a \geq a_n$ для любого $n \ll \infty$. Таким образом, случайная величина x суммируема на этих атомах. Так как $\varepsilon \gg 0$ произвольно, это завершает доказательство. $\square$

Обсуждая противоположную импликацию, предположим, что случайная величина x суммируема на *каждом* атоме. Если число $a \simeq \infty$, то $E_{\mathcal{A}}|x - x^{(a)}| \simeq 0$ всюду, так что $E|x - x^{(a)}| \simeq 0$. Таким образом, если x суммируема на каждом атоме $\mathcal{A}$, то x суммируема. Это самое большее, что мы можем сказать в общем случае, потому что всегда можно изменить суммируемую случайную величину в одной точке, имеющей бесконечно малую вероятность, и получить случайную несуммируемую величину.

Теорема 8.4 имеет

Следствие (Фубини). *Если случайная величина x суммируема на $\langle \Omega_1 \times \Omega_2, \mathrm{pr}_1 \times \mathrm{pr}_2 \rangle$, то случайная величина x_{ω_1}, заданная на $\langle \Omega_2, \mathrm{pr}_2 \rangle$ равенством $x_{\omega_1}(\omega_2) = x(\omega_1, \omega_2)$, суммируема для п. в. $\omega \in \Omega_1$.*

Глава 9
Разложение стохастических процессов

Будем изучать стохастические процессы ξ, параметризованные элементами конечного подмножества $T \subset \mathbf{R}$. Напомним общие понятия, введенные в начале гл. 6. Типичными случаями будут те, в которых $T = \{1, \ldots, \nu\}$, где ν — недоступное натуральное число, или те, в которых T — около-интервал. Таким образом, хотя мы требуем, чтобы множество T было конечным, будем изучать классические темы, связанные с «бесконечными» последовательностями случайных величин и стохастическими процессами с «непрерывным» временем.

Пусть $\mathcal{P} : t \mapsto \mathcal{P}_t$ — возрастающая функция из множества T в множество всех алгебр случайных величин на $\langle \Omega, \mathrm{pr} \rangle$. Она называется *фильтрацией.* Вместо $\mathbf{E}_{\mathcal{P}_t}$ будем сокращенно писать $\mathbf{E}_t$. По определению $\mathcal{P}$*-процесс,* или *процесс, адаптированный к* $\mathcal{P}$, — это стохастический процесс ξ, параметризованный элементами множества T и такой, что $\xi(t) \in \mathcal{P}_t$ для любого $t \in T$. Так как $\mathcal{P}_s \subseteq \mathcal{P}_t$ при $s \leq t$, если ξ — это $\mathcal{P}$-процесс, то $\xi(s) \in \mathcal{P}_t$ для всех $s \leq t$. Произвольный стохастический процесс ξ, параметризованный элементами множества T, будет $\mathcal{P}$-процессом, если мы определим каждую алгебру $\mathcal{P}_t$ как алгебру, порожденную $\xi(s)$ с $s \leq t$. Однако удобно учесть возможность того, что $\mathcal{P}_t$ шире. Алгебра $\mathcal{P}_t$ представляет прошлое в момент времени t, и если y — случайная величина, то $\mathbf{E}_t y$ — это наилучшее предсказание для случайной величины y, которое можно сделать, зная прошлое в момент времени t.

$\mathcal{P}$-процесс ξ называется *мартингалом*, если $\mathbf{E}_s \xi(t) = \xi(s)$ для любого $s \leq t$, *субмартингалом*, если $\mathbf{E}_s \xi(t) \geq \xi(s)$ для любого $s \leq t$, и *супермартингалом*, если $\mathbf{E}_s \xi(t) \leq \xi(s)$ для любого $s \leq t$. Таким образом, в тривиальном случае, когда пространство Ω состоит из единственной точки, мартингал сводится к постоянной функции от t, субмартингал — к возрастающей функции, а супермартингал — к убывающей функции.

Заметим, что если ξ — мартингал, то

$$\xi(t) = \mathbf{E}_t \xi(b)$$

для любого t. Обратно, если заданы фильтрация $\mathcal{P}$ и некоторая случайная величина x, то процесс ξ, определенный равенством $\xi(t) = \mathbf{E}_t x$, — мартингал.

Пусть ξ — это некоторый $\mathcal{P}$-процесс. Определим $D\xi$, $d\widehat{\xi}$ и σ_ξ^2 следующим образом:

$$D\xi(t)dt = \mathbf{E}_t d\xi(t),$$

$$d\xi(t) = D\xi(t)dt + d\widehat{\xi}(t),$$

$$\sigma_\xi^2(t)dt = \mathbf{E}_t d\widehat{\xi}(t)^2.$$

Таким образом, $D\xi(t)dt$ и $\sigma_\xi^2(t)dt$ — это соответственно условное среднее и дисперсия приращения $d\xi(t)$. Отметим, что $D\xi$ и σ_ξ^2 суть $\mathcal{P}$-процессы, параметризованные элементами множества T', тогда как $d\xi$ и $d\widehat{\xi}$, вообще говоря, $\mathcal{P}$-процессами не являются.

Отметим, что $D\xi = 0$, если ξ — мартингал, $D\xi \geq 0$, если ξ — субмартингал, и $D\xi \leq 0$, если ξ — супермартингал. (Конечно, общий процесс $D\xi$ вовсе не обязан иметь постоянный знак ни по t, ни по ω.) Покажем, что эти условия не только необходимы, но и достаточны. Имеем

$$\xi(t) = \xi(s) + \sum_{s \leq r < t} D\xi(r)dr + \sum_{s \leq r < t} d\widehat{\xi}(r), \quad s \leq t. \tag{9.1}$$

Итак, $\mathbf{E}_r d\widehat{\xi}(r) = 0$. Так как $\mathcal{P}_s \subseteq \mathcal{P}_r$ при $s \leq r$, имеет место равенство $\mathbf{E}_s = \mathbf{E}_s \mathbf{E}_r$ при $s \leq r$. Следовательно, $\mathbf{E}_s d\widehat{\xi}(r) = 0$ при $s \leq r$. Значит, если мы применим $\mathbf{E}_s$ к (9.1), то получим

$$\mathbf{E}_s \xi(t) = \xi(s) + \mathbf{E}_s \sum_{s \leq r < t} D\xi(r)dr, \quad s \leq t. \tag{9.2}$$

Поэтому случайный процесс ξ является мартингалом, если и только если $D\xi = 0$, субмартингалом, если и только если $D\xi \geq 0$, и супермартингалом, если и только если $D\xi \leq 0$.

Будем называть $D\xi$ *трендом* процесса ξ, и если $d\widehat{\xi} = 0$, то говорить, что ξ — *предсказуемый процесс.* Таким образом, ξ является предсказуемым процессом тогда и только тогда, когда $\sigma_\xi^2 = 0$ или, что эквивалентно, $d\xi$ — это $\mathcal{P}$-процесс. Пусть

$$\widetilde{\xi}(t) = \sum_{s<t} D\xi(s)ds,$$

так что $\widetilde{\xi}(a) = 0$. Тогда $d\widetilde{\xi}(t) = D\xi(t)dt \in \mathcal{P}_t$ и, значит, $\widetilde{\xi}$ — предсказуемый процесс. Мы называем его *предсказуемым процессом, ассоциированным с* ξ. Заметим, что если известна алгебра $\mathcal{P}_t$, то известно и $d\widetilde{\xi}(t)$; мы достоверно можем предсказать приращение. Однако чтобы достоверно предсказать следующее приращение, нужно знать алгебру $\mathcal{P}_{t+dt}$, а она в общем случае не порождается $\mathcal{P}_t$ и $d\widetilde{\xi}(t)$. Положим

$$\widehat{\xi}(t) = \xi(a) + \sum_{s<t} d\widehat{\xi}(s).$$

Тогда $\widehat{\xi}$ — возрастающий $\mathcal{P}$-процесс с приращениями $d\widehat{\xi}(t)$, как и подсказывают обозначения. Так как $D\widehat{\xi} = 0$, процесс $\widehat{\xi}$ — мартингал. Будем называть его *мартингалом, ассоциированным с процессом* ξ. Отметим, что если ξ уже является мартингалом, то $\widehat{\xi} = \xi$. Имеет место разложение

$$\xi = \widetilde{\xi} + \widehat{\xi}$$

произвольного $\mathcal{P}$-процесса ξ на предсказуемый процесс $\widetilde{\xi}$ и мартингал $\widehat{\xi}$, и это разложение однозначно, если наложено условие нормировки $\widetilde{\xi}(a) = 0$.

Если величины $d\xi(t)$ независимы и $\mathcal{P}_t$ — алгебра, порожденная $\xi(s)$ с $s \le t$, то $D\xi(t) = \mathbf{E}_t d\xi(t) = \mathbf{E} d\xi(t)$. Следовательно, частичные суммы $\xi(t) = \sum_{s<t} d\xi(s)$ независимых случайных величин $d\xi(s)$ с нулевым средним образуют мартингал. Чтобы выделить такой процесс с точностью до эквивалентности, необходимо только задать вероятностное распределение приращений. Приведем два примера. В первом, назовем его *винеровским блужданием*, положим

$$d\xi(t) = \begin{cases} \sqrt{dt} \text{ с вероятностью } \frac{1}{2}, \\ -\sqrt{dt} \text{ с вероятностью } \frac{1}{2}, \end{cases}$$

а во втором, который назовем *пуассоновским блужданием*, предполагая, что $dt \le 1$ при любом t, определим

$$d\xi(t) = \begin{cases} \ \ 1 \text{ с вероятностью } \frac{1}{2}dt, \\ \ \ 0 \text{ с вероятностью } 1 - dt, \\ -1 \text{ с вероятностью } \frac{1}{2}dt. \end{cases}$$

Пусть ξ — мартингал. Если $r_1 \le r_2$, то $d\xi(r_1) \in \mathcal{P}_{r_2}$. Так как $\mathbf{E}_{r_2}$ $\mathcal{P}_{r_2}$-линейно, получаем, что $\mathbf{E}_{r_2} d\xi(r_1) d\xi(r_2) = d\xi(r_1)\mathbf{E}_{r_2} d\xi(r_2) = 0$. Но $\mathbf{E}_s \mathbf{E}_{r_2} = \mathbf{E}_s$ при $s \le r_2$, следовательно,

$$\mathbf{E}_s d\xi(r_1) d\xi(r_2) = 0, \quad r_1 \ne r_2, \quad s \le \max\{r_1, r_2\}.$$

Из равенства $\xi(t) - \xi(s) = \sum_{s \le r < t} d\xi(r)$ для $s \le t$ следует, что

$$\mathbf{E}_s(\xi(t) - \xi(s))^2 = \mathbf{E}_s \sum_{s \le r < t} \sigma_\xi^2(r) dr, \quad s \le t.$$

Взяв математические ожидания абсолютных величин, видим, что процесс ξ имеет ортогональные приращения и

$$\|\xi(t) - \xi(s)\|_2^2 = \sum_{s \le r < t} \|d\xi(r)\|_2^2, \quad s \le t.$$

Пусть теперь η — произвольный $\mathcal{P}$-процесс. Рассмотрим $\mathcal{P}$-процесс ζ, определенный равенством

$$\zeta(t) = \sum_{s < t} \eta(s) d\xi(s).$$

Тогда $\zeta(a) = 0$ и $d\zeta(t) = \eta(t) d\xi(t)$, так что $D\zeta(t) = \eta(t) D\xi(t) = 0$. Таким образом, случайный процесс ζ — тоже мартингал, и $\sigma_\zeta^2(t) = \eta(t)^2 \sigma_\xi^2(t)$. Мы можем представлять себе процесс ξ как описание честной игры. В любой момент времени t математическое ожидание выиграть следующую ставку равно $\mathbf{E}_t d\xi(t) = 0$. При этом процесс η представляет игровую стратегию. В произвольный момент t игрок на основании знания прошлого решается умножить свою ставку на $\eta(t)$. Тогда $\zeta(t)$ представляет ее выигрыш в момент t. никаких причин, подсказывающих умному игроку, как ему выбирать стратегию. Любой возможный выигрыш в момент времени t компенсироваться Не имеет значения, сколь умно игрок выбирает свою стратегию, любой возможный выигрыш в момент t будет сведен на нет возможным проигрышем в этот момент, так как $\mathbf{E}_\zeta(t) = 0$. Это факт вдохновил на поэму «Owed to a Martingale».

Случайная величина x называется *нормированной*, если $\mathbf{E}x = 0$ и $\operatorname{Var} x = 1$. Если x — это не постоянная случайная величина, то $x = \mu + \sigma\breve{x}$, где μ — математическое ожидание, σ — стандартное отклонение случайной величины x, а $\breve{x} = (x - \mu)/\sigma$ — нормированная случайная величина. Такое представление возможно, даже если x — постоянная случайная величина (т. е. даже при $\sigma = 0$), если принять за $\breve{x}$ произвольную нормированную случайную величину. Существует одно исключение: если множество Ω состоит из единственной точки, то оно не допускает никакой нормированной случайной величины и должно быть заменено каким-либо бóльшим конечным вероятностным пространством.

Эти тривиальности перечислены с тем, чтобы быть путеводителем в рассуждениях о более интересном случае стохастического процесса ξ. Стохастический процесс ξ называется *нормированным*, если $D\xi = 0$ и $\sigma_\xi^2 = 1$. В частности, если случайный процесс ξ нормирован, то он является мартингалом. И винеровское блуждание, и пуассоновское блуждание суть нормированные мартингалы. Стохастический процесс ξ называется *невырожденным*, если $\sigma_\xi^2 > 0$ на $T' \times \Omega$. Если ξ — невырожденный стохастический процесс, то определим $\breve{\xi}$ равенством

$$\breve{\xi} = \sum_{s<t} \sigma_\xi^{-1}(s) d\widehat{\xi}(s),$$

где $\sigma_\xi^{-1} = 1/\sqrt{\sigma_\xi^2}$ и $\widehat{\xi}$ — ассоциированный мартингал. Теперь процесс $\breve{\xi}$ нормирован, и мы имеем представление

$$\xi(t) = \xi(a) + \sum_{s<t} D\xi(s)ds + \sum_{s<t} \sigma_\xi(s) d\breve{\xi}(s). \tag{9.3}$$

Даже если процесс ξ вырожден, мы можем добиться такого представления, перейдя к большему конечному вероятностному пространству. Пусть η — произвольный нормированный мартингал, параметризованный элементами множества T и определенный на конечном вероятностном пространстве $\langle \Omega', \mathrm{pr}' \rangle$. Положим $\sigma_\xi^{-1}(s) = 1/\sqrt{\sigma_\xi^2}$, если

$\sigma_\xi^2(s) > 0$, и 0 в противном случае. Пусть, далее, $\breve{\xi}$ — стохастический процесс, определенный на $\langle \Omega \times \Omega', \mathrm{pr} \times \mathrm{pr}' \rangle$ равенствами $\breve{\xi}(a) = 0$ и

$$d\breve{\xi}(s) = \sigma_\xi^{-1}(s) d\widehat{\xi}(s) + \chi_{\{\sigma_\xi^2(s)=0\}} d\eta(s).$$

Тогда процесс $\breve{\xi}$ нормирован, и мы имеем представление (9.3).

Глава 10
Полная вариация процесса

В предыдущей главе обсуждение было полностью внутренним. Нас, в основном, будут интересовать процессы, для которых $D\xi$ и σ_ξ^2 доступны в соответствующих нормах. При этом приращения $d\widetilde{\xi}(t) = D\xi(t)dt$ ассоциированного предсказуемого процесса имеют порядок dt, тогда как приращения ассоциированного мартингала будут порядка $\sqrt{dt}$. Представляющие интерес локальные флуктуации стохастического процесса осуществляются в ассоциированном мартингале. Тем не менее, в этой главе мы начинаем изучать флуктуации стохастических процессов, оценивая $\sum |d\xi(t)|$. Этот метод слишком груб, чтобы представлять интерес для мартингалов. Мы увидим, что если для процесса ξ разностное отношение $d\xi/dt$ суммируемо и по ω, и по t (в смысле, который будет уточнен), то процесс ξ абсолютно непрерывен п. н. (т. е. почти наверное выборочные траектории абсолютно непрерывны) и что почти наверное ассоциированный мартингал остается бесконечно близким к своему первоначальному значению. Последнее свойство выполняется также для возрастающего процесса с бесконечно малыми приращениями.

Заметим, что определение

$$\mathrm{pr}_T(t) = \frac{dt}{b-a}$$

превращает $\langle T', \mathrm{pr}_T\rangle$ в конечное вероятностное пространство. Обозначим математическое ожидание в этом пространстве через $\mathbf{E}_T$. Если ξ — фиксированная функция на T, то она имеет доступную вариацию тогда и только тогда, когда $\mathbf{E}_T|d\xi/dt| \ll \infty$. По теореме 8.1 процесс ξ абсолютно непрерывен в том и только в том случае, если $d\xi/dt$ суммируемо на $\langle T', \mathrm{pr}_T\rangle$.

Пусть теперь ξ — стохастический процесс, параметризованный элементами множества T, адаптированный к фильтрации $\mathcal{P}$. Если $T = \{1, \ldots, \nu\}$, мы иногда используем обозначение ξ_n вместо $\xi(n)$.

Теорема 10.1. (i) *Если* $\sum \|d\xi(t)\|_1 \ll \infty$, *то* $\sum |d\xi(t)| \ll \infty$ *п. н.* (*это значит, что п. н. процесс* ξ *имеет доступную вариацию*).

(ii) *Если* $T = \{1, \ldots, \nu\}$ *и* $\sum \|d\xi_n\|$ *сходится, то* ξ_n *сходится п. н.*

(iii) *Если величина* $d\xi/dt$ *суммируема на* $T' \times \Omega$, *то п. н. процесс* ξ *абсолютно непрерывен.*

Доказательство. Для случая (i) положим $c = \sum \|d\xi(t)\|_1$ и $x = \sum |d\xi(t)|$, так что $c = \mathbf{E}x$. Тогда в (i) утверждается, что если $\mathbf{E}|x| \ll \infty$, то п. н. $|x| \ll \infty$. Это в общем случае вытекает из неравенства Чебышёва.

В случае (ii) в силу теоремы 7.2 нам нужно только показать, что для любого $\lambda \gg 0$ и недоступного $n < \nu$

$$\Pr\left\{\sum_{i=n}^{\nu-1} |d\xi_i| \geq \lambda\right\} \simeq 0.$$

Но левая часть не больше чем $\sum_{i=n}^{\nu-1} \|d\xi_i\|_1/\lambda \simeq 0$.

Утверждение (iii) сразу вытекает из теоремы Фубини (следствие теоремы 8.4). □

Утверждение (iii) было сформулировано для общих стохастических

процессов, но, как мы увидим ниже (в теореме 10.4), условие в нем весьма ограничительно, и из него следует, что ξ — около предсказуемый процесс в том смысле, что для ассоциированного мартингала почти наверное $\widehat{\xi}(t) \simeq \widehat{\xi}(a)$ для любого t.

Теорема 10.2. *Пусть* ξ — *стохастический процесс такой, что приращение* $d\xi/dt$ *суммируемо на* $T' \times \Omega$. *Тогда величина* $D\xi$ *суммируема на* $T' \times \Omega$ *и, следовательно, приращение* $d\widehat{\xi}/dt = d\xi/dt - D\xi$ *суммируемо на* $T' \times \Omega$.

Доказательство. Пусть $\mathcal{P}'$ — алгебра всех случайных величин y на $T' \times \Omega$ таких, что $y(t, \cdot) \in \mathcal{P}_t$ для каждого t. Тогда

$$D\xi = \mathbf{E}_{\mathcal{P}'} \frac{d\xi}{dt} \ ,$$

откуда согласно теореме 8.3 следует наше утверждение. □

Следующий результат — это лемма о срезке. Она утверждает, что при сделанных ниже предположениях мы можем слегка модифицировать процесс (считая, что почти наверное траектории все время остаются бесконечно близкими) так, что модифицированный процесс имеет бесконечно малые приращения.

Теорема 10.3. *Пусть ξ — мартингал, параметризованный около-интервалом T, такой, что разностное отношение $d\xi/dt$ суммируемо на $T' \times \Omega$. Для $\alpha > 0$ пусть $\xi_{(\alpha)}$ — мартингал, $\xi_{(\alpha)}(a) = \xi(a)$ и*

$$d\xi_{(\alpha)}(t) = d\xi(t)^{(\alpha)} - \mathbf{E}_t d\xi(t)^{(\alpha)},$$

так что $|d\xi_{(\alpha)}(t)| \leq 2\alpha$ для любого t и любого ω. Тогда существует бесконечно малое α такое, что п. н. $\xi_{(\alpha)}(t) \simeq \xi(t)$ для каждого t.

Доказательство. По теореме 10.1 (iii) случайный процесс ξ п. н. абсолютно непрерывен, а потому непрерывен. Так как T — около-интервал, то $\max_t |d\xi(t)| \simeq 0$ п. н. По теореме 7.1 для любой достаточно большой инфинитезимали α будет

$$\Pr\{\max_t |d\xi(t)| \geq \alpha\} \simeq 0.$$

Значит, п. н. выполнено равенство

$$\xi_{(\alpha)}(t) = \xi(a) + \sum_{s<t} d\xi(s)^{(\alpha)}.$$

Следовательно, по определению $\xi_{(\alpha)}$, для того чтобы показать, что $\xi_{(\alpha)}(t) \simeq \xi(t)$ п. н. для любого t, достаточно проверить, что п. в. $\sum |\mathbf{E}_t d\xi(t)^{(\alpha)}| \simeq 0$, а для этого достаточно убедиться, что математическое ожидание бесконечно мало. Но

$$\mathbf{E}\sum |\mathbf{E}_t d\xi(t)^{(\alpha)}| = \mathbf{E}\sum |\mathbf{E}_t (d\xi(t) - d\xi(t)^{(\alpha)})|$$

(поскольку ξ — мартингал, то $\mathbf{E}_t d\xi(t) = 0$) и, далее,

$$\leq \mathbf{E}\sum \mathbf{E}_t |(d\xi(t) - d\xi(t)^{(\alpha)}| = \mathbf{E}\sum \left| \frac{d\xi}{dt} - \left(\frac{d\xi}{dt}\right)^{(\alpha/dt)} \right| dt$$

$$\leq \mathbf{E}\sum \left| \frac{d\xi}{dt} - \left(\frac{d\xi}{dt}\right)^{(c)} \right| dt,$$

где $c = \alpha/\max dt$. Если $\alpha \gg 0$, то $c \simeq \infty$; а так как $d\xi/dt$ суммируема на $T' \times \Omega$, интересующее нас математическое ожидание бесконечно мало. Следовательно, множество всех α, для которых

$$\mathbf{E}\sum|\mathbf{E}_t d\xi(t)^{(\alpha)}| \leq \alpha,$$

содержит все $\alpha \gg 0$, и потому согласно принципу переполнения содержит и каждую достаточно большую инфинитезималь α. $\square$

В доказательствах двух следующих теорем мы апеллируем к результату (теорема 11.1), который будет доказан позже.

