Московский государственный университет им. М.В. Ломоносова
Физический факультет

Кафедра квантовой статистики и теории поля

А. Н. Соболевский

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ДЛЯ ФИЗИКОВ

Учебное пособие по курсу лекций

Москва
Физический факультет МГУ
2007

УДК 519.2
С54

Соболевский А. Н.

С54 Теория вероятностей и основы математической статистики для физиков. — М.: Физический факультет МГУ им. М. В. Ломоносова, 2007. — 46 с.

Учебное пособие содержит изложение основ теории вероятностей и математической статистики для студентов-физиков теоретической специализации. Наряду с классическим материалом (схема независимых испытаний Бернулли, конечные однородные цепи Маркова, диффузионные процессы), значительное внимание уделено таким темам, как теория больших уклонений, понятие энтропии в его различных вариантах, устойчивые законы и распределения вероятности со степенным убыванием, стохастическое дифференциальное исчисление.

Учебное пособие предназначено для студентов 3 года обучения, специализирующихся по различным разделам теоретической и математической физики.

УДК 519.2

Подписано к печати ____________

Тираж ________. Заказ ________

Отпечатано в отделе оперативной печати физического факультета МГУ им. М. В. Ломоносова

Оглавление

Предисловие

С 1970-х годов в Московском университете теория вероятностей и математическая статистика читается студентам-физикам как отдельный курс, закладывающий математические основания для обработки результатов эксперимента и для важнейшего раздела теоретической физики — статистической механики. С годами актуальность такого курса для физиков возрастает благодаря все более важной роли, которую вероятностные модели и методы играют в современных исследованиях неравновесных физических процессов в хаотических, турбулентных и наносистемах.

Данное пособие основано на варианте общего курса теории вероятностей и математической статистики, прочитанном в 2005/6 и 2006/7 уч. г.

студентам кафедры квантовой статистики и теории поля физического факультета МГУ.

Курс предназначен для овладения наиболее употребительными вероятностными моделями и методами вычислений на «физическом» уровне строгости. Как правило, вместо строгих доказательств в полной общности даны их наброски или разбор типичных примеров. Ряд необходимых для изложения или интересных результатов приводится без доказательства или с частичным доказательством (необходимое и достаточное условие сходимости по вероятности, закон повторного логарифма, парадокс Тарского, усиленный закон больших чисел, теоремы Бохнера, Карлемана, Фишера–Типпета–Гнеденко, Марцинкевича). Читатель-математик может рассматривать данное пособие как набор задач на доказательство или развернутых мотивировок строгих построений, изложенных в математической литературе.

Особое внимание в курсе уделено ряду вопросов, которые традиционно считаются материалом «повышенного уровня» в учебниках, но широко встречаются в журнальной и монографической литературе по статистической физике, нелинейной и стохастической динамике и численным методам Монте–Карло. К таким вопросам относятся большие уклонения, энтропия и другие теоретико-информационные понятия, степенные законы и статистика экстремальных значений, стохастические дифференциальные уравнения и исчисление Ито.

В пособии опущены выкладки и большинство мотивировок, примеров и обсуждений, приводимых на лекциях. В дальнейшем пособие планируется дополнить отдельным задачником.

Автор благодарен В. П. Маслову и У. Фришу, беседы с которыми оказали большое влияние на отбор материала и характер изложения, а также А. В. Леонидову, А. М. Чеботареву и А. Б. Шаповалу, которые прочитали предварительные варианты текста и высказали многочисленные ценные замечания. Особую признательность автор хотел бы выразить Б. И. Садовникову за понимание и поддержку на всех этапах подготовки данного учебного пособия. Конечно, ответственность за любые оставшиеся ошибки и недостатки изложения целиком лежит на авторе.

1 Целочисленные случайные величины

1.1. Будем называть **случайным испытанием** эксперимент, который можно повторять много раз при фиксированных условиях, получая при этом различные, непредсказуемые заранее (т. е. случайные) количественные результаты.

1.2. Числовая величина, измеренная в результате случайного испытания, называется **случайной величиной** (сл. вел.).

1.3. В этой лекции рассматриваются сл. вел., принимающие значения $0, 1, 2, 3, \ldots$

1.4. Будем обозначать целочисленные сл. вел. прописными латинскими буквами $M, N, \ldots$, а их значения — соответствующими строчными буквами $m, n, \ldots$

1.5. **Вероятность** того, что сл. вел. N примет значение n, обозначается $\mathsf{P}(N = n)$ или $\mathsf{P}(n)$, если из контекста ясно, о какой сл. вел. идет речь. Набор чисел $\mathsf{P}(n)$, $n = 0, 1, 2, \ldots$, называется **распределением вероятности** сл. вел. на множестве ее возможных значений — неотрицательных целых чисел.

1.6. Если сл. вел. никогда не принимает некоторое значение n, то $\mathsf{P}(n) = 0$. Например, если N представляет число очков, выпавших на игральной кости, то $\mathsf{P}(N = 0) = \mathsf{P}(N = 7) = \mathsf{P}(N = 8) = \ldots = 0$.

1.7. Распределение вероятности $\mathsf{P}(n)$ обязано удовлетворять двум условиям: (a) $\mathsf{P}(n) \geqslant 0$ при всех $n = 0, 1, 2, \ldots$; (b) $\sum_{n \geqslant 0} \mathsf{P}(n) = 1$.

1.8. **Математическое ожидание** сл. вел. N: $\langle N \rangle = \sum_{n \geqslant 0} n\, \mathsf{P}(n)$; вообще мат. ожидание любой функции $f(N)$ сл. вел. N определяется как $\langle f(N) \rangle = \sum_{n \geqslant 0} f(n)\, \mathsf{P}(n)$.

В математической литературе используются и другие обозначения мат. ожидания: $\mathsf{M}N$ (англ. <u>m</u>ean, фр. <u>m</u>oyenne) или $\mathsf{E}N$ (англ. <u>e</u>xpectation, фр. <u>e</u>spérance).

1.9. Рассматривают следующие характеристики сл. вел. N: **момент** k-го порядка: $\left\langle N^k \right\rangle$; **центральный момент** k-го порядка: $\left\langle (N - \langle N \rangle)^k \right\rangle$; **факториальный момент** k-го порядка: $\left\langle N^{\underline{k}} \right\rangle$, где k-я факториальная степень числа n определяется как $n^{\underline{k}} = n(n-1)\ldots(n-k+1)$.

1.10. Центральный момент второго порядка называется **дисперсией** и обозначается $\mathsf{D}N$.

1.11. $\mathsf{D}N = \left\langle N^2 \right\rangle - \langle N \rangle^2$.

1.12. Дисперсия всегда неотрицательна и равна нулю только для «неслучайной» величины, принимающей единственное значение с вероятностью 1. Этот факт есть частный случай неравенства Иенсена (п. B.15).

1.13. Совместное распределение вероятности пары сл. вел. M и N: $\mathsf{P}(M = m, N = n)$ или $\mathsf{P}(m, n)$. **Маргинальные вероятности**: $\mathsf{P}(M = m) = \sum_{n\geqslant 0} \mathsf{P}(m, n)$, $\mathsf{P}(N = n) = \sum_{m\geqslant 0} \mathsf{P}(m, n)$; **условные вероятности**: $\mathsf{P}_{M|N}(\,m \mid n\,) = \frac{\mathsf{P}(m,n)}{\mathsf{P}(N=n)}$ есть вероятность того, что $M = m$ при условии $N = n$ (где $\mathsf{P}(N = n) > 0$).

1.14. $\langle M + N\rangle = \langle M\rangle + \langle N\rangle$.

1.15. Сл. вел. M, N **независимы**, если $\mathsf{P}(m, n) = \mathsf{P}(M = m)\,\mathsf{P}(N = n)$.

1.16. У независимых сл. вел. условные распределения совпадают с соответствующими маргинальными.

1.17. $\langle MN\rangle = \langle M\rangle\ \langle N\rangle$ и $\mathsf{D}(M + N) = \mathsf{D}M + \mathsf{D}N$ для независимых сл. вел.

1.18. Производящей функцией распределения вероятности сл. вел. N называется функция $\Psi_N(z) = \left\langle z^N\right\rangle = \sum_{n\geqslant 0} z^n\,\mathsf{P}(N = n)$.

1.19. $\Psi'_N(1) = \langle N\rangle$, $\Psi''_N(1) = \langle N(N-1)\rangle$, $\mathrm{d}^k\Psi_N/\mathrm{d}z^k\big|_{z=1} = \left\langle N^{\underline{k}}\right\rangle$.

1.20. Производящей функцией моментов сл. вел. N называется функция $\Phi_N(s) = \Psi_N(\mathrm{e}^s) = \left\langle \mathrm{e}^{sN}\right\rangle = \sum_{n\geqslant 0} \mathrm{e}^{sn}\,\mathsf{P}(N = n)$.

1.21. $\mathrm{d}^k\Phi_N/\mathrm{d}s^k\big|_{s=0} = \left\langle N^k\right\rangle$.

1.22. Если сл. вел. M, N независимы, то $\Psi_{M+N}(z) = \Psi_M(z)\,\Psi_N(z)$.

1.23. Распределение вероятности суммы независимых сл. вел. выражается формулой $\mathsf{P}(M + N = k) = \sum_{0\leqslant n\leqslant k} \mathsf{P}(M = k - n)\mathsf{P}(N = n)$.

Благодаря формуле п. 1.22 распределение вероятности суммы независимых сл. вел. часто удается вычислить гораздо проще, чем при непосредственном применении формулы п. 1.23.

1.24. Пример. Пусть n раз производятся независимые случайные испытания, каждое из которых может завершиться успехом с вероятностью p и неудачей с вероятностью $\bar{p} = 1-p$. Число успехов в n испытаниях распределено по **биномиальному закону** $\mathsf{P}(k) = \binom{n}{k} p^k \bar{p}^{\,n-k}$, где $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ и $0 \leqslant k \leqslant n$; производящая функция $\Psi(z) = (1 + p(z - 1))^n$.

1.25. Пример. При $n \to \infty$, $p \to 0$, $np = \mu = \mathrm{const}$ производящая функция биномиального распределения сходится к производящей функции **распределения Пуассона** $\Psi(z) = \mathrm{e}^{\mu(z-1)}$, $\mathsf{P}(k) = \mu^k \mathrm{e}^{-\mu}/k!$.

1.26. Пример. Частица совершает **случайное блуждание** по бесконечному одномерному кристаллу. На каждом шаге частица может перепрыгнуть на один узел вправо с вероятностью p или на один узел влево с вероятностью $\bar{p}$ независимо от ее движения на предыдущих шагах. Смещение частицы за n шагов распределено как $2M_n - n$, где M_n — биномиальная сл. вел.

В последнем примере результатом случайного испытания можно считать не только смещение блуждающей частицы, но и всю **реализацию** случайного блуждания за n шагов. Возникает распределение вероятности не на множестве целых чисел, а на более абстрактном множестве всех возможных реализаций. Отдельные точки множества («пространства»), на котором задано распределение вероятности, называют **элементарными событиями**. Подмножества пространства элементарных событий называются **событиями**. Более подробно этот абстрактный подход рассматривается в лекции 8, а пространство реализаций случайного блуждания — в лекции 4.

2 Скалярные непрерывные случайные величины

2.1. В этой лекции рассматриваются скалярные сл. вел., которые принимают значения на всей непрерывной числовой прямой $(-\infty, \infty)$.

2.2. Будем обозначать скалярные непрерывные сл. вел. прописными латинскими буквами $X, Y, Z, \ldots$, а их значения — строчными буквами $x, y, z, \ldots$

2.3. Пример. Пусть сл. вел. X равномерно распределена на отрезке $(0, 1)$ (п. А.6). Тогда $\mathsf{P}(X = x) = 0$ при $0 < x < 1$, но $\mathsf{P}(a < X < b) = b - a$ при $0 \leqslant a < b \leqslant 1$.

Как показывает этот пример, вероятность каждого отдельного значения непрерывной сл. вел. может обращаться в нуль (ср. п. 1.6), но полная вероятность при этом остается равной единице. Поэтому распределение вероятности непрерывной сл. вел. характеризуется не вероятностями отдельных ее значений, как в п. 1.5, а **дифференциалами вероятности** $\mathsf{P}(x < X < x + \mathrm{d}x)$ или $\mathsf{P}(\mathrm{d}x)$.

2.4. Полезно выделять три типа дифференциалов вероятности. Если $\mathsf{P}(|X - x| < \frac{1}{2}\Delta)/\Delta \to p(x)$ при $\Delta \to 0$, будем писать $\mathsf{P}(\mathrm{d}x) = p(x)\,\mathrm{d}x$ и называть такой дифференциал вероятности **гладким**; если $\mathsf{P}(|X - x| < \frac{1}{2}\Delta)/\Delta^\alpha \to w(x)$ при $\Delta \to 0$, будем писать $\mathsf{P}(\mathrm{d}x) = w(x)\,|\mathrm{d}x|^\alpha$ и называть такой дифференциал вероятности **сингулярным** с показателем α; если $\mathsf{P}(|X - x| < \frac{1}{2}\Delta) \to p_0 > 0$ при $\Delta \to 0$, будем говорить, что в точке x расположен **атом** вероятности массы p_0.

2.5. Пример. Пусть $\mathcal{M}_0$ — отрезок $[0, 1]$, $\mathcal{M}_1$ — пара отрезков $[0, \frac{1}{3}]$ и $[\frac{2}{3}, 1]$, получаемая выбрасыванием из $\mathcal{M}_0$ его средней трети, и вообще $\mathcal{M}_{i+1}$ — совокупность отрезков, получаемых выбрасыванием средней трети из каждого отрезка, входящего в $\mathcal{M}_i$; пусть P_i — равномерное распределение вероятности на $\mathcal{M}_i$. В пределе при $i \to \infty$ получается распределение вероятности P_∞, в котором для разных дифференциалов вероятности либо $\mathsf{P}_\infty(\mathrm{d}x) = 0$, либо $\mathsf{P}_\infty(\mathrm{d}x) \sim |\mathrm{d}x|^\alpha$. (Каков порядок сингулярности α?)

«Гладкий» и атомарный типы можно понимать как предельные случаи сингулярного типа при $\alpha = 1$ и $\alpha = 0$. Целочисленные сл. вел. являются частным случаем атомарных. Приведенная классификация неполна: возможны и более сложные типы асимптотики $\mathsf{P}(\mathrm{d}x)$, чем степенной. Придать строгий математический смысл п. 2.4 — непростая задача, требующая использования средств теории меры.

2.6. Распределение вероятности $\mathsf{P}(\mathrm{d}x)$ должно удовлетворять двум условиям (ср. п. 1.7): (a) $\int_a^b \mathsf{P}(\mathrm{d}x) \geqslant 0$ для любых $a < b$; (b) $\int_{-\infty}^{\infty} \mathsf{P}(\mathrm{d}x) = 1$.

2.7. Распределение вероятности скалярной сл. вел. X однозначно задается **кумулятивной функцией распределения** (к. ф. р.) $F_X(x)$ или $F(x) = \mathsf{P}(X \leqslant x)$.

2.8. К. ф. р. не убывает и удовлетворяет условиям $F(-\infty) = 0$, $F(\infty) = 1$; если распределение гладкое, то $F(x) = \int_{-\infty}^{x} p(x)\,\mathrm{d}x$. «Атомы» вероятности соответствуют скачкам к. ф. р., в которых она согласно данному определению непрерывна слева.

2.9. **Мат. ожидание** сл. вел. X: $\langle X \rangle = \int_{-\infty}^{\infty} x\,\mathsf{P}(\mathrm{d}x)$ (ср. п. 1.8).

2.10. **Моменты** и **центральные моменты** (в частности, **дисперсию**) непрерывной сл. вел. определяют так же, как в дискретном случае (ср. п. 1.9): например, $\langle X^k \rangle = \int x^k\,\mathsf{P}(\mathrm{d}x)$. Квадратный корень из дисперсии называется **стандартным отклонением**: $\sigma_X = \sqrt{\langle (X - \langle X \rangle)^2 \rangle}$.

2.11. **Медиана** x_{med}: $F(x_{\mathrm{med}}) = \frac{1}{2}$ (или максимальное значение x, при котором $F(x) \leqslant \frac{1}{2}$, если в x_{med} находится атом вероятности). **Мода** $x_{\max}$: $p(x_{\max}) = \max_x p(x)$, если существует всюду конечная непрерывная ф. п. в. $p(x)$.

2.12. **Совместное распределение вероятности** двух сл. вел. X, Y задается к. ф. р. $F(x, y) = \mathsf{P}(X \leqslant x, Y \leqslant y)$, а для гладких распределений — ф. п. в. $p(x, y)$. **Маргинальное распределение** сл. вел. X определяется к. ф. р. $F_X(x) = F(x, \infty)$ или ф. п. в. $p_X(x) = \int_{-\infty}^{\infty} p(x, y)\,\mathrm{d}y$.

2.13. Сл. вел. X, Y **независимы**, если $F(x, y) = F_X(x)\,F_Y(y)$ или для гладкого распределения $p(x, y) = p_X(x)\,p_Y(y)$ (ср. п. 1.15).

Подробнее совместные распределения рассматриваются в лекции 6.

2.14. Свойства пп. 1.14 и 1.17 (аддитивность мат. ожидания и дисперсии, если сл. вел. независимы) выполнены и в непрерывном случае.

2.15. **Характеристической функцией** (х. ф.) распределения вероятности сл. вел. X называется функция $\varphi_X(s) = \langle \mathrm{e}^{\mathrm{i}sX} \rangle = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}sx}\,\mathsf{P}(\mathrm{d}x)$, которая для гладкого распределения вероятности совпадает с преобразование Фурье ф. п. в. $p(x)$.