Теорема 10.4 *Пусть ξ — стохастический процесс, параметризованный элементами около-интервала T, такой, что величина $d\xi(t)/dt$ суммируема на $T' \times \Omega$, и пусть $\widehat{\xi}$ — ассоциированный мартингал. Тогда п. н. $\widehat{\xi}(t) = \widehat{\xi}(a)$ для любого t.*

Доказательство. Не теряя общности, предположим, что $\xi(a) = 0$. Согласно теоремам 10.2 и 10.3 можно считать, что $\xi = \widehat{\xi}$, а также что он является мартингалом, причем таким, что для некоторого бесконечно малого числа α выполнено неравенство $|d\xi(t)| \leq \alpha$ для любых t и ω. Тогда

$$\|d\xi(t)\|_2^2 = \mathbf{E}d\xi(t)^2 \leq \alpha\mathbf{E}|d\xi(t)|,$$

а так как приращение $d\xi(t)/dt$ суммируемо на $T' \times \Omega$, то

$$\sum \mathbf{E}|d\xi(t)| = \mathbf{E}\sum\left|\frac{d\xi}{dt}\right| dt \ll \infty.$$

Поэтому

$$\|\xi(b)\|_1^2 \leq \|\xi(b)\|_2^2 = \sum\|d\xi(t)\|_2^2 \leq \alpha\sum\mathbf{E}|d\xi(t)| \simeq 0.$$

Согласно теоремам 7.1 и 11.1 получаем, что п. н. $\max_t|\xi(t)| \simeq 0$. $\square$

Теорема 10.5. *Пусть ξ — возрастающий стохастический процесс, все приращения которого бесконечно малы всюду, и предположим, что $\mathbf{E}(\xi(b) - \xi(a)) \ll \infty$. Пусть $\widehat{\xi}$ — ассоциированный мартингал. Тогда п. н. $\widehat{\xi}(t) \simeq \widehat{\xi}(a)$ для любого t.*

Доказательство. И снова, не теряя общности, предположим, что $\xi(a) = 0$. Пусть $\alpha = \max_{t,\omega}|d\xi(t)|$, так что по условию α бесконечно мало. Тогда

$$\|\widehat{\xi}(b)\|_1^2 \leq \|\widehat{\xi}(b)\|_2^2 = \sum\|d\widehat{\xi}(t)\|_2^2 \leq \sum\|d\xi(t)\|_2^2$$

$$= \sum\mathbf{E}d\xi(t)^2 \leq \alpha\sum\mathbf{E}d\xi(t) = \alpha\mathbf{E}(\xi(b) - \xi(a)) \,|\simeq 0.$$

Согласно теоремам 7.1 и 11.1 имеем п. н. $\max_t|\widehat{\xi}(t)| \simeq 0$. $\square$

Глава 11
Сходимость мартингалов

В этой главе мы получим оценку максимума супермартингала или субмартингала и используем ее для изучения свойств сходимости и непрерывности в фиксированный момент времени.

На фондовой бирже цена одной акции компании “Nonstandard Oil" — стохастический процесс ξ. «Бык»*) покупает акцию в момент времени t и решает держать ее до тех пор, пока цена не возрастет по крайней мере на λ долларов, а в этот момент он забирает свою прибыль и продает акцию. Тогда доходы инвестора к моменту времени t составляют

$$\zeta_1(t) = \sum_{s<t} \eta_1(s) d\xi(s),$$

где $\eta_1(s) = 1$, пока он владеет акцией, и $\eta_1(s) = 0$ после этого. Будет ли величина $\eta_1(s)$ равняться 0 или 1, зависит только от значений $\xi(r)$ для $r \leq s$, так что $\eta_1(s)$ — это $\mathcal{P}$-процесс. Вспомним обсуждение таких сумм («стохастические интегралы») в гл. 9. Пусть событие Λ состоит в том, что стратегия инвестора успешна:

$$\Lambda = \{\max(\xi(t) - \xi(a)) \geq \lambda\}.$$

В случае успеха окончательный доход инвестора по крайней мере равен λ, иначе доход равен $\xi(b) - \xi(a)$. Таким образом,

$$\zeta_1(b) \geq \lambda\chi_\Lambda + (\xi(b) - \xi(a))\chi_{\Lambda^c}. \tag{11.1}$$

---

*) «Бык» (bullish investor) — биржевой спекулянт, играющий на повышение.

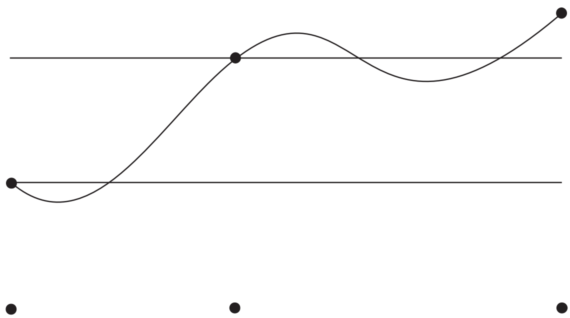


Рис. 11.1. Две инвестиционные стратегии

Теперь предположим, что, к несчастью для инвестора, ξ — супермартингал (падающий спрос). Так как η_1 — положительный $\mathcal{P}$-процесс, имеем

$$D\zeta_1(t) = \eta_1(t) D\xi(t) \leq 0,$$

так что ζ_1 — тоже супермартингал. Так как $\zeta_1(a) = 0$, получаем, что $\mathbf{E}\zeta_1(b) \leq 0$. Отсюда согласно (11.1) вытекает, что если ξ — супермартингал и $\lambda > 0$, то

$$\Pr\{\max(\xi(t) - \xi(a)) \geq \lambda\} \leq \frac{1}{\lambda} \|\xi(b) - \xi(a)\|_1. \tag{11.2}$$

Кажется, что приведенные соображения не дают никакой информации, если ξ — субмартингал. Но рассмотрим «медведицу»*), которая придерживается противоположной стратегии: она внимательно выжидает момента, когда цена акции возрастает по крайней мере на λ долларов, и как только это происходит, она покупает акцию и придерживает ее. Доход инвестора в момент времени t равен

$$\zeta_2(t) = \sum_{s<t} \eta_2(s) d\xi(s),$$

*) «Медведь» (bearish investor) — биржевой спекулянт, играющий на понижение.

где $\eta_2(s) = 1 - \eta_1(s)$, таким образом, η_2 — также положительный $\mathcal{P}$-процесс. Ее окончательный доход равен 0, пока курс растет (происходит событие Λ), и в этом случае она не больше, чем $\xi(b) - \xi(a) - \lambda$, т. е.

$$\zeta_2(b) \leq (\xi(b) - \xi(a) - \lambda)\chi_\Lambda. \tag{11.3}$$

Отметим, что неравенство (11.3) — это то же самое, что и неравенство (11.1), так как $\zeta_1 + \zeta_2 = \xi - \xi(a)$. Если теперь ξ — субмартингал, то и ζ_2 — тоже субмартингал, поскольку $D\zeta_2(t) = \eta_2(t) D\xi(t) \geq 0$. Так как $\zeta_2(a) = 0$, получаем $\mathbf{E}\zeta_2(b) \geq 0$. Согласно (11.3), если ξ — субмартингал, а $\lambda > 0$, то неравенство (11.2) выполнено.

Таким образом, неравенство (11.2) выполняется как для супермартингалов, так и для субмартингалов (и, конечно же, для мартингалов, потому что мартингал — это или супермартингал, или субмартингал). Так как процесс $-\xi$ является супермартингалом тогда и только тогда, когда ξ — субмартингал, и наоборот, справедливо также неравенство

$$\Pr\{\max(\xi(a) - \xi(t)) \geq \lambda\} \leq \frac{1}{\lambda}\|\xi(b) - \xi(a)\|_1,$$

выполняющееся как для субмартингалов, так и для супермартингалов. Мы доказали следующую теорему.

Теорема 11.1. *Пусть ξ — супермартингал или субмартингал, и пусть $\lambda > 0$. Тогда*

$$\Pr\{\max(\xi(t) - \xi(a)) \geq \lambda\} \leq \frac{1}{\lambda}\|\xi(b) - \xi(a)\|_1,$$

$$\Pr\{\max|(\xi(t) - \xi(a)| \geq \lambda\} \leq \frac{2}{\lambda}\|\xi(b) - \xi(a)\|_1. \tag{11.4}$$

Если ξ — мартингал, то согласно следующей теореме мы можем доказать результат, который вдвое лучше, чем (11.4).

Теорема 11.2. (i) *Если ξ — мартингал, а f — выпуклая функция, то $f \circ \xi$ — субмартингал.*

(ii) *Если ξ — субмартингал, а f — выпуклая возрастающая функция, то $f \circ \xi$ — субмартингал.*

Доказательство. Из релятивизации неравенства Йенсена следует, что если f выпукла, то $f(\mathbf{E}_s\xi(t)) \leq \mathbf{E}_s f(\xi(t))$. Если ξ — мартингал, то $\mathbf{E}_s\xi(t) = \xi(s)$ при $s \leq t$, таким образом, $f(\xi(s)) \leq \mathbf{E}_s f(\xi(t))$ при

$s \leq t$, а $f \circ \xi$ — субмартингал. Если ξ — субмартингал, то $\xi(s) \leq \mathbf{E}_s\xi(t)$ при $s \leq t$. Если f — возрастающая выпуклая функция, то $f(\xi(s)) \leq f(\mathbf{E}_s\xi(\mathrm{e})) \leq \mathbf{E}_s f(\xi(t))$ при $s \leq t$ (вновь в силу все той же релятивизации неравенства Йенсена), а $f \circ \xi$ — субмартингал. □

Если ξ — мартингал, то и $\xi - \xi(a)$ — тоже мартингал; согласно (ii) $|\xi - \xi(a)|$ — субмартингал, и тогда из (11.2) следует неравенство (11.4) с заменой $1/\lambda$ на $2/\lambda$.

Множество $\mathbf{R}^\Omega$ всех случайных величин на $\langle \Omega, \mathrm{pr} \rangle$ является метрическим пространством с метрикой $\|x - y\|_p$, где $1 \leq p \leq \infty$. Мы обозначаем это метрическое пространство через $\mathbf{L}^p$. Когда p меняется, пространства $\mathbf{L}^p$ — это одно и то же множество $\mathbf{R}^\Omega$, но с различными метриками. Будем говорить, что последовательность $x_1, \ldots, x_\nu$ (около-)*сходится к y в пространстве* $\mathbf{L}^p$, или (около-)$\mathbf{L}^p$*-сходится к* y, если $\|x_n - y\|_p \simeq 0$ для любого недоступного $n \leq \nu$. Если ξ — стохастический процесс, параметризованный конечным подмножеством T из $\mathbf{R}$, а $t \in T$, то говорят, что процесс ξ (около-)$\mathbf{L}^p$*-непрерывен в точке* t, или (около-)*непрерывен в точке t в пространстве* $\mathbf{L}^p$, если $s \simeq t$ влечет $\|\xi(s) - \xi(t)\|_p \simeq 0$. В противном случае говорят, что процесс ξ (строго) $\mathbf{L}^p$*-разрывен в точке* t.

$\mathbf{L}^p$-непрерывность — это аналитическое свойство, которое легко подтвердить или опровергнуть, однако непрерывность почти наверное выборочных траекторий является интересным и тонким вероятностным свойством. Многие теоремы теории стохастических процессов утверждают, что при соответствующих аналитических предположениях некоторое свойство выборочных траекторий выполняется почти наверное.

Теорема 11.3. (i) *Пусть $x_1, \ldots, x_\nu$ — супермартингал или субмартингал, сходящийся в* $\mathbf{L}^1$. *Тогда он сходится п. н.*

(ii) *Пусть ξ — супермартингал или субмартингал,* $\mathbf{L}^1$*-непрерывный в точке t. Тогда он п. н. непрерывен в точке t.*

Доказательство. Это утверждение немедленно вытекает из теоремы 11.1, из теоремы 7.2 и ее следствия. □

Следующая теорема устанавливает частично обратный результат.

Теорема 11.4. (i) *Пусть $x_1, \ldots, x_\nu$ — мартингал такой, что случайная величина x_ν суммируема. Тогда $x_1, \ldots, x_\nu$ сходится в* $\mathbf{L}^1$ *в том и только в том случае, если он сходится п. н.*

(ii) *Пусть ξ — такой мартингал, что случайная величина $\xi(b)$ суммируема. Тогда мартингал ξ будет* $\mathbf{L}^1$*-непрерывным в точке* t

(*непрерывным в точке* t *в пространстве* $\mathbf{L}^1$) *тогда и только тогда, когда он непрерывен п. н. в точке* t.

Доказательство. На основании теоремы 8.3 случайная величина $\xi(t) = \mathbf{E}_t\xi(b)$ суммируема при любом t. По теореме Лебега (теорема 8.2) сходимость п. н. или непрерывность влечет $\mathbf{L}^1$-сходимость или непрерывность. Импликации в другом направлении были установлены в теореме 11.3. □

Предположение о том, что ξ — мартингал, было использовано в этом доказательстве лишь для того, чтобы установить, что все случайные величины $\xi(t)$ суммируемы. Значит, результат остается верным, если предположить последнее.

Будем говорить, что t — (сильная) *неподвижная точка разрыва* стохастического процесса ξ, если неверно, что процесс ξ п. н. непрерывен в точке t. Теперь теорема 11.4 (ii) утверждает, что если ξ — мартингал, для которого случайная величина $\xi(b)$ суммируема, то t — неподвижная точка разрыва стохастического процесса ξ тогда и только тогда, когда t — точка $\mathbf{L}^1$-разрыва процесса ξ. Следующий контрпример показывает, что может происходить, если случайная величина $\xi(b)$ не суммируема.

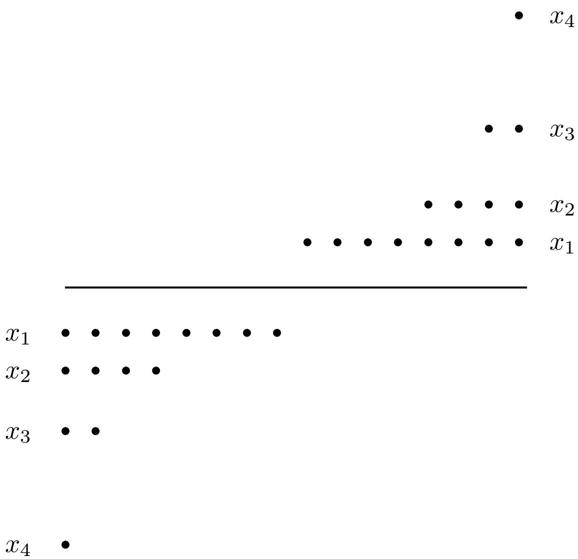


Рис. 11.2. Плохой мартингал

Пусть Ω — упорядоченное множество из 2^ν точек, где $\nu \simeq \infty$, каждая из которых имеет вероятность $2^{-\nu}$, и пусть x_n, $n = 1, \ldots, \nu$, — случайная величина, принимающая значение -2^{n-1} на первых $2^{\nu-n}$ точках, значение 2^{n-1} на последних $2^{\nu-n}$ точках и значение 0 на промежуточных точках. (Ненулевые значения показаны на рис. 11.2, где для экономии бумаги взято $\nu = 4$ вместо $\nu \simeq \infty$.) Тогда x_n — мартингал относительно алгебры $\mathcal{P}$, порожденной случайными величинами $x_1, \ldots, x_n$. При любом n имеем $\mathbf{E}x_n = 0$, $\|x\|_1 = 1$ и $\|x_n - x_m\|_1 \geq 1$, если $n \neq m$. Пусть k и $\nu - k$ недоступны, и пусть T — около-интервал, содержащий $\nu - k$ точек. Определим процесс ξ равенством $\xi(t) = x_{n+k}$, где t — n-я точка T. Тогда процесс ξ всюду $\mathbf{L}^p$-разрывен, но так как п. н. $\xi(t) = 0$ для любого t, он п. н. непрерывен при любом t.

Оставшаяся часть этой главы посвящена доказательству того факта, что имеется очень мало точек разрыва мартингала ξ, для которых случайная величина $\xi(b)$ суммируема.

Пусть $\varepsilon > 0$. Будем говорить что функция $\xi : T \to M$, где $\langle M, \rho \rangle$ — метрическое пространство (такое, как $\mathbf{L}^1$ или $\mathbf{R}$), (около-) *ε-непрерывна* в точке t, если $\rho(\xi(t)_1, \xi(t)_2) \leq \varepsilon$, как только $t_1 \simeq t$ и $t_2 \simeq t$; в противном случае говорят, что функция ξ (сильно) *ε-разрывна* в точке t и что t есть точка (сильного) *ε-разрыва* функции ξ. Отметим, что если $s \simeq t$, то s является точкой ε-разрыва функции ξ тогда и только тогда, когда точкой ε-разрыва является t. Если t — точка ε-разрыва функции ξ, то найдутся точки t_1 и t_2, бесконечно близкие к t и такие, что $\rho(\xi(t_1), \xi(t_2)) > \varepsilon$. Тогда по неравенству треугольника существует точка t' (или t_1, или t_2), которая бесконечно близка к t, причем $\rho(\xi(t'), \xi(t)) > \varepsilon/2$. Отметим также, что точка t является точкой разрыва функции ξ тогда и только тогда, когда t является точкой ε-разрыва для некоторого $\varepsilon \gg 0$.

Теорема 11.5. *Пусть ξ — такой мартингал, что случайная величина $\xi(b)$ суммируема, и пусть $\varepsilon \gg 0$. Тогда найдется доступное число точек $t_1, \ldots, t_n$, никакие две из которых не бесконечно близки друг к другу, такие, что t является точкой ε-разрыва ξ в пространстве $\mathbf{L}^1$ тогда и только тогда, когда $t \simeq t_i$ для некоторого $i = 1, \ldots, n$.*

Доказательство. Множество всех чисел c, для которых выполнено неравенство $\|\xi(b) - \xi(b)^{(c)}\|_1 \leq \varepsilon/4$, содержит все $c \simeq \infty$, а следовательно, содержит некоторое $c \ll \infty$. Пусть $\xi_c(t) = \mathbf{E}_t \xi(b)^{(c)}$. Как отмечено в гл. 9, ξ_c — мартингал. Так как условные математические ожидания уменьшают $\mathbf{L}^1$-нормы, $\|\xi_c(t) - \xi(t)\|_1 \leq \varepsilon/4$ при любом t.

Предположим, что найдутся n точек $t_1, \ldots, t_n$, никакие две из которых не бесконечно близки друг к другу, такие, что $\|\xi(s_i) - \xi(r_i)\|_1 > \varepsilon$ для некоторых s_i и r_i, $s_i \simeq r_i \simeq t_i$. Обозначим их так, чтобы $r_i < s_i$. Тогда интервалы $[r_i, s_i]$ не пересекаются. По неравенству треугольника $\|\xi_c(s_i) - \xi_c(r_i)\|_1 > \varepsilon/2$, а также $\|\xi_c(s_i) - \xi_c(r_i)\|_2 > \varepsilon/2$. Так как мартингал имеет ортогональные приращения, имеем

$$n\frac{\varepsilon^2}{4} \le \sum_{i=1}^{n} \|\xi_c(s_i) - \xi_c(r_i)\|_2^2 \le \|\xi_c(b) - \xi_c(a)\|_2^2$$

$$= \|\xi_c(b) - \mathbf{E}_a\xi_c(b)\|_2^2 \le \|\xi_c(b)\|_2^2 \le c^2.$$

Поэтому $n \le 4c^2/\varepsilon^2 \ll \infty$. Доказательство завершается ссылкой на внешний принцип наименьшего числа. □

* **Теорема 11.6.** *Пусть ξ — мартингал такой, что случайная величина $\xi(b)$ суммируема. Тогда либо*

(i) *существует лишь доступное число точек $t_1, \ldots, t_n$, никакие две из которых не бесконечно близки друг к другу, таких, что t — неподвижная точка разрыва ξ тогда и только тогда, когда $t \simeq t_i$ для некоторого i,*

либо

(ii) *существует бесконечная последовательность, состоящая из точек t_i, никакие две из которых со стандартными номерами не являются бесконечно близкими друг к другу, и такая, что t — неподвижная точка разрыва ξ тогда и только тогда, когда $t \simeq t_i$ для некоторого стандартного i.*

Доказательство. По теореме 11.4 (ii) точка t — неподвижная точка разрыва процесса ξ в том и только в том случае, когда t — точка $(1/k)$-разрыва мартингала ξ в пространстве $\mathbf{L}^1$ для некоторого стандартного числа k. Пусть число k стандартно. По теореме 11.5 найдется множество

$$E_k = \{t_{k1}, \ldots, t_{kn_k}\},$$

где $n_k \ll \infty$ и $|t_{k_i} - t_{k_j}| \gg 0$ при $i \ne j$, такое, что ни для какого $l < k$ точка t, будучи точкой $(1/k)$-разрыва в $\mathbf{L}^1$, не является точкой $(1/l)$-разрыва в пространстве $\mathbf{L}^1$, если и только если точка t бесконечно близка к некоторому (единственному) элементу из E_k. Согласно принципу последовательности найдется последовательность $k \mapsto E_k$ подмножеств множества T такая, что эти свойства выполнены для каждого стандартного числа k. Если существует такое стандартное число

j, что множества E_k пусты для всех стандартных $k > j$, то определим $t_1, \dots, t_n$ как $\bigcup_{k \le j} E_k$. Тогда последовательность $t_1, \dots, t_n$ имеет все свойства, установленные в (i). В противном случае, если j нестандартно, определим $t_1, \dots, t_\nu$ как $\bigcup_{k \le j} E_k$, где t_i задается произвольно при $i > \nu$. Тогда последовательность t_i имеет все свойства, сформулированные в (ii). $\square$

Проверим, что в (ii) действительно утверждается, что существует мало точек разрыва.

Теорема 11.7. *Пусть T — около-интервал. Тогда не существует последовательности t_i такой, что каждый элемент множества T бесконечно близок к некоторому t_i со стандартным номером i.*

Доказательство проведем от противного. Не теряя общности, предположим, что $a = 0$ и $b = 1$. Пусть $x \in [0, 1]$ и t — наибольший элемент конечного множества T' такой, что $t \le x$. Тогда $t \le x \le t + dt$, так что x бесконечно близко к $t \in T$ и, следовательно, согласно предположению, к некоторому числу t_i со стандартным номером i. Но диагональный метод Кантора порождает некоторый элемент x в $[0, 1]$, который в i-м десятичном знаке отличается от t_i; это означает, что $\|x - t_i\| \ge 10^{-i}$ для любого номера i, включая стандартные. Получили противоречие. $\square$

Мы видели, что существует мало неподвижных точек разрыва в смысле «количества». Покажем теперь, что существует мало неподвижных точек разрыва в смысле «меры».

***Теорема 11.8.** *Пусть ξ — мартингал такой, что случайная величина $\xi(b)$ суммируема, и функция $f : \mathbf{R}^+ \to \mathbf{R}^+$ такова, что $f(h) \simeq 0$ при любом $h \simeq 0$. Для любого числа $\varepsilon \gg 0$ найдутся натуральное число ν и такие интервалы $[r_i, s_i]$, $i = 1, \dots, \nu$, что $\sum_{i=1}^{\nu} f(s_i - r_i) \le \varepsilon$ и каждая неподвижная точка разрыва мартингала ξ принадлежит одному из интервалов $[r_i, s_i]$.*

Доказательство. Пусть число $\alpha > 0$ бесконечно мало. Для любого $i \in \mathbf{N}$ пусть h_i — наибольшее число, кратное α, которое меньше 1 и такое, что $f(h_i) \le \varepsilon/2^i$, или, если такого числа не существует, $h_i = 0$. Тогда $h_i \gg 0$ и $f(h_i) \le \varepsilon/2^i$ для любого $i \ll \infty$. Пусть t_i — такая конечная или бесконечная последовательность, как в теореме 11.6. Пусть $\nu = n$ в случае (i), $\nu \simeq \infty$ в случае (ii), и пусть $r_i = t_i - h_i/2$, а $s_i = t_i + h_i/2$. Тогда $\sum_{i=1}^{\nu} f(s_i - r_i) \le \varepsilon$, и каждая неподвижная точка разрыва ξ принадлежит одному из интервалов $[r_i, s_i]$. $\square$

В частности, пусть ξ — мартингал, параметризованный около-интервалом T и такой, что случайная величина $\xi(b)$ суммируема. Выберем

$f(h) = h$. Тогда почти всюду относительно pr_T («нормированной меры Лебега») точка t не является неподвижной точкой разрыва мартингала ξ.

Наконец, отметим, что если ξ — мартингал такой, что случайная величина $\xi(b)$ суммируема, то случайная величина $\xi(a) = \mathbf{E}_a\xi(b)$ суммируема, а значит, случайная величина $\xi(b) - \xi(a)$ суммируема. Теоремы 11.4–11.8 остаются верными в предположении,что случайная величина $\xi(b) - \xi(a)$ суммируема, так как мартингал ξ тривиально отличается от мартингала $\xi - \xi(a)$. В следующей главе мы намерены изучать мартингалы при более слабом предположении $\|\xi(b) - \xi(a)\|_1 \ll \infty$.

Глава 12

Флуктуации мартингалов

Если процесс ξ — винеровское или пуассоновское блуждание, то $\mathbf{E}d\xi(r)^2 = dr$, так что $\|\xi(t) - \xi(s)\|_2^2 = |t - s|$. Таким образом, в обоих примерах процессы непрерывны в пространстве $\mathbf{L}^2$, а следовательно, и в пространстве $\mathbf{L}^1$ при любом t. По теореме 11.3 (ii) для любой точки t процесс п. н. непрерывен в точке t, неподвижных точек разрыва нет. Однако отсюда не следует, что п. н. процесс непрерывен в точке t для всех t, потому что исключительные множества могут зависеть от t. Рассмотрим пуассоновское блуждание, параметризованное околоинтервалом. Оно принимает только целые значения, так что единственная возможность для него быть непрерывным во всех точках t — оставаться тождественным нулем. Вероятность скачка в точке t равна dt, так что вероятность отсутствия скачков вследствие независимости приращений равна $\prod_{t\in T'}(1 - dt)$. Но по теореме 5.3

$$\log \prod (1 - dt) = \sum \log(1 - dt) \sim -\sum dt = -(b - a),$$

так как $\log(1 - dt) \sim -dt$. Поэтому вероятность отсутствия скачков $\sim e^{-(b-a)}$, что строго меньше 1. В следующей главе мы увидим, что винеровский процесс на самом деле п. н. непрерывен в любой точке t. Типичные траектории двух обсуждаемых процессов показаны на рис. 12.1.

Таким образом, возникает вопрос, насколько плохими могут быть разрывы траекторий мартингалов. В традиционной терминологии у ограниченной функции существует два типа разрывов: разрыв первого рода (скачок) и разрыв второго рода (осцилляция), как на рис. 12.2. (На рис. 12.2,b подразумевается бесконечно много колебаний, как у функции $\sin \frac{1}{x}$ в начале координат.) Понятие функции доступного колебания, введенное в гл. 6, является внешним аналогом внутреннего

Пуассоновское блуждание *Винеровское блуждание*

Рис. 12.1. Две траектории

понятия функции (определенной в ограниченной области), имеющей только разрывы первого рода. Мы можем воспринимать рис. 12.2,a как иллюстрацию функции доступного колебания, а рис. 12.2,b — как иллюстрацию функции недоступного колебания.

Основной результат этой главы заключается в том, что при очень слабых предположениях траектории супермартингала или субмартингала почти наверное имеют доступное колебание. Сначала мы установим техническую лемму, а затем построим инвестиционную стратегию, которая приведет к оценке числа выбросов за интервал вверх.

Теорема 12.1. *Пусть ξ — стохастический процесс, параметризованный элементами конечного подмножества T множества $\mathbf{R}$. Процесс ξ п. н. имеет доступное колебание тогда и только тогда, когда для всех $\varepsilon \gg 0$ и $k \simeq \infty$ имеет место соотношение*

$$\Pr\{\xi \textit{ допускает } k \textit{ колебаний порядка } \varepsilon\} \simeq 0.$$

Доказательство. Ясно, что условие необходимо. Наоборот, допустим, что условие на колебание выполнено. Пусть событие $M(\varepsilon, k)$ состоит в том, что процесс ξ допускает k колебаний порядка ε, и пусть $\delta \gg 0$. Возьмем в качестве k_n наименьшее натуральное число такое, что

$$\Pr M\left(\frac{1}{n}, k_n\right) \leq \frac{\delta}{2^n},$$

a. Первого рода *b. Второго рода*

Рис. 12.2. Разрывы

и пусть

$$N = \bigcup_{n=1}^{\infty} M\left(\frac{1}{n}, k_n\right).$$

Тогда $\Pr N \leq \delta$. По предположению, число k_n доступно, если доступно число n, так что ξ имеет доступное колебание на N^c. Так как число $\delta \gg 0$ произвольно, это завершает доказательство. $\square$

Спекулянтка на фондовой бирже надеется урвать куш на дико колеблющемся рынке, придерживаясь правила «купить подешевле — продать подороже». Она решает дождаться первого момента времени s_1, когда цена акции упадет ниже уровня λ_1, прежде чем ее купить, а затем продать акцию в первый из последующих моментов времени t_1, когда цена станет выше λ_2 (где $\lambda_2 > \lambda_1$). Затем она повторяет эту процедуру настолько часто, насколько это возможно. Ее доход в момент времени t выражается формулой

$$\zeta_3(t) = \sum_{s<t} \eta_3(s) d\xi(s),$$

где η_3 принимает значение 1, когда она придерживает акцию (это значит, пока для некоторого номера k верно, что $s_k \leq s$, но не $s \geq t_k$), и значение 0 в другие моменты времени. Тогда η_3 — положительный $\mathcal{P}$-процесс. Спекулянтка временно зарабатывает по крайней мере $\beta(\lambda_2 - \lambda_1)$ долларов, где β — число выбросов вверх из интервала $[\lambda_1, \lambda_2]$ для процесса ξ (т. е. β — это наибольшее из k, для которых t_k определено), но она может потерять в конце, если за последним s_k не следует

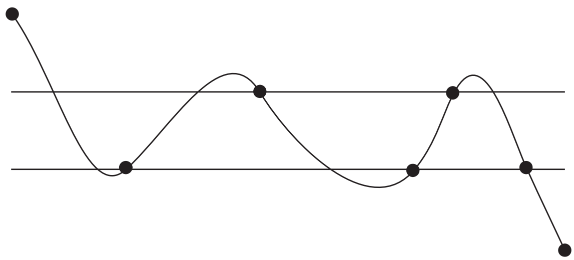


Рис. 12.3. Два выброса вверх

t_k. Однако эта потеря составляет самое большее $(\lambda_1 - \xi(b))^+$, где x^+ обозначает $\max\{x, 0\}$. Таким образом,

$$\zeta_3(b) \geq \beta(\lambda_2 - \lambda_1) - (\lambda_1 - \xi(b))^+. \tag{12.1}$$

Предположим, что ξ — супермартингал, так что $0 \geq \mathbf{E}\zeta_3(b)$. Тогда согласно (12.1) выполнено

$$(\lambda_2 - \lambda_1)\mathbf{E}\beta \leq \mathbf{E}(\lambda_1 - \xi(b))^+ \leq \|\xi(b)\|_1 + |\lambda_1|.$$

Мы доказали следующую теорему.

Теорема 12.2. *Пусть ξ — супермартингал, $\lambda_1 < \lambda_2$ и β — число выбросов вверх из интервала $[\lambda_1, \lambda_2]$ для процесса ξ. Тогда*

$$\mathbf{E}\beta \leq \frac{\|\xi(b)\|_1 + |\lambda_1|}{\lambda_2 - \lambda_1}.$$

Хотя мы и не будем использовать этот факт, интересно отметить, что теорема 12.2 верна также и для субмартингала ξ. Чтобы убедиться в этом, рассмотрим консервативного инвестора, который хочет найти надежные акции фонда для вложения средств и придерживается стратегии, противоположной стратегии спекулянтки, т. е. он покупает акцию в момент времени a, пока $\xi(a) \leq \lambda_1$. В момент времени s_1 он продает акцию, учитывает свои потери для нужд налогообложения и дожидается момента времени t_1, чтобы купить акцию снова, а затем повторяет эту процедуру столь часто, сколь это необходимо. Его доход в момент времени t задается формулой

$$\zeta_4(t) = \sum_{s<t} \eta_4(s) d\xi(s),$$

где $\eta_4(s) = 1 - \eta_3(s)$. При этом η_4 — также положительный $\mathcal{P}$-процесс. Я утверждаю, что

$$\zeta_4(b) \leq (\xi(b) - \lambda_1)^+ - \beta(\lambda_2 - \lambda_1). \tag{12.2}$$

Чтобы доказать это, заметим, что если $\xi(a) \geq \lambda_1$, то он покупает по цене $\xi(a)$ и теряет по меньшей мере $\lambda_2 - \lambda_1$ при каждом выбросе вверх. Таким образом,

$$\zeta_4(b) \leq \xi(b) - \xi(a) - \beta(\lambda_2 - \lambda_1) \leq (\xi(b) - \lambda_1)^+ - \beta(\lambda_2 - \lambda_1).$$

При $\xi(a) < \lambda_1$ и $\beta = 0$ он никогда не купит, так что $\zeta_4(b) = 0$ и (12.2) тривиально верно. Наконец, если $\xi(a) < \lambda_1$, а $\beta > 0$, то он сначала покупает по цене $\geq \lambda_2$ после первого выброса вверх, так что

$$\zeta_4(b) \leq (\xi(b) - \lambda_2) - (\beta - 1)(\lambda_2 - \lambda_1) \leq (\xi(b) - \lambda_1)^+ - \beta(\lambda_2 - \lambda_1).$$

Итак, (12.2) выполняется во всех случаях, и теорема (12.2) верна для субмартингала ξ (так как $0 \leq \mathbf{E}\zeta_4(b)$).

Теорема 12.3. *Пусть ξ — супермартингал или субмартингал, для которого*

$$\|\xi(b) - \xi(a)\|_1 \ll \infty.$$

Тогда процесс ξ п. н. имеет доступное колебание.

Доказательство. Так как процесс $-\xi$ является супермартингалом тогда и только тогда, когда ξ — субмартингал, достаточно доказать теорему для супермартингалов. Не теряя общности, можно дополнительно предположить, что $\xi(a) = 0$. Пусть $\delta \gg 0$ и

$$\lambda = \frac{2\|\xi(b)\|_1}{\delta},$$

так что $\lambda \ll \infty$. По теореме 11.1

$$\Pr\{\max|\xi(t)| \geq \lambda\} \leq \frac{2\|\xi(b)\|_1}{\lambda} \leq \delta. \tag{12.3}$$

Пусть теперь $\varepsilon \gg 0$ таково, что $n = \lambda/\varepsilon$ — целое число. Заметим, что $n \ll \infty$. Разделим $[-\lambda, \lambda]$ на $2n$ подынтервалов длины ε.

Теперь предположим, что $\max|\xi(t)| < \lambda$ и что процесс ξ допускает $2n + 2k$ колебаний порядка 2ε, где k — недоступное число, кратное $2n$. Каждое 2ε-колебание порождает или выброс вниз, или выброс вверх из одного из $2n$ подынтервалов, так что из некоторого подынтервала имело место по крайней мере $1 + k/n$ выбросов. Пусть β — число его выбросов вверх из такого подынтервала. Число выбросов вверх может отличаться от числа выбросов вниз самое большее на 1, значит, $\beta \geq k/2n$. Но по неравенству Чебышёва и по теореме 12.2

$$\Pr\left\{\beta \geq \frac{k}{2n}\right\} \leq \frac{2n}{k}\mathbf{E}\beta \leq \frac{n}{k\varepsilon}(\|\xi(b)\|_1 + \lambda).$$

Так как всего подынтервалов $2n$, вероятность того, что $\max\|\xi(t)\| < \lambda$ и что процесс ξ допускает $2n + 2k$ колебаний порядка 2ε, не более чем

$$\frac{2n^2}{k\varepsilon}(\|\xi(b)\|_1 + \lambda) \simeq 0.$$

Вместе с (12.3) это показывает, что

$$\Pr\{\xi \text{ допускает } 2n + 2k \text{ колебаний порядка } 2\varepsilon\} \lesssim \delta.$$

Ввиду произвольности $\delta \gg 0$ доказательство завершается ссылкой на теорему 12.1. $\square$

Глава 13
Разрывы мартингалов

Мы только что видели, что при весьма слабых предположениях почти каждая траектория супермартингала или субмартингала имеет доступное колебание. Это дает некоторую информацию о природе возможных разрывов траектории, но требует более детального анализа. Пусть $\varepsilon \gg 0$. Тогда п. н. найдется только доступное число точек, для которых $|d\xi(t)| \geq \varepsilon$. Может ли случиться, что две такие точки расположены бесконечно близко друг к другу? Может ли случиться, что каждое приращение $d\xi(t)$ бесконечно мало, но при этом возникнуть разрыв (путем складывания недоступного числа бесконечно малых приращений до того, пока сумма не станет больше, чем ε, в течение бесконечно малого промежутка времени)? Может ли траектория не быть непрерывной ни слева, ни справа в некоторой точке t? Короче говоря, могут ли случиться два ε-колебания в течение бесконечно малого интервала времени? Конечно, некоторые из этих локальных ужасов могут иметь место для функций доступного колебания, но мы увидим, что ответ на эти вопросы — *нет* для почти каждой траектории супермартингала или субмартингала ξ, для которого $0 \ll b - a$ и случайная величина σ_ξ^2 суммируема на $T' \times \Omega$. Вспомним, что случайная величина σ_ξ^2 суммируема на $T' \times \Omega$, если каждая из случайных величин $\sigma_\xi^2(t)$ суммируема на Ω (см. замечание после доказательства теоремы 8.4).

Для стохастического процесса ξ, параметризованного конечным подмножеством T из $\mathbf{R}$, определим *собственное время* τ_ξ равенством

$$\tau_\xi(t) = \sum_{s<t} \sigma_\xi^2(s) ds.$$

Собственное время возрастает на каждой траектории, но темп, с которым оно прирастает, может меняться от траектории к траектории.

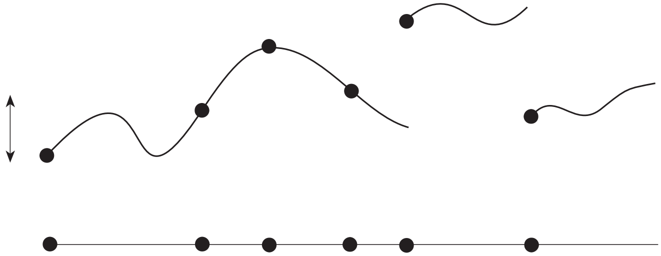


Рис. 13.1. Другая стратегия капиталовложений

Если $s \le t$, мы называем $\tau_\xi(t) - \tau_\xi(s)$ *продолжительностью собственного времени* интервала $[s,t]$.

Цена акции некоторого фонда является мартингалом ξ, удовлетворяющим неравенству $\|\xi(b) - \xi(a)\|_1 \ll \infty$. Инвестор озабочен тем, чтобы его капитал не был омертвлен в течение долгого времени (важно, конечно, собственное время — ожидание быстро колеблющейся цены делает время медленно текущим для инвестора), и он считает подходящим временем покупки момент, когда фонд демонстрирует признаки активности. Он выбирает натуральное число j, $\varepsilon \gg 0$ и бесконечно малое $\alpha > 0$. Пусть $s_0 = a$, s_1 — первый момент времени, для которого $|\xi(s_1) - \xi(a)| \ge \varepsilon/4$, и s_2, $s_2 > s_1$, — первый момент времени, для которого $|\xi(s_2) - \xi(s_1)| \ge \varepsilon/4$, и т. д.

Инвестор покупает акцию фонда в момент s_j и продает ее в момент s_{j+1}, если $\tau_\xi(s_{j+1}) - \tau_\xi(s_j) \le \alpha$, в противном случае он продает ее в первый момент времени s, для которого $\tau_\xi(s+ds) - \tau_\xi(s_j) > \alpha$. Он никогда не покупает снова. Его доход в момент времени t выражается формулой

$$\zeta_j(t) = \sum_{s<t} \eta_j(s) d\xi(s),$$

где η_j — $\mathcal{P}$-процесс, принимающий значения 0 или 1. (Заметим, что $\tau_\xi(s+ds) = \tau_\xi(s) + \sigma_\xi^2(s)ds \in \mathcal{P}_s$, так что η_j — это действительно $\mathcal{P}$-процесс.)

Имеем

$$\|\zeta_j(b)\|_2^2 = \mathbf{E}\biggl(\sum_{t\in T'} \eta_j(t) d\xi(t)\biggr)^2 = \mathbf{E}\sum_{t\in T'} \eta_j(t)^2 \sigma_\xi^2(t) dt \le \alpha \simeq 0,$$

и по неравенству Чебышёва получаем, что п. н. $\zeta_j(b) \simeq 0$.

По теореме 12.3 процесс ξ п. н. имеет доступное колебание, так что п. н. существует только доступное число точек s_j. Поэтому

$$\max_j |\zeta_j(b)| \simeq 0 \text{ п. н.} \tag{13.1}$$

Пусть событие $M(\varepsilon, \alpha)$ заключается в том, что внутри некоторого интервала, продолжительность собственного времени которого $\leq \alpha$, найдутся по крайней мере два ε-колебания . Отметим, что если $|\xi(t_2) - \xi(t_1)| \geq \varepsilon$ и $|\xi(t_4) - \xi(t_3)| \geq \varepsilon$, где $t_1 \leq t_2 \leq t_3 \leq t_4$, то найдутся по крайней мере две точки s_j между t_1 и t_4, так как должны быть хоть одна точка в интервале $[t_1, t_2]$ и хоть одна точка в интервале $[t_3, t_4]$. Поэтому если происходит событие $M(\varepsilon, \alpha)$, то $\tau_\xi(s_{j+1}) - \tau_\xi(s_j) \leq \alpha$ для некоторого номера j, и следовательно,

$$\max_j |\zeta_j(b)| \geq \varepsilon/4.$$

Поэтому согласно (13.1) если $\varepsilon \gg 0$ и $\alpha \simeq 0$, то $\Pr M(\varepsilon, \alpha) \simeq 0$.

Пусть $\delta \gg 0$ и α_n — наибольшее число (в конечном множестве всех возможных значений разности $\tau_\xi(t) - \tau_\xi(s)$), для которого

$$\Pr M\left(\frac{1}{n}, \alpha_n\right) \leq \frac{\delta}{2^n}.$$

Тогда $\alpha_n \gg 0$ при $n \ll \infty$, и если

$$N = \bigcup_{n=1}^{\infty} M\left(\frac{1}{n}, \alpha_n\right),$$

то $\Pr N \geq \delta$. Но на N^c ни для какого $\varepsilon \gg 0$ ни на каком интервале бесконечно малой продолжительности собственного времени не происходит двух ε-колебаний. Так как $\delta \gg 0$ произвольно, мы доказали следующий результат.