2.16. Х. ф. равномерно непрерывна при $-\infty < s < \infty$.

2.17. $|\varphi_X(s)| \leqslant 1$ всюду, причем $\varphi_X(0) = 1$.

2.18. $\operatorname{Re}\varphi_X(-s) = \operatorname{Re}\varphi_X(s)$, $\operatorname{Im}\varphi_X(-s) = -\operatorname{Im}\varphi_X(s)$; если распределение сл. вел. X симметрично, то $\operatorname{Im}\varphi_X(s) \equiv 0$.

2.19. Функция $\varphi(s)$ является х. ф. некоторой сл. вел. тогда и только тогда, когда она непрерывна и положительно определена: для любых наборов (s_i) и комлексных чисел (ξ_i), $1 \leqslant i \leqslant k$, выполнено неравенство $\sum_{i,j} \varphi(s_i - s_j)\xi_i\xi_j^* \geqslant 0$, где ξ^* обозначает комплексное сопряжение (**теорема Бохнера**).

2.20. $\varphi_{aX+b}(s) = \mathrm{e}^{ibs}\,\varphi_X(as)$.

2.21. $\mathrm{d}^k\varphi_X/\mathrm{d}s^k\big|_{s=0} = \mathrm{i}^k\left\langle X^k\right\rangle$; $\varphi_X(s) = \sum_{k\geqslant 0}\left\langle X^k\right\rangle(\mathrm{i}s)^k/k!$, если данный ряд сходится.

Интегралы, выражающие моменты сл. вел. достаточно высоких порядков (п. 2.10), могут расходиться, если распределение вероятности слишком медленно убывает на бесконечности. В таких случаях х. ф. не имеет производных соответствующих порядков в нуле. Трудность другого типа возникает в том случае, если все моменты существуют, но ряд Тейлора х. ф. расходится; в этом случае распределение вероятности не может быть восстановлено однозначно по совокупности $\left\langle X^k\right\rangle$ (т. е. по своему ряду Тейлора).

2.22. Пример. Все моменты распределения Коши $p(x) = (\pi(1+x^2))^{-1}$, $\varphi(s) = \mathrm{e}^{-|s|}$ (см. п. А.8) расходятся.

2.23. Пример. Моменты логнормального распределения (см. п. А.13)

$$p(x) = \mathrm{e}^{-(\ln x)^2/2}/\sqrt{2\pi\sigma^2 x^2}$$

растут настолько быстро: $\left\langle X^k\right\rangle = \mathrm{e}^{k^2/2}$, что ряд Тейлора его х. ф. расходится. Такие же моменты имеет распределение $\tilde{p}(x) = p(x)(1+\sin(2\pi\ln x))$.

2.24. Теорема Карлемана. Если ряд $\sum_{k\geqslant 0}\left\langle X^{2k}\right\rangle^{-1/2k}$ расходится (т. е. моменты растут не очень быстро), то х. ф. аналитична при $s = 0$ и может быть восстановлена по своему ряду Тейлора.

2.25. Сравните определение х. ф. и формулы п. 2.21 с определением производящей функции моментов (пп. 1.20 и 1.21). Зачем в определение х. ф. введена мнимая единица?

2.26. В силу пп. 2.16 и 2.17 в окрестности $s = 0$, где $\varphi_X(s) > 0$, может быть определен **характеристический показатель** $\eta_X(s) = \ln\varphi_X(s)$ (однозначно, если положить $\eta_X(0) = 0$).

2.27. При сложении независимых сл. вел.

$$p_{X+Y}(u) = \int_{-\infty}^{\infty} p_Y(u-x)p_X(x)\,\mathrm{d}x \text{ (ср. п. 1.23)},$$
$$\varphi_{X+Y}(s) = \varphi_X(s)\varphi_Y(s) \text{ (ср. п. 1.22)}, \quad \eta_{X+Y}(s) = \eta_X(s) + \eta_Y(s).$$

2.28. Пример. Сл. вел. Y имеет **обобщенное распределение Пуассона**, если она является суммой случайного числа N независимых одинаково распределенных слагаемых X_i, причем N распределено по Пуассону с параметром μ. Если х. ф. сл. вел. X_i есть $\varphi(s)$, то х. ф. сл. вел. Y имеет вид $\varphi_Y(s) = \sum_k (\mu\varphi(s))^k \mathrm{e}^{-\mu}/k! = \mathrm{e}^{\mu(\varphi(s)-1)}$.

2.29. Коэффициенты $\langle\!\langle X^k \rangle\!\rangle$ разложения $\eta_X(s) = \sum_{n\geqslant 0} \langle\!\langle X^n \rangle\!\rangle (\mathrm{i}s)^n/n!$ в ряд по степеням s называются **кумулянтами** сл. вел. X (ср. п. 2.21).

2.30. $\langle\!\langle X \rangle\!\rangle = \langle X \rangle$, $\langle\!\langle X^2 \rangle\!\rangle = \mathsf{D}X$, $\langle\!\langle X^3 \rangle\!\rangle = \left\langle (X - \langle X \rangle)^3 \right\rangle$, $\langle\!\langle X^4 \rangle\!\rangle = \left\langle (X - \langle X \rangle)^4 \right\rangle - 3\langle\!\langle X^2 \rangle\!\rangle^2$.

В силу п. 2.27 при сложении независимых сл. вел. их кумулянты складываются.

Вероятности и х. ф. безразмерны; если сл. вел. размерна, то ее стандартное отклонение имеет такую же размерность, а ф. п. в. и аргумент ее х. ф. — обратную размерность.

2.31. «Обезразмеренные» кумулянты 3 и 4 порядков: **асимметрия** (англ. skewness) $\mathsf{S} = \langle\!\langle X^3 \rangle\!\rangle/\sigma^3$ и **эксцесс** (англ. kurtosis) $\mathsf{K} = \langle\!\langle X^4 \rangle\!\rangle/\sigma^4$. Величину $\mathsf{F} = \mathsf{K} + 3 = \left\langle (X - \langle X \rangle)^4 \right\rangle/\sigma^4$ называют **пологостью** (англ. flatness).

3 Закон больших чисел и центральная предельная теорема

В лекциях 3–5 рассматривается статистика совокупности независимых одинаково распределенных сл. вел. $X_1, X_2, \ldots, X_n$ при больших n (формально при $n \to \infty$).

3.1. Пусть каждая из сл. вел. $X_1, X_2, \ldots, X_n$ описывается к. ф. р. F или ф. п. в. p; говорят, что эти сл. вел. являются **независимыми, одинаково распределенными** (н. о. р.), когда $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = F(x_1)\ldots F(x_n)$ или $p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = p(x_1)\ldots p(x_n)$. Конкретные значения $x_1,\ldots,x_n$, полученные в результате однократного осуществления данной **последовательности независимых случайных испытаний**, называются **выборкой** из распределения, заданного к. ф. р. F.

3.2. Выборочным средним совокупности сл. вел. $X_1,\ldots,X_n$ называется сл. вел. $\frac{1}{n}S_n$, где $S_n = X_1 + \ldots + X_n$.

3.3. Если сл. вел. X обладает конечными мат. ожиданием $\langle X\rangle$ и дисперсией $\mathsf{D}X = \sigma^2$, то для любого $\varepsilon > 0$ выполнено **неравенство Чебышёва** $\mathsf{P}(|X - \langle X\rangle\,| > \varepsilon\sigma) \leqslant 1/\varepsilon^2$.

3.4. Если $\langle X_i\rangle = \mu$, $\mathsf{D}X_i = \sigma^2$, то $\left\langle \frac{1}{n}S_n\right\rangle = \mu$, $\mathsf{D}(\frac{1}{n}S_n) = \sigma^2/n$ и по неравенству Чебышёва $\mathsf{P}(|\frac{1}{n}S_n - \mu| > \varepsilon\sigma) \leqslant \frac{1}{n\varepsilon^2} \to 0$ при $n \to \infty$.

Таким образом, распределение выборочного среднего при $n \to \infty$ сосредотачивается около неслучайного значения μ (ср. п. 1.12). Говорят, что для выборочного среднего выполнен **закон больших чисел**.

3.5. Если сл. вел. X_i не имеют мат. ожидания, закон больших чисел может нарушаться: например, выборочное среднее н. о. р. сл. вел., распределенных по Коши с параметрами μ, Γ (п. A.8), само распределено по Коши с теми же параметрами при любом n, т. е. не сосредотачивается.

3.6. Говорят, что последовательность сл. вел. Y_k **сходится к сл. вел.** Y **по вероятности**, если $\mathsf{P}(\alpha < Y_k \leqslant \beta) \to \mathsf{P}(\alpha < Y \leqslant \beta)$ при $k \to \infty$ для любых $\alpha < \beta$. В частности, $F_{Y_k}(y) \to F_Y(y)$ в любой точке y, где F_Y непрерывна.

3.7. Сходимость выборочного среднего к мат. ожиданию в законе больших чисел — пример сходимости по вероятности.

3.8. Последовательность сл. вел. Y_k сходится к сл. вел. Y по вероятности тогда и только тогда, когда $\lim_{k\to\infty} \varphi_{Y_k}(s) = \varphi(s)$ при всех s, где функция $\varphi(s)$ непрерывна и является х. ф. сл. вел. Y.

3.9. Пример. Пусть $X_1, \ldots, X_n$ — независимые сл. вел., распределенные по Коши с одинаковыми параметрами μ, Γ и $S_n = X_1 + \ldots + X_n$. Тогда $\varphi_{S_n}(s) \to 0$ при $s \neq 0$, в то время как $\varphi_{S_n}(0) = 1$. Следовательно, предельная функция φ разрывна при $s = 0$. При этом сл. вел. S_n не сходятся по вероятности, поскольку ширина их распределения при $n \to \infty$ неограниченно возрастает: вероятность как бы «утекает на бесконечность».

3.10. Нормальное распределение, или **распределение Гаусса** с мат. ожиданием 0 и дисперсией 1: ф. п. в. $p(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}$, х. ф. $\varphi(s) = \mathrm{e}^{-\frac{1}{2}s^2}$ (случай мат. ожидания μ и дисперсии σ^2 см. в п. A.11).

3.11. Центральная предельная теорема: если $X_1, \ldots, X_n$ — н. о. р. сл. вел. с мат. ожиданием μ и дисперсией σ^2, то при $n \to \infty$ последовательность сл. вел. $Y_n = (X_1 + \ldots + X_n - \mu n)/(\sigma\sqrt{n})$ сходится по вероятности к гауссовой сл. вел. с мат. ожиданием 0 и дисперсией 1.

Действительно, $\langle Y_n\rangle = 0$, $\mathsf{D}Y_n = 1$ и $\varphi_{Y_n}(s) = (\varphi(\frac{s}{\sigma\sqrt{n}}))^n$, где φ — х. ф. сл. вел. $X_i - \mu$. Поскольку $\langle X_i - \mu\rangle = 0$, эта х. ф. при малых s имеет вид

$\varphi(s) = 1 - \frac{1}{2}\sigma^2 s^2 + o(s^2)$, и при $n \to \infty$

$$\varphi_{Y_n}(s) = \left(\varphi\left(\tfrac{s}{\sigma\sqrt{n}}\right)\right)^n = \left(1 - \tfrac{s^2}{2n} + o(n^{-1})\right)^n \to \mathrm{e}^{-\frac{1}{2}s^2}.$$

3.12. Более общий случай **закона больших чисел**: если х. ф. $\varphi(s)$ сл. вел. X дифференцируема в нуле, то выборочное среднее $\frac{1}{n}(X_1 + \ldots + X_n)$ сходится по вероятности к неслучайному значению $-\mathrm{i}\varphi'(0)$.

Центральная предельная теорема показывает, что $S_n - \mu n = Y_n \, \sigma\sqrt{n}$, где случайный коэффициент Y_n при $n \to \infty$ распределен асимптотически нормально и, в частности, может принимать сколь угодно большие значения. Данная оценка роста может быть уточнена следующим образом:

3.13. Закон повторного логарифма: $\limsup_{n\to\infty} \frac{S_n - \mu n}{\sigma\sqrt{2n \ln\ln n}} = 1$.

3.14. Пример. Пусть сл. вел. X_i распределены по показательному закону с параметром λ (п. A.9). Тогда S_n имеет гамма-распределение с параметрами λ, n (п. A.10). Сл. вел. $Y_n = (\lambda S_n - n)/\sqrt{n}$, определенные аналогично п. 3.11, сходятся по вероятности к симметрично распределенной гауссовой сл. вел., ф. п. в. которой убывает при $|y| \to \infty$ как $\mathrm{e}^{-\frac{1}{2}y^2}$. Тем не менее при каждом конечном n распределение сл. вел. Y_n асимметрично, причем его левый «хвост» нулевой: $\mathsf{P}(Y_n < -\sqrt{n}) = 0$, а правый убывает экспоненциально.

На практике вместо предельного перехода $n \to \infty$ рассматривают большие, но конечные значения n. В этом случае центральная предельная теорема описывает статистику лишь «типичных» значений выборочного среднего (находящихся на расстоянии порядка $1/\sqrt{n}$ от мат. ожидания). При выполнении условий центральной предельной теоремы (*конечность* дисперсии) значения, уклоняющиеся от мат. ожидания на расстояние порядка n, появляются очень редко. Они называются «большими уклонениями» и рассматриваются в лекции 4.

Пример сл. вел., распределенной по Коши (п. 3.5) показывает, что для сл. вел., дисперсия которой *бесконечна*, сходимость к предельному распределению возможна, но при иной нормировке выборочного среднего. Возникающие при этом предельные распределения не являются гауссовыми; они изучаются в лекции 5. Значения, далеко отклоняющиеся от медианы или моды, появляются в выборках из таких распределений с большой вероятностью и дают основной вклад в выборочное среднее.

4 Энтропия и большие уклонения

4.1. Пусть смещения X_i на каждом шаге случайного блуждания (ср. п. 1.26) являются н. о. р. сл. вел.: $\mathsf{P}(X_i = 1) = p_+$, $\mathsf{P}(X_i = 0) = p_0$, $\mathsf{P}(X_i = -1) = p_-$.

4.2. Пусть N_+ (N_-, N_0) есть число шагов, на которых блуждающая частица смещалась вправо (влево, оставалась на месте): $N_+ + N_0 + N_- = n$, $S_n = X_1 + \ldots + X_n = N_+ - N_-$. Согласно закону больших чисел, сл. вел. $\frac{1}{n} S_n$ (**выборочное среднее**) и $\nu_+ = \frac{1}{n} N_+$, $\nu_0 = \frac{1}{n} N_0$, $\nu_- = \frac{1}{n} N_-$ (**выборочные частоты**) сходятся по вероятности к константам $p_+ - p_-$, p_+, p_0 и p_-.

4.3. Теорема Макмиллана. Пусть $0 < \varepsilon \ll 1$; будем называть **типичными реализациями** такие реализации случайного блуждания, для которых $\max(|\nu_+ - p_+|, |\nu_0 - p_0|, |\nu_- - p_-|) < \varepsilon$. Пусть $N_{\text{тип}}$ — число типичных реализаций, а P_0 — вероятность одной из них. Тогда при достаточно больших $n > n(\varepsilon)$ справедливы оценки $|\frac{1}{n} \ln N_{\text{тип}} - \mathsf{H}_X| < \varepsilon$ и $|\frac{1}{n} \ln \mathsf{P}_0 - (-\mathsf{H}_X)| < \varepsilon$. Здесь $\mathsf{H}_X = -p_+ \ln p_+ - p_0 \ln p_0 - p_- \ln p_- > 0$ есть **энтропия** распределения вероятности (p_+, p_0, p_-).

Действительно, вероятность любой типичной реализации равна

$$p_+^{N_+} p_0^{N_0} p_-^{N_-} = \mathrm{e}^{N_+ \ln p_+ + N_0 \ln p_0 + N_- \ln p_-} \approx \mathrm{e}^{-n \mathsf{H}_X},$$

где приближенное равенство означает, что показатели экспонент совпадают с относительной ошибкой порядка ε.

Таким образом, с относительной точностью ε пространство элементарных событий случайного блуждания можно заменить на меньшее пространство, состоящее только из типичных реализаций, число которых определяется энтропией и между которыми вероятность распределена «почти равномерно».

4.4. В случае дискретного распределения вероятности энтропия имеет вид $\mathsf{H}_X = -\sum_i p_i \ln p_i$, где по определению $0 \ln 0 = 0$.

4.5. Энтропия H_X как функция вероятностей p_i строго выпукла вверх (пп. В.5, В.11).

4.6. $\mathsf{H}_X \geqslant 0$; $\mathsf{H}_X = 0$ тогда и только тогда, когда $p_{i_0} = 1$ и $p_i = 0$ при $i \neq i_0$ (детерминированная величина).

Следующие два свойства являются следствиями неравенства Иенсена (п. В.15).

4.7. Если число возможных исходов (положительных вероятностей p_i) конечно и равно k, то $\mathsf{H}_X \leqslant \ln k$; $\mathsf{H}_X = \ln k$ тогда и только тогда, когда $p_i = \frac{1}{k}$ при $1 \leqslant i \leqslant k$ (равномерное распределение).

4.8. Пусть X, Y — две дискретных сл. вел. с совместным распределением p_{mn} и маргинальными распределениями q_m, r_n. Если

$$\mathsf{H}_{X,Y} = -\sum_{m,n} p_{mn} \ln p_{mn}, \quad \mathsf{H}_X = -\sum_m q_m \ln q_m, \quad \mathsf{H}_Y = -\sum_n r_n \ln r_n,$$

то $\mathsf{H}_{X,Y} \leqslant \mathsf{H}_X + \mathsf{H}_Y$, причем равенство достигается тогда и только тогда, когда сл. вел. X, Y независимы (т. е. $p_{mn} = q_m r_n$).