Теорема 13.1. *Пусть ξ — мартингал и $\|\xi(b) - \xi(a)\|_1 \ll \infty$. Тогда п. н. для любого $\varepsilon \gg 0$ не происходит двух ε-колебаний ни в каком интервале бесконечно малой продолжительности собственного времени.*

Отметим, что для нормированного мартингала ξ собственное время тривиальным образом отличается от обычного времени: $\tau_\xi(t) = t - a$, а также что

$$\|\xi(b) - \xi(a)\|_1 \le \|\xi(b) - \xi(a)\|_2 = \sqrt{b-a},$$

таким образом, $\|\xi(b) - \xi(a)\|_1 \ll \infty$, если $b - a \ll \infty$. Тогда и для него остается в силе заключение теоремы 13.1. Если всюду на $T' \times \Omega$ каждое приращение $d\xi$ бесконечно мало, отсюда следует, что почти каждая траектория непрерывна в любой момент времени, так как если $|\xi(t_2) - \xi(t_1)| \gg 0$, то должен существовать такой момент t между t_1 и t_2, что $|\xi(t_2) - \xi(t)| \gg 0$ и $|\xi(t) - \xi(t_1)| \gg 0$. Тем самым теорема 13.1 имеет

Следствие. *Почти каждая траектория винеровского блуждания непрерывна во всех точках t.*

Преобразуем информацию о собственном времени в теореме 13.1 в информацию об обычном времени.

Теорема 13.2. *Пусть ξ — стохастический процесс с $b - a \gg 0$, для которого дисперсия σ_ξ^2 суммируема на $T' \times \Omega$. Тогда п. н. собственное время τ_ξ абсолютно непрерывно. В частности, п. н. для каждого интервала бесконечно малой длины продолжительность собственного времени бесконечно мала.*

Доказательство. По теореме Фубини для почти каждого ω функция $t \mapsto \sigma_\xi^2(t)$ суммируема на T'. По теореме Лебега отсюда следует, что п. н. собственное время τ_ξ абсолютно непрерывно (так как $d\tau_\xi(t)\, dt = \sigma_\xi^2(t)$ и $b - a \gg \infty$). Заключительное утверждение теоремы просто означает, что абсолютно непрерывная функция непрерывна. □

Пусть T — конечное подмножество множества $\mathbf{R}$ и $\xi : T \to \mathbf{R}$. Точку $s \in T'$ назовем (точкой) *ε-скачка* процесса ξ, если $|d\xi(s)| \ge \varepsilon$, и (точкой) *скачка* процесса ξ, если для некоторого $\varepsilon \gg 0$ точка $s \in T'$ — это *ε-скачок* процесса ξ. Если s — точка скачка процесса ξ, $ds \simeq 0$, и если $t \simeq s$, то t — точка разрыва процесса ξ. Мы говорим, что t — точка *разрыва со скачком* процесса ξ, или точка разрыва первого рода процесса ξ, если существует единственная точка скачка s, для которой $s \simeq t$.

Теорема 13.3. *Пусть T — конечное подмножество множества $\mathbf{R}$ и процесс $\xi : T \to \mathbf{R}$ таков, что для любого числа $\varepsilon \gg 0$ ни в каком бесконечно малом интервале не происходит двух ε-колебаний. Тогда каждая точка разрыва t процесса ξ есть точка разрыва первого рода.*

Если s — скачок, $s \simeq t$, то процесс ξ непрерывен слева в точке t тогда и только тогда, когда $t \leq s$, и ξ непрерывен справа в t тогда и только тогда, когда $t > s$. Если все приращения процесса ξ бесконечно малы, то процесс ξ непрерывен. Никакие две точки скачка процесса ξ не являются бесконечно близкими друг другу. Если $b - a \ll \infty$, то для любого $\varepsilon \gg 0$ найдется только доступное число ε-скачков.

Доказательство. Пусть t — точка разрыва процесса ξ. Это значит, что существуют t_1 и t_2, где $t_1 \leq t_2$, бесконечно близкие к t и такие, что $|d\xi(t_2) - d\xi(t_1)| \geq \delta$ для некоторого $\delta \gg 0$. Пусть s — первый момент времени $\geq t_1$, для которого $|\xi(s+ds) - \xi(t_1)| \geq \delta/2$. По предположению относительно ξ имеем $\|\xi(s+ds) - \xi(t_1)\| \gtrsim \delta$ и $|\xi(s) - \xi(t_1)| \simeq 0$. Поэтому $|\xi(s+ds) - \xi(s)| \gtrsim \delta$, так что s — это точка $(\delta/2)$-скачка. Следовательно, существует точка скачка s такая, что $s \simeq t$. Такая точка, очевидно, единственна и $ds \leq t_2 - t_1 \simeq 0$, так что t — точка разрыва первого рода. Теперь утверждение теоремы очевидно. □

Из теорем 13.1–13.3 вытекает следующий результат.

Теорема 13.4. *Пусть $b - a \gg 0$ и ξ — такой мартингал, что дисперсия σ_ξ^2 суммируема на $T' \times \Omega$. Тогда для почти каждой траектории выполнены утверждения теоремы* 13.3.

Используя принцип последовательности, мы (при сделанных предположениях и при условии $b - a \ll \infty$) можем показать также, что п. н. найдется последовательность точек $s_i \in T$ такая, что при $s \in T$ точка s будет точкой скачка процесса ξ в том и только в том случае, если $s = s_i$ для некоторого стандартного номера i.

Глава 14
Условие Линдеберга

Будем говорить, что процесс ξ удовлетворяет (*около-*)*условию Линдеберга*, если для любого числа $\varepsilon \gg 0$

$$\mathbf{E}\sum d\xi(t)^2 \simeq \mathbf{E}\sum d\xi(t)^{(\varepsilon)2}. \tag{14.1}$$

Для нормированного мартингала левая часть выражения (14.1) равна $b-a$. Винеровское блуждание удовлетворяет условию Линдеберга, потому что обе части выражения (14.1) равны, но для пуассоновского блуждания правая часть равна 0 при $\varepsilon < 1$.

Один из способов получения нормированного мартингала состоит в следующем. Пусть $x_1, \ldots, x_\nu$, $\nu \simeq \infty$, — независимые случайные величины со средним 0, причем $0 < \mathbf{E}x_k^2 \ll \infty$, $1 \leq k \leq \nu$, и пусть

$$s_n = \sqrt{\sum_{k=1}^{n} \mathbf{E}x_k^2}.$$

Предположим, что $s_\nu \simeq \infty$. Пусть T — множество всех чисел вида s_n^2/s_ν^2, $n = 0, \ldots, \nu$. Тогда T — около-интервал, для которого $a = 0$ и $b = 1$. Для $t \in T$, где $t = s_n^2/s_\nu^2$, положим

$$\xi(t) = \frac{1}{s_\nu}\sum_{k=1}^{n} x_k.$$

Тогда ξ — нормированный мартингал, который будем называть мартингалом, *ассоциированным* с $x_1, \ldots, x_\nu$. Он удовлетворяет условию Линдеберга тогда и только тогда, когда для любого $\varepsilon \gg 0$

$$\frac{1}{s_\nu^2}\sum_{k=1}^{\nu} \mathbf{E}x_k^{(\varepsilon s_\nu)2} \simeq 1. \tag{14.2}$$

Соотношение (14.2) очень похоже на принятую форму условия Линдеберга, но оно много легче для восприятия, когда выражено в форме (14.1) для ассоциированного нормированного мартингала.

Теорема 14.1. *Пусть ξ — стохастический процесс, удовлетворяющий условию Линдеберга. Тогда п. н. каждое приращение бесконечно мало.*

Доказательство. Для $\varepsilon \gg 0$ имеем:

$$0 \simeq \mathbf{E} \sum (d\xi(t)^2 - d\xi(t)^{(\varepsilon)2}) = \sum_t \sum_{|\lambda|>\varepsilon} \lambda^2 \mathrm{pr}_{d\xi(t)}(\lambda)$$

$$\geq \varepsilon^2 \sum_t \Pr\{|d\xi(t)| > \varepsilon\} \geq \varepsilon^2 \Pr\{\max |d\xi(t)| > \varepsilon\}.$$

Из того, что $\varepsilon \gg 0$, следует, что $\Pr\{\max |d\xi(t)| > \varepsilon\} \simeq 0$. Ввиду произвольности $\varepsilon \gg 0$ согласно теореме 7.1 это означает, что $\max |d\xi(t)| \simeq 0$ п. н. □

Теорема 14.2. *Пусть ξ — мартингал с $0 \ll b - a$, удовлетворяющий условию Линдеберга, такой, что дисперсия σ_ξ^2 суммируема на $T' \times \Omega$. Тогда п. н. ξ непрерывен на T.*

Доказательство вытекает из теорем 14.1 и 13.4. □

Для $\xi : T \to \mathbf{R}$ положим

$$q_\xi(t) = \sum_{s<t} d\xi(s)^2.$$

Будем называть q_ξ *квадратичной вариацией* процесса ξ. Если T — около-интервал, а ξ — гладкая функция, то квадратичная вариация бесконечно мала. Но для мартингала ξ справедливо равенство $\mathbf{E}q_\xi(t) = \|\xi(t) - \xi(a)\|_2^2$, так что вполне естественно, что он имеет не инфинитезимальную квадратичную вариацию.

Вспомним типичные траектории пуассоновского и винеровского блужданий, изображенные на рис. 12.1. Отметим, что для винеровского блуждания, в отличие от пуассоновского, квадратичная вариация не зависит от траектории. Чтобы показать, что этот феномен является результатом условия Линдеберга, мы сначала установим лемму о срезке. аналогичную теореме 10.3.

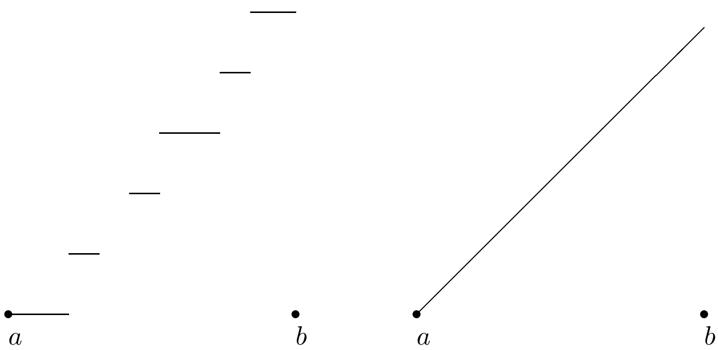
Теорема 14.3. *Пусть ξ — мартингал, удовлетворяющий условию Линдеберга. Пусть $\alpha > 0$ и ξ_α — мартингал, для которого $\xi_\alpha(a) = \xi(a)$, а*

$$d\xi_\alpha(t) = d\xi(t)^{(\alpha)} - \mathbf{E}_t d\xi(t)^{(\alpha)},$$

так что $|d\xi_\alpha(t)| \leq 2\alpha$ *при любых* α *и* ω. *Тогда существует такое бесконечно малое число* α, *что п. н.* $\xi_\alpha(t) \simeq \xi(t)$, $q_{\xi_\alpha}(t) \simeq q_\xi(t)$, *а* $\tau_{\xi_\alpha} \simeq \tau_\xi(t)$ *для всех* t.

Доказательство. По теореме (14.1) $\max|d\xi(t)| \simeq 0$, поэтому из теоремы 7.1 следует, что $\Pr\{\max|d\xi(t) \geq \alpha\} \simeq 0$ для любого достаточно большого инфинитезимального числа α. Тогда п. н.

$$\xi(t) = \xi(a) + \sum_{s<t} d\xi(s)^{(\alpha)}$$



Пуассоновское блуждание *Винеровское блуждание*

Рис. 14.1. Квадратичная вариация

Поэтому для доказательства того, что п. н. $\xi_\alpha(t) \simeq \xi(t)$ при любом t, достаточно показать, что

$$\sum \left|\mathbf{E}_t d\xi(t)^{(\alpha)}\right| \simeq 0 \text{ п. н.},$$

а для этого достаточно убедиться, что математическое ожидание рассматриваемой величины бесконечно мало. Но так как $\mathbf{E}_t d\xi(t) = 0$, имеем

$$\mathbf{E}\sum|\mathbf{E}_t d\xi(t)^{(\alpha)}| = \mathbf{E}\sum|\mathbf{E}_t(d\xi(t) - d\xi(t)^{(\alpha)})|$$

$$\leq \mathbf{E}\sum \mathbf{E}_t|d\xi(t) - d\xi(t)^{(\alpha)}| \leq \frac{1}{\alpha}\mathbf{E}\sum(d\xi(t)^2 - d\xi(t)^{(\alpha)2}).$$

Если $\alpha \gg 0$, это выражение бесконечно мало по условию Линдеберга. Поэтому множество всех α таких, что $\mathbf{E}\sum|\mathbf{E}_t d\xi(t)^{(\alpha)}| \leq \alpha$, содержит все $\alpha \gg 0$ и по принципу переполнения содержит все достаточно большие инфинитезимали α.

Такое же рассуждение дает, что $\sum(d\xi(t)^2 - d\xi(t)^{(\alpha)2}) \simeq 0$ п. н. Но $d\xi_\alpha(t)^2$ отличается от $d\xi(t)^{(\alpha)2}$ на величину

$$-2d\xi(t)^{(\alpha)}\mathbf{E}_t d\xi(t)^{(\alpha)} + (\mathbf{E}_t d\xi(t)^{(\alpha)})^2,$$

которая по абсолютной величине меньше, чем $3\alpha\left|\mathbf{E}_t d\xi(t)^{(\alpha)}\right|$. Следовательно, п. н. $q_{\xi_\alpha}(t) \simeq q_\xi(t)$ для любого t. Наконец, применяя то же рассуждение в случае, когда перед каждым членом стоит $\mathbf{E}_t$, получаем, что п. н. $\tau_{\xi_\alpha}(t) \simeq \tau_\xi(t)$ для любого t. □

Теорема 14.4. *Пусть мартингал ξ удовлетворяет условию Линдеберга, $\|\xi(b) - \xi(a)\|_2^2 \ll \infty$. Тогда п. н. $q_\xi(t) \simeq \tau_\xi(t)$ для любой точки t.*

Доказательство. Согласно предыдущей теореме мы можем предполагать, что для самогó мартингала ξ найдется такое число $\alpha \simeq 0$, что $|d\xi(t)| \le \alpha$ для любых t и ω. Отметим, что $q_\xi - \tau_\xi$ — мартингал с начальным значением 0. Имеем

$$\|q_\xi(b) - \tau_\xi(b)\|_1^2 \le \|q_\xi(b) - \tau_\xi(b)\|_2^2 = \sum \|dq_\xi(t) - d\tau_\xi(t)\|_2^2$$

$$= \sum \|d\xi(t)^2 - \mathbf{E}_t d\xi(t)^2\|_2^2 \le \sum \|d\xi(t)^2\|_2^2 = \sum \mathbf{E} d\xi(t)^4$$

$$\le \alpha^2 \sum \mathbf{E} d\xi(t)^2 = \alpha^2 \|\xi(b) - \xi(a)\|_2^2 \simeq 0.$$

По теореме 11.1 $\Pr\{\max|q_\xi(t) - \tau_\xi(t)| \ge \lambda\} \simeq 0$ для любого $\lambda \gg 0$, поэтому по теореме 7.1 п. н. $|q_\xi(t) - \tau_\xi(t)| \simeq 0$. □

Глава 15
Максимум мартингала

Вернемся к доказательству теоремы 11.1. Если ξ — субмартингал, то согласно (11.3)

$$0 \leq \mathbf{E}(\xi(b) - \xi(a) - \lambda)\chi_{\{\max(\xi(t) - \xi(a)) \geq \lambda\}}.$$

Мы можем релятивозировать этот результат для алгебры $\mathcal{P}_a$, заменив $\mathbf{E}$ на $\mathbf{E}_a$, а λ — на любой элемент из $\mathcal{P}_a$. В частности, если заменить λ на $\lambda - \xi(a)$, то получаем

$$0 \leq \mathbf{E}_a(\xi(b) - \lambda)\chi_{\{\max \xi(t) \geq \lambda\}}.$$

Взяв абсолютное математическое ожидание, имеем

$$\lambda \Pr\{\max \xi(t) \geq \lambda\} \leq \mathbf{E}\xi(b)\chi_{\{\max \xi(t) \geq \lambda\}} \leq \|\xi(b)\|_1. \tag{15.1}$$

Мы доказали следующую теорему.

Теорема 15.1. *Пусть ξ — субмартингал. Тогда выполнено* (15.1).

Приведем другое доказательство этой теоремы. Пусть

$$\Lambda = \{\max \xi(t) \geq \lambda\}$$

и

$$\Lambda_t = \{\xi(t) \geq \lambda \text{ и } \xi(s) < \lambda \text{ для любого } s < t\}.$$

Тогда Λ — объединение попарно не пересекающихся Λ_t, а $\chi_{\Lambda_t} \in \mathcal{P}_t$. Имеем

$$\|\xi(b)\|_1 \geq \mathbf{E}\xi(b)\chi_\Lambda = \mathbf{E}\sum \xi(b)\chi_{\Lambda_t} = \mathbf{E}\sum \mathbf{E}_t\xi(b)\chi_{\Lambda_t}$$

$$= \mathbf{E}\sum \chi_{\Lambda_t}\mathbf{E}_t\xi(b) \geq \mathbf{E}\sum \chi_{\Lambda_t}\xi(t) \geq \mathbf{E}\sum \chi_{\Lambda_t}\lambda = \lambda\,\mathrm{Pr}\,\Lambda,$$

что доказывает (15.1). □

Если ξ — мартингал, то по теореме 11.2 $|\xi|^p$ для $1 \leq p < \infty$ — субмартингал. Так как

$$\{\max|\xi(t)| \geq \lambda\} = \{\max|\xi(t)|^p \geq \lambda^p\}$$

при $\lambda > 0$, из теоремы 11.1 вытекает, что

$$\mathrm{Pr}\{\max|\xi(t)| \geq \lambda\} \leq \frac{1}{\lambda^p}\mathbf{E}|\xi(b)|^p. \tag{15.2}$$

Если бы $\max|\xi(t)|$ имел $\mathbf{L}^p$-норму, меньшую, чем $\xi(b)$, то (15.2) вытекало бы из неравенства Чебышёва. В общем случае это неверно, но мы имеем замечательный результат

$$\|\max|\xi(t)|\,\|_p \leq p'\|\xi(b)\|_p, \tag{15.3}$$

где $1 < p \leq \infty$, а p' — сопряженный показатель. Этот результат является следствием теоремы 15.1 и следующего утверждения.

Теорема 15.2. *Пусть x и y — такие положительные случайные величины, что при всех $\lambda > 0$*

$$\mathrm{Pr}\{y \geq \lambda\} \leq \frac{1}{\lambda}\mathbf{E}x\chi_{\{y\geq x\}}.$$

Тогда $\|y\|_p \leq p'\|x\|_p$ для любого p, $1 < p \leq \infty$.

Доказательство. При $1 < p < \infty$ имеем

$$\|y\|_p^p = \mathbf{E}y^p = \sum \lambda^p \mathrm{pr}_y(\lambda) = -\int_0^\infty \lambda^p\, d\,\mathrm{Pr}\{y \geq \lambda\}$$

$$= \int_0^\infty p\lambda^{p-1}\,\mathrm{Pr}\{y \geq \lambda\}d\lambda \leq \int_0^\infty p\lambda^{p-2}\mathbf{E}x\chi_{\{y\geq\lambda\}}d\lambda$$

$$= \mathbf{E}\left(x\int_0^y p\lambda^{p-2}d\lambda\right) = \frac{p}{p-1}\mathbf{E}xy^{p-1} \leq p'\|x\|_p\|y^{p-1}\|_{p'}$$

$$= p'\|x\|_p\|y^{p/p'}\|_{p'} = p'\|x\|_p\|y\|_p^{p-1}.$$

Случай $p = \infty$ доказывается переходом к пределу при $p \to \infty$. □

Глава 16

Закон больших чисел

По некоторым причинам специалисты по теории вероятностей пользуются термином «закон больших чисел» вместо более точного оборота «закон средних арифметических». Но проблема заключается в изучении поведения средних $y_n = (x_1 + \cdots + x_n)/n$ последовательности случайных величин.

Пусть $x_1, \ldots, x_\nu$ — независимые случайные величины с нулевыми математическими ожиданиями и единичными дисперсиями, $\mathcal{P}$ — алгебра, порожденная случайными величинами $x_1, \ldots, x_n$, и пусть $y_n = (x_1 + \cdots + x_n)/n$. Тогда

$$dy_n = \frac{x_1 + \cdots + x_n + x_{n+1}}{n+1} - \frac{x_1 + \cdots + x_n}{n} = -\frac{y_n}{n+1} + \frac{x_{n+1}}{n+1}. \tag{16.1}$$

Поскольку случайные величины x_n независимы, получаем, что $\mathbf{E}_n x_{n+1} = \mathbf{E} x_{n+1} = 0$. Следовательно,

$$Dy_n = -\frac{y_n}{n+1}, \quad d\widehat{y}_n = \frac{x_{n+1}}{n+1}.$$

Для $n \simeq \infty$ имеем $\|y_n\|_2^2 = 1/n \simeq 0$, откуда по неравенству Чебышёва следует, что $y_n \simeq 0$ п. н. Значит, последовательность $y_1, \ldots, y_\nu$ сходится к 0 по вероятности. Это — слабый закон больших чисел. Мы можем также доказать усиленный закон больших чисел: последовательность случайных величин $y_1, \ldots, y_\nu$ сходится к 0 п. н. Чтобы убедиться в этом, заметим, что

$$\|Dy_n\|_1 \le \|Dy_n\|_2 = \frac{1}{n+1} \frac{1}{\sqrt{n}},$$

так что сумма $\sum \|Dy_n\|_1$ сходится. По теореме 10.1 (ii) предсказуемый процесс, ассоциированный с y_n, сходится п. н. Сходится также сумма

$$\sum \|d\widehat{y}_n\|_2^2 = \sum \frac{1}{(n+1)^2},$$

так что по теореме 11.3 (i) мартингал, ассоциированный с y_n, сходится п. н. Таким образом, последовательность y_n сходится п. н., и согласно слабому закону больших чисел она сходится к 0 п. н. Отсюда вытекает усиленный закон больших чисел для случая, обсуждавшегося в начале гл. 4, однако мы докажем несколько более сильных результатов. Отметим, что на самом деле в доказательстве использовалась не независимость, а только то, что случайные величины x_n суть приращения мартингала.