4.9. Условной энтропией сл. вел. Y относительно сл. вел. X называется величина $\mathsf{H}_{Y|X} = \mathsf{H}_{X,Y} - \mathsf{H}_X$. **Взаимной информацией** сл. вел. X и Y называется величина $\mathsf{I}_{X,Y} = \mathsf{H}_X + \mathsf{H}_Y - \mathsf{H}_{X,Y}$.

С точки зрения теории информации энтропия случайного испытания — это мера неопределенности, существующей до испытания, а также количество информации, получаемое в его результате. Так, неопределенность $\mathsf{H}_{X,Y} - \mathsf{H}_X$ измерения сл. вел. Y после измерения сл. вел. X, вообще говоря, не превосходит неопределенность H_Y измерения Y без учета значения X.

4.10. Событие вида $a < \frac{1}{n} S_n < b$ называется **большим уклонением**, если интервал (a, b) не содержит значения $\left\langle \frac{1}{n} S_n \right\rangle = p_+ - p_-$.

4.11. Принцип больших уклонений для случайного блуждания. Логарифмическая асимптотика при $n \to \infty$ вероятности большого уклонения $P_n(a, b) = \mathsf{P}(a < \frac{1}{n} S_n < b)$ имеет вид

$$\lim_{n\to\infty} \frac{\ln P_n(a,b)}{n} = -\min_{\substack{\nu_+ + \nu_- + \nu_0 = 1, \\ a < \nu_+ - \nu_- < b}} \mathsf{I}(\,\nu : p\,),$$

где $\mathsf{I}(\,\nu : p\,) = \nu_+ \ln \frac{\nu_+}{p_+} + \nu_0 \ln \frac{\nu_0}{p_0} + \nu_0 \ln \frac{\nu_0}{p_0}$ — мера отклонения распределения (ν_+, ν_0, ν_-) от распределения вероятности (p_+, p_0, p_-).

Действительно, поскольку

$$P_n(a, b) = \sum_{a < \frac{N_+ - N_-}{n} < b} \frac{n!}{N_+!\, N_0!\, N_-!} p_+^{N_+} p_0^{N_0} p_-^{N_-}$$

из формулы Стирлинга $\ln n! = n(\ln n - 1) + o(n)$ следует, что

$$P_n(a, b) = \sum_{a < \frac{N_+ - N_-}{n} < b} \mathrm{e}^{n(\nu_+ \ln \frac{p_+}{\nu_+} + \nu_0 \ln \frac{p_0}{\nu_0} + \nu_- \ln \frac{p_-}{\nu_-}) + o(1)} = \sum_{a < \frac{N_+ - N_-}{n} < b} \mathrm{e}^{-n\mathsf{I}(\,\nu : p\,) + o(1)},$$

Из-за присутствия в показателе экспоненты большого множителя n определяющий вклад в вероятность большого уклонения вносят те совместимые с условием $a < \frac{1}{n} S_n < b$ реализации, для которых распределение выборочных частот (ν_+, ν_0, ν_-) обладает наименьшим отклонением $\mathsf{I}(\,\nu : p\,)$

от распределения (p_+, p_0, p_-) (ср. это свойство «типичности» с теоремой Макмиллана п. 4.3).

4.12. В случае дискретных распределений вероятности мерой отклонения распределения вероятности (ν_i) от распределения (p_i) служит величина $\mathsf{I}(\,\nu : p\,) = \sum_i p_i \frac{\nu_i}{p_i} \ln \frac{\nu_i}{p_i}$, которая может принимать значение $+\infty$, если $\nu_i > 0$ при $p_i = 0$.

4.13. Взаимная информация дискретных сл. вел. X и Y может быть представлена в виде $\mathsf{I}_{X,Y} = \sum_{m,n} q_m r_n \frac{p_{mn}}{q_m r_n} \ln \frac{p_{mn}}{q_m r_n} = \mathsf{I}(\,p : q \otimes r\,)$.

4.14. Если число возможных исходов дискретной сл. вел. X равно k, то данная мера отклонения от равномерного распределения $(u_i = 1/k)$ с точностью до замены знака и сдвига начала отсчета совпадает с энтропией: $\mathsf{I}(\,\nu : u\,) = \ln k - \mathsf{H}_X$. По этой причине данную меру отклонения называют **относительной энтропией** или **энтропией Кульбака–Лейблера–Санова**.

4.15. Относительная энтропия $\mathsf{I}(\,\nu : p\,)$ как функция вероятностей (ν_i) строго выпукла вниз (пп. B.5, B.11).

4.16. $\mathsf{I}(\,\nu : p\,) \geqslant 0$ для любых (ν_i), (p_i); $\mathsf{I}(\,\nu : p\,) = 0$ тогда и только тогда, когда распределения вероятности (ν_i) и (p_i) совпадают.

4.17. Вообще говоря, $\mathsf{I}(\,\nu : p\,) \neq \mathsf{I}(\,p : \nu\,)$.

4.18. Если $\nu_{kl} = \lambda_k \mu_l$, $p_{kl} = q_k r_l$, то $\mathsf{I}(\,\nu : p\,) = \mathsf{I}(\,\lambda : q\,) + \mathsf{I}(\mu : r\,)$.

4.19. Если распределения вероятности непрерывных сл. вел. X, Y удовлетворяют $\mathsf{P}_Y(\mathrm{d}x) = q(x)\mathsf{P}_X(\mathrm{d}x)$, то относительная энтропия определяется формулой $\mathsf{I}_{Y:X} = \int q(x) \ln q(x)\, \mathsf{P}_X(\mathrm{d}x)$ (ср. п. 4.12). В этом случае справедливы аналоги свойств пп. 4.15–4.18. Функция $q(x) = \mathsf{P}_Y(\mathrm{d}x)/\mathsf{P}_X(\mathrm{d}x)$ называется **производной Радона–Никодима** меры P_Y по мере P_X.

4.20. Относительная энтропия инвариантна относительно гладких взаимнооднозначных замен переменных: $\mathsf{I}_{X,Y} = \mathsf{I}_{f(X),f(Y)}$.

4.21. Взятая с обратным знаком относительная энтропия по отношению к (ненормируемому) равномерному распределению на всей числовой прямой $\mathsf{H}_X = -\int p(x) \ln p(x)\, \mathrm{d}x$ называется **дифференциальной энтропией**.

4.22. В отличие от энтропии дискретной сл. вел., дифференциальная энтропия может быть отрицательна.

4.23. В отличие от относительной энтропии, дифференциальная энтропия неинвариантна относительно замен переменных: $\mathsf{H}_{f(X)} \neq \mathsf{H}_X$.

4.24. Нормальное распределение (п. A.11) имеет максимальную дифференциальную энтропию среди всех распределений с фиксированными мат. ожиданием и дисперсией (ср. п. 4.7).

Многие распределения вероятности могут быть получены при помощи максимизации энтропии или дифференциальной энтропии при подходящих ограничениях.

4.25. Пусть X_i — н. о. р. непрерывные сл. вел. с распределением, заданным ф. п. в. $p(x)$ и вероятность большого уклонения $P_n(a,b)$ определена так же, как в п. 4.11. Тогда $\ln P_{n'+n''}(a,b) \geqslant \ln P_{n'}(a,b) + \ln P_{n''}(a,b)$, где $P_n(a,b) = \mathsf{P}(a < \frac{1}{n}S_n < b)$.

4.26. Принцип больших уклонений для последовательности н. о. р. сл. вел.: $\lim_{n\to\infty} \frac{1}{n} \ln P_n(a,b) = -\min_{a<x<b} I_X(x)$ при $n \to \infty$, где I_X называется **функцией роста** (rate function) или **функцией Крамера**.

4.27. $I_X(x) \geqslant 0$; $I_X(\langle X \rangle) = 0$.

4.28. Если ф. п. в. p достаточно быстро убывает на бесконечности, так что производящая функция моментов $\Phi_X(s) = \left\langle \mathrm{e}^{sX} \right\rangle = \int_{-\infty}^{\infty} \mathrm{e}^{sx} p(x)\,\mathrm{d}x$ конечна для значений s, лежащих в некотором интервале, функция Крамера и производящая функция моментов связаны преобразованием Лежандра (п. B.6): $\ln \Phi_X(s) = \max_x(sx - I_X(x))$ и $I_X(x) = \max_s(sx - \ln \Phi_X(s))$.

Как правило, функция Крамера является гладкой и в окрестности минимума $I_X(\langle X \rangle) = 0$ ведет себя квадратично, что приводит в центральной предельной теореме к появлению нормального распределения «малых» (порядка $\sqrt{n}$) уклонений выборочного среднего от мат. ожидания.

Принципами больших уклонений называются возникающие при изучении больших совокупностей сл. вел. утверждения типа $\lim_{n\to\infty} \frac{1}{n} P_n(A) = -\min_{x\in A} I(x)$, где A — некоторое (маловероятное) событие, а $I(x) > 0$ — подходящая функция Крамера (ср. пп. 4.11, 4.26). Смысл всевозможных принципов больших отклонений можно неформально охарактеризовать словами «даже невероятные события происходят наиболее вероятным образом».

5 Степенные законы и экстремальные значения

5.1. В данной лекции рассматриваются «степенные законы», т. е. распределения вероятности с ф. п. в., убывающей на бесконечности степенным образом: $p(x) \sim |x|^{-(1+\alpha)}$ при $|x| \to \infty$, где $0 < \alpha < 2$. Параметр α называется **показателем** степенного закона.

5.2. $F(x) \sim |x|^{-\alpha}$ при $x \to -\infty$; $F(x) \sim 1 - x^{-\alpha}$ при $x \to \infty$.

5.3. Дисперсия степенного закона при $0 < \alpha \leqslant 2$ бесконечна.

5.4. Пример. Пусть $X_1, X_2, \ldots$ — н. о. р. сл. вел. с симметричной ф. п. в. и х. ф. вида

$$\bar{p}(x) = \begin{cases} A/2, & |x| \leqslant 1, \\ A/(2|x|^{\alpha+1}), & |x| > 1, \end{cases} \quad \text{где } A = \tfrac{\alpha}{\alpha+1};$$

$$\bar{\varphi}(s) = 1 - A|s|^{\alpha} \int_s^{\infty} \tfrac{1-\cos\xi}{\xi^{\alpha+1}}\, \mathrm{d}\xi + A\left(\tfrac{\sin s}{s} - 1\right);$$

поэтому $\bar{\varphi}(s) = 1 - \Gamma|s|^{\alpha} + O(s^2)$ при малых $|s|$, если $0 < \alpha < 2$ (при $\alpha = 2$ интеграл логарифмически расходится в нуле). Тогда при $0 < \alpha < 2$ х. ф. сл. вел. $(X_1 + \ldots + X_n)/n^{1/\alpha}$ имеет вид $\varphi(s) = \left(1 - \Gamma|s|^{\alpha}/n + O(s^2/n^{2/\alpha})\right)^n$ и при $n \to \infty$ стремится к $\varphi_{\alpha}(s) = \mathrm{e}^{-\Gamma|s|^{\alpha}}$. При $\alpha = 2$ аналогичный результат может быть получен для нормирующего множителя $1/(n \ln n)^{1/2}$ вместо $1/n^{1/2}$.

5.5. Распределение вероятности с х. ф. $\varphi_{\alpha}(s) = \mathrm{e}^{-\Gamma|s|^{\alpha}}$ называется **симметричным распределением Леви–Парето** с параметром $\Gamma > 0$ и показателем α ($0 < \alpha < 2$).

5.6. Если $X_1, X_2, \ldots$ — н. о. р. сл. вел. с асимметричной ф. п. в.

$$\bar{p}(x) = \begin{cases} A/2, & |x| \leqslant 1, \\ A(1+\beta)/(2x^{\alpha+1}), & x > 1, \\ A(1-\beta)/(2|x|^{\alpha+1}), & x < -1, \end{cases}$$

где $|\beta| \leqslant 1$, то х. ф. сл. вел. $(X_1 + \ldots + X_n)/n^{1/\alpha}$ при $n \to \infty$ стремится к х. ф. **асимметричного распределения Леви–Парето**

$$\varphi_{\alpha,\beta}(s) = \mathrm{e}^{-\Gamma|s|^{\alpha}(1 + \mathrm{i}\beta\,\mathrm{tg}(\pi\alpha/2)\,\mathrm{sign}\, s)} \quad \text{при } 0 < \alpha < 2,\ \alpha \neq 1,$$

$$\varphi_{1,\beta}(s) = \mathrm{e}^{-\Gamma|s|(1 + \mathrm{i}(2\beta/\pi)\ln|s|\,\mathrm{sign}\, s)} \quad \text{при } \alpha = 1.$$

Сравнивая п. 5.4 с доказательством центральной предельной теоремы в п. 3.11, можно заметить, что распределения Леви–Парето играют среди степенных законов такую же роль, как нормальное распределение — среди распределений с конечной дисперсией: к ним сходятся распределения сумм н. о. р. сл. вел. при подходящей нормировке (**обобщенная центральная предельная теорема**).

Явный вид ф. п. в. симметричного распределения Леви–Парето известен только при $\alpha = 1$ (распределение Коши, п. A.8). При $\alpha = 2$ получается распределение Гаусса (п. A.11). Вполне асимметричное ($\beta = 1$) распределение Леви–Парето при $\alpha = \frac{1}{2}$ имеет ф. п. в. $p(x) = \frac{1}{\sqrt{2\pi x^3}}\mathrm{e}^{-\frac{1}{2x}}$ ($x > 0$; п. A.19).

5.7. Две сл. вел. X, Y называются **равными по распределению** (обозначение: $X \overset{\Delta}{=} Y$), если их распределения вероятности совпадают.

5.8. Распределение вероятности сл. вел. X называется **устойчивым**, если $X_1+\ldots+X_n \overset{\Delta}{=} c_n X + b_n$ $(c_n > 0)$ для набора н. о. р. сл. вел. $X_1, X_2, \ldots$, закон распределения которых совпадает с законом сл. вел. X (будем кратко называть такие сл. вел. **реализациями** сл. вел. X). Распределение называется **строго устойчивым**, если $b_n \equiv 0$.

5.9. Пусть X', X'' — н. о. р. реализации сл. вел. X со строго устойчивым распределением; тогда $c_{m+n}X \overset{\Delta}{=} c_m X' + c_n X''$, $c_{mn}X \overset{\Delta}{=} c_m c_n X$.

5.10. Общим решением уравнения $c_{mn} = c_m c_n$ является $c_n = 1/n^{1/\alpha}$.

5.11. Все устойчивые распределения, кроме нормального, являются распределениями Леви–Парето (доказательство аналогично пп. 5.4, 5.6).

5.12. Распределение вероятности сл. вел. X называется **безгранично делимым**, если $X \overset{\Delta}{=} Y_{n,1} + \ldots + Y_{n,n}$ при любом $n > 1$, где $Y_{n,1}, \ldots, Y_{n,n}$ — н. о. р. сл. вел.

5.13. Все устойчивые распределения безгранично делимы.

5.14. Пример. Обобщенное распределение Пуассона (п. 2.28) с параметром μ безгранично делимо: в качестве $Y_{n,i}$ можно выбрать обобщенные пуассоновские сл. вел. с параметром μ/n.

5.15. Все безгранично делимые распределения являются пределами последовательностей обобщенных пуассоновских распределений.

Пусть $X_1, X_2, \ldots$ — н. о. р. сл. вел., распределенные по степенному закону с показателем $0 < \alpha < 1$, Z_α — сл. вел., распределенная по соответствующему закону Леви–Парето. Последовательность выборочных средних $Y_n = \frac{1}{n}(X_1 + \ldots + X_n)$ при больших n ведет себя как $n^{\frac{1}{\alpha}-1} Z_\alpha$, т. е. эффективная ширина распределения выборочного среднего *увеличивается* с ростом n (**явление Мандельброта**). Причина этого в том, что определяющий вклад в значение выборочного среднего вносит наибольшее слагаемое $X_{(n)} = \max(X_1, \ldots, X_n)$ (ср. ниже п. 5.17).

5.16. Характеристическим наибольшим значением выборки из n реализаций сл. вел. с непрерывной к. ф. р. $F(x)$, называется решение $x_{(n)}$ уравнения $\mathsf{P}(X > x_{(n)}) = 1 - F(x_{(n)}) = \frac{1}{n}$ (ср. с определением медианы в п. 2.11).

В типичной последовательности из n случайных испытаний событие $X > x_{(n)}$, имеющее вероятность $1/n$, реализуется один раз. Поэтому характеристическое наибольшее значение является «типичной оценкой снизу» сл. вел. $X_{(n)}$.

5.17. Пример. Если «хвост» к. ф. р. имеет вид $F(x) = 1 - Ax^{-\alpha}$, где $\alpha > 0$, то $x_{(n)} = (An)^{\frac{1}{\alpha}}$.

5.18. Пример. Если сл. вел. ограничена сверху ($X \leqslant 0$) и в окрестности нуля ее к. ф. р. имеет вид $F(x) = 1 - A|x|^{\alpha}$ при $x \leqslant 0$, $F(x) = 1$ при $x > 0$, где $\alpha > 0$, то $x_{(n)} = -(An)^{-\frac{1}{\alpha}}$. Данный случай сводится к предыдущему заменой $x \mapsto -1/x$.

5.19. Пример. Если «хвост» к. ф. р. имеет вид $F(x) = 1 - \mathrm{e}^{-\lambda x}$, то $x_{(n)} = \frac{1}{\lambda} \ln n$.