Теорема 16.1. *Пусть $x_1, \ldots, x_\nu$ — приращения мартингала.*

(i) *Если сумма $\sum\limits_{i=1}^{\nu} \frac{\mathbf{E}x_i^2}{i^2}$ сходится, то последовательность $(x_1 + \cdots + x_n)/n$ сходится к 0 п. н.*

(ii) *Если сумма $\sum\limits_{i=1}^{\nu} \frac{\mathbf{E}x_i^2}{i^2} \ll \infty$, то п. н. последовательность $(x_1 + \cdots + x_n)/n$ имеет доступное колебание.*

Доказательство. Пусть

$$y_n = \frac{i}{n}\sum_{i=1}^{n} x_i, \quad z_n = \sum_{i=1}^{n} \frac{x_i}{i}.$$

Тогда

$$y_n = -\frac{1}{n}\sum_{j=1}^{n} z_j + \frac{n+1}{n} z_n, \tag{16.2}$$

что можно увидеть, собирая коэффициенты при x_i в обеих частях. Таким образом, каждый из z_n — мартингал, причем $z_0 = 0$ и

$$\|z_\nu\|_2^2 = \sum_{i=1}^{\nu} \frac{\mathbf{E}x_i^2}{i^2} \ll \infty.$$

Поэтому из теоремы 11.1 следует, что п. н. z_n доступны при всех n. В случае (i) из теоремы 11.3(i) вытекает, что z_n сходится п. н., а в случае

(ii) из теоремы 12.3 — что z_n п. н. имеет доступное колебание. Поэтому доказательство будет завершено, если мы установим следующие два утверждения.

(I) *Если последовательность случайных величин z_n доступна и сходится, то последовательность случайных величин y_n сходится к* 0.

(II) *Если последовательность случайных величин z_n доступна и имеет доступное колебание, то последовательность случайных величин y_n имеет доступное колебание.*

Чтобы доказать (I), положим $n \simeq \infty$. Так как случайная величина z_n доступна, имеем
$$\frac{n+1}{n} z_n \simeq z_n \simeq z_\nu.$$
Пусть $\varepsilon \gg 0$ и k — наибольшее натуральное число $\leq \nu$, для которого $|z_k - z_\nu| > \varepsilon$ (если такого числа нет, то полагаем $k = 0$). Тогда $k \ll \infty$, каждое $|z_j| \ll \infty$, но $n \simeq \infty$, и поэтому
$$\frac{1}{n} \sum_{j=1}^{k} z_j \simeq 0.$$
Тем самым
$$\left| \frac{1}{n} \sum_{j=1}^{n} z_j - z_\nu \right| < \varepsilon.$$
Так как $\varepsilon \gg 0$ произвольно, это означает, что правая часть выражения (16.2) бесконечно близка к $-z_\nu + z_\nu = 0$, что и доказывает (I).

Чтобы доказать (II), заметим, что $z_n/n \simeq 0$, если $n \simeq \infty$, так что последовательность z_n/n сходится к 0 и потому имеет доступное колебание. По неравенству треугольника сумма двух последовательностей с доступным колебанием обладает доступным колебанием. Поэтому последовательность
$$\frac{n+1}{n} z_n = z_n + \frac{1}{n} z_n$$
имеет доступное колебание. Пусть
$$w_n = \frac{1}{n} \sum_{j=1}^{n} z_j.$$

Остается показать, что последовательность w_n имеет доступное колебание. Но, как и в доказательстве теоремы 12.3, видно, что если последовательность w_n имеет недоступное колебание, то найдутся $\lambda_1 \ll \lambda_2$ такие, что w_n имеет недоступное число выбросов вверх из интервала $[\lambda_1, \lambda_2]$. Определим $s_1, t_1, s_2, t_2, \ldots$ для последовательности w_n, как в гл. 12 (см. рис. 12.3).

Для того чтобы w_n вышло вверх за $[\lambda_1, \lambda_2]$ между точками s_1 и t_1, должно существовать z_{j_1} с j_1 между s_1 и t_1 такое, что $z_{j_1} \geq \lambda_2$. Аналогично, для того чтобы был выброс вверх снова в более поздний момент, должно найтись z_{k_1} с k_1 между t_1 и s_2 такое, что $z_{k_1} \leq \lambda_1$, и т. д. Таким образом, z_n должен иметь по крайней мере $\beta - 1$ выбросов вверх за интервал $[\lambda_1, \lambda_2]$, если w_n имеет β таких выбросов за $[\lambda_1, \lambda_2]$. Но это невозможно для $\beta \simeq \infty$, потому что последовательность z_n имеет доступное колебание. Это доказывает (II). □

Приведем пример, к которому примени́м случай (ii), но не применим случай (i). Пусть $x_n = 0$, если $n < \nu$, $x_\nu = \nu$ с вероятностью $\frac{1}{2}$ и $x_\nu = -\nu$ с вероятностью $\frac{1}{2}$, так что $\sum \mathbf{E} x_i^2 / i^2 = 1$. Тогда средние y_n равны 0 для $n < \nu$, но $y_\nu = \pm 1$, каждое с вероятностью $\frac{1}{2}$.

Таким образом, в случае (ii) y_n с недоступным номером n — не обязательно инфинитезималь. Тем не менее, y_n с доступным n становится и остается меньше любого $\varepsilon \gg 0$. Чтобы показать это, заметим, что если

$$\sum_{i=1}^{\nu} \frac{\mathbf{E} x_i^2}{i^2} \ll \infty,$$

то по теореме 6.1 найдется недоступное число μ, $\mu \leq \nu$, такое, что

$$\sum_{i=1}^{\mu} \frac{\mathbf{E} x_i^2}{i^2} \text{ сходится.}$$

Поэтому случай (i) относится и к последовательности $x_1, \ldots, x_\mu$.

Теорема 16.1 обобщается на другие показатели p, но доказательство использует прием срезки, который сводит ситуацию к случаю $p = 2$.

Теорема 16.2. *Пусть $x_1, \ldots, x_\nu$ — приращения мартингала и $0 \ll p \leq 2$.*

(i) *Если сумма $\sum\limits_{i=1}^{\nu} \mathbf{E}|x_i|^p / i^p$ сходится, то $(x_1 + \cdots + x_n)/n$ сходится к 0 п. н.*

(ii)*Если сумма* $\sum\limits_{i=1}^{\nu} \mathbf{E}|x_i|^p/i^p \ll \infty$, *то* $(x_1+\dots+x_n)/n$ *п. н. имеет доступное колебание.*

Доказательство. В обоих случаях я утверждаю, что п. н. найдется $c \ll \infty$ такое, что $|x_n| \leq cn$ при всех n. Для доказательства возьмем произвольное $\varepsilon \gg 0$, положим $a = \sum \mathbf{E}|x_i|^p/i^p$ и возьмем $c = (a/\varepsilon)^{1/p}$. Тогда $c \ll \infty$, так как $a \ll \infty$ и $p \gg 0$. По неравенству Чебышёва

$$\Pr\{|x_n| > cn \text{ для некоторого } n\} \leq \sum \Pr\{|x_n| > cn\}$$

$$\leq \sum \frac{\mathbf{E}|x_n|^p}{(cn)^p} = \frac{a}{c^p} = \varepsilon.$$

Поскольку число $\varepsilon \gg 0$ произвольно, последнее доказывает утверждение. □

Будем говорить, что последовательность случайных величин *ведет себя правильно*, если она сходится к нулю п. н. в случае (i), и имеет доступное колебание п. н. в случае (ii). Как мы только что показали, для доказательства этого достаточно проверить, что при $c \ll \infty$ последовательность $(x_1^{(c1)} + \dots + x_n^{(cn)})/n$ ведет себя правильно. Теперь, вообще говоря, $x_n^{(cn)}$ не являются приращениями мартингала; однако $x_n^{(cn)} - \mathbf{E}_{n-1}x_n^{(cn)}$ таковы. Имеем

$$\sum \left\| \mathbf{E}_{n-1}\frac{x_n^{(cn)}}{n} \right\|_1 = \sum \frac{1}{n}\big\|\mathbf{E}_{n-1}\big(x_n - x_n^{(cn)}\big)\big\|_1$$

$$= \sum \frac{1}{n} \sum_{|\lambda|>cn} |\lambda| \mathrm{pr}_{x_n}(\lambda) \leq \sum \sum_{|\lambda|>cn} \frac{|\lambda|^p}{c^{p-1}n^p} \mathrm{pr}_{x_n}(\lambda)$$

$$\leq \frac{1}{c^{p-1}} \sum \frac{\mathbf{E}|x_n|^p}{n^p}.$$

По теореме 10.2

$$\sum \frac{\mathbf{E}_{n-1}x_n^{(cn)}}{n}$$

сходится п. н. в случае (i), п. н. имеет доступное колебание в случае (ii) и в обоих случаях п. н. все члены доступны. Согласно утверждениям

(I) и (II) из доказательства теоремы 16.1 это означает, что последовательность

$$\big(\mathbf{E}_0 x_1^{(c1)} + \dots + \mathbf{E}_{n-1} x_n^{(cn)}\big)/n$$

ведет себя правильно, так что осталось доказать, что последовательность, образованная из средних арифметических величин

$$x_n^{(cn)} - \mathbf{E}_{n-1} x_n^{(cn)},$$

ведет себя правильно. Для этого нам нужно только проверить предположения теоремы 16.1:

$$\begin{aligned}\sum \frac{\mathbf{E}\big(x_n^{(cn)} - \mathbf{E}_{n-1} x_n^{(cn)}\big)^2}{n^2} &\le \sum \frac{\mathbf{E} x_n^{(cn)2}}{n^2} \\ &= \sum \frac{1}{n^2} \sum_{|\lambda| \le cn} |\lambda|^2 \mathrm{pr}_{x_n}(\lambda),\end{aligned}$$

а поскольку $p \le 2$, последнее оценивается далее как

$$\le \sum \frac{1}{n^2} \sum_{|\lambda| \le cn} \frac{|\lambda|^p}{(cn)^{p-2}} \mathrm{pr}_{x_n}(\lambda) = \frac{1}{c^{p-2}} \sum \frac{1}{n^p} \sum_{|\lambda| \le cn} |\lambda|^p \mathrm{pr}_{x_n}(\lambda)$$

$$\le \frac{1}{c^{p-2}} \sum \frac{\mathbf{E}|x_n|^p}{n^p}. \ \Box$$

Следствие. *Пусть $x_1, \dots, x_\nu$ — приращения мартингала и $p \gg 1$. Если каждая из норм $\|x_n\|_p$ доступна, то последовательность $(x_1 + \dots + x_n)/n$ сходится к 0 п. н.*

Доказательство. Не теряя общности, предположим, что $p \le 2$. Пусть $M = \max \|x_n\|_p$. Тогда $M \ll \infty$, так как $M = \|x_n\|_p$ для некоторого номера n. Поэтому условие (i) теоремы 16.2 выполнено. $\Box$

Отметим, что теорема 16.2 тривиальна при $p = 1$: утверждение следует из теоремы 10.2, (I) и (II), даже без предположения, что x_n — приращения мартингала.

Однако, как показывает следующий контрпример, следствие неверно при $p = 1$. Пусть случайные величины x_n, где $n = 3, 4, \dots, \nu$, независимы и удовлетворяют условию

$$x_n = \begin{cases} n & \text{с вероятностью } \frac{1}{2}\frac{1}{n \log n}, \\ 0 & \text{с вероятностью } \frac{1}{2} - \frac{1}{2}\frac{1}{n \log n}, \\ -\frac{1}{\log n} & \text{с вероятностью } \frac{1}{2}. \end{cases}$$

Тогда $\mathbf{E}x_n = 0$ и $\|x_n\|_1 = 1/\log n$. Отметим, что каждая случайная величина x_n суммируема и, более того, норма $\|x_n\|_1$ бесконечно мала при $n \simeq \infty$. Но по кардинальной версии теоремы Бореля — Кантелли (теорема 7.4 (ii)) $x_n = n$ п. н. для недоступного числа номеров n. Поэтому п. н. последовательность y_n удовлетворяет неравенству

$$y_n = \frac{x_3 + \cdots + x_n}{n-2} \geq 1 - \frac{1}{\log 3}$$

для недоступного числа номеров n, и значит, не сходится к 0. На самом деле, используя (16.1), можно легко показать, что п. н. последовательность y_n имеет недоступное колебание.

Следующая теорема показывает, что существует важный случай повторных независимых наблюдений случайной величины, в котором усиленный закон больших чисел выполняется в предположении суммируемости.

Теорема 16.3. *Пусть x — случайная величина и $x_1, \dots, x_\nu$ — независимые случайные величины, которые имеют такие же распределения вероятностей, как и x.*

(i) *Если случайная величина x суммируема, то последовательность $(x_1 + \cdots + x_n)/n$ сходится к $\mathbf{E}x$ п. н.*

(ii) *Если $\|x\|_1 \ll \infty$, то п. н. последовательность $(x_1+\cdots+x_n)/n$ имеет доступное колебание.*

Доказательство. Стратегия доказательства та же, что и в теореме 16.2, но тактика несколько отличается. Не теряя общности, предположим, что $\mathbf{E}x = 0$.

Я утверждаю, что п. н. найдется число $c \ll \infty$ такое, что $|x_n| \leq cn$ для всех n. Действительно, для $\varepsilon \gg 0$ возьмем $c = \mathbf{E}|x|/\varepsilon$ и

$$E_k = \{ck < |x| \leq c(k+1)\}.$$

Заметим, что

$$\Pr\{|x| > cn \text{ для некоторого } n\} \leq \sum_{n=1}^{\nu} \Pr\{|x_n| > cn\}$$

$$= \sum_{n=1}^{\nu} \Pr\{|x| > cn\} = \sum_{n=1}^{\nu} \sum_{k=n}^{\infty} \Pr E_k$$

$$= \sum_{k=0}^{\infty} \sum_{n=1}^{\min\{\nu,k\}} \Pr E_k \le \sum_{k=0}^{\infty} k \Pr E_k \le \frac{1}{c} \sum_{\lambda} \mathrm{pr}_x(\lambda) = \frac{1}{c} \mathbf{E}|x| = \varepsilon.$$

Это и доказывает мое утверждение.

Значит, достаточно показать, что при $c \ll \infty$ средние случайных величин $x_n^{(cn)}$ ведут себя правильно. Итак, $\mathbf{E}x_n^{(cn)} = \mathbf{E}x^{(cn)} = \mathbf{E}\left(x - x^{(cn)}\right)$, а это по абсолютной величине меньше чем

$$\sum_{|\lambda|>cn} |\lambda| \mathrm{pr}_x(\lambda).$$

Аналогично, для $n < m$ имеем

$$\left|\mathbf{E}x_n^{(cn)} - \mathbf{E}x_m^{(cm)}\right| = \left|\mathbf{E}x^{(cn)} - \mathbf{E}x^{(cm)}\right| \le \sum_{cn<|\lambda|\le cm} |\lambda| \mathrm{pr}_x(\lambda).$$

Таким образом, математическое ожидание $\mathbf{E}x_n^{(cn)}$ всегда доступно и ведет себя правильно. Отсюда вытекает, как и в доказательстве (I) и (II), что средние величин $\mathbf{E}x_n^{(cn)}$ ведут себя правильно. (В этом рассуждении мы использовали независимость — было существенно, что мы имели дело с абсолютными математическими ожиданиями, а не с условными математическими ожиданиями.) Следовательно, нам нужно только показать, что случайная величина $x_n^{(cn)} - \mathbf{E}x_n^{(cn)}$ удовлетворяет условиям теоремы 16.1. Но

$$\sum_{n=m}^{\nu} \frac{\mathbf{E}\big(x_n^{(cn)} - \mathbf{E}x_n^{(cn)}\big)^2}{n^2} \le \sum_{n=m}^{\nu} \frac{\mathbf{E}x_n^{(cn)2}}{n^2}$$

$$= \sum_{n=m}^{\nu} \frac{1}{n^2} \sum_{|\lambda|\le cn} \lambda^2 \mathrm{pr}_x(\lambda) = \sum_{|\lambda|} \lambda^2 \mathrm{pr}_x(\lambda) \sum_{n\ge\max\{m,|\lambda|/c\}} \frac{1}{n^2}$$

$$\le \sum_{|\lambda|\le\sqrt{m}} \lambda^2 \mathrm{pr}_x(\lambda) \sum_{n=m}^{\nu} \frac{1}{n^2} + \sum_{|\lambda|>\sqrt{m}} \lambda^2 \mathrm{pr}_x(\lambda) \sum_{n=|\lambda|/c}^{\nu} \frac{1}{n^2}$$

$$\le \sum_{|\lambda|\le\sqrt{m}} \lambda^2 \frac{1}{m-1} + \sum_{|\lambda|>\sqrt{m}} \lambda^2 \mathrm{pr}_x(\lambda) \frac{1}{(|\lambda|/c)-1}$$

$$\leq \frac{\sqrt{m}}{m-1} \sum_{|\lambda| \leq \sqrt{m}} |\lambda| \mathrm{pr}_x(\lambda) + \sum_{|\lambda| > \sqrt{m}} 2c|\lambda|$$

при $\sqrt{m}/c \geq 2$. В случае (i) это выражение бесконечно мало для $m \simeq \infty$ (таким образом, сумма сходится), а в случае (ii) оно доступно. □

Пусть $x_1, \ldots, x_\nu$ и x — случайные величины, определенные не обязательно на одном и том же конечном вероятностном пространстве. Говорят, что последовательность случайных величин $x_1, \ldots, x_\nu$ *мажорируется по распределению* случайной величиной x, если найдется такое число $a \ll \infty$, что для любого $n = 1, \ldots, \nu$ и всех натуральных чисел k выполнено неравенство

$$\Pr\{k < |x_n| \leq k+1\} \leq a \Pr\{k < |x| \leq k+1\}.$$

Так, конечно, бывает, когда все $x_1, \ldots, x_\nu$ имеют то же самое распределение вероятностей, что и x. В контрпримере, приведенном ранее в этой главе, последовательность $x_3, x_4, \ldots, x_\nu$ не мажорируется по распределению никакой случайной величиной.

Пусть $x_1, \ldots, x_\nu$ — независимые случайные величины с нулевыми математическими ожиданиями, мажорируемые по распределению случайной величиной x. Если проследить доказательство теоремы 16.3, то можно обнаружить, что все переносится на этот более общий случай, за исключением оценки величины $\left|\mathbf{E}x_n^{(cn)} - \mathbf{E}x_m^{(cm)}\right|$, которая была использована в случае (ii). Таким образом, мы получаем следующий результат.

Теорема 16.4. *Пусть x — случайная величина и $x_1, \ldots, x_\nu$ — независимые случайные величины с нулевыми математическими ожиданиями, мажорируемые по распределению случайной величиной x. Если случайная величина x суммируема, то $(x_1 + \cdots + x_n)/n$ сходится к 0 п. н.*

Ниже приведен контрпример к случаю (ii). Пусть $\nu \simeq \infty$, μ таково, что $\mu/\nu \simeq \infty$, и x равна 1 с вероятностью $1 - 1/\mu$ и $-\mu + 1$ с вероятностью $1/\mu$. Тогда случайная величина x не суммируема, но $\|x\|_1 = 2 - 2/\mu \ll \infty$, а $\mathbf{E}x = 0$. Пусть $x'_1, \ldots, x'_\nu$ — независимые наблюдения случайной величины x. Тогда п. н. $x'_n = 1$ для всех $n \leq \nu$. Пусть $x_1, \ldots, x_\nu$ — те же самые случайные величины, что и $x'_1, \ldots, x'_\nu$, за исключением того, что случайная величина x'_1 заменена на 0, следующие 2! случайных величин x'_n не изменены, следующие 3! случайных величин заменены на 0, следующие 4! случайных величин x'_n не изменены, и так далее до ν. Тогда $x_1, \ldots, x_\nu$ — независимые случайные величины с нулевыми математическими ожиданиями, которые мажорируются по распределению случайной величиной x, причем $\|x\|_1 \ll \infty$, однако п. н. последовательность случайных величин $(x_1 + \cdots + x_n)/n$ имеет недоступное колебание.

Смысл усиленного закона больших чисел в том, что, в конце концов, средние $(x_1 + \cdots + x_n)/n$ повторных независимых наблюдений случайной величины x располагаются вблизи $\mathbf{E}x$. Это верно по теореме 16.3(i), если x суммируема. Если мы предположим только, что $\|x\|_1 \ll \infty$, то утверждение теоремы 16.3(ii) становится весьма неинформативным: оно просто говорит нам, что средние могут значительно колебаться только доступное число раз. Тем не менее, если $\|x\|_1 \ll \infty$, то по теореме 6.1 (примененной к последовательности $\mathbf{E}\big|x_n^{(a)}\big|$) найдется такое недоступное число a, что $x^{(a)}$ суммируема. Если мы выберем $\mu \leq \nu$ недоступным, но достаточно малым, то можем обеспечить, что п. н. $x_n = x_n^{(a)}$ для всех $n \simeq \mu$, так что по теореме 16.3(i) последовательность случайных величин $(x_1 + \cdots + x_n)/n$, $n \leq \mu$, п. н. сходится к $\mathbf{E}x^{(a)}$.

Будем говорить, что случайная величина x *около-суммируема*, если существует такая суммируемая случайная величина y, что $x \simeq y$ п. в.; в этом случае мы называем $\mathbf{E}y$ *приведенным математическим ожиданием* случайной величины x. По теореме Лебега, если z также суммируемая случайная величина, причем $z \simeq x$, то $\mathbf{E}z \simeq \mathbf{E}y$. Таким образом, приведенное математическое ожидание около-суммируемой случайной величины доступно и однозначно определяется с точностью до бесконечно малой.