5.20. Пусть $X_1, X_2, \ldots$ — н. о. р. сл. вел. с непрерывной к. ф. р. $F(x)$. Поскольку $\mathsf{P}(X_{(n)} \leqslant x) = \mathsf{P}(X_1 \leqslant x, \ldots, X_n \leqslant x)$, к. ф. р. сл. вел. $X_{(n)}$ равна $F_{(n)}(x) = (F(x))^n$.

5.21. В случае п. 5.17 при $n \to \infty$ к. ф. р. сл. вел. $Y = X_{(n)}/x_{(n)}$ стремится к пределу $F(y) = \mathrm{e}^{-y^{-\alpha}}$ при $y \geqslant 0$, $F(y) = 0$ при $y < 0$ (**распределение Фреше**, п. А.14); в случае п. 5.18 аналогичная замена дает предел $F(y) = \mathrm{e}^{-|y|^{\alpha}}$ при $y \leqslant 0$, $F(y) = 1$ при $y > 0$ (**распределение Вейбулла**, п. А.15); в случае п. 5.19 к. ф. р. сл. вел. $Y = X_{(n)} - x_{(n)}$ стремится к пределу $F(y) = \mathrm{e}^{-\mathrm{e}^{-\lambda y}}$ (**распределение Гумбеля**, п. А.16).

5.22. Распределения предельных значений, перечисленные в предыдущем пункте, обладают свойстом «устойчивости» относительно максимизации: $\max(X_1, \ldots, X_n) \stackrel{\Delta}{=} c_n X + b_n$.

5.23. Распределения п. 5.21 исчерпывают все возможные предельные распределения наибольших значений так же, как распределения Леви–Парето и нормальное — все возможные типы предельных распределений сумм (**теорема Фишера–Типпета–Гнеденко**).

6 Случайные векторы и корреляции случайных величин

6.1. Векторные сл. вел. (сл. векторы) обозначаются прописными полужирными латинскими буквами $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \ldots$, а их значения — соответствующими полужирными строчными буквами $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots$

6.2. Распределение вероятности случайного n-мерного вектора $\boldsymbol{X}$ — это **совместное распределение вероятности** его компонент, сл. вел. $X_1, \ldots, X_n$, задаваемое к. ф. р. $F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathsf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n)$ или для гладкого распределения ф. п. в. $p_{\boldsymbol{X}}(\boldsymbol{x})$ (ср. п. 2.12; нижний индекс $\boldsymbol{X}$ часто можно опустить).

6.3. К. ф. р. сл. вектора обладает следующими свойствами:

$$F(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = 0 \text{ при любом } 1 \leqslant i \leqslant n;$$
$$F(\infty, \infty, \dots, \infty) = 1;$$
$$\textstyle\sum_{(\varepsilon_1, \dots, \varepsilon_n)} (\varepsilon_1 1) \dots (\varepsilon_n 1) F(x_1^{\varepsilon_1}, \dots, x_n^{\varepsilon_n}) \geqslant 0,$$

где суммирование проводится по 2^n наборам знаков $\varepsilon_i = \pm$.

По последнему свойству $\mathsf{P}(x_1^- < X_1 \leqslant x_1^+, \dots, x_n^- < X_n \leqslant x_n^+) \geqslant 0$. Данное свойство легче понять на примере $n = 2$: должно выполняться неравенство $F(x_1^+, x_2^+) - F(x_1^+, x_2^-) - F(x_1^-, x_2^+) + F(x_1^-, x_2^-) \geqslant 0$.

6.4. Обратно, если функция $F(\boldsymbol{x})$ удовлетворяет трем свойствам п. 6.3, она однозначно определяет распределение вероятности некоторого сл. вектора.

6.5. Маргинальные распределения для отдельных компонент сл. вектора имеют ф. п. в. $p_{X_i}(x_i) = \int p(x_1, \dots, x_n)\, \mathrm{d}x_1 \dots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \dots \mathrm{d}x_n$ и к. ф. р. $F_{X_i}(x_i) = F(\infty, \dots, x_i, \dots, \infty)$. Маргинальные распределения можно также определять для любых совокупностей компонент сл. вектора; их также называют **проекциями** распределения сл. вектора $\boldsymbol{X}$ на соответствующие координатные подпространства.

6.6. Функция условной плотности вероятности сл. вектора $\boldsymbol{X}$ при условии, что $y_1 < Y_1 < y_1 + \mathrm{d}y_1$, $y_2 < Y_2 < y_2 + \mathrm{d}y_2$, ...: $p_{\boldsymbol{X}|\boldsymbol{Y}}(\, \boldsymbol{x} \mid \boldsymbol{y} \,) = p(\boldsymbol{x}, \boldsymbol{y}) / p_{\boldsymbol{Y}}(\boldsymbol{y})$.

6.7. Пример. Пусть (X, Y) — сл. вектор, равномерно распределенный на единичном квадрате $(0,1) \times (0,1)$. Условная плотность $p(\, x \mid X = Y \,)$ имеет разный вид при разной интерпретации условия $X = Y$: если $Z_1 = X - Y$, то $p_{X|Z_1}(\, x \mid 0 \,) = 1$, а если $Z_2 = X/Y$, то $p_{X|Z_2}(\, x \mid 1 \,) = 2x$.

Причина этого парадокса в том, что в действительности условная плотность определяется относительно событий не нулевой вероятности, таких как $X = Y$, а бесконечно малой вероятности ($0 < X - Y < \mathrm{d}z_1$ или $1 < X/Y < 1 + \mathrm{d}z_2$).

6.8. Компоненты $X_1, \dots, X_n$ сл. вектора $\boldsymbol{X}$ называются **независимыми в совокупности**, если их совместное распределение вероятности распадается в произведение индивидуальных распределений отдельных компонент: $F(\boldsymbol{x}) = F_{X_1}(x_1) \dots F_{X_n}(x_n)$, $p(\boldsymbol{x}) = p_{X_1}(x_1) \dots p_{X_n}(x_n)$ (ср. пп. 2.13, 3.1).

6.9. Пример. Пусть сл. вел. X_1, X_2, X_3 принимают значения 0 или 1 с вероятностями $p_{000} = pq$, $p_{110} = \bar{p}q$, $p_{101} = \bar{p}\bar{q}$, $p_{011} = p\bar{q}$, $p_{001} = p_{010} = p_{100} = p_{111} = 0$, где $0 < p, q < 1$, $p + \bar{p} = q + \bar{q} = 1$. Тогда X_1, X_2, X_3 не являются независимыми в совокупности, но маргинальное распределение любых двух компонент распадается в произведение независимых распределений $(p, \bar{p})$ и $(q, \bar{q})$.

Данный пример показывает, что попарной независимости набора сл. вел. (т. е. независимости любой пары сл. вел. в соответствующем маргинальном распределении) недостаточно для их независимости в совокупности.

6.10. Моменты сл. вектора определяются как мат. ожидания произведений его компонент: $\langle X_{i_1} \dots X_{i_k} \rangle$, где среди индексов $i_1, \dots, i_k$ могут быть повторяющиеся. Число k сомножителей в этом произведении называется **порядком** момента. Аналогично определяются **центральные моменты**: $\langle (X_{i_1} - \langle X_{i_1} \rangle) \dots (X_{i_k} - \langle X_{i_k} \rangle) \rangle$.

6.11. Аналогом дисперсии для сл. вектора является **матрица коэффициентов ковариации** $\Gamma_{ij} = \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle$ (справедлива формула $\Gamma_{ij} = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle$).

6.12. Матрица ковариации обладает свойствами симметрии: $\Gamma_{ij} = \Gamma_{ji}$ для всех i, j и положительной определенности: $\sum_{ij} \Gamma_{ij} \xi_i \xi_j \geqslant 0$ для любого вектора $\boldsymbol{\xi}$ (ср. п. 2.19). Последнее следует из равенства $\sum_{i,j} \Gamma_{ij} \xi_i \xi_j = \langle \left(\sum_{ij} \xi_i (X_i - \langle X_i \rangle) \right)^2 \rangle$.

6.13. Матрица ковариации n-мерного сл. вектора $\boldsymbol{X}$ имеет n взаимно ортогональных собственных векторов с неотрицательными собственными значениями. Проекции сл. вектора $\boldsymbol{X}$ на направления этих собственных векторов называются его **главными компонентами**.

6.14. Коэффициентами корреляции компонент сл. вектора называются безразмерные величины $\rho_{ij} = \Gamma_{ij} / \sqrt{\Gamma_{ii} \Gamma_{jj}}$. Если $\rho_{ij} > 0$ ($\rho_{ij} < 0$), говорят, что сл. вел. X_i, X_j **коррелированы** (**антикоррелированы**).

6.15. Коэффициенты корреляции не превосходят по модулю единицы.

6.16. Х. ф. сл. вектора: $\varphi_{\boldsymbol{X}}(\boldsymbol{s}) = \left\langle \mathrm{e}^{\mathrm{i}(\boldsymbol{s} \cdot \boldsymbol{X})} \right\rangle = \int \mathrm{e}^{\mathrm{i} \sum_i s_i X_i} \, \mathsf{P}(\mathrm{d}\boldsymbol{x})$.

6.17. $\varphi_{A\boldsymbol{X}+\boldsymbol{b}}(\boldsymbol{s}) = \left\langle \mathrm{e}^{\mathrm{i}(\boldsymbol{s} \cdot (A\boldsymbol{X}+\boldsymbol{b}))} \right\rangle = \mathrm{e}^{\mathrm{i}(\boldsymbol{s} \cdot \boldsymbol{b})} \varphi_{\boldsymbol{X}}(A^\mathsf{T} \boldsymbol{s})$ (ср. п. 2.20; заметим, что матрица в аргументе х. ф. $\varphi_{\boldsymbol{X}}$ транспонирована).

6.18. Сложение независимых сл. векторов соответствует умножению их х. ф. (ср. п. 2.27).

6.19. $\partial^k \varphi_{\boldsymbol{X}} / \partial s_{i_1} \dots \partial s_{i_k} \big|_{\boldsymbol{s}=\boldsymbol{0}} = \mathrm{i}^k \langle X_{i_1} \dots X_{i_k} \rangle$ (ср. п. 2.21). Аналогичное представление для кумулянтов $\langle\!\langle X_{i_1} \dots X_{i_k} \rangle\!\rangle$ возникает при дифференцировании характеристического показателя $\eta_{\boldsymbol{X}} = \ln \varphi_{\boldsymbol{X}}$ (ср. п. 2.29).

6.20. $\Gamma_{ij} = \langle\!\langle X_i X_j \rangle\!\rangle$.

6.21. Обозначим $\langle i_1, \dots, i_k \rangle = \langle X_{i_1} \dots X_{i_k} \rangle$, $\langle\!\langle i_1, \dots, i_k \rangle\!\rangle = \langle\!\langle X_{i_1} \dots X_{i_k} \rangle\!\rangle$. Тогда $\langle 1 \rangle = \langle\!\langle 1 \rangle\!\rangle$, $\langle 1, 2 \rangle = \langle\!\langle 1, 2 \rangle\!\rangle + \langle\!\langle 1 \rangle\!\rangle \langle\!\langle 2 \rangle\!\rangle$, $\langle 1, 2, 3 \rangle = \langle\!\langle 1, 2, 3 \rangle\!\rangle + \langle\!\langle 1, 2 \rangle\!\rangle \langle\!\langle 3 \rangle\!\rangle + \langle\!\langle 1, 3 \rangle\!\rangle \langle\!\langle 2 \rangle\!\rangle + \langle\!\langle 2, 3 \rangle\!\rangle \langle\!\langle 1 \rangle\!\rangle + \langle\!\langle 1 \rangle\!\rangle \langle\!\langle 2 \rangle\!\rangle \langle\!\langle 3 \rangle\!\rangle$; вообще момент $\langle i_1, \dots, i_k \rangle$ выражается через сумму произведений кумулянтов, соответствующих всевозможным разбиениям набора индексов $(i_1, \ldots, i_k)$ на произвольные подмножества (**кластерное разложение**).

Доказательство следует из тождества $\mathrm{d}(\exp \eta)/\mathrm{d}s = \exp \eta \, \mathrm{d}\eta/\mathrm{d}s$. Если при этом все индексы выбирать совпадающими, получатся формулы, обратные соотношениям из п. 2.30.

6.22. Сл. вектор называется **гауссовым**, или распределенным по **нормальному закону**, если его проекция на любое направление является гауссовой сл. вел..

6.23. Образ гауссова сл. вектора под действием любого линейного преобразования остается гауссовым.

6.24. Если $\boldsymbol{X}$ — гауссов сл. вектор с матрицей ковариации Γ и мат. ожиданием $\langle \boldsymbol{X} \rangle = \boldsymbol{m}$, то $\varphi_{\boldsymbol{X}}(\boldsymbol{s}) = \mathrm{e}^{\mathrm{i}(\boldsymbol{s}\cdot\boldsymbol{m}) - \frac{1}{2}\sum_{i,j}\Gamma_{ij}s_i s_j}$.

6.25. Матрица ковариации n-мерного гауссова сл. вектора $\boldsymbol{X}$ может быть сингулярна ($\det \Gamma = 0$), если $\boldsymbol{X}$ с вероятностью 1 принимает значения в некотором подпространстве размерности, меньшей n. В этом случае гауссов сл. вектор называется **вырожденным**.

6.26. Если $\det \Gamma > 0$, то ф. п. в. n-мерного гауссова сл. вектора $\boldsymbol{X}$ из п. 6.24 имеет вид

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det \Gamma}} \mathrm{e}^{-\frac{1}{2}\sum_{i,j}\hat{\Gamma}_{ij}(x_i - m_i)(x_j - m_j)},$$

где $\hat{\Gamma}$ — матрица, обратная к Γ.

6.27. Пример. Пусть σ_1, σ_2, ρ — стандартные отклонения и коэффициент корреляции сл. вел. X_1, X_2, образующих (если $\rho \neq \pm 1$, невырожденный) двумерный гауссов сл. вектор с нулевым мат. ожиданием. Тогда

$$\Gamma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \hat{\Gamma} = \frac{1}{1-\rho^2}\begin{pmatrix} \sigma_1^{-2} & -\rho/(\sigma_1\sigma_2) \\ -\rho/(\sigma_1\sigma_2) & \sigma_2^{-2} \end{pmatrix};$$

$$\varphi(s_1, s_2) = \mathrm{e}^{-\frac{1}{2}(\sigma_1^2 s_1^2 + \sigma_2^2 s_2^2 + 2\rho\sigma_1\sigma_2 s_1 s_2)};$$

$$p(x_1, x_2) = \tfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\tfrac{1}{2(1-\rho^2)}\left((\tfrac{x_1}{\sigma_1})^2 + (\tfrac{x_2}{\sigma_2})^2 - 2\rho\tfrac{x_1 x_2}{\sigma_1\sigma_2}\right)\right);$$

$$p_{X_1|X_2}(\, x_1 \mid x_2 \,) = \tfrac{1}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}} \exp\left(-\tfrac{1}{2\sigma_1^2(1-\rho^2)}\left(x - \rho\tfrac{\sigma_1}{\sigma_2}y\right)^2\right).$$

7 Статистические оценки и критерии

7.1. Пусть $X_1, \ldots, X_n$ — н. о. р. сл. вел., распределенные по закону $\mathsf{P}(\mathrm{d}x)$, который в математической статистике принято называть распределением **генеральной совокупности** (population).

7.2. **Выборкой** (sample) объема n называются значения $x_1, \ldots, x_n$, которые данные сл. вел. приняли при однократной реализации последовательности n независимых случайных испытаний.

7.3. Атомарное распределение вероятности с атомами веса $\frac{1}{n}$ в каждой из точек $x_1, \ldots, x_n$ называется **распределением выборки**. В скалярном случае целесообразно определять к. ф. р. выборки $\bar{F}(x) = \#\{\, i \mid x_i < x \,\}/n$.

7.4. Мат. ожидание относительно распределения выборки обозначается $\overline{f(x)} = \frac{1}{n} \sum_i f(x_i)$ (ср. с мат. ожиданием относительно генеральной совокупности $\langle f(X) \rangle = \int f(x) \, \mathsf{P}(\mathrm{d}x)$).

7.5. Любая функция выборки $g(x_1, \ldots, x_n)$ называется **статистикой** (statistic).

7.6. Совместное распределение сл. вел. $X_1, \ldots, X_n$ (произведение n копий распределения генеральной совокупности, ср. п. 6.8) называется **выборочным распределением**. Для любой статистики g распределение вероятности сл. вел. $g(X_1, \ldots, X_n)$ называется **выборочным распределением** этой статистики.

Одной из основных задач математической статистики является **оценивание параметров**.

7.7. **Пример.** Если распределение генеральной совокупности обладает мат. ожиданием μ, его оценкой может служить среднее арифметическое выборки $\overline{x}$. По закону больших чисел (п. 3.12) выборочное распределение соответствующей сл. вел. (выборочного среднего, п. 3.2) сосредотачивается на μ.

7.8. **Пример.** Если распределение генеральной совокупности обладает дисперсией σ^2, ее оценкой может служить статистика $\overline{(x - \overline{x})^2}$, распределение которой сосредоточивается на σ^2, причем $\left\langle \overline{(X - \overline{X})^2} \right\rangle = \frac{n-1}{n} \sigma^2$. **Дисперсией выборки** называется статистика $s_n^2 = n \, \overline{(x - \overline{x})^2}/(n-1)$.

7.9. Пусть распределение генеральной совокупности зависит от параметра θ. Статистика $\hat{\theta}(x_1, \ldots, x_n)$ (точнее, семейство таких статистик с $n = 1, 2, \ldots$) называется **состоятельной оценкой** (consistent estimator) этого параметра, если ее выборочное распределение сосредотачивается на θ при $n \to \infty$.

7.10. Статистика $\hat{\theta}$ называется **несмещенной оценкой** (unbiased estimator) параметра θ генеральной совокупности, если $\langle \hat{\theta} \rangle = \theta$.

7.11. Для распределения, обладающего ф. п. в. $p(\, x \mid \theta \,)$, **логарифмической функцией правдоподобия** (logarithmic likelihood function) называется $L(\theta \mid x) = \ln p(x \mid \theta)$, а величина $\partial L(\theta \mid x)/\partial\theta$ — **информантом**. Здесь θ является независимой переменной, а x — параметром.

7.12. Информант инвариантен относительно замены переменных (ср. п. 4.20).

7.13. Оценкой наибольшего правдоподобия (maximal likelihood) параметра θ называется такая статистика $\hat\theta_{\mathrm{ml}}(x_1, \ldots, x_n)$, что $L(\theta \mid x_1) + \ldots + L(\theta \mid x_n)$ достигает максимума по θ при $\theta = \hat\theta_{\mathrm{ml}}(x_1, \ldots, x_n)$.

7.14. Пусть $\boldsymbol{x} = (x_1, \ldots, x_n)$, $p(\boldsymbol{x} \mid \theta) = p(x_1 \mid \theta) \ldots p(x_n \mid \theta)$. Для произвольной несмещенной оценки $\hat\theta(\boldsymbol{x})$ выполнено **неравенство Рао–Крамера**:

$$\mathsf{D}\hat\theta \int\left(\sum_i \frac{\mathrm{d}L(\theta \mid x_i)}{\mathrm{d}\theta}\right)^2 p(\boldsymbol{x} \mid \theta)\,\mathrm{d}\boldsymbol{x} = n\,\mathsf{D}\hat\theta \int\left(\frac{\mathrm{d}L(\theta \mid x)}{\mathrm{d}\theta}\right)^2 p(x \mid \theta)\,\mathrm{d}x \geqslant 1.$$

7.15. Величина $\langle(\partial L/\partial\theta)^2\rangle = \int(\partial L/\partial\theta)^2\, p(x \mid \theta)\,\mathrm{d}x$ называется **информацией Фишера**. Имеет место следующая связь информации Фишера с относительной энтропией:

$$\int\left(\frac{\mathrm{d}L(\theta \mid x)}{\mathrm{d}\theta}\right)^2 p(x \mid \theta)\,\mathrm{d}x = \left.\frac{\mathrm{d}^2 \mathsf{I}_{\theta',\theta}}{\mathrm{d}\theta'^2}\right|_{\theta'=\theta}$$

7.16. Оценка, обращающая неравенство Рао–Крамера в равенство, называется **эффективной** (efficient estimator); она обладает наименьшей возможной дисперсией.

7.17. Если эффективная оценка существует, то она является оценкой наибольшего правдоподобия.

Другой основной задачей математической статистики является проверка статистических гипотез. Пусть фиксирована некоторая **нулевая гипотеза** (null hypothesis) и необходимо выяснить, может ли она быть опровергнута выборкой $x_1, \ldots, x_n$. Общий метод состоит в вычислении некоторой статистики $d(x_1, \ldots, x_n)$, измеряющей степень «несовпадения» выборки с предсказанием гипотезы, и оценке вероятности превышения этой статистикой заданного порога.

7.18. Пусть $0 < \varepsilon < 1$; **ε-порог значимости** (significance level) для статистики d есть такое число D_ε, что $\mathsf{P}(d(X_1, \ldots, X_n) > D_\varepsilon) = \varepsilon$.

7.19. Если для данной выборки $d(x_1, \ldots, x_n) > D_\varepsilon$, говорят, что нулевая гипотеза отвергнута **на уровне значимости** ε.

Важно понимать, что если нулевая гипотеза не отвергается выборкой, это не может служить доказательством ее правильности.

Пусть нулевая гипотеза состоит в том, что неизвестная генеральная совокупность, из которой получена выборка, имеет распределение $\mathsf{P}_0(\mathrm{d}x)$. В этом случае используется статистика, выражающая некоторую меру отклонения распределения выборки от распределения P_0.

7.20. Пример. Пусть сл. вел. X скалярна, $\bar{F}(x)$ — к. ф. р. выборки, $F_0(x)$ — к. ф. р. распределения P_0. Статистика $d_n = \max_x |\bar{F}(x) - F_0(x)|$ называется **статистикой Колмогорова–Смирнова**. Для определения ε-порога значимости при достаточно больших n используется распределение п. A.18, к которому, если нулевая гипотеза верна, сходится при $n \to \infty$ выборочное распределение сл. вел. $\sqrt{n}\, d_n$ для любого непрерывного распределения P_0.

7.21. Пример. Пусть по выборке $x_1, \ldots, x_n$ построена гистограмма с k классами, N_j $(j = 1, \ldots, k)$ — число точек выборки в j-м классе $(N_1 + \ldots + N_k = n)$, p_j — вероятность попадания сл. вел., распределенной по закону P_0, в j-й класс $(p_1 + \ldots + p_k = 1)$. Статистика $\chi^2 = \sum_j (N_j - np_j)^2/(np_j)$ называется **статистикой χ^2 Пирсона**. Вектор сл. вел. $(N_1, \ldots, N_k)$ распределен по мультиномиальному распределению с вероятностями $p_1, \ldots, p_k$ (п. A.2); при $n \to \infty$ распределение вектора перенормированных сл. вел. $X_j = (N_j - np_j)/\sqrt{np_j}$ сходится к нормальному распределению с единичной матрицей ковариации и $k - 1$ независимой компонентой, поскольку условие $\sum_j X_j \sqrt{p_j} = 0$ тождественно выполнено по построению. Для определения ε-порога значимости при достаточно больших n используется **распределение χ^2 с $k - 1$ степенью свободы** (п. A.10), к которому, если нулевая гипотеза верна, сходится при $n \to \infty$ выборочное распределение статистики χ^2. Величину n и классы рекомендуется выбирать так, чтобы все $np_j \geqslant 5$.

Пусть нулевая гипотеза состоит в том, что выборка получена из нормально распределенной генеральной совокупности с мат. ожиданием μ и *неизвестной* дисперсией. В этом случае критерий значимости отклонения среднего выборки от μ обеспечивает **статистика Стьюдента**:

7.22. $t_n(x_1, \ldots, x_n) = (\overline{x} - \mu)/(s_n/\sqrt{n})$, где $\overline{x}$, s_n^2 — среднее и дисперсия выборки. Для определения ε-порога значимости используется **распределение Стьюдента с $n - 1$ степенью свободы** (п. A.17), которым обладает статистика t_n, если нулевая гипотеза верна.

При проверке статистических гипотез важно различать *статистически значимые* и *практически значимые* отличия. С одной стороны, реальное и существенное на практике отличие генеральной совокупности, из которой извлечена выборка, от нулевой гипотезы может не быть установлено на заданном уровне статистической значимости, если объем выборки недостаточен; с другой стороны, при достаточно большом объеме выборок статистически значимым с очень высоким уровнем значимости (например, $\varepsilon < 0{,}01\%$) может оказаться ничтожное, несущественное на практике отличие.

8 Вероятностные пространства и события

Всякая сл. вел. принимает значения в некотором *множестве*, а ее статистика задается *распределением вероятности* на этом множестве. Целесообразно отделить понятие распределения вероятности от понятия сл. вел. и стать на более абстрактную точку зрения, развитую А. Н. Колмогоровым («колмогоровская схема теории вероятностей»).

8.1. Пространство элементарных событий Ω — множество произвольной природы, на котором рассматривается распределение вероятности. Отдельные точки этого множества $\omega \in \Omega$ называются **элементарными событиями**, а подмножества $A \subset \Omega$ — **событиями**.

8.2. Алгебраические операции над событиями: **объединение** $A + B$ (или $A \cup B$), **пересечение** AB (или $A \cap B$) и **дополнение** $\overline{A}$.

8.3. Пустое множество: $\varnothing = \overline{\Omega}$.

8.4. События A и B **несовместны**, если $AB = \varnothing$.

8.5. Алгеброй событий $\mathfrak{A}$ называется любой набор событий (подмножеств Ω), в который входит само множество Ω и, вместе с любыми событиями A, B — события $\overline{A}$ и $A + B$ (а следовательно, также пустое множество $\varnothing$ и событие AB).

8.6. Вероятностным пространством называется совокупность трех объектов: Ω — пространства элементарных событий, $\mathfrak{A}$ — алгебры событий в нем и P — **вероятностной меры** этих событий, обладающей следующими свойствами (*аксиомы А. Н. Колмогорова*): $\mathsf{P}(A) \geqslant 0$ для любого $A \in \mathfrak{A}$; $\mathsf{P}(\Omega) = 1$; если $AB = \varnothing$, то $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B)$.

8.7. $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$; $\mathsf{P}(\varnothing) = 0$.

8.8. $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB)$.

С точки зрения теории вероятностей наиболее важны понятия условной вероятности и независимости.

8.9. Условная вероятность события B при условии, что произошло событие A положительной вероятности: $\mathsf{P}(\, B \mid A\,) = \mathsf{P}(AB)/\mathsf{P}(A)$.

8.10. $\mathsf{P}(AB) = \mathsf{P}(\, B \mid A\,)\,\mathsf{P}(A)$; вообще
$\mathsf{P}(A_1 \ldots A_n) = \mathsf{P}(A_1)\,\mathsf{P}(\, A_2 \mid A_1\,)\,\mathsf{P}(\, A_3 \mid A_1 A_2\,) \ldots \mathsf{P}(\, A_n \mid A_1 \ldots A_{n-1}\,)$.

8.11. Пространство элементарных событий Ω и алгебра $\mathfrak{A}$ образуют вероятностное пространство относительно $\mathsf{P}(\,\cdot \mid A\,)$.

8.12. **Формула полной вероятности**: если $A_1 + \ldots + A_n = \Omega$ и события A_i несовместны друг с другом, то $\mathsf{P}(B) = \sum_i \mathsf{P}(\,B \mid A_i\,)\,\mathsf{P}(A_i)$.

8.13. **Формулы Байеса**: $\mathsf{P}(\,A \mid B\,) = \mathsf{P}(\,B \mid A\,)\,\mathsf{P}(A)/\mathsf{P}(B)$;

$$\mathsf{P}(\,A_i \mid B\,) = \mathsf{P}(\,B \mid A_i\,)\,\mathsf{P}(A_i)\Big/\Bigl(\sum_j \mathsf{P}(\,B \mid A_j\,)\,\mathsf{P}(A_j)\Bigr).$$

В последней формуле $\mathsf{P}(A_i)$ иногда называют **априорной** вероятностью события A_i, а $\mathsf{P}(\,A_i \mid B\,)$ — **апостериорной** вероятностью события A после наступления события B.

8.14. События $A_1, \ldots, A_n$ называются **независимыми в совокупности**, если $\mathsf{P}(A_{i_1} \ldots A_{i_k}) = \mathsf{P}(A_{i_1}) \ldots \mathsf{P}(A_{i_k})$ для любого набора несовпадающих индексов $1 \leqslant i_1, \ldots, i_k \leqslant n$.

8.15. Попарная независимость ($\mathsf{P}(A_iA_j) = \mathsf{P}(A_i)\,\mathsf{P}(A_j)$ для любых i, j) не влечет за собой независимости в совокупности (ср. п. 6.9).

8.16. Если события A, B независимы и $\mathsf{P}(A) > 0$, то $\mathsf{P}(\,B \mid A\,) = \mathsf{P}(B)$.

8.17. Две подалгебры $\mathfrak{B}, \mathfrak{C}$ алгебры событий $\mathfrak{A}$ в вероятностном пространстве называются **независимыми**, если любые два события $B \in \mathfrak{B}$, $C \in \mathfrak{C}$ независимы. Аналогичное определение можно дать для любого набора подалгебр $\mathfrak{A}_1, \ldots, \mathfrak{A}_n$.

Для большинства применений теории вероятностей требуется распространить данную схему на бесконечные совокупности событий.

8.18. Алгебра событий называется **σ-алгеброй**, если объединение любой бесконечной последовательности непересекающихся событий $A_1, A_2, \ldots$ из алгебры $\mathfrak{A}$ также принадлежит $\mathfrak{A}$.

8.19. Чтобы множество Ω с σ-алгеброй событий $\mathfrak{A}$ образовывало вероятностное пространство, вероятностная мера должна удовлетворять дополнительной аксиоме (ср. п. 8.6): $\mathsf{P}(A_1 + A_2 + \ldots) = \mathsf{P}(A_i) + \mathsf{P}(A_2) + \ldots$, где $A_1, A_2, \ldots$ — любая бесконечная последовательность *несовместных* событий из $\mathfrak{A}$.

8.20. Если в σ-алгебре $A_1 \supset A_2 \supset \ldots$ и пересечение данной последовательности событий пусто, то $\lim_{i\to\infty} \mathsf{P}(A_i) = 0$ (**непрерывность вероятности**).

Связь старого и нового подходов к теории вероятностей обеспечивается следующими конструкциями.

8.21. (Скалярной) **случайной величиной** на вероятностном пространстве $(\Omega, \mathfrak{A}, \mathsf{P})$ называется числовая функция $X(\omega)$, $\omega \in \Omega$, для которой событие $\{\,\omega \in \Omega \mid X(\omega) < x\,\}$ (сокращенно обозначаемое $X < x$) измеримо, т. е. входит в σ-алгебру $\mathfrak{A}$, каково бы ни было число x.

8.22. Каждому событию $A \in \mathfrak{A}$ можно сопоставить сл. вел. — его **индикаторную функцию**:

$$\chi_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

8.23. Для любой индикаторной функции $\chi_A^2(\omega) \equiv \chi_A(\omega)$. Наоборот, любая сл. вел. $\chi(\omega)$, для которой $\chi^2 = \chi$, определяет событие $\chi(\omega) = 1$. (Почему оно принадлежит алгебре $\mathfrak{A}$?)

8.24. Каждая сл. вел. X **порождает** σ-алгебру $\mathfrak{A}_X \subset \mathfrak{A}$, содержащую все события вида $X(\omega) < x$ («множества уровня» функции X), а также их дополнения и их всевозможные конечные и счетные объединения.

8.25. Сл. вел. X, Y, определенные на одном и том же вероятностном пространстве, независимы, если независимы порожденные ими σ-алгебры (ср. п. 6.8).

Почему вероятностную меру определяют не для всех подмножеств Ω, а лишь для тех событий, которые входят в алгебру $\mathfrak{A}$? Следующий пример показывает, что в пространстве элементарных событий, состоящем из *континуума* точек, слишком «экзотические» множества невозможно измерить, не впадая в противоречие.

8.26. Пример. Пусть Ω — шар единичного объема в трехмерном евклидовом пространстве, а вероятностная мера P любого подмножества шара совпадает с объемом этого подмножесттва. Оказывается, что шар Ω можно разделить на конечное число непересекающихся частей (подмножеств) и, перемещая эти части в пространстве как твердые тела, собрать из них два сплошных шара такого же диаметра, как исходный (**парадокс Тарского**).

8.27. Если пространство элементарных событий Ω конечно или счетно, вероятности могут быть заданы для каждого элементарного события по отдельности и все подмножества пространства Ω являются измеримыми (ср. лекцию 1).

8.28. Если $\mathsf{P}(A) = 1$, говорят, что событие A происходит **почти наверное**. Противоположное событие $\overline{A}$ имеет вероятность нуль, но не обязательно является пустым множеством (ср. п. 2.3).

8.29. Пример. Пусть $\Omega = [0,1]$ с равномерным распределением вероятности. Множество рациональных чисел на этом отрезке непусто, но $\mathsf{P}(x \text{ рационально}) = 0$ (ср. п. 2.3).

8.30. Говорят, что последовательность сл. вел. $X_1(\omega), X_2(\omega), \ldots$, определенных на одном и том же вероятностном пространстве, сходится к пределу $X(\omega)$: **по вероятности**, если $\mathsf{P}(|X_n(\omega) - X(\omega)|) \to 0$ (п. 3.6) и **почти наверное**, если $X_n(\omega) \to X(\omega)$ на множестве вероятности 1.

8.31. Первая лемма Бореля–Кантелли. Пусть $A_1, A_2, \ldots$ — произвольная последовательность событий, для которой $\sum_i \mathsf{P}(A_i) < \infty$. Тогда для любой бесконечной подпоследовательности $A_{i_1}, A_{i_2}, \ldots$ выполнено $\mathsf{P}(A_{i_1} A_{i_2} \ldots) = 0$, т. е. почти наверное произойдет лишь конечное число событий последовательности (A_i).

8.32. Пример: усиленный закон больших чисел. Пусть $X_1(\omega)$, $X_2(\omega), \ldots$ — последовательность н. о. р. сл. вел. с $\langle X \rangle = \mu$ и $\mathsf{D}X = \sigma^2$, определенных на одном и том же вероятностном пространстве. Закон больших чисел в форме п. 3.4 утверждает, что $\frac{1}{n}(X_1(\omega) + \ldots + X_n(\omega)) \to \mu$ по вероятности. Можно показать, что $\mathsf{P}(|\frac{1}{n}(X_1 + \ldots + X_n) - \mu| > \varepsilon) < u_n$, где $\sum_n u_n < \infty$; отсюда следует, что $X_n \to X$ почти наверное.