***Теорема 16.5.** *Пусть x — случайная величина. Тогда следующие условия эквивалентны:*

(i) $x^{(a)}$ *около-суммируема*;

(ii) *для некоторого $a \simeq \infty$ верны соотношения* $\Pr\{|x| > a\} \simeq 0$ *и* $\mathbf{E}|x^{(a)}| \ll \infty$;

(iii) *для некоторого $a \simeq \infty$ имеет место соотношение* $\Pr\{|x| > a\} \simeq 0$ *и величина $x^{(a)}$ суммируема.*

Доказательство. Как уже отмечалось, из теоремы 6.1 следует, что (ii) $\Rightarrow$ (iii). Импликация (iii) $\Rightarrow$ (i) очевидна. Предположим, что выполнено условие (i). Тогда ясно, что $\Pr\{|x| > a\} \simeq 0$ для любого $a \simeq \infty$. Пусть величина y суммируема и такова, что $x \simeq y$ п. в., и K — такое число, что $\mathbf{E}|y| \ll K \ll \infty$. Предположим, что $\mathbf{E}|x^{(a)}| \simeq \infty$ для каждого $a \simeq \infty$. Тогда множество всех a, для которых $\mathbf{E}|x^{(a)}| \geq K$, содержит все $a \simeq \infty$, и поэтому содержит некоторое $a \ll \infty$. Но $|x^{(a)}| \leq |x| \simeq |y|$ п. в., а так как ясно, что величина $|x^{(a)}|$ суммируема, из теоремы Лебега следует, что $\mathbf{E}|x^{(a)}| \lesssim \mathbf{E}|y| \leq K$. Получили противоречие. $\square$

Имеют место импликации

$$x \text{ суммируема} \Rightarrow \|x\|_1 \ll \infty \Rightarrow x \text{ около-суммируема},$$

но ни одна из обратных импликаций, вообще говоря, не имеет места.

Следующая теорема является следствием теорем 16.3(i) и 16.5.

****Теорема 16.6.** *Пусть случайная величина x около-суммируема и $x_1, \ldots, x_\nu$, $\nu \simeq \infty$, — независимые случайные величины, имеющие те же распределения вероятностей, что и x. Тогда найдется недоступное число $\mu \leq \nu$ такое, что последовательность $(x_1 + \cdots + x_n)/n$, $n = 1, \ldots, \mu$, сходится к приведенному математическому ожиданию случайной величины x.*

Глава 17
Около-эквивалентные процессы

В гл. 3 мы определили внутреннее понятие эквивалентности двух стохастических процессов. Теперь определим внешнее понятие — около-эквивалентность. Интуитивное содержание этого понятия заключается в том, что два стохастических процесса около-эквивалентны, если их нельзя различить в наблюдениях, не способных различить инфинитезималей.

Напомним, что траектории стохастического процесса ξ, параметризованного конечным множеством T и определенного на конечном вероятностном пространстве $\langle \Omega, \mathrm{pr} \rangle$, являются элементами конечного подмножества $\Lambda_\xi \subseteq \mathbf{R}^T$. Если $\Lambda \subseteq \mathbf{R}^T$ — конечное подмножество, а $F : \Lambda \to \mathbf{R}$, то мы называем F *функционалом*, определенным на Λ. Если F — функционал, определенный на Λ, а $\Lambda_\xi \subseteq \Lambda$, то пишем $F(\xi)$ для случайной величины, значения которой в каждой точке $\omega \in \Omega$ есть $F(\xi(\omega))$. Пространство Λ является метрическим пространством относительно метрики

$$\rho(\lambda, \mu) = \max_{t \in T} |\lambda(t) - \mu(t)|,$$

а функционал (около-)*непрерывен*, когда из равенства $\rho(\lambda, \mu) \simeq 0$ (эквивалентного тому, что $\lambda(t) \simeq \mu(t)$ для всех t) следует $F(\lambda) \simeq F(\mu)$. Будем говорить, что функционал F *доступен*, если $|F(\lambda)| \ll \infty$ для каждого $\lambda \in \Lambda$. Так как функционал F достигает своего максимума, это то же самое, что $\max |F| \ll \infty$.

Пусть ξ и η — два стохастических процесса, параметризованных элементами одного и того же конечного множества T, но определенных, возможно, на различных конечных вероятностных пространствах. Отметим, что процессы ξ и η эквивалентны в том и только в том случае, если $\mathbf{E}F(\xi) = \mathbf{E}F(\eta)$ для всех функционалов F (определенных на $\Lambda_\xi \cup \Lambda_\eta$; но тогда мы должны иметь $\Lambda_\xi = \Lambda_\eta$). Будем говорить, что процессы ξ и η *около-эквивалентны*, если $\mathbf{E}F(\xi) \simeq \mathbf{E}F(\eta)$ для всех доступных непрерывных функционалов F на $\Lambda_\xi \cup \Lambda_\eta$.

Теорема 17.1. *Пусть ξ — стохастический процесс, определенный на $\langle \Omega, \mathrm{pr} \rangle$. Предположим также, что Ω — конечное вероятностное пространство относительно pr' и $\sum |\mathrm{pr}(\omega) - \mathrm{pr}'(\omega)| \simeq 0$. Пусть ξ' — та же самая функция, что и ξ, но рассматриваемая как стохастический процесс на $\langle \Omega, \mathrm{pr}' \rangle$. Тогда процессы ξ и ξ' около-эквивалентны.*

Доказательство. Пусть F — доступный функционал. Тогда

$$|\mathbf{E}F(\xi) - \mathbf{E}'F(\xi')| \le \sum |F(\xi(\omega))| \, |\mathrm{pr}(\omega) - \mathrm{pr}'(\omega)|$$

$$\le \max |F| \sum |\mathrm{pr}(\omega) - \mathrm{pr}'(\omega)| \simeq 0. \ \square$$

Теорема 17.2. *Пусть ξ и η — стохастические процессы, параметризованные элементами множества T и определенные на одном и том же вероятностном пространстве. Если п. н. $\xi(t) \simeq \eta(t)$ для всех t, то процессы ξ и η около-эквивалентны.*

Доказательство. Если F — доступный функционал, то $F(\xi)$ и $F(\eta)$ суммируемы, а если F непрерывен, то $F(\xi) \simeq F(\eta)$ п. н., так что по теореме Лебега $\mathbf{E}F(\xi) \simeq \mathbf{E}F(\eta)$. $\square$

В следующей теореме пусть $\mathbf{A}$ — любая формула, внутренняя или внешняя. Эта теорема показывает, что наше определение около-эквивалентности имеет желаемое интуитивное содержание.

Теорема 17.3. *Предположим, что для любых λ и μ из того, что $\rho(\lambda, \mu) \simeq 0$, следует, что формула $\mathbf{A}(\lambda)$ выполняется в том и только в том случае, если выполняется формула $\mathbf{A}(\mu)$. Пусть ξ и η — около-эквивалентные стохастические процессы, а $\Lambda_\xi \cup \Lambda_\eta \subseteq \Lambda$. Тогда формула $\mathbf{A}(\xi)$ выполняется п. н. тогда и только тогда, когда п. н. выполняется формула $\mathbf{A}(\eta)$.*

Доказательство. Не теряя общности, предположим, что процессы ξ и η определены на Λ_ξ и Λ_η. Предположим, что $\mathbf{A}(\xi)$ выполняется п. н. Нам нужно доказать, что п. н. выполняется $\mathbf{A}(\eta)$.

Пусть $\varepsilon \gg 0$. Тогда существует множество $\Phi \subseteq \Lambda_\xi$ такое, что $\Pr_\xi \Phi \ge 1 - \varepsilon$ и формула $\mathbf{A}(\lambda)$ выполняется для всех $\lambda \in \Phi$. Для $\beta > 0$ пусть

$$F_\beta(\lambda) = \left(1 - \frac{\rho(\lambda, \Phi)}{\beta}\right)^+.$$

Тогда $0 \le F_\beta(\lambda) \le 1$, поэтому функционал F_β доступен. Если $\beta \gg 0$, то функционал F_β непрерывен. Пусть $\Phi_\beta = \{\lambda : \rho(\lambda, \Phi) \le \beta\}$. Заметим, что

$$\chi_{\Phi_\beta} \ge F_\beta \ge \chi_\Phi.$$

Следовательно, при $\beta \gg 0$

$$\mathrm{Pr}_\eta \Phi_\beta \ge \mathbf{E} F_\beta(\eta) \simeq \mathbf{E} F_\beta(\xi) \ge \mathrm{Pr}_\xi \Phi \ge 1 - \varepsilon,$$

так что $\mathrm{Pr}_\eta \Phi_\beta \ge 1 - 2\varepsilon$. Так как это неравенство верно при всех $\beta \gg 0$, по принципу переполнения оно верно и для некоторого $\beta \simeq 0$. Но для любого $\beta \simeq 0$ имеем $\mathbf{A}(\lambda)$ для всех $\lambda \in \Phi_\beta$. Так как $\varepsilon \gg 0$ произвольно, формула $\mathbf{A}(\eta)$ выполняется п. н. $\square$

Следствие 1. *Пусть ξ и η — около-эквивалентные стохастические процессы и F — функционал. Если F непрерывен на почти каждой траектории процесса ξ, то F непрерывен и на почти каждой траектории процесса η.*

Следствие 2. *Пусть ξ и η — около-эквивалентные стохастические процессы, параметризованные элементами конечного подмножества $T \subseteq \mathbf{R}$. Если почти каждая траектория процесса ξ непрерывна, то и почти каждая траектория процесса η непрерывна.*

Теорема 17.4. *Пусть ξ и η — около-эквивалентные стохастические процессы и F — доступный функционал, непрерывный почти на каждой траектории процесса ξ. Тогда $\mathbf{E}F(\xi) \simeq \mathbf{E}F(\eta)$.*

Доказательство. Без потери общности можно считать, что процессы ξ и η определены на Λ_ξ и Λ_η и что $1 \le F \le \varepsilon \ll \infty$. Пусть $\varepsilon \gg 0$. Согласно следствию 1 найдется подмножество $\Phi \subseteq \Lambda_\xi \cup \Lambda_\eta$ такое, что $\mathrm{Pr}_\xi(\Phi \cap \Lambda_\xi) \ge 1 - \varepsilon$ и $\mathrm{Pr}_\eta(\Phi \cap \Lambda_\eta) \ge 1 - \varepsilon$, а функционал F непрерывен на каждом элементе из Φ. Определим теперь функционал G следующим образом: если $\lambda \in \Phi$, то $G(\lambda) = F(\lambda)$, а если $\lambda \notin \Phi$, то

$$G(\lambda) = \frac{\min_{\mu \in \Phi} \rho(\lambda, \mu) F(\mu)}{\rho(\lambda, \Phi)}.$$

Для любого λ из $\Lambda = \Lambda_\xi \cup \Lambda_\eta$ пусть λ^* — такой элемент из Φ, что $\rho(\lambda, \Phi) = \rho(\lambda, \lambda^*)$, и пусть λ' будет тем элементом $\mu \in \Phi$, на котором $\rho(\lambda, \mu) F(\mu)$ достигает своего минимума на Φ. Тогда $1 \le G \le \varepsilon$. Если $\lambda \in \Phi$, это очевидно, а если $\lambda \notin \Phi$, то имеем

$$1 \le \frac{\rho(\lambda, \lambda') F(\lambda')}{\rho(\lambda, \Phi)} = G(\lambda) \le \frac{\rho(\lambda, \lambda^*) F(\lambda^*)}{\rho(\lambda, \Phi)} = F(\lambda^*) \le c. \qquad (17.1)$$

Я утверждаю, что функционал G непрерывен всюду. Чтобы убедиться в этом, предположим, что $\rho(\lambda_1, \lambda_2) \simeq 0$. Если $\lambda_1, \lambda_2 \in \Phi$, то $G(\lambda_1) - G(\lambda_2) = F(\lambda_1) - F(\lambda_2) \simeq 0$. Если $\lambda \notin \Phi$, но $\rho(\lambda_1, \Phi) \simeq 0$, то $\rho(\lambda_1, \lambda_1') \simeq 0$ по определению λ_1', так как $1 \leq F$. Поэтому $\rho(\lambda_1', \lambda_1^*) \simeq 0$, а также $F(\lambda_1') \simeq F(\lambda_1^*)$. Согласно (17.1)

$$1 \leq \frac{\rho(\lambda_1, \lambda_1')}{\rho(\lambda, \Phi)} \simeq 1,$$

а также $G(\lambda_1) \simeq F(\lambda_1^*)$. Поскольку $G(\lambda_2) \simeq F(\lambda_2^*)$, имеем $G(\lambda_1) \simeq G(\lambda_2)$. Наконец, если $\rho(\lambda_1, \Phi) \gg 0$, то $G(\lambda_1) \simeq G(\lambda_2)$, что доказывает мое утверждение.

Следовательно, по определению около-эквивалентности, $\mathbf{E}G(\xi) = \mathbf{E}G(\eta)$. Но $\mathbf{E}|F(\xi) - G(\xi)| \leq c\varepsilon$ и $\mathbf{E}|F(\eta) - G(\eta)| \leq c\varepsilon$, а так как $\varepsilon \gg 0$ произвольно, то $\mathbf{E}F(\xi) \simeq \mathbf{E}F(\eta)$. $\square$

Глава 18

Теорема Муавра — Лапласа — Линдеберга — Феллера — Винера — Ле́ви — Дуба — Эрдёша — Каца — Донскера — Прохорова

Пусть ξ — стохастический процесс, параметризованный около-интервалом T (например, нормированный мартингал, ассоциированный с последовательностью случайных величин из гл. 14). Говорят, что ξ — (около-)*винеровский процесс,* если ξ около-эквивалентен винеровскому блужданию на T. Следующая теорема представляет собой вариант центральной предельной теоремы Муавра — Лапласа, который содержит теорему Линдеберга о достаточности условия Линдеберга, теорему Феллера о необходимости условия, теорему Винера о непрерывности траекторий винеровского процесса, характеризацию Ле́ви — Дуба этого процесса, как единственного нормированного мартингала с непрерывными траекториями и принцип инвариантности Эрдёша и Каца в обобщенной форме Донскера и Прохорова.

Теорема 18.1. *Пусть ξ — нормированный мартингал, параметризованный около-интервалом, $\xi(a) = 0$. Тогда следующие утверждения эквивалентны*:

(i) ξ *— винеровский процесс*;

(ii) ξ *непрерывен п. н. при всех t, и $\xi(b)$ — это L^2-суммируемая случайная величина*;

(iii) *процесс ξ удовлетворяет условию Линдеберга.*

Доказательство Покажем, что (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i).

Пусть выполнено (i). Согласно следствию теоремы 13.1 почти каждая траектория винеровского блуждания непрерывна при всех t, поэтому по следствию 2 теоремы 17.3 траектория процесса ξ непрерывна при всех t. Нам нужно доказать, что величина $\xi(b)$ —L^2-случайная величина. Положим

$$f_{(c)}(\lambda) = \begin{cases} \lambda^2, & |\lambda| \le c, \\ c^2, & |\lambda| > c, \end{cases}$$

и заметим, что y — это L^2-суммируемая случайная величина тогда и только тогда, когда для любого $\varepsilon \gg 0$ найдется такое $c \ll \infty$, что $\mathbf{E} f_{(c)}(y) \ge \mathbf{E} y^2 - \varepsilon$. Если $c \ll \infty$, то функция $f_{(c)}$ доступна и непрерывна. Поэтому если положить $F_{(c)}(\xi) = f_{(c)}(\xi(b))$, то возникающий $F_{(c)}$ — доступный непрерывный функционал. Пусть w — винеровское блуждание. Легко видеть, что $w(b)$ — это L^2-суммируемая величина, например, потому, что

$$\mathbf{E} w(b)^4 = \mathbf{E} \left(\sum dw(t) \right)^4 = \mathbf{E} \sum dw(t_1) dw(t_2) dw(t_3) dw(t_4)$$

$$= 3 \sum_{t \ne s} \mathbf{E} dw(t)^2 dw(s)^2 + \sum dw(t)^4$$

$$= 3 \sum_{t \ne s} dt ds + \sum dt^2 \simeq 3(b-a)^2 \ll \infty.$$

Поэтому для $\varepsilon \gg 0$ найдется $c \ll \infty$ такое, что $\mathbf{E} f_c(w(b)) \ge b - a - \varepsilon$, а также

$$\mathbf{E} f_c(\xi(b)) \gtrsim b - a - \varepsilon = \mathbf{E} \xi(b)^2 - \varepsilon.$$

Следовательно, $\xi(b)$ — L^2-суммируемая случайная величина, и потому (i) $\Rightarrow$ (ii).

Предположим, что выполнено (ii). В равенстве

$$\xi(b)^2 = 2 \sum \xi(t) d\xi(t) + \sum d\xi(t)^2$$

левая часть суммируема по предположению. Первое слагаемое правой части суммируемо потому, что математическое ожидание его квадрата равно

$$4\mathbf{E} \sum \xi(t)^2 dt = 4 \sum (t-a) dt \simeq 2(b-a)^2 \ll \infty.$$

Поэтому второе слагаемое правой части также суммируемо и, следовательно, $\sum d\xi(t)^{(\varepsilon)2}$ суммируема. Но если $\varepsilon \gg 0$, то п. н. $\sum d\xi(t)^2 =$

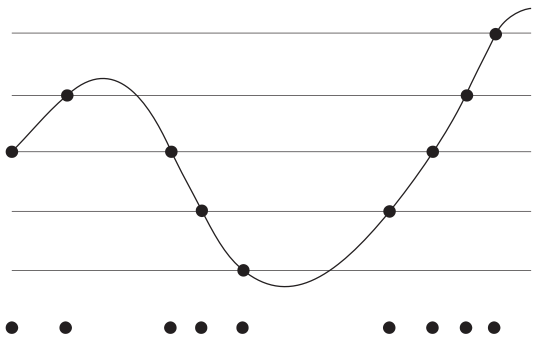


Рис. 18.1. Винеровский процесс

$\sum d\xi(t)^{(\varepsilon)2}$, так как траектории непрерывны. По теореме Лебега их математические ожидания бесконечно близки, а это и есть условие Линдеберга. Таким образом, (ii) $\Rightarrow$ (iii).

Предположим, что выполнено условие (iii). Идея доказательства импликации (iii) $\Rightarrow$ (i) проста. Рассмотрим процесс ξ в моменты времени t_n, как на рис 18.1, где ε — громадное инфинитезимальное число. Тогда он почти наверное идет вверх или вниз примерно на $\sqrt{\varepsilon}$ и в соответствии со свойством мартингала такое случается примерно с равной вероятностью. Моменты времени t_n — случайные величины, но так как квадратичная вариация процесса (которая примерно равна $n\varepsilon$) примерно равна истекшему времени $t - a$, они ведут себя так, как будто размещены на расстоянии ε от среднего (что не показано на рис. 18.1), и процесс выглядит подобно винеровскому процессу.

Докажем, что (iii) $\Rightarrow$ (i) при более слабом предположении, что ξ *около-нормирован.* Под этим мы подразумеваем, что п. н. $\tau_\xi(t) \simeq t - a$ для всех t. Согласно теоремам 14.3 и 17.2 не уменьшает общности предположение, что для некоторого числа $\alpha \simeq 0$ справедливо $|d\xi(t)| \leq \alpha$ для всех t и для всех ω.

Пусть $\overline{b} - b \simeq \infty$ и $\overline{T}$ — объединение множества T и всех чисел вида

$$b + k\frac{\overline{b} - b}{m},\ k = 1, \ldots, m,\ \text{где}\ \frac{\overline{b} - b}{m} \simeq 0.$$

Продолжим процесс ξ на $\overline{T}$, полагая $\xi(t) = \xi(b) + \overline{w}(t)$, где $\overline{w}$ — винеровское блуждание на $\overline{T} \setminus T$. Такое продолжение сохраняет свойства процесса ξ, указанные выше. Смысл этого продолжения состоит в том, чтобы избежать волнений, когда моменты времени t_n становятся неопределенными для исходного процесса.

Обозначим предшественника числа t через $t - d_* t$. Будем говорить, что процесс ξ *пересекает* уровень λ *в момент времени* t, если или $\xi(t - d_* t) < \lambda$ и $\xi(t) \geq \lambda$, или если $\xi(t - d_* t) > \lambda$ и $\xi(t) \leq \lambda$. Пусть $\varepsilon > 0$, $\sqrt{\varepsilon} > 2\alpha$ и $\nu = [(b-a)/\varepsilon]$. Пусть $t_0 = a$, а t_n для $n = 1, \dots, \nu$ определяется индуктивно как первый момент времени, следующий за t_{n-1}, в который ξ пересекает некоторый уровень $k_n\sqrt{\varepsilon}$, где k_n — целое число той же четности, что и n (если такого момента времени нет, считаем, что t_n равно $\overline{b}$), см. рис. 18.1.

Пусть $\widetilde{\mathcal{P}}_n$ — алгебра, порожденная случайными величинами $\xi(t_1)$, $\dots, \xi(t_n)$. Покажем, что $\xi(t_n)$ — мартингал относительно этой фильтрации. Чтобы убедиться в этом, обозначим через A событие, характеристическая функция которого χ_A принадлежит $\widetilde{\mathcal{P}}_n$. Тогда

$$\chi_A(\xi(t_{n+1}) - \xi(t_n)) = \sum \eta(s) d\xi(s),$$

где η — это некоторый $\mathcal{P}$-процесс: $\eta(s) = 1$, если событие A произошло и $t_n \leq s < t_{n+1}$, в противном случае $\eta(s) = 0$. Следовательно, $\mathbf{E}\chi_A(\xi(t_{n+1}) - \xi(t_n)) = 0$, а так как это верно для любого события A, $\chi_A \in \mathcal{P}_n$, отсюда вытекает равенство

$$\mathbf{E}_{\widetilde{\mathcal{P}}_n}(\xi(t_{n+1}) - \xi(t_n)) = 0,$$

которое и доказывает утверждение.

Предположим, что $0 \ll \varepsilon \ll \infty$. Тогда п. н.

$$\left|(\xi(t_{n+1}) - \xi(t_n)) \mp \sqrt{\varepsilon}\right| \leq 2\alpha$$

для всех $n < \nu$, так как этого может не быть лишь только если t_{n+1} равно $\overline{b}$, но поскольку $\overline{b} - b \simeq \infty$, а $\nu \ll \infty$, если $\varepsilon \gg 0$, легко видеть, что п. н. этого не происходит. Отметим также, что $|\xi(t_{n+1}) - \xi(t_n)| \leq \sqrt{\varepsilon} + 2\alpha \ll \infty$ всюду для всех $n < \nu$. Так как $\xi(t_{n+1}) - \xi(t_n)$ имеет среднее 0, получаем

$$\Pr\left\{\left|(\xi(t_{n+1}) - \xi(t_n)) - \sqrt{\varepsilon}\right| \leq 2\alpha\right\} \simeq \frac{1}{2},$$

$$\Pr\left\{\left|(\xi(t_{n+1})-\xi(t_n))+\sqrt{\varepsilon}\right|\le 2\alpha\right\}\simeq\frac{1}{2}.$$

То же самое верно для условной вероятности относительно $\widetilde{\mathcal{P}}_n$. Поскольку $\nu \ll \infty$ при $\varepsilon \gg 0$, это значит, что

$$\sum_{\pi}\left|\Pr\{|(\xi(t_{n+1})-\xi(t_n))-(-1)^{\pi(n)}\sqrt{\varepsilon}|\le 2\alpha \text{ для всех } n\}-\frac{1}{2^{\nu}}\right|\simeq 0, \tag{18.1}$$

где суммирование производится по всем 2^{ν} отображениям π множества $\{0,\dots,\nu-1\}$ в $\{0,1\}$.