8.33. Вторая лемма Бореля–Кантелли. Пусть $A_1, A_2, \ldots$ — произвольная последовательность *независимых* событий, для которой $\sum_i \mathsf{P}(A_i)$ расходится. Тогда почти наверное произойдет бесконечно много событий из последовательности (A_i).

9 Конечные однородные цепи Маркова

9.1. Однородной цепью Маркова называется последовательность случайных испытаний, в которой условное распределение результата i-го испытания X_i зависит от результата непосредственно предшествующего испытания X_{i-1}, но не от результатов более ранних испытаний (*марковское свойство*) и не от номера i (*однородность*): $\mathsf{P}_i(\,\mathrm{d}x \mid X_0, \ldots, X_{i-1}\,) = \mathsf{P}(\,\mathrm{d}x \mid X_{i-1}\,)$. Распределение вероятности $\mathsf{P}(\,\mathrm{d}x \mid y\,)$ называется **распределением вероятности перехода**.

9.2. Цепь Маркова называется **конечной** (**счетной**), если сл. вел. X_i могут принимать лишь конечное (счетное) число значений.

Конечные и счетные однородные цепи Маркова удобно описывать как случайные блуждания по ориентированным графам.

9.3. Ориентированным графом называется диаграмма, состоящая из конечного или счетного числа точек (**вершин**) и соединяющих их стрелок (**дуг**). Вершина, из которой исходит (в которую входит) каждая дуга, называется ее **начальной** (**концевой**) вершиной. Между одними и теми же начальной и концевой вершинами не может проходить более одной дуги. Будем рассматривать лишь такие ориентированные графы, в которых из каждой вершины исходит по крайней мере одна дуга.

9.4. Пусть для каждой вершины k графа задано распределение вероятностей p_{kl} на исходящих из нее дугах: $p_{kl} > 0$ тогда и только тогда, когда вершины k и l соединены дугой и $\sum_l p_{kl} = 1$.

9.5. Последовательность $(k_0, k_1, \ldots, k_n)$, где $p_{k_i k_{i+1}} > 0$ для всех $i = 0, 1, \ldots, n-1$, называется **реализацией** случайного блуждания на графе; число $n > 0$ называется **длиной** этой реализации.

9.6. Пусть π_k — некоторое распределение вероятности на вершинах графа. Множество всех реализаций случайного блуждания фиксированной длины n, имеющих вероятности $\mathsf{P}(k_0, \ldots, k_n) = \pi_{k_0} p_{k_0 k_1} \ldots p_{k_{n-1} k_n}$, образует конечное вероятностное пространство (ср. п. 8.27).

9.7. Пусть $\pi_k(0) = \pi_k$, $\pi_k(i) = \mathsf{P}(k_i = k) = \sum_{k_0, \ldots, k_{i-1}} \pi_{k_0} p_{k_0 k_1} \ldots p_{k_{i-1} k}$ при $i > 0$. Распределение $\pi_k(i)$ называется распределением **маргинальных вероятностей** в момент $i \geqslant 0$.

9.8. По определению $p_{kl} = 0$, если в графе нет дуги, ведущей из k в l. Если граф имеет m вершин, вероятности перехода образуют квадратную матрицу $P = (p_{kl})$ порядка m, а маргинальные вероятности при каждом i — m-мерный вектор-строку $\pi(i) = (\pi_k(i))$.

9.9. $p_{kl} \geqslant 0$; $\sum_l p_{kl} = 1$ для всех k, l. Любая квадратная матрица $P = (p_{kl})$, обладающая этими свойствами, называется **стохастической** и определяет случайное блуждание на некотором графе.

9.10. Все степени $P^i = (p^i_{kl})$ стохастической матрицы P являются стохастическими матрицами.

9.11. $\pi(i) = \pi(i-1)P = \ldots = \pi(0)P^i$ в смысле матричной алгебры.

9.12. Стационарным распределением случайного блуждания на графе называется распределение маргинальных вероятностей $\bar{\pi}$, для которого $\bar{\pi} = \bar{\pi} P$.

У любой стохастической матрицы P существует правый собственный вектор $(1, \ldots, 1)^{\mathsf{T}}$, принадлежащий собственному значению 1 и, следовательно, также по крайней мере один левый собственный вектор $u = Pu$. Чтобы вектор u выражал стационарное распределение вероятности, его элементы должны быть неотрицательны.

9.13. Стохастическая матрица $P = (p_{kl})$ называется **бистохастической**, если $\sum_k p_{kl} = 1$ для любого l (ср. п. 9.9).

9.14. Равномерное распределение является стационарным для любой бистохастической (в частности, **симметричной**) матрицы.

9.15. Число $\pi_k(i)p_{kl}$ называется **потоком вероятности** из вершины k в вершину l на i-м шаге.

9.16. Стохастическая матрица P удовлетворяет **принципу детального равновесия**, если существует распределение маргинальных вероятностей π, в котором все потоки вероятности компенсированы: $\pi_k p_{kl} = \pi_l p_{lk}$ для всех k, l. Такое распределение называется **равновесным**.

9.17. Стохастическая матрица P удовлетворяет принципу детального равновесия тогда и только тогда, когда

$$p_{k_1 k_2} \dots p_{k_{s-1} k_s} p_{k_s k_1} = p_{k_1 k_s} p_{k_s k_{s-1}} \dots p_{k_2 k_1} \text{ для всех наборов } k_1, \dots, k_s.$$

9.18. Равновесное распределение является стационарным.

9.19. Случайное блуждание в равновесном состоянии **обратимо** во времени в следующем смысле:

$$\mathsf{P}(\, k_i = l \mid k_{i+1} = k\,) = \tfrac{\mathsf{P}(k_i = l, k_{i+1} = k)}{\mathsf{P}(k_{i+1} = k)} = \tfrac{\pi_l p_{lk}}{\pi_k} = p_{kl} = \mathsf{P}(\, k_{i+1} = l \mid k_i = k\,).$$

Теперь выясним вопрос о существовании и единственности стационарного распределения в общей конечной однородной цепи Маркова.

9.20. Состояние l **достижимо** из состояния k, если $p_{kl}^i > 0$ для некоторого $i > 0$. Если состояния k, l достижимы друг из друга, они называются **сообщающимися**.

9.21. Все состояния конечной цепи Маркова можно разбить на непересекающиеся **классы сообщающихся состояний** (в теории графов они называются **сильно связными компонентами** графа цепи Маркова).

9.22. Класс сообщающихся состояний называется **поглощающим**, если из его состояний недостижимы состояния других классов.

9.23. Цепь Маркова называется **неприводимой**, если все ее состояния сообщаются друг с другом и образуют один класс.

9.24. **Периодом** $q(k)$ состояния k называется наибольший общий делитель таких i, что $p_{kk}^i > 0$ (если $p_{kk}^i = 0$ для всех $i \geqslant 0$, по определению $q(k) = \infty$).

9.25. Если состояния k и l сообщаются, их периоды равны.

9.26. Цепь Маркова называется **периодической** с периодом q (**непериодической**), если период всех ее состояний равен q (соответственно 1).

9.27. Если P определяет периодическую цепь Маркова с периодом q, то P^q определяет непериодическую цепь Маркова с q изолированными классами сообщающихся состояний.

9.28. Если конечная цепь Маркова — неприводимая и непериодическая, то существуют такие $i_0 > 0$ и $\lambda > 0$, что $p_{kl}^i \geqslant \lambda$ для всех k, l и всех $i \geqslant i_0$.

Любой поглощающий класс сообщающихся состояний можно рассматривать как самостоятельную неприводимую цепь Маркова. Покажем, что конечная неприводимая непериодическая цепь Маркова имеет ровно одно стационарное распределение; в общем случае все стационарные распределения конечной цепи Маркова получаются как выпуклые комбинации (п. B.4) стационарных распределений, сосредоточенных на ее поглощающих классах.

9.29. Пусть матрица P такова, что $p_{kl} \geqslant \lambda > 0$ для всех k, l. В этом случае цепь Маркова неприводима и непериодична.

9.30. Расстояние полной вариации между распределениями π и σ маргинальных вероятностей: $\|\pi - \sigma\| = \sum_k |\pi_k - \sigma_k| = \max_{|\xi_k| \leqslant 1} \sum_k (\pi_k - \sigma_k)\xi_k$ (максимум достигается на векторе ξ, все компоненты которого равны ± 1).

9.31. Если $\xi_k = \pm 1$ и среди чисел ξ_k есть как положительные, так и отрицательные, то $|(P\xi)_k| = |\sum_l p_{kl}\xi_l| \leqslant (1 - 2\lambda)$.

9.32. Поскольку $\sum_k (\pi_k - \sigma_k) = 0$, среди разностей $\pi_k - \sigma_k$ должны быть как положительные, так и отрицательные числа. Поэтому

$$\|\pi P - \sigma P\| = \max_{|\xi_l| \leqslant 1} \sum_{k,l} (\pi_k - \sigma_k) p_{kl} \xi_l = \max_{|\xi_l| \leqslant 1} \sum_k (\pi_k - \sigma_k)(P\xi)_k \leqslant (1 - 2\lambda)\|\pi - \sigma\|.$$

9.33. Принцип сжимающих отображений. Если для некоторого $0 < \rho < 1$ выполнено $\|\pi P - \sigma P\| \leqslant \rho\|\pi - \sigma\|$, то итерации любого начального распределения $\pi(0)$ экспоненциально быстро сходятся к единственному пределу $\bar{\pi}$.

9.34. Если цепь Маркова с матрицей P неприводима и непериодична, то все $p^i_{kl} \geqslant \lambda > 0$ при $i > i_0$ (п. 9.28). Тогда $\|\pi P^i - \sigma P^i\| \leqslant (1 - 2\lambda)\|\pi - \sigma\|$ для всех $i > i_0$ и, следовательно, $\|\pi P^i - \sigma P^i\| \leqslant (1 - 2\lambda)^k \|\pi - \sigma\|$ при всех $i > k i_0$. Существование и единственность стационарного распределения $\bar{\pi}$ доказывается аналогично п. 9.33.

9.35. Маргинальные вероятности в конечной однородной цепи Маркова удовлетворяют уравнению $\pi_k(i+1) - \pi_k(i) = \sum_j \pi_j(i) p_{jk} - \sum_l \pi_k(i-1) p_{kl}$.

9.36. Пусть i-й шаг цепи Маркова совершается в момент $t = i\Delta t$. Полагая $p_{kl} = \Delta t\, w_{kl}$ при $k \neq l$, $p_{kk} = 1 - \Delta t \sum_{l \neq k} p_{kl}$, в пределе $\Delta t \to 0$ получаем уравнение для случайного блуждания в непрерывном времени: $\dot{\pi}_k(t) = \sum_{j \neq k} \pi_j(t) w_{jk} - \sum_{l \neq k} \pi_k(t) w_{kl}$, где w_{kl} — **интенсивность перехода** из состояния k в l.

Уравнения пп. 9.35 и 9.36 называются **уравнениями марковской эволюции** (master equations).

9.37. Случайный процесс Пуассона интенсивности μ можно описать как случайное блуждание в непрерывном времени с интенсивностями перехода $w_{k,k+1} = \mu$, $w_{kl} = 0$ при $l - k \neq 1$. Число **скачков** (переходов из состояния k в $k+1$ за время t в процессе Пуассона распределено по закону Пуассона с параметром μt.

Промежутки времени между скачками процесса Пуассона распределены по показательному закону (п. А.9). Из характеристического свойства показательного распределения следует, что время ожидания очередного скачка при $\tau > t$ зависит лишь от числа скачков, произошедших к моменту t, т. е. от состояния процесса в момент t, но не от величины промежутка времени, протекшего с момента последнего скачка (марковское свойство).

10 Диффузионные процессы

10.1. Модель симметричного случайного блуждания по прямой является однородной цепью Маркова с вероятностями перехода $p_{k,k-1} = p_{k,k+1} = \frac{1}{2}$, $p_{kl} = 0$ при $k = l$ или $|k - l| > 1$ $(k, l = 0, \pm 1, \pm 2, \ldots)$.

10.2. Пусть шаги случайного блуждания равны Δx, Δt. Обозначая $p(t,x)\,\Delta x$ вероятность находиться в точке $x = k\Delta x$ в момент $t = i\Delta t$, запишем уравнение марковской эволюции в виде $p(t + \Delta t, x) - p(t, x) = \frac{1}{2}\big(p(t, x - \Delta x) - 2p(t, x) + p(t, x + \Delta x)\big)$ и разложим его в ряд Тейлора по $\Delta x, \Delta t$: $\partial p(t,x)/\partial t\, \Delta t + O((\Delta t)^2) = \frac{1}{2}\partial^2 p(t,x)/\partial x^2\, (\Delta x)^2 + O((\Delta x)^3)$.

10.3. При $\Delta t, \Delta x \to 0$, $(\Delta x)^2/\Delta t = \varkappa = \mathrm{const}$ возникает случайный процесс **броуновского движения** $B_\varkappa(t)$, описываемый **уравнением диффузии** вероятности $\partial p/\partial t = \frac{\varkappa}{2}\partial^2 p/\partial x^2$, где $\varkappa$ — **коэффициент диффузии**.

10.4. Блуждание по регулярной решетке в d-мерном пространстве рассматривается аналогично и, если $\varkappa = (\Delta x)^2/(d\,\Delta t)$, приводит к d-мерному броуновскому движению $\boldsymbol{B}_\varkappa(t)$ с уравнением диффузии $\partial p/\partial t = \frac{\varkappa}{2}\nabla^2 p$.

10.5. Уравнение диффузии вероятности имеет вид «закона сохранения»: $\partial p/\partial t = -\nabla \cdot \boldsymbol{j}$, где $\boldsymbol{j} = -\frac{\varkappa}{2}\nabla p$ — **поток вероятности**. Заметим, что диффузия вероятности в броуновском движении изотропна.

10.6. Решение $p_B(\,t, x \mid s, y\,)$ при $t > s$ одномерного уравнения диффузии на всей вещественной прямой имеет смысл условной ф. п. в. координаты блуждающей частицы в момент времени t, если при $t = s$ она была в точке y (**ф. п. в. перехода**): $\mathsf{P}(\,x < B_\varkappa(t) < x + \mathrm{d}x \mid B_\varkappa(s) = y\,) = p_B(\,t, x \mid s, y\,)\,\mathrm{d}x = (2\pi\varkappa(t-s))^{-\frac{1}{2}} \exp\big(-\frac{1}{2\varkappa(t-s)}(x-y)^2\big)\,\mathrm{d}x$.

10.7. $p_B(\,t, x \mid s, y\,) = \int p_B(\,t, x \mid u, z\,)\, p_B(\,u, z \mid s, y\,)\,\mathrm{d}z$, где $s < u < t$.

Данное соотношение можно рассматривать как выражение *марковского свойства* броуновского движения: начиная с любого момента времени $u > s$ оно как бы начинается заново и не зависит от своей истории в интервале времени (s, t).

10.8. Смещения $B_\varkappa(t_2) - B_\varkappa(t_1)$, $B_\varkappa(t_4) - B_\varkappa(t_3)$ броуновского движения за неперекрывающиеся интервалы времени ($t_1 < t_2 \leqslant t_3 < t_4$) являются независимыми гауссовыми сл. вел.

10.9. Если $B_\varkappa(0) = 0$, то $\langle B_\varkappa(t)\rangle = 0$, $\langle B_\varkappa(s)B_\varkappa(t)\rangle = \varkappa\min(t, s)$.

10.10. Одномерное броуновское движение внутри (может быть, бесконечного) интервала $0 \leqslant x \leqslant L \leqslant \infty$ с **непроницаемыми границами** описывается уравнением диффузии п. 10.3 с краевыми условиями исчезновения потока вероятности: $j|_{x=0,L} = -\frac{\varkappa}{2}(\partial p/\partial x)\big|_{x=0,L} = 0$.

10.11. Краевое условие $p(t, 0) = 0$ соответствует броуновскому движению с **поглощающей границей**: при достижении координаты $x = 0$ движение обрывается.

10.12. Решение уравнения диффузии на полупрямой $x \geqslant 0$ с поглощающей границей при $x = 0$:

$$p(\,t, x \mid 0, y\,) = (2\pi\varkappa t)^{-\frac{1}{2}}\left(\mathrm{e}^{-\frac{1}{2\varkappa t}(x-y)^2} - \mathrm{e}^{-\frac{1}{2\varkappa t}(x+y)^2}\right) \quad (y > 0).$$

10.13. Распределение **времени выхода** T_y на поглощающую границу $x = 0$ определяется формулой $1 - F_{T_y}(t) = \mathsf{P}(T_y > t) = \int_0^\infty p(\,t, x \mid 0, y\,)\,\mathrm{d}x$ (выражение для соответствующей ф. п. в. см. в п. А.19).

10.14. Если случайное блуждание таково, что $p_{k,k+1} = p$, $p_{k-1,k} = \bar{p}$ ($p + \bar{p} = 1$), то разложение уравнения марковской эволюции в ряд Тейлора будет иметь вид $\partial p(t,x)/\partial t\,\Delta t + O((\Delta t)^2) = (\bar{p} - p)\,\partial p(t,x)/\partial x\,\Delta x + \frac{1}{2}\partial^2 p(t,x)/\partial x^2\,(\Delta x)^2 + O((\Delta x)^3)$. Предел такого случайного блуждания при $\Delta t, \Delta x \to 0$ нетривиален, если $\frac{(\Delta x)^2}{\Delta t} = \varkappa = \mathrm{const}$ и $\frac{(\bar{p}-p)}{\Delta x} = v = \mathrm{const}$: в пределе возникает **уравнение диффузии с переносом** $\partial p/\partial t + v\,\partial p/\partial x = \frac{\varkappa}{2}\partial^2 p/\partial x^2$ (если $\bar{p} - p = O(1)$, возникает перенос без диффузии).

При диффузии с переносом как мат. ожидание, так и дисперсия смещения $\mathrm{d}B_\varkappa(t)$ за время $\mathrm{d}t$ имеют одинаковый порядок малости $O(\mathrm{d}t)$. Учет этого обстоятельства приводит к своеобразному варианту дифференциального исчисления.

10.15. Пример. Определим интеграл $I(t) = \int_0^t B(s)\,\mathrm{d}B(s)$ как предел интегральных сумм $S = \sum_i B(\tau_i)\,(B(t_{i+1}) - B(t_i))$, где $0 = t_0 < t_1 < \ldots < t_n = t$, $t_i \leqslant \tau_i \leqslant t_{i+1}$ (здесь и ниже коэффициент диффузии броуновского движения $\varkappa = 1$). Данное определение зависит от способа выбора моментов времени τ_i: $\langle S\rangle = \sum_i(\tau_i - t_i)$ (см. п. 10.9). Выберем $\tau_i = t_i$ (**правило Ито**), чтобы гарантировать независимость приращений $I(t)$ за неперекрывающиеся интервалы времени $\Delta t_i = t_{i+1} - t_i$; при этом $\langle I(t)\rangle = \lim\langle S\rangle = 0$. Обозначая $B_i = B(t_i)$, $\Delta B_i = B(t_{i+1}) - B(t_i)$, представим интегральную сумму в виде $\sum_i B_i\Delta B_i = \frac{1}{2}\sum_i(B_{i+1}^2 - B_i^2 - (\Delta B_i)^2) = \frac{1}{2}(B^2(t) - B^2(0)) - \frac{1}{2}\sum_i(\Delta B_i)^2$. Поскольку $\left\langle\sum_i(\Delta B_i)^2\right\rangle = \sum_i \Delta t_i = t$, $\mathsf{D}\big(\sum_i(\Delta B_i)^2\big) = \sum_i(\Delta t_i)^2$, при измельчении разбиения сл. вел. $\sum_i(\Delta B_i)^2$ сходится по вероятности к неслучайному значению t. Следовательно, $\int_0^t B(s)\,\mathrm{d}B(s) = \frac{1}{2}(B^2(t) - B^2(0)) - \frac{1}{2}t$.

10.16. Интеграл $\int_0^t f(s, B(s))\,\mathrm{d}B(s)$, вычисленный по правилу Ито как предел интегральных сумм вида $\sum_i f(t_i, B(t_i))\Delta B_i$, называется **стохастическим интегралом Ито**.

10.17. Аналогично примеру п. 10.15 можно установить формулу

$$\left\langle \int_0^t f(u)\,\mathrm{d}B(u) \int_0^s g(v)\,\mathrm{d}B(v)\right\rangle = \int_0^{\min(t,s)} f(u)v(u)\,\mathrm{d}u.$$

10.18. Из пп. 10.15 и 10.17 следуют формальные правила **дифференциального исчисления Ито**: $(\mathrm{d}B(t))^2 = \mathrm{d}t$, $(\mathrm{d}B(t))^m = 0$ при $m > 2$, $\langle \mathrm{d}B(t)\mathrm{d}B(s)\rangle = \delta(t-s)\,\mathrm{d}t$, дополняющие основное правило классического дифференциального исчисления: $(\mathrm{d}t)^n = 0$ при $n > 1$.

10.19. Формула Ито: для любой достаточно гладкой функции $f(t,x)$

$$\mathrm{d}f(t, B(t)) = (\partial f/\partial t + \tfrac{1}{2}\partial^2 f/\partial x^2)\,\mathrm{d}t + \partial f/\partial x\,\mathrm{d}B(t).$$

10.20. Диффузионным процессом называется решение $X(t)$ **стохастического дифференциального уравнения** $\mathrm{d}X = a(t, X(t))\,\mathrm{d}t + b(t, X(t))\,\mathrm{d}B(t)$ (при $b(t,x) \equiv 0$ оно переходит в обыкновенное дифференциальное уравнение $\mathrm{d}X/\mathrm{d}t = a(t, X(t))$).

10.21. Пример. Уравнение $\mathrm{d}X = cX\,\mathrm{d}B(t)$ с **мультипликативным шумом** решается нелинейной заменой переменной:

$$Y = \ln X, \quad \mathrm{d}Y = X^{-1}\,\mathrm{d}X - \tfrac{1}{2}X^{-2}(\mathrm{d}X)^2 = c\,\mathrm{d}B(t) - \tfrac{c^2}{2}\,\mathrm{d}t,$$

откуда $Y(t) - Y(0) = cB(t) - \frac{c^2}{2}t$, $X(t) = X(0)\mathrm{e}^{cB(t) - \frac{c^2}{2}t}$.

10.22. Пример. Пусть $X(t)$ — диффузионный процесс. Тогда для любой гладкой функции $\varphi(x)$

$$\mathrm{d}\varphi(X(t)) = \big(\varphi'(X)a(t,X) + \tfrac{1}{2}\varphi''(X)b^2(t,X)\big)\,\mathrm{d}t + \varphi'(X)b(t,X)\,\mathrm{d}B.$$

Учитывая, что $\langle \mathrm{d}B\rangle = 0$, получим для математических ожиданий равенство $\langle \mathrm{d}\varphi(X(t))/\mathrm{d}t\rangle = \langle\varphi'(X)a(t,X)\rangle + \frac{1}{2}\left\langle\varphi''(X)b^2(t,X)\right\rangle$. Фигурирующие здесь мат. ожидания можно выразить через ф. п. в. перехода:

$$\langle \mathrm{d}\varphi(X(t))/\mathrm{d}t \rangle = \textstyle\int \varphi(x)\,(\partial p(\,t,x \mid 0,X(0)\,)/\partial t)\,\mathrm{d}x,$$
$$\langle \varphi'(X)a(t,X) \rangle = \int \varphi'(x)a(t,x)p(\,t,x \mid 0,X(0)\,)\,\mathrm{d}x,$$
$$\left\langle \varphi''(X)b^2(t,X) \right\rangle = \int \varphi''(x)b^2(t,x)p(\,t,x \mid 0,X(0)\,)\,\mathrm{d}x.$$

Интегрируя по частям, чтобы снять производные с $\varphi(x)$, и учитывая произвольность выбора φ, получим **уравнение Фоккера–Планка** $\partial p/\partial t + \partial(a(t,x)p)/\partial x = \frac{1}{2}\partial^2(b^2(t,x)p)/\partial x^2$, которому удовлетворяет распределение вероятности диффузионного процесса (ср. уравнение диффузии с переносом в п. 10.14).

Описания диффузионного процесса при помощи стохастического дифференциального уравнения п. 10.20 и уравнения Фоккера–Планка для ф. п. в. перехода являются эквивалентными.

10.23. Если смещения $X(t_2) - X(t_1)$, $X(t_4) - X(t_3)$ случайного процесса $X(t)$ за неперекрывающиеся интервалы времени $(t_1 < t_2 \leqslant t_3 < t_4)$ независимы, процесс $X(t)$ называется процессом с **независимыми приращениями**.

10.24. Если распределение вероятности перехода $\mathsf{P}(\,t, \mathrm{d}x \mid s, y\,)$ случайного процесса $X(t)$ зависит только от разности $t-s$, так что выполнено равенство $\mathsf{P}(\,t, \mathrm{d}x \mid s, y\,) = \mathsf{P}(\,t+\tau, \mathrm{d}x \mid s+\tau, y\,)$ при любом τ, процесс $X(t)$ называется процессом со **стационарными приращениями**.

10.25. Процессы со стационарными независимыми приращениями называются **процессами Леви**; распределения вероятности их приращений безгранично делимы (п. 5.12).

10.26. Случайный процесс называется **стационарным**, если совместные распределения вероятности двух наборов значений $X(t_1), \ldots, X(t_k)$ и $X(t_1+\tau), \ldots, X(t_k+\tau)$ совпадают при любом выборе $t_1, \ldots, t_k$ и τ.

10.27. Стационарное распределение вероятности $\bar{p}(x)$ значения $X(t)$ одномерного диффузионного процесса, удовлетворяющего уравнению Фоккера–Планка п. 10.22 с независящими от времени коэффициентами $a(x)$, $b(x)$, удовлетворяет уравнению $j = -a(x)p + \frac{1}{2}\partial(b^2(x)p)/\partial x = 0$ и выражается формулой $\bar{p}(x) = \mathrm{const}\, b^{-2}(x) \exp\big(\int_0^x 2(a(\xi)/b^2(\xi))\,\mathrm{d}\xi\big)$, если интеграл $\int \bar{p}(x)\,\mathrm{d}x$ сходится.

10.28. Если распределение вероятности значения $X(s)$ диффузионного процесса стационарно, то при $t > s$ он является стационарным.

10.29. Пример. Стационарный диффузионный процесс, удовлетворяющий уравнению $\mathrm{d}X = -\Gamma X\,\mathrm{d}t + \mathrm{d}B$, где $\Gamma = \mathrm{const}$, называется **процессом Орнштейна–Уленбека**. Его стационарное распределение нормально.

10.30. Автокорреляционная функция случайного процесса $X(t)$: $F_X(t,s) = \langle X(t)X(s)\rangle$.

10.31. Пример. Автокорреляционная функция процесса Орнштейна–Уленбека: $F(t,s) = \mathrm{e}^{-\Gamma|t-s|}$.

10.32. Спектральное разложение стационарного процесса:

$$X(t) = \int \mathrm{e}^{\mathrm{i}\omega t}\hat{X}_\omega\,\mathrm{d}\omega, \quad \hat{X}_\omega = \tfrac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}\omega t}X(t)\,\mathrm{d}t.$$

10.33. Теорема Бохнера–Хинчина: $\left\langle (\hat{X}_\omega)^*\hat{X}_{\omega'}\right\rangle = \delta(\omega-\omega')J_X(\omega)$, где $J_X(\omega) = \frac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}\omega t}F_X(t)\,\mathrm{d}t$ — **спектральная плотность мощности** процесса $X(t)$: $\left\langle X^2(t)\right\rangle = \int J_X(\omega)\,\mathrm{d}\omega$.

10.34. Непрерывная функция $F(t)$ есть автокорреляционная функция некоторого стационарного процесса только и только тогда, когда она положительно определена в смысле п. 2.19.

10.35. Стационарный процесс $X(t)$ имеет **среднеквадратичный показатель Гельдера** H, если $\left\langle (X(t+s)-X(t))^2\right\rangle = O(s^2H)$ при $|s|\to 0$.

10.36. $F(0) - F(s) = \mathrm{const}\,|s|^{2H}$

10.37. Если $J(\omega) = O(|\omega|^{-(\alpha+1)})$, то $H = \alpha/2$ (ср. п. 5.4).

A «Зоопарк» распределений вероятности

В каждом из пунктов данного раздела символы M или X обозначают сл. вел., определяемую в данном пункте, а S и K — ее асимметрию и эксцесс; значения целочисленных величин везде неотрицательны.

A.1. Биномиальное распределение с параметрами $0 < p, \bar{p} < 1$, $p + \bar{p} = 1$, $n \geqslant 1$ (п. 1.24):

$$\mathsf{P}(m) = \tbinom{n}{m}p^m\bar{p}^{n-m}, \quad G(z) = (\bar{p} + pz)^n,$$

где $\binom{n}{m} = \frac{n!}{m!\,(n-m)!} = \frac{\Gamma(n+1)}{\Gamma(m+1)\,\Gamma(n-m+1)}$ при $0 \leqslant m \leqslant n$ (иначе $\binom{n}{m} = 0$); $\langle M\rangle = np$, $\mathsf{D}M = np\bar{p}$, $\mathsf{S} = (\bar{p}-p)/\sqrt{np\bar{p}}$, $\mathsf{K} = (1-p\bar{p})/np\bar{p}$.

A.2. Мультиномиальное распределение с параметрами $p_1, \ldots, p_k$ $(0 < p_i < 1$, $p_1 + \ldots + p_k = 1)$ и $n \geqslant 1$:

$$\mathsf{P}(m_1, \ldots, m_k) = \tfrac{n!}{m_1!\ldots m_k!}p_1^{m_1}\ldots p_k^{m_k}, \quad G(z_1, \ldots, z_k) = \Bigl(\textstyle\sum_{1\leqslant i\leqslant k} z_i p_i\Bigr)^n;$$

$\langle M_i \rangle = np_i$, $\mathsf{D}M_i = \Gamma_{ii} = np_i(1-p_i)$, $\langle M_i M_j \rangle = (n^2-n)p_i p_j$, $\Gamma_{ij} = -np_i p_j$ при $i \neq j$. Мультиномиальное распределение обобщает биномиальное на случай более двух исходов; отрицательность коэффициентов ковариации (п. 6.11) объясняется тем, что сумма сл. вел. $M_1, \ldots, M_k$ фиксирована и равна n. Мультиномиальное распределение возникает, например, при выводе критерия χ^2 (п. 7.21).

А.3. Распределение Пуассона с параметром $\lambda > 0$:

$$\mathsf{P}(m) = \tfrac{\lambda^m}{m!}\mathrm{e}^{-\lambda}, \quad G(z) = \mathrm{e}^{\lambda(z-1)};$$

$\langle M \rangle = \mathsf{D}M = \lambda$, $\mathsf{S} = 1/\sqrt{\lambda}$, $\mathsf{K} = 1/\lambda$. Распределение Пуассона возникает из биномиального в пределе $n \to \infty$, $p \to 0$, $np = \lambda = \text{const}$. Число событий, происходящих за время t в **пуассоновском сл. процессе** интенсивности μ (п. 9.37), распределено по Пуассону с параметром μt.

А.4. Геометрическое распределение с параметром $0 < p < 1$:

$$\mathsf{P}(m) = p\bar{p}^{m-1} \ (m \geqslant 1), \quad G(z) = \tfrac{pz}{1-\bar{p}z};$$

$\langle M \rangle = 1/p$, $\mathsf{D}M = \bar{p}/p^2$, $\mathsf{S} = (1+\bar{p})/\sqrt{\bar{p}}$, $\mathsf{K} = (1+4\bar{p}+\bar{p}^2)/\bar{p}$. Геометрическому распределению подчинено число независимых испытаний, происходящих до первого успеха включительно (ср. п. 1.24); название связано с тем, что вероятности образуют геометрическую прогрессию.

А.5. Отрицательное биномиальное распределение с параметрами $0 < p < 1$, $k > 0$:

$$\mathsf{P}(m) = \tbinom{m-1}{k-1} p^k \bar{p}^{m-k} \ (m \geqslant k), \quad G(z) = \left(\tfrac{pz}{1-\bar{p}z}\right)^k;$$

$\langle M \rangle = m/p$, $\mathsf{D}M = m\bar{p}/p^2$, $\mathsf{S} = (1+\bar{p})/\sqrt{m\bar{p}}$, $\mathsf{K} = (1+4\bar{p}+\bar{p}^2)/(m\bar{p})$. Отрицательному биномиальному распределению подчинено число независимых испытаний, происходящих до k-го успеха включительно; геометрическое распределение является его частным случаем.

А.6. Равномерное распределение на отрезке $(\mu - \Gamma, \mu + \Gamma)$, $\Gamma > 0$:

$$p(x) = \begin{cases} 1/2\Gamma & \text{при } |x-\mu| < \Gamma, \\ 0 & \text{в противном случае;} \end{cases} \qquad \varphi(s) = \mathrm{e}^{\mathrm{i}\mu s}\, \tfrac{\sin \Gamma s}{\Gamma s};$$

$\langle X \rangle = \mu$, $\mathsf{D}X = \Gamma^2/3$, $\mathsf{S} = 0$, $\mathsf{K} = -1\tfrac{1}{5}$.

А.7. Треугольное распределение с параметрами μ, $\Gamma > 0$:

$$p(x) = \max\left(\tfrac{1}{\Gamma} - \tfrac{|x-\mu|}{\Gamma^2}, 0\right), \quad \varphi(s) = \tfrac{\sin^2(\Gamma s/2)}{(\Gamma s/2)^2};$$

$\langle X\rangle = \mu$, $\mathsf{D}X = \Gamma^2/6$, $\mathsf{S} = 0$, $\mathsf{K} = -\frac{3}{5}$. При сложении независимых равномерно распределенных сл. вел. с параметрами μ, Γ получается сл. вел., распределенная по треугольному закону с параметрами $2\mu, 2\Gamma$. Поскольку х. ф. треугольного распределения неотрицательна, функция $\Gamma p(x)$ сама является х. ф. по отношению к $\Gamma\varphi(s)$, рассматриваемой как ф. п. в.

A.8. Распределение Коши (Брейта–Вигнера, Лоренца) с параметрами μ и $\Gamma > 0$:

$$p(x) = \tfrac{1}{\pi}\tfrac{\Gamma}{\Gamma^2+(x-\mu)^2}, \quad \varphi(s) = \mathrm{e}^{\mathrm{i}\mu s-\Gamma|s|}.$$

Моменты всех порядков этого распределения выражаются расходящимися интегралами: в частности, его мат. ожидание не определено, а дисперсия бесконечна. Параметр μ является медианой и модой распределения Коши. Это распределение устойчиво (п. 5.8) с показателем $\alpha = 1$.