Воспользуемся принципом переполнения. Так как множество всех ε, для которых левая часть (18.1) $\le \varepsilon$ (т. е. все, но $\simeq 0$), содержит все $\varepsilon \gg 0$, оно содержит все достаточно большие $\varepsilon \simeq 0$. Зафиксируем ε («громадное инфинитезимальное число») так, что $\varepsilon \simeq 0$, $\alpha/\varepsilon \simeq 0$ и условие (18.1) выполнено.

Пусть случайная величина $\nu(t)$ — наибольшее число n, $n \le \nu$, такое, что $t_n \le t$. Пусть для $t \in T$

$$\zeta(t)=\sum_{n=1}^{\nu(t)}(\xi(t_n)-\xi(t_{n-1}))^2+(\xi(t)-\xi(t_{\nu(t)}))^2. \tag{18.2}$$

Тогда

$$d\zeta(t)=d\xi(t)^2+2(\xi(t)-\xi(t_{\nu(t)}))d\xi(t),$$

так что $\|d\zeta(t)\| \le \gamma\|d\xi(t)\|$ для всех $t \in T'$ и для всех ω, где $\gamma=\alpha+2(\sqrt{\varepsilon}+\alpha)\simeq 0$. Следовательно,

$$\left\|\sum d\zeta(t)-\sum \mathbf{E}_t d\zeta(t)\right\|_2^2 \le \sum\|d\zeta(t)\|_2^2 \le \gamma^2\sum\|d\xi(t)\|_2^2 \simeq 0.$$

Из теорем 11.1 и 7.1 следует, что п. н.

$$\zeta(t)\simeq\sum_{s<t}\mathbf{E}_s d\zeta(s)=\sum_{s<t}\mathbf{E}_s d\xi(s)^2\simeq\tau_\xi(t)\simeq t-a$$

для всех $t \in T$. Так как $\alpha/\varepsilon \simeq 0$, п. н. каждый член в сумме (18.2) $\simeq \varepsilon$. Последний член в (18.2) $\simeq 0$, так что п. н. $\zeta(t)\simeq\nu(t)\varepsilon$. Таким образом, п. н.

$$\nu(t)\varepsilon \simeq t-a \tag{18.3}$$

для всех $t \in T$.

Пусть Π — конечное вероятностное пространство всех отображений π множества $\{1, \ldots, \nu\}$ в $\{0, 1\}$ с вероятностями $\mathrm{pr}(\pi) = 1/2^\nu$ для каждого $\pi \in \Pi$. Пусть w_ε — стохастический процесс, параметризованный элементами множества T и определенный на $\langle \Pi, \mathrm{pr} \rangle$ равенством

$$w_\varepsilon(t) = \sum_{n \le (t-a)/\varepsilon} (-1)^{\pi(n)} \sqrt{\varepsilon}.$$

Из (18.1), (18.3) и теорем 17.1 и 17.2 видно, что процесс ξ около-эквивалентен процессу w_ε. Если ξ' — любой другой процесс, удовлетворяющий нашим предположениям, в частности, если ξ' — винеровское блуждание на T, то для достаточно больших $\varepsilon \simeq 0$ процессы ξ и ξ' около-эквивалентны процессу w_ε, а также около-эквивалентны друг другу. Следовательно, ξ — винеровский процесс, и поэтому (iii) $\Rightarrow$ (i). $\square$

Мы остановились в произвольном месте. Можно сделать больше. Я надеюсь, что кто-нибудь напишет подлинно элементарную книгу о стохастических процессах в духе этих строк, дополнив ее упражнениями и приложениями.

Приложение

Введение

Цель этого приложения — продемонстрировать, как теоремы традиционной теории стохастических процессов могут быть выведены из их элементарных аналогов с помощью соображений того типа, который обычно квалифицируют как обобщенную чепуху; в этом приложении нет вероятностных рассуждений. Тем самым элементарная нестандартная теория стохастических процессов может быть использована для получения традиционных результатов; с другой стороны, этим демонстрируется, что ни тщательно разработанный аппарат традиционной теории, ни приемы подробной теории нестандартного анализа, необходимые для доказательства эквивалентности элементарных результатов в их традиционной форме, не добавляют ничего существенного: элементарная теория имеет то же самое научное содержание, что и традиционная теория. Следующий текст задуман как саморазрушающееся приложение.

Мы предполагаем знакомство с обычной теорией меры и нестандартным анализом в форме внутренней теории множеств (IST), см. [1].

Элементарные аксиомы гл. 4 являются теоремами теории (IST). Аксиомы (1) и (2) вытекают из принципа переноса, аксиома (3) вытекает из принципа идеализации, а внешняя индукция (4) вытекает из принципа стандартизации. Принцип последовательности (*5) — это частный случай принципа насыщения, см. [2]. Таким образом, все теоремы в книге представляют собой теоремы теории (IST).

Около-элементарные процессы

Теоремы традиционной математики являются внутренними, поэтому чтобы доказать подобную теорему о стохастическом процессе, по принципу переноса мы можем предполагать, что процесс является стандартным. Отсюда вытекает, что множество, которое параметризует процесс, и вероятностное пространство, на котором процесс определен, стандартны.

Пусть ξ_0 — стандартный стохастический процесс, параметризованный элементами множества T_0 и определенный на пространстве $\langle \Omega_0, S_0, \mathrm{Pr}_0 \rangle$. Под *около-элементарным процессом* (для ξ_0) мы подразумеваем стохастический процесс ξ, параметризованный элементами конечного подмножества $T \subset T_0$, содержащего все стандартные элементы множества T_0, и такой, что ξ определен на пространстве $\langle \Omega_0, S_0, \mathrm{Pr}_0 \rangle$, но принимает лишь конечное множество значений, причем для него выполнено соотношение

$$\sum_{t \in T} |\xi(t) - \xi_0(t)| \simeq 0 \tag{A.1}$$

за исключением множества бесконечно малой меры, и условие, что если ξ_0 принимает значения в L^p, где число з, $1 \leq p \leq \infty$, стандартно, то

$$\sum_{t \in T} \|\xi(t) - \xi_0(t)\|_p \simeq 0. \tag{A.2}$$

Отметим, что если процесс ξ_0 принимает L^p-суммируемые значения для некоторого стандартного p, то по неравенству Чебышёва (A.1) следует из (A.2). Пусть $\mathcal{S}$ — это σ-алгебра, порожденная процессом $\xi(t)$ для $t \in T$; тогда $\mathcal{S}$ — конечная булева подалгебра алгебры $\mathcal{S}_0$. Пусть Ω — множество, составленное из всех атомов алгебры $\mathcal{S}$ строго положительной меры. Определим вероятность pr на пространстве Ω равенством $\mathrm{pr}(\omega) = \mathrm{Pr}_0(\omega)$. Тогда $\langle \Omega, \mathrm{pr} \rangle$ — конечное вероятностное пространство. Мы можем рассматривать процесс ξ или как стохастический процесс, определенный на пространстве $\langle \Omega, \mathrm{pr} \rangle$, как в основном тексте, или, с целью сравнения с процессом ξ_0, как определенный на пространстве $\langle \Omega_0, \mathcal{S}_0, \mathrm{Pr}_0 \rangle$; мы не будем делать различия в обозначениях для этих двух понятий.

Чтобы получить около-элементарный процесс, нужно всего лишь взять достаточно большое недоступное число членов десятичного разложения процесса $\xi_0(t)$ для $t \in T$. Для произвольного положительного вещественного числа x и натурального числа n будем считать $x_{[n]}$ наибольшим числом $\leq x$ вида

$$\sum_{k=-n}^{n} a_k 2^{-k},$$

где каждое из чисел a_k принимает значение 0 или 1, при этом для $x < 0$ полагаем $x_{[n]} = -(-x)_{[n]}$.

Теорема А.1. *Пусть ξ_0 — стандартный стохастический процесс. Тогда существует около-элементарный процесс для ξ_0.*

Доказательство. Существование множества T вытекает из принципа идеализации. Пусть $\varepsilon > 0$ — бесконечно малое число. Пусть P — множество всех $p \in [1, \infty]$ таких, что процесс ξ_0 принимает значения в L^p. Если множество P непусто, то по принципу идеализации оно содержит элемент p_0 такой, что $p_0 \geq p$ для любого стандартного числа $p \in P$. Тогда по теореме Лебега о мажорированной сходимости если число n достаточно велико, то

$$\sum_{t \in T} \|\xi_0(t)_{[n]} - \xi_0(t)\|_{p_0} \leq \varepsilon.$$

Пусть $\xi(t) = \xi_0(t)_{[n]}$ для $t \in T$. Тогда (А.2) выполняется для всех стандартных $p \in P$. Как уже отмечалось, отсюда вытекает (А.1). Если множество P пусто, положим

$$N_n = \{\max_{t \in T} |\xi_0(t)| \geq 2^n\}.$$

Выберем n столь большим, чтобы $\Pr_0 N_n \leq \varepsilon$ (это возможно, так как N_n убывает и сходится к пустому множеству), а также чтобы n было больше, чем мощность множества T, а $n 2^{-n} \leq \varepsilon$. Пусть снова $\xi(t) = \xi_0(t)_{[n]}$ для $t \in T$. Так как $|\xi(t) - \xi_0(t)| \leq 2^{-n}$, соотношение (А.1) будет выполнено, за исключением точек из N_n. $\square$

Ясно, что так определенные случайные величины $\xi(t)$ независимы, если независимы случайные величины $\xi_0(t)$. Если ξ_0 — мартингал, мы должны несколько модифицировать приведенную конструкцию для того, чтобы получить элементарный мартингал.

Теорема А.2 *Пусть стандартный стохастический процесс ξ_0, параметризованный элементами подмножества T_0 множества* $\mathbf{R}$ *и* *принимающий значения в L^1, т. е. супермартингал, субмартингал или мартингал. Тогда существует около-элементарный процесс с тем же самым свойством.*

Доказательство. Выберем конечное подмножество T множества T_0, содержащее все его стандартные элементы, и используем обозначения $(T', dt, a, \text{etc.})$ основного текста. Мы предполагаем, что процесс ξ_0 приспособлен к некоторой фильтрации $\mathcal{P}_0$. Выберем число n, и пусть $\mathcal{P}_t$ — конечная булева алгебра, порожденная $\xi_0(s)_{[n]}$ для $s \le t$, $s \in T$, так что $\mathcal{P}_t \subseteq \mathcal{P}_{0t}$. Затем определим процесс ξ равенством

$$d\xi(t) = d(\xi_0(t)_{[n]}) + \mathbf{E}\{d\xi_0(t) - d(\xi_0(t)_{[n]}) \mid \mathcal{P}_t\}, \quad t \in T',$$

$$\xi(t) = \xi_0(a)_{[n]} + \sum_{s<t} d\xi(s), \quad t \in T.$$

Тогда

$$\mathbf{E}\{d\xi(t)|\mathcal{P}_t\} = \mathbf{E}\{d\xi_0(t)|\mathcal{P}_t\}.$$

Так как $\mathcal{P}_t \subseteq \mathcal{P}_{0t}$, здесь стоит нужный знак, т. е. если процесс ξ_0 — супермартингал, субмартингал или мартингал, то таковым же будет и процесс ξ. Пусть $\varepsilon > 0$ — бесконечно малое число, и p_0 такое, как в предыдущем доказательстве. Если число n достаточно велико, то по теореме Лебега о мажорированной сходимости

$$\sum_{t \in T} \|\xi(t)_{[n]} - \xi_0(t)\|_{p_0} \le \varepsilon. \quad \Box$$

Эти доказательства показывают, что бесконечно малое число, неявно участвующее в определении около-элементарного процесса, может быть выбрано произвольно малым.

Эквивалентность аналитических свойств

Типичная теорема теории стохастических процессов утверждает, что при некоторых аналитических предположениях некоторые вероятностные свойства выборочных траекторий выполняются почти наверное. Чтобы получить традиционную теорему из ее элементарного аналога, нам нужно показать, что внутренние аналитические предположения влекут аналогичные внешние аналитические предположения и что внешнее вероятностное утверждение влечет аналогичное внутренее вероятностное утверждение. Но чтобы убедиться, что две формулировки говорят об одном и том же на различных языках, мы должны установить эквивалентность этих двух форм предположений и этих двух форм утверждений.

В этом разделе мы хотим проверить, что некоторые внутренние аналитические свойства стандартного стохастического процесса эквивалентны соответствующим внешним аналитическим свойствам около-элементарного процесса. Остаются в силе следующие соглашения: ξ_0 — стандартный стохастический процесс, параметризованный элементами множества T_0 и определенный на пространстве $\langle \Omega_0, S_0, \mathrm{Pr}_0 \rangle$; ξ — около-элементарный процесс, параметризованный элементами множества T и определенный на пространстве $\langle \Omega, \mathrm{pr} \rangle$, как выше (являющийся супермартингалом и т. д. , подобно процессу ξ_0); a и b — первый и последний элементы множества T, действует соглашение: $T' = T \setminus \{b\}$, для $t \in T'$ его преемник есть $t + dt$, и для любой функции f на множестве T мы полагаем $df(t) = f(t + dt) - f(t)$ для $t \in T'$; если $T_0 = \mathbf{N}^+ = \mathbf{N} \setminus \{0\}$, то в качестве T выбираем множество $\{1, \ldots, \nu\}$, где ν — недоступное натуральное число. Отметим, что если T_0 — замкнутый интервал $[a, b]$, то T — около-интервал с первым элементом a и последним элементом b.

Теорема A.3. *Пусть процесс ξ_0 принимает значения в пространстве L^1. Если $T_0 = \mathbf{N}^+$, то*

(1) $\sum_{n=1}^{\infty} \|\xi_0(n)\|_1$ *сходится* $\Longleftrightarrow$ $\sum_{n=1}^{\nu} \|\xi(n)\|_1$ *(около-)сходится;*

(2) *процесс $\xi_0(n)$ сходится в L^1* $\Longleftrightarrow$ *процесс $\xi(n)$ (около-)сходится в $\mathbf{L}^1$.*

Если t_0 — стандартная точка множества T_0 и $T_0 \subseteq \mathbf{R}$, то

(3) *процесс ξ_0 непрерывен в точке t_0 в пространстве $\mathbf{L}^1$* $\Longleftrightarrow$ *процесс ξ непрерывен в точке t_0 в пространстве L^1.*

Если $T_0 \subseteq \mathbf{R}$, то

(4) *процесс ξ_0 имеет ограниченную вариацию в пространстве L^1* $\Longleftrightarrow \sum_{t \in T} \|d\xi_t(n)\|_1 \leq \infty$.

Если T_0 — замкнутый интервал $[a, b]$, то

(5) $\xi_0 \in L^1(T_0 \times \Omega_0) \Longleftrightarrow \xi \in L^1$ *на* $(T' \times \Omega)$.

(6) *процесс ξ_0 абсолютно непрерывен в L^1* $\Longleftrightarrow$ *случайная величина $d\xi/dt$ суммируема на $(T' \times \Omega)$ для некоторого около-элементарного процесса ξ_0.*

Доказательство. Заключение (1) эквивалентно утверждению

$$\sum_{n=1}^{\nu} \|\xi_0(n)\|_1 \text{ (около-)сходится,} \tag{A.3}$$

так как ξ — около-элементарный процесс. Таким образом, в (1) утверждается, и это легко проверяется, что стандартный ряд сходится тогда и только тогда, когда его частичная сумма до некоторого недоступного ν около сходится; утверждения (2) и (3) доказываются аналогично.

Предположим, что выполнено условие (4). По принципу переноса найдется такое число K, что для любого конечного подмножества множества T_0, в частности T, будет $\sum_{t\in T'} \|d\xi_0(t)\|_1 \leq K$. Поэтому заключение (4) выполнено. Обратно, предположим, что заключение (4) выполнено. Так как T содержит все стандартные элементы множества T_0, существует фиксированная стандартная оценка сверху вариации процесса ξ_0 на любом конечном подмножестве множества T_0, поэтому по принципу переноса предположение (4) выполнено.

Предположение (5) эквивалентно тому, что

$$\left\|\xi_0^{(n)} - \xi_0^{(m)}\right\|_1 \to 0 \ \text{ при } m, n \to \infty. \tag{A.4}$$

Так как процесс ξ_0 стандартен, (A.4) эквивалентно тому, что $\|\xi_0^{(n)} - \xi_0^{(m)}\|_1 \simeq 0$ для $m, n \simeq \infty$, а так как ξ — около-элементарный процесс, релятивизованное соотношение эквивалентно заключению (5).

Для любого конечного набора непересекающихся подынтервалов $[a, b]$ обозначим через $|I|$ их общую длину, и через $\mathrm{var}_0(I)$ — сумму $\sum_i \|\xi_0(b_i) - \xi_0(a_i)\|_1$, где $[a_i, b_i]$ — интервалы из I. Тогда предположение (6) состоит в следующем:

$$\forall\varepsilon\, \exists\delta \forall I (|I| \leq \delta \Rightarrow \mathrm{var}_0(I) \leq \varepsilon). \tag{A.5}$$

Но (A.5) эквивалентно тому, что

$$\forall I\, (|I| \simeq 0 \Rightarrow \mathrm{var}_0(I) \simeq 0). \tag{A.6}$$

Чтобы убедиться в этом, применим к (A.6) алгоритм редукции[*] следующим образом. Запишем (A.6) в виде

$$\forall I \left(\forall^{\mathrm{st}}\delta |I| \leq \delta \Rightarrow \forall^{\mathrm{st}}\varepsilon\ \mathrm{var}_0(I) \leq \varepsilon\right).$$

---

[*] Его часто называют алгоритмом Нельсона. — Прим ред.

Перепишем последнее так:

$$\forall^{\mathrm{st}}\varepsilon\forall I\exists^{\mathrm{st}}\delta(|I|\le\delta\Rightarrow \mathrm{var}_0(I)\le\varepsilon).$$

Используя принцип идеализации, приведем к виду

$$\forall^{\mathrm{st}}\varepsilon\exists^{\mathrm{stfin}}\delta'\forall I\exists\delta\in\delta'(|I|\le\delta\Rightarrow \mathrm{var}_0(I)\le\varepsilon),$$

а затем, выберем наименьшее $\delta\in\delta'$ и применим принцип переноса, чтобы получить (A.5).

Будем называть интервал I *хорошим*, если концевые точки интервалов из I принадлежат T. Я утверждаю, что (A.6) эквивалентно утверждению

$$\forall^{\mathrm{good}}I(|I|\simeq 0\Rightarrow \mathrm{var}_0(I)\simeq 0). \tag{A.7}$$

Ясно, что из (A.6) следует (A.7). Если мы применим алгоритм редукции к импликации (A.7), то обнаружим, что она эквивалентна утверждению

$$\forall^{\mathrm{st}}\varepsilon\exists^{\mathrm{st}}\delta\forall^{\mathrm{good}}I(|I|\le\delta\Rightarrow \mathrm{var}_0(I)\le\varepsilon).$$

Поскольку T содержит все стандартные точки интервала $[a,b]$, отсюда следует, что

$$\forall^{\mathrm{st}}\varepsilon\exists^{\mathrm{st}}\delta\forall^{\mathrm{st}}I(|I|\le\delta\Rightarrow \mathrm{var}_0(I)\le\varepsilon),$$

что по принципу переноса эквивалентно (A.5), а следовательно, (A.6). Это и доказывает утверждение. □

Таким образом, (A.5) и (A.7) эквивалентны. Для хорошего интервала I через $\mathrm{var}(I)$ обозначим $\sum_i\|\xi(b_i)-\xi(a_i)\|_1$. Так как ξ — околопроцесс, импликация (A.7) эквивалентна импликации

$$\forall^{\mathrm{good}}I(|I|\simeq 0\Rightarrow \mathrm{var}(I)\simeq 0).$$

Но это означает, что для множеств вида $M=I\times\Omega$, если M имеет бесконечно малую вероятность в $T'\times\Omega$, то

$$\mathbf{E}\frac{d\xi}{dt}\chi_M\simeq 0,$$

и по теореме 8.1 это выполняется, если случайная величина $d\xi/dt$ суммируема на $T'\times\Omega$.

Для доказательства противоположной импликации заметим, что предположение (6) эквивалентно существованию стандартной функции $\eta_0 \in \mathbf{L}(T_0 \times \Omega_0)$ такой, что для всех $t \in T_0$

$$\xi_0(t, \omega_0) = \xi_0(a, \omega_0) + \int_a^t \eta_0(s, \omega_0) ds$$

для п. в. $\omega_0 \in \Omega_0$. Итак, пусть η — около-элементарный процесс для процесса η_0, и, значит, согласно (5), процесс η суммируем на $T' \times \Omega_0$. Выберем около-элементарную случайную величину $\xi(a)$ для $\xi_0(a)$ (т. е. найдем около-элементарный процесс для стандартного стохастического процесса с множеством параметров, состоящим из единственной точки a, и со случайной величиной $\xi_0(a)$) и определим ξ равенством

$$\xi(t) = \xi(a) + \sum_{a < s \leq t} \eta(s).$$

Тогда легко проверить, что ξ — около-элементарный процесс для ξ_0, а случайная величина $d\xi/dt = \eta$ суммируема на $T' \times \Omega$. □

Теорема A.4. *Пусть процесс ξ_0 принимает значения в L^1, и предположим, что $T_0 \subseteq \mathbf{R}$. Если ξ_0 — мартингал или положительный субмартингал, то $\|\xi(a)\|_1 \ll \infty$, и $\|\xi(b)\|_1 \ll \infty$ тогда и только тогда, когда функция $t \mapsto \|\xi_0(t)\|_1$ ограничена. Если ξ_0 — положительный супермартингал, то $\|\xi(b)\|_1 \ll \infty$, и $\|\xi(a)\|_1 \ll \infty$ тогда и только тогда, когда функция $t \mapsto \|\xi_0(t)\|_1$ ограничена .*

Доказательство. При первом предположении $\|\xi_0(t)\|_1$ и $\|\xi(t)\|_1$ возрастают по t, поэтому $\|\xi(a)\|_1 \ll \infty$. Если $\|\xi(b)\|_1 \ll \infty$, то найдется такое стандартное число K, что $\|\xi(t)\|_1 \leq K$ для всех $t \in T$, отсюда $\|\xi_0(t)\|_1 \leq K + 1$ для всех стандартных t и, следовательно, по принципу переноса, для всех t. Обратно, если $t \mapsto \|\xi_0(t)\|_1$ ограничена, то имеется стандартное ограниченное число K такое, что $\|\xi_0(b)\|_1 \leq K$ и $\|\xi(b)\|_1 \ll \infty$. Доказательство при втором предположении (когда нормы уменьшаются) совершенно аналогично. □

Регулярные вероятностные меры

Пусть η_0 — стохастический процесс, параметризованный элементами множества T_0. Если множество T_0 несчетно, то в нашей теории возникает много сложностей, связанных с теорией меры. Чтобы избежать большинства этих сложностей, мы всегда будем рассматривать *каноническую версию* ξ_0 процесса: см. [3]: процесс параметризован элементами того же самого множества T_0, эквивалентен процессу η_0, имеет те же самые конечные совместные распределения. Он определен на пространстве траекторий

$$\Omega_0 = \prod_{t \in T_0} \dot{\mathbf{R}},$$

где $\dot{\mathbf{R}}$ — одноточечная компактификация пространства $\mathbf{R}$. При этом Ω_0 — компактное хаусдорфово пространство в топологии произведения. Через обозначим S_0 σ-алгебру всех борелевских подмножеств пространства Ω_0, а через Pr_0 — единственную регулярную вероятностную меру такую, что стохастический процесс ξ_0, определенный на $\langle \Omega_0, S_0, \mathrm{Pr}_0 \rangle$ равенством $\xi_0(t, \omega_0) = \omega_0(t)$ для $t \in T_0$ и $\omega_0 \in \Omega_0$, имеет те же самые конечные совместные распределения, что и процесс η_0. Если процесс η_0 стандартен, то таков и процесс ξ_0.

Одно из преимуществ канонической версии заключается в том, что многие интересные подмножества пространства траекторий являются борелевскими множествами, а потому автоматически являются измеримыми.

Нам потребуются два важных внешних результата о регулярных вероятностных мерах. Напомним, что *направленное множество* — это множество $\mathcal{D}$ с заданным на нем транзитивным бинарным отношением $\prec$ таким, что каждое конечное подмножество множества $\mathcal{D}$ имеет нижнюю границу, а *сеть* — это функция $F \mapsto \Phi_F$, определенная на $\mathcal{D}$. Если значениями этой функции являются множества, то она называется *возрастающей*, если $F_1 \prec F_2$ влечет $\Phi_{F_1} \subseteq \Phi_{F_2}$, и *убывающей*, если $F_1 \prec F_2$ влечет $\Phi_{F_1} \supseteq \Phi_{F_2}$. Суть следующей теоремы заключается в том, что она приложима к несчетным семействам множеств.

Теорема A.5. *Пусть* Pr_0 *— регулярная вероятностная мера на компактном хаусдорфовом пространстве* Ω_0. *Пусть* $F \mapsto \Phi_F$ *— убывающая сеть замкнутых подмножеств пространства* Ω_0 *и* $G \mapsto \Gamma_G$ *— возрастающая сеть открытых подмножеств пространства* Ω_0. *Тогда*

$$\mathrm{Pr}_0 \bigcap_F \Phi_F = \inf_F \mathrm{Pr}_0 \Phi_F, \tag{A.8}$$

$$\mathrm{Pr}_0 \bigcup_G \Gamma_G = \sup_G \mathrm{Pr}_0 \Gamma_G. \tag{A.9}$$

Доказательство. Пусть $\varepsilon > 0$. По определению регулярности найдется компактное множество Φ, содержащееся в $\bigcup_G \Gamma_G$, такое, что

$$\mathrm{Pr}_0\Big(\bigcup_G \Gamma_G \setminus \Phi\Big) \le \varepsilon.$$

Так как Φ компактно, найдется конечное множество $\{G_1, \ldots, G_n\}$ такое, что

$$\bigcup_{i=1}^n \Gamma_{G_i} \supseteq \Phi.$$

Обозначим через G_0 нижнюю границу этого конечного множества. Тогда $\Gamma_{G_0} \supseteq \Phi$, так что

$$\mathrm{Pr}_0\Big(\bigcup_G \Gamma_G \setminus \Gamma_{G_0}\Big) \le \varepsilon.$$

Так как ε произвольно, это доказывает (A.9), а (A.8) вытекает из соображений двойственности. $\square$

Теорема A.6. *Пусть* Pr_0 *— регулярная вероятностная мера на компактном хаусдорфовом пространстве* Ω_0. *Пусть* Φ_{jkF} *— замкнутые подмножества пространства* Ω_0, *а* Γ_{kjG} *— открытые подмножества пространства* Ω_0, *где* j *и* k *пробегают* $\mathbf{N}$, *а* F *и* G *пробегают направленное множество* $\mathcal{D}$, *и пусть они убывают по* j *и* F *и возрастают по* k *и* G. *Тогда*

$$\mathrm{Pr}_0 \bigcap_j \bigcup_k \bigcap_F \Phi_{jkF} = \sup_{\tilde{k}} \inf_{j,F} \mathrm{Pr}_0\, \Phi_{j\tilde{k}(j)F}, \tag{A.10}$$

где $\tilde{k}$ *пробегает множество всех функций из* $\mathbf{N}$ *в* $\mathbf{N}$, *и*

$$\mathrm{Pr}_0 \bigcup_k \bigcap_j \bigcup_G \Gamma_{kjG} = \sup_{k,\widetilde{G}} \inf_{j} \mathrm{Pr}_0\, \Gamma_{kj\widetilde{G}(j)}, \tag{A.11}$$

где $\widetilde{G}$ *пробегает множество всех функций из* $\mathbf{N}$ *в* $\mathcal{D}$.

Доказательство. Пусть

$$p = \mathrm{Pr}_0 \bigcap_j \bigcup_k \bigcap_F \Phi_{jkF}$$

и $\varepsilon > 0$. Тогда

$$\forall j \exists k \operatorname{Pr}_0 \bigcap_F \Phi_{jkF} \geq p - \varepsilon,$$

поэтому найдется функция $\tilde{k}$ такая, что

$$\forall j \operatorname{Pr}_0 \bigcap_F \Phi_{j\tilde{k}(j)F} \geq p - \varepsilon.$$

По предыдущей теореме

$$\inf_{j,F} \operatorname{Pr}_0 \Phi_{j\tilde{k}(j)F} \geq p - \varepsilon.$$

Так как ε произвольно, это доказывает неравенство $\leq$ в (A.10). Но поскольку

$$\bigcap_j \bigcup_k \bigcap_F \Phi_{jkF} = \bigcup_{\tilde{k}} \bigcap_{j,F} \Phi_{j\tilde{k}(j)F},$$

обратное неравенство тривиально.

Пусть теперь

$$p = \operatorname{Pr}_0 \bigcup_k \bigcap_j \bigcup_G \Gamma_{kjG}$$

и $\varepsilon > 0$. Тогда по предыдущей теореме

$$\exists k \forall j \exists G \operatorname{Pr}_0 \Gamma_{kjG} \geq p - \varepsilon,$$

тем самым существуют k и $\widetilde{G}$ такие, что для всех j

$$\operatorname{Pr}_0 \Gamma_{kj\widetilde{G}(j)} \geq p - \varepsilon.$$

Ввиду произвольности ε это доказывает неравенство $\leq$ в (A.11), и снова обратное неравенство тривиально. $\square$

Эквивалентность вероятностных свойств

Пусть $\xi_0 : T_0 \to \mathbf{R}$ — стандартная функция. Понятие около-элементарной функции ξ (для ξ_0) ясно: нужно просто применить определение около-элементарного процесса к случаю, когда основное вероятностное пространство состоит из единственной точки. Мы часто сталкиваемся с такой парой свойств, внутренним свойством **A** и внешним свойством **B**, что свойство **A** выполняется для ξ_0 тогда и только тогда, когда **B** выполняется для ξ. Например, свойство **A** выражает непрерывность в стандартной точке t_0, а свойство **B** выражает около-непрерывность в t_0. Отметим, что если процесс ξ_0 нестандартен, то эквивалентность не обязательно имеет место. Пусть теперь ξ_0 — стандартный стохастический процесс и ξ — около-элементарный процесс. Зададимся вопросом, эквивалентно ли выполнение п. н. свойства **A** для процесса ξ_0 выполнению п. н. свойства **B** для процесса ξ. Во-первых, мы должны убедиться в том, что множество точек основного вероятностного пространства, для которых выполнено свойство **A**, измеримо; иначе вопрос не имеет смысла. Но даже теперь мы еще не можем рассуждать по-траекторно. Имеется конечное множество, содержащее все траектории, а для большинства интересных стохастических процессов любое конечное множество траекторий имеет нулевую вероятность (т. е., вообще говоря, выборочные траектории стандартного стохастического процесса нестандартны!), так что в общем случае эквивалентность свойства **A** для стандартной функции и свойства **B** для около-элементарной функции не является ответом на наш вопрос.

Теорема A.7. *Пусть ξ_0 — стандартный стохастический процесс в канонической версии. Если $T_0 = \mathbf{N}^+$, то*

(1) $\sum_{n=1}^{\infty} |\xi_0(n)| < \infty$ *п. н.* $\iff$ $\sum_{n=1}^{\nu} |\xi(n)| \ll \infty$ *п. н.*,

(2) $\sum_{n=1}^{\infty} |\xi_0(n)| < \infty$ *п. н.* $\iff$ $\sum_{n=1}^{\nu} |\xi(n)|$ *(около-)сходится п. н.*

Если t_0 — стандартная точка множества T_0, а $T_0 \subseteq \mathbf{R}$, то

(3) ξ_0 *непрерывен в* t_0 *п. н.* $\iff \xi$ *(около-)непрерывен в* t_0 *п. н.*

Если T_0 — компактное подмножество множества $\mathbf{R}$, то

(4) ξ_0 *непрерывен п. н.* $\iff \xi$ *(около-)непрерывен п. н.*

(5) ξ_0 *не имеет разрывов второго рода п. н.* $\iff \xi$ *имеет доступное колебание п. н.*

Если $T_0 \subseteq \mathbf{R}$, то

(6) ξ_0 *имеет ограниченную вариацию п. н.* $\iff \xi$ *имеет ограниченную вариацию п. н.*

Для любого T_0

(7) $\xi_0 = 0$ *п. н.* $\iff \xi \simeq 0$ *п. н.*

Доказательство. Для $i = 1, \dots, 7$ эквивалентность (i) имеет вид $\mathbf{A}_i \Leftrightarrow \mathbf{B}_i$, где $\mathbf{A}_i$ — внутренняя формула, а $\mathbf{B}_i$ — внешняя формула.

По определению существует подмножество N пространства траекторий Ω_0 такое, что $\mathrm{Pr}_0\, N \simeq 0$ и на N^c

$$\sum_{t\in T} |\xi_0(t) - \xi(t)| \simeq 0.$$

Предположим, ложным то, что утверждение $\mathbf{A}_1$ выполнено п. н., и пусть

$$\Phi^1_{jk} = \left\{ \sum_{n=1}^{k} |\xi_0(n)| \geq j \right\},$$

$$\Phi^1 = \bigcap_j \bigcup_k \Phi^1_{jk}.$$

Тогда $\mathrm{Pr}_0\, \Phi^1 > 0$, и потому по принципу переноса существует стандартное $\varepsilon > 0$ такое, что $\mathrm{Pr}_0\, \Phi^1 > \varepsilon$. По теореме **A**.6 и по принципу переноса существует стандартная функция $\tilde{k}$ такая, что $\mathrm{Pr}_0\, \Phi^1_{\tilde{k}} > \varepsilon$, где

$$\Phi^1_{\tilde{k}} = \bigcap_j \Phi^1_{j\tilde{k}(j)}.$$

Если j стандартно, то $\tilde{k}(j)$ стандартно и $\leq \nu$ и, следовательно,

$$\sum_{n=1}^{\nu} |\xi_0(n)| \simeq \infty$$

на $\Phi^1_{\tilde{k}}$, так как эта сумма больше каждого стандартного j. Поэтому

$$\sum_{n=1}^{\nu} |\xi(n)| \simeq \infty$$

на $N^c \cap \Phi^1_{\tilde{k}}$, а $\mathrm{Pr}_0(N^c \cap \Phi^1_{\tilde{k}}) > \varepsilon$. Так как это верно для некоторого стандартного $\varepsilon > 0$, значит ложно то, что утверждение $\mathbf{B}_1$ выполнено п. н. Таким образом, справедливость $\mathbf{B}_1$ п. н. влечет справедливость $\mathbf{A}_1$ п. н.

A fortiori справедливость $\mathbf{B}_2$ п. н. влечет справедливость п. н. утверждения $\mathbf{A}_2$ (которое идентично с $\mathbf{A}_1$).

Пусть неверно, что $\mathbf{A}_3$ выполняется п. н., и пусть для конечного подмножества G множества T_0

$$\Gamma^3_{kjG} = \bigcup_{\substack{s\in G\\ |s-t_0|\le 1/j}} \left\{|\xi_0(s)-\xi_0(t_0)| > \frac{1}{k}\right\},$$

$$\Gamma^3 = \bigcup_k \bigcap_j \bigcup_G \Gamma^3_{kjG}.$$

Тогда Γ^3 — это множество тех траекторий, которые разрывны в точке t_0. Заметим, что это борелевское множество, так как несчетное объединение по G есть объединение открытых множеств. Значит, существует стандартное $\varepsilon > 0$ такое, что $\mathrm{Pr}_0\,\Gamma^3 > \varepsilon$. По теореме **A**.6 и по принципу переноса существуют стандартное k и $\widetilde{G}$ такие, что $\mathrm{Pr}_0\,\Gamma^3_{k\widetilde{G}} > \varepsilon$, где

$$\Gamma^3_{k\widetilde{G}} = \bigcap_j \Gamma^3_{kj\widetilde{G}(j)}.$$

Если j стандартно, то $\widetilde{G}(j)$ — стандартное множество, так что каждый его элемент стандартен и, следовательно, $\widetilde{G}(j) \subseteq T$. Поэтому на $\Gamma^3_{k\widetilde{G}}$ для любого стандартного j существует $s \in T$, для которого $|s-t_0| \le 1/j$ и $|\xi_0(s)-\xi_0(t_0)| > 1/k$, поэтому по принципу переполнения существует $s \in T$, $s \simeq t_0$, такое, что $|\xi_0(s)-\xi_0(t_0)| > 1/k$. Следовательно, на $N^c \cap \Gamma^3_{k\widetilde{G}}$ процесс ξ не около непрерывен в t_0, и ложно утверждение, что $\mathbf{B}_3$ выполняется п. в. Таким образом, выполнение $\mathbf{B}_3$ п. в. влечет выполнение $\mathbf{A}_3$ п. в.

Доказательство того, что выполнение $\mathbf{B}_4$ п. в. влечет выполнение $\mathbf{A}_4$ п. в., совершенно аналогично: нужно только заменить фиксированное t_0 переменным t.

Доказательство того, что выполнение $\mathbf{B}_5$ п. в. влечет выполнение $\mathbf{A}_5$ п. в., аналогично. Для конечного подмножества G множества T_0, содержащего j точек $t_1 < \cdots < t_j$, положим

$$\Gamma^5_{kjG} = \bigcup_{i=1}^{j-1} \left\{|\xi_0(t_i)-\xi_0(t_{i+1})| > \frac{1}{k}\right\},$$

$$\Gamma^5 = \bigcup_k \bigcap_j \bigcup_G \Gamma^5_{kjG}$$

и используем те же соображения, что и прежде. Отметим, что борелевское множество Γ^5 есть множество траекторий, имеющих разрывы второго рода. Доказательство того, что выполнение $\mathbf{B}_6$ п. в. влечет выполнение $\mathbf{A}_6$ п. в., также аналогично. Для конечного подмножества G множества T_0 пусть G' получается из G удалением последнего элемента, а для $t \in G'$ пусть $t + dt$ — его преемник в G. Положим

$$\Gamma^6_{kjG} = \left\{ \sum_{t \in G'} |\xi_0(t + dt) - \xi_0(t)| > j \right\},$$

$$\Gamma^6 = \bigcap_j \bigcup_G \Gamma^6_{kjG},$$

и проведем те же рассуждения, что и прежде.

Допустим ложность утверждения, что $\mathbf{A}_7$ выполняется п. в. Для конечного подмножества G множества T_0 пусть

$$\Gamma^7_{kG} = \left\{ \max_{t \in G} |\xi_0(t)| > \frac{1}{k} \right\},$$

$$\Gamma^7 = \bigcup_k \bigcup_G \Gamma^7_{kG}.$$

Тогда существует стандартное $\varepsilon > 0$ такое, что $\mathrm{Pr}_0\, \Gamma^7 > \varepsilon$, а также стандартное k и G такие, что $\mathrm{Pr}_0\, \Gamma^7_{kG} > \varepsilon$. Тогда $G \subseteq T$, и на Γ^7_{kG} процесс ξ не бесконечно мал на T, тем самым ложно утверждение, что п. в. выполняется $\mathbf{B}_7$. Итак, выполнение $\mathbf{B}_7$ п. в. влечет выполнение $\mathbf{A}_7$ п. в.

Теперь рассмотрим импликации в обратном направлении.

Предположим, что утверждение $\mathbf{A}_2$ выполнено п. в., и пусть

$$\Phi^2_{jk} = \left\{ \sum_{n=k}^{\infty} |\xi_0(n)| \le \frac{1}{j} \right\},$$

$$\Phi^2 = \bigcap_j \bigcup_k \Phi^2_{jk}.$$

Тогда $\mathrm{Pr}_0\, \Phi^2 = 1$. Пусть $\varepsilon > 0$ стандартно. По теореме **A.6** и по принципу переноса существует стандартная функция $\tilde{k}$ такая, что $\mathrm{Pr}_0\, \Phi^2_{\tilde{k}} \ge 1 - \varepsilon$, где

$$\Phi^2_{\tilde{k}} = \bigcap_j \Phi^2_{j\tilde{k}(j)}.$$

На $\Phi^2_{\tilde{k}}$ для всех стандартных j имеем

$$\sum_{n=\tilde{k}(j)}^{\infty} |\xi_0(n)| \leq \frac{1}{j},$$

а на $N^c \cap \Phi^2_{\tilde{k}}$ будет

$$\sum_{n=\tilde{k}(j)}^{\nu} |\xi(n)| \lesssim \frac{1}{j}.$$

Так как $\tilde{k}(j)$ стандартно для всех стандартных j, для всех недоступных $\mu \leq \nu$ имеем

$$\sum_{n=\mu}^{\nu} |\xi(n)| \simeq 0,$$

так что $\sum_{n=1}^{\nu} |\xi(n)|$ около-сходится. Следовательно, $\mathbf{B}_2$ выполняется п. н., и поэтому выполнение $\mathbf{A}_2$ п. н. влечет выполнение $\mathbf{B}_2$ п. н.

A fortiori выполнение $\mathbf{A}_1$ п. н. влечет выполнение $\mathbf{B}_1$ п. н.

Предположим, что утверждение $\mathbf{A}_3$ выполнено п. в., и пусть

$$\Phi^3_{jk} = \bigcap_{|s-t_0|\leq 1/k} \left\{ |\xi_0(s) - \xi_0(t_0)| \leq \frac{1}{j} \right\},$$

$$\Phi^3 = \bigcap_j \bigcup_k \Phi^3_{jk}.$$

Тогда $\mathrm{Pr}_0\, \Phi^3 = 1$. Пусть $\varepsilon > 0$ стандартно. Тогда существует стандартная функция $\tilde{k}$ такая, что $\mathrm{Pr}_0\, \Phi^3_{\tilde{k}} \geq 1 - \varepsilon$, где

$$\Phi^3_{\tilde{k}} = \bigcap_j \Phi^3_{j\tilde{k}(j)}.$$

На $N^c \cap \Phi^3_{\tilde{k}}$ процесс ξ около-непрерывен в t_0, и, таким образом, выполнение $\mathbf{A}_3$ п. н. влечет выполнение $\mathbf{B}_3$ п. н. Остальные случаи рассматриваются аналогично. $\square$

Литература

1. **Nelson E.** Internal set theory: A new approach to nonstandard analysis // Bull. Amer. Math. Soc. 1977. V. 83. P. 1165–1198.

2. **Nelson E.** The syntax of nonstandard analysis // Ann. Pure and Appl. Logic. 1988. V. 38.P. 123–134

3. **Nelson E.** Regular probability measures on function space // Ann. of Math. 1959. V. 69. P. 630–643.

Указатель

Указатель обозначений

Издание подготовлено с использованием макро-пакета $\mathcal{AMS}$-TEX,
разработанного Американским математическим обществом.

This publication was typeset using $\mathcal{AMS}$-TEX,
the American Mathematical Society's TEX macro system.

Подписано в печать 30.1.95. Формат $60\times84\,{}^1/_{16}$. Бумага тип. № 2. Печать офсетная. Усл. печ. л. 8. Усл. кр.-отт. 8. Уч.-изд. л. 8. Тираж 400 экз. Заказ № 9.

Лицензия ЛР № 020633 от 18 сентября 1992 г.

Издательство Института математики им. С. Л. Соболева СО РАН.

630090, Новосибирск, Университетский пр., 4.

Отпечатано на полиграфическом участке

Института математики им. С. Л. Соболева СО РАН.

630090, Новосибирск, Университетский пр., 4.