A.9. Показательное распределение с параметром $\lambda > 0$:

$$p(x) = \begin{cases} \lambda\mathrm{e}^{-\lambda x}, & x > 0, \\ 0, & x < 0, \end{cases} \quad \varphi(s) = \tfrac{1}{1-\mathrm{i}s/\lambda};$$

$\langle X\rangle = 1/\lambda$, $\mathsf{D}X = 1/\lambda^2$, $\mathsf{S} = 2$, $\mathsf{K} = 6$. Показательное распределение можно рассматривать как «непрерывный аналог» геометрического распределения (см. выше); оно часто возникает как распределение времени ожидания некоторого события. Например, в **пуассоновском случайном процессе** интенсивности μ (п. 9.37) промежутки времени между последовательными событиями распределены по показательному закону с параметром μ. Показательно распределенная сл. вел. X обладает следующим **характеристическим свойством**: $\mathsf{P}(\,X > t+\tau \mid X > t\,) = \mathsf{P}(X > \tau)$ при $t, \tau \geqslant 0$.

A.10. Гамма-распределение с параметрами $\lambda > 0$, $\nu > 0$:

$$p(x) = \tfrac{\lambda^\nu}{\Gamma(\nu)}x^{\nu-1}\mathrm{e}^{-\lambda x}, \quad \varphi(s) = \tfrac{1}{(1-\mathrm{i}s/\lambda)^\nu};$$

$\langle X\rangle = \nu/\lambda$, $\mathsf{D}X = \nu/\lambda^2$, $\mathsf{S} = 2/\sqrt{\nu}$, $\mathsf{K} = 6/\nu$. Гамма-распределение можно рассматривать как «непрерывный аналог» отрицательного биномиального распределения; оно часто возникает при описании непрерывных неотрицательных сл. вел. различной природы. Частным случаем гамма-распределения является **распределение χ^2 с n степенями свободы** ($\lambda = \frac{1}{2}$, $\nu = \frac{n}{2}$, см. п. 7.21):

$$p(x) = \tfrac{(x/2)^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}}{2\Gamma(n/2)}, \quad \varphi(s) = \tfrac{1}{(1-2\mathrm{i}s)^{\frac{n}{2}}},$$

возникающее при сложении квадратов n независимых сл. вел., распределенных по нормальному закону.

А.11. Скалярное **нормальное распределение**, или распределение Гаусса, с мат. ожиданием μ и дисперсией σ^2:

$$p(x) = \tfrac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad \varphi(s) = \mathrm{e}^{\mathrm{i}\mu s - \frac{\sigma^2 s^2}{2}};$$

$\langle X \rangle = \mu$, $\mathsf{D}X = \sigma^2$, $\mathsf{S} = 0$, $\mathsf{K} = 0$. Характеристический показатель гауссовой сл. вел. является многочленом 2 порядка, так что все кумулянты $\langle\!\langle X^n \rangle\!\rangle = 0$ при $n \geqslant 3$; не существует сл. вел., характеристические показатели которых были бы многочленами более высокого порядка (**теорема Марцинкевича**).

А.12. Многомерное **нормальное распределение**, или распределение Гаусса, задается вектором мат. ожидания $\boldsymbol{m}$, имеющим n компонент, и квадратной $n \times n$ положительно определенной матрицей ковариации Γ:

$$p(\boldsymbol{x}) = \tfrac{1}{\sqrt{(2\pi)^n \det \Gamma}} \mathrm{e}^{-\frac{1}{2}\sum_{i,j} \widehat{\Gamma}_{ij}(x_i - m_i)(x_j - m_j)},$$
$$\varphi(\boldsymbol{s}) = \mathrm{e}^{\mathrm{i}\boldsymbol{s}\cdot\boldsymbol{m} - \frac{1}{2}\sum_{i,j} s_i s_j},$$

где $\hat{\Gamma}$ — матрица, обратная к матрице Γ; $\langle X_i \rangle = m_i$, $\langle\!\langle X_i X_j \rangle\!\rangle = \Gamma_{ij}$, кумулянты более высоких порядков обращаются в нуль. Можно рассматривать **вырожденное** гауссово распределение, матрица ковариации которого неотрицательно определена, но $\det \Gamma = 0$. Такое распределение сосредоточено в пространстве переменной $\boldsymbol{x}$ на линейном подпространстве, размерность которого определяется числом положительных собственных значений матрицы Γ.

А.13. Логнормальное распределение с параметрами μ, σ:

$$p(x) = \tfrac{1}{\sqrt{2\pi}\sigma\, x} \mathrm{e}^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} \text{ при } x > 0;$$

$\langle X \rangle = \mathrm{e}^{\mu + \frac{1}{2}\sigma^2}$, $\mathsf{D}X = \mathrm{e}^{2\mu}\xi(\xi - 1)$, $\mathsf{S} = \sqrt{\xi - 1}(\xi + 2)$, $\mathsf{K} = (\xi - 1)(\xi^3 + 2\xi^2 + 6\xi + 6)$, где $\xi = \mathrm{e}^{\sigma^2}$. Это распределение описывает сл. вел., логарифм которой распределен нормально (например, среднее геометрическое большого числа случайных сомножителей, ср. лекцию 3l); из-за множителя $\frac{1}{x}\mathrm{e}^{\frac{\mu}{\sigma^2}\ln x} = x^{-(1 - \frac{\mu}{\sigma^2})}$ в ф. п. в. выборка из логнормального распределения при $\mu < 0$ трудно отличима от выборки из распределения-степенного закона с показателем $\alpha = |\mu|/\sigma^2$ (п. 5.1).

А.14. Распределение Фреше с параметрами $\alpha > 0, \mu, \sigma > 0$:

$$p(x) = \tfrac{\alpha}{\sigma}\mathrm{e}^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}} \left(\tfrac{x-\mu}{\sigma}\right)^{-\alpha-1} \quad (x > 0), \quad F(x) = \begin{cases} \mathrm{e}^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}}, & x > \mu, \\ 0, & x < \mu. \end{cases}$$

А.15. Распределение Вейбулла с параметрами $\alpha > 0, \mu, \sigma > 0$:

$$p(x) = \tfrac{\alpha}{\sigma}\mathrm{e}^{-\left(\frac{\mu-x}{\sigma}\right)^\alpha}\left(\tfrac{\mu-x}{\sigma}\right)^{\alpha-1} \quad (x < \mu), \quad F(x) = \begin{cases} \mathrm{e}^{-\left(\frac{\mu-x}{\sigma}\right)^\alpha}, & x < \mu, \\ 1, & x > \mu. \end{cases}$$

А.16. Распределение Гумбеля с параметрами $\mu, \sigma > 0$:

$$p(x) = \tfrac{1}{\sigma}\mathrm{e}^{-\frac{x-\mu}{\sigma}-\mathrm{e}^{-\frac{x-\mu}{\sigma}}}, \quad F(x) = \mathrm{e}^{-\mathrm{e}^{-\frac{x-\mu}{\sigma}}}.$$

Распределения трех последних пунктов характеризуют наибольшие значения выборок из n н. о. р. сл. вел., ф. п. в. которых имеют различный характер убывания на бесконечности (подробнее см. пп. 5.17–5.19). Параметр μ выбирается равным характеристическому наибольшему значению $x_*^{(n)}$ соответствующей выборки (п. 5.16).

А.17. Распределение Стьюдента с n степенями свободы:

$$p(x) = \tfrac{1}{\sqrt{\pi n}}\tfrac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}\tfrac{1}{(1+\frac{x^2}{n})^{\frac{n+1}{2}}};$$

$\langle X\rangle = 0$ при $n > 1$, $\mathsf{D}X = \frac{n}{n-2}$ при $n > 2$, $\mathsf{S} = 0$ при $n > 3$, $\mathsf{K} = \frac{6}{n-4}$ при $n > 4$. Данное распределение возникает как распределение отношения гауссовой сл. вел. (п. А.11) к независимой от нее сл. вел., распределенной по закону χ^2 с n степенями свободы (п. А.10) и имеет степенные хвосты с $\alpha = n$; при $n = 1$ оно совпадает с распределением Коши (п. А.8), а при $n \to \infty$ стремится к нормальному.

А.18. Предельное распределение статистики Колмогорова–Смирнова (п. 7.20) задается к. ф. р.

$$F(x) = 1 - 2\textstyle\sum_{k\geqslant 1}(-1)^{k-1}\mathrm{e}^{-2k^2x^2}.$$

А.19. Распределение времени выхода с параметром $\Gamma > 0$:

$$p(x) = \tfrac{\sqrt{\Gamma}}{\sqrt{2\pi x^3}}\mathrm{e}^{-\frac{\Gamma}{2x}}, \quad \varphi(s) = \mathrm{e}^{-\Gamma\sqrt{s}\,(1+\mathrm{i}\,\mathrm{sign}\,s)};$$

моменты всех порядков расходятся. Это распределение устойчиво с параметром $\alpha = \frac{1}{2}$; оно возникает как распределение времени достижения границы $x = \Gamma$ в одномерном броуновском движении с коэффициентом диффузии $\varkappa = 1$ (п. 10.13).

B Выпуклые функции и некоторые неравенства

B.1. Выпуклое тело в пространстве можно описать как пересечение всех содержащих его полупространств. Очевидно, достаточно рассматривать только такие полупространства, граничные плоскости которых проходят через граничные точки выпуклого тела. Такие плоскости называются **опорными**.

B.2. Куб и шар являются выпуклыми телами в трехмерном пространстве. Любая опорная плоскость к шару является касательной; среди опорных плоскостей к кубу касательными являются только те шесть, которые целиком содержат в себе его грани.

B.3. Из определения п. B.1 следует, что выпуклое тело вместе с любыми двумя точками $\boldsymbol{x}, \boldsymbol{y}$ целиком содержит соединяющий их отрезок, который состоит из точек вида $p\boldsymbol{x} + \bar{p}\boldsymbol{y}$, где $0 \leqslant p \leqslant 1$, $p + \bar{p} = 1$.

B.4. Вместе с любыми точками $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ выпуклое тело содержит любую их **выпуклую комбинацию** $p_1\boldsymbol{x}_1 + \ldots + p_n\boldsymbol{x}_n$, где $(p_1, \ldots, p_n)$ — произвольное распределение вероятности.

B.5. Функция $f(\boldsymbol{x})$ называется **выпуклой вниз**, если множество точек в пространстве $(\boldsymbol{x}, f)$, лежащих выше ее графика, является выпуклым. Если выпукло множесто точек, лежащих ниже графика функции, она называется **выпуклой вверх**. Для определенности будем рассматривать выпуклые вниз функции и называть их **выпуклыми**.

B.6. Выпуклая функция $f(\boldsymbol{x})$ может быть представлена в виде

$$f(\boldsymbol{x}) = \max_{\boldsymbol{y}}(\boldsymbol{y} \cdot \boldsymbol{x} - f^c(\boldsymbol{y})),$$

где выражение $\boldsymbol{y} \cdot \boldsymbol{x} - f^c(\boldsymbol{y})$ задает опорную плоскость к множеству точек, лежащих выше графика функции $f(\boldsymbol{x})$. Наклон этой опорной плоскости выражается вектором $\boldsymbol{y}$. График выпуклой функции целиком лежит выше любой своей опорной плоскости.

B.7. Функция $f^c(\boldsymbol{y})$, определяющая возвышение опорной плоскости наклона $\boldsymbol{y}$, называется **преобразованием Лежандра** функции $f(\boldsymbol{x})$.

B.8. Из формулы п. B.6 следует **неравенство Юнга**: при любых $\boldsymbol{x}, \boldsymbol{y}$

$$f(\boldsymbol{x}) + f^c(\boldsymbol{y}) \geqslant \boldsymbol{x} \cdot \boldsymbol{y},$$

причем равенство выполняется только тогда, когда опорная плоскость с наклоном $\boldsymbol{y}$ касается графика функции f в точке $\boldsymbol{x}$.

B.9. Неравенство Юнга симметрично относительно замены $\boldsymbol{x}$ на $\boldsymbol{y}$ и f на f^c. Поэтому $f^c(\boldsymbol{y}) = \max_{\boldsymbol{x}}(\boldsymbol{x} \cdot \boldsymbol{y} - f(\boldsymbol{x}))$: двукратное применение преобразования Лежандра восстанавливает исходную выпуклую функцию.

B.10. Не все выпуклые функции дифференцируемы (например, функция $|\boldsymbol{x}|$ не дифференцируема в начале координат). Если функция f выпукла и дифференцируема, то неравенство Юнга обращается в равенство при $\boldsymbol{y} = \nabla_{\boldsymbol{x}} f(\boldsymbol{x})$. Аналогично, $\boldsymbol{x} = \nabla_{\boldsymbol{y}} f^c(\boldsymbol{y})$, если f^c дифференцируема.

B.11. Выпуклая функция f называется **строго выпуклой**, если любая опорная плоскость пересекает график функции лишь в одной точке. В частности, строго выпуклая вниз функция имеет не более одной точки минимума.

B.12. Для строгой выпуклости дважды дифференцируемой функции f достаточно, чтобы матрица ее вторых производных $(\partial^2 f(\boldsymbol{x})/\partial x_i \partial x_j)$ была всюду положительно определена. В частности, если x — скалярная переменная, то для выпуклости достаточно, чтобы $f''(x) > 0$ для всех x.

B.13. Пример. Показательная функция e^x выпукла, и ее график целиком лежит выше опорной прямой при $x = 0$:

$$\mathrm{e}^x > \mathrm{e}^0 + \left.\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^x\right|_{x=0} \cdot x = 1 + x.$$

Преобразование Лежандра показательной функции: $y(\ln y - 1)$.

B.14. Из п. B.3 следует, что любая выпуклая функция удовлетворяет неравенству $f(p\boldsymbol{x} + \bar{p}\boldsymbol{y}) \leqslant pf(\boldsymbol{x}) + \bar{p}f(\boldsymbol{y})$. Если $\boldsymbol{x} \neq \boldsymbol{y}$ и функция f строго выпукла, это неравенство является строгим.

B.15. Пусть функция f выпукла, а (p_i) — дискретное распределение вероятности. Из неравенства предыдущего пункта по индукции нетрудно вывести, что

$$f\left(\sum_i p_i \boldsymbol{x}_i\right) \leqslant \sum_i p_i f(\boldsymbol{x}_i).$$

Если $\mathsf{P}(\mathrm{d}\boldsymbol{x})$ — распределение вероятности сл. вел. $\boldsymbol{X}$, предельным переходом можно получить неравенство

$$f(\langle \boldsymbol{X} \rangle) \leqslant \langle f(\boldsymbol{X}) \rangle .$$

Любое из этих неравенств называется **неравенством Иенсена**.

C Вопросы экзаменационного минимума

1. Случайное испытание, случайные величины, распределения вероятности.

2. Математическое ожидание, моменты, кумулянты случайной величины; медиана и мода.

3. Функция плотности вероятности, кумулятивная функция распределения в одномерном и многомерном случаях.

4. Производящие функции распределения вероятности и моментов целочисленной случайной величины.
5. Характеристическая функция в одномерном и многомерном случаях, ее связь с моментами распределения вероятности.
6. Моменты случайных векторов, матрица ковариации и коэффициенты корреляции.
7. Определение и характеристическая функция многомерного нормального распределения.
8. Сходимость по вероятности.
9. Закон больших чисел для суммы независимых одинаково распределенных случайных величин.
10. Центральная предельная теорема для суммы независимых одинаково распределенных случайных величин.
11. Энтропия дискретного распределения вероятности и теорема Макмиллана.
12. Относительная энтропия в дискретном и непрерывном случаях.
13. Принцип больших уклонений для случайного блуждания.
14. Симметричные распределения Леви–Парето и обобщение центральной предельной теоремы.
15. Характеристическое наибольшее значение и распределение наибольшего значения выборки.
16. Вероятностное пространство и аксиомы А. Н. Колмогорова.
17. Условная вероятность события, условная функция плотности распределения.
18. Формула полной вероятности, формулы Байеса.
19. Цепи Маркова и распределение вероятности перехода.
20. Стационарное распределение вероятности конечной цепи Маркова.
21. Принцип детального равновесия и обратимость конечной цепи Маркова.
22. Стохастический интеграл Ито, формула Ито.
23. Стохастическое дифференциальное уравнение и его связь с уравнением Фоккера–Планка. Диффузионный процесс.
24. Процессы с независимыми и стационарными приращениями.
25. Стационарный процесс.
26. Автокорреляционная функция и спектральная плотность мощности стационарного процесса.
27. Неравенство Иенсена и примеры его применения.
28. Распределения вероятности (формулы для распределения или ф. п. в., характеристические функции, моменты первого и второго порядков): биномиальное распределение; распределение Пуассона; геометрическое распределение; показательное распределение, гамма-распределение, распределение Коши, нормальное распределение.
29. Случайный процесс Пуассона.
30. Броуновское движение в одномерном и многомерном случае.

D Литература

1. Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х тт. Пер. с англ. Ю.В. Прохорова с предисл. А.Н. Колмогорова. — М.: Мир, 1984.

2. Крамер Г. Математические методы статистики. Пер. с англ. под ред. акад. А.Н. Колмогорова. — М.: Мир, 1976.

3. Колмогоров А. Н. Основные понятия теории вероятностей. — М.: Наука, 1974 (2-е из.); М.: Фазис, 1998 (3-е изд.).

4. Вероятность и математическая статистика. Энциклопедия/Гл. ред. Ю.В. Прохоров. — М.: Большая Российская энциклопедия, 1999.

5. Гнеденко Б. В. Курс теории вероятностей. — М.: Физматгиз, 1969.

6. Ширяев А. Н. Вероятность. В 2-х кн. — М.: МЦНМО, 2004.

7. Пытьев Ю. П., Шишмарев И. А. Курс теории вероятностей и математической статистики для физиков: Учеб. пособие. — М.: Изд-во Моск. университета, 1983. — 256 с.

8. Худсон Д. Статистика для физиков. — М.: Мир, 1970.