КОНКРЕТНАЯ ТЕОРИЯ ВЕРОЯТНОСТЕЙ

А. Н. СОБОЛЕВСКИЙ

Аннотация. Вариант курса, читавшийся в НМУ осенью 2010 г.

1. Распределения вероятности на целых числах

Будем исходить из интуитивного представления о **случайном испытании**, т. е. таком эксперименте по измерению некоторой величины, который можно повторять много раз при фиксированных условиях, получая при этом случайные, т. е. различные и **непредсказуемые** заранее результаты.

Из-за непредсказуемости доступными для теоретического изучения остаются лишь множество значений, которое может принимать результат такого эксперимента, и распределение вероятности по этому множеству. Распределение вероятности можно понимать либо субъективно (как количественное выражение наших ожиданий относительно результата случайного эксперимента), либо объективно (как распределение относительных частот различных результатов в серии повторных экспериментов).

Не вдаваясь в эти мета-вероятностные тонкости, наметим пока «рабочее» понятие **случайной величины** как пары из (1) множества значений и (2) распределения вероятности по этому множеству. С математической точки зрения распределение вероятности мыслится как **мера**, а следовательно множество значений случайной величины должно быть **измеримым пространством**.

Поскольку в результате измерения обычно получается число, ограничимся пока наиболее просто устроенными подмножествами числовой прямой: счетными множествами в этой лекции и интервалами — в лекциях 2 и 3. На счетных множествах теория меры тривиальна, так что в этой лекции никакие упоминания о теории меры вообще не потребуются.

1.1. Множеством значений **целочисленной случайной величины** по определению будем считать множество натуральных чисел с нулем $\mathbf{N}_0 = \{0, 1, 2, \dots\}$. Элементы этого множества будем также называть **исходами**. Целочисленные случайные величины обозначаются прописными латинскими буквами

$$M, N, \dots, M', M'', \dots, N_1, N_2, \dots$$

и т. п., а исходы — строчными буквами $m, n, \dots$

1.2. Распределение вероятности случайной величины N есть совокупность чисел, обозначаемых $p_N(n)$ (или просто $p(n)$, если из контекста ясно, о какой случайной величине идет речь), занумерованных элементами множества $\mathbf{N}_0$ и подчиненных двум условиям:

$$p(n) \geqslant 0 \text{ при всех } n = 0, 1, 2, \dots;$$

$$\sum_{n \geqslant 0} p(n) = 1.$$

С использованием только что введенного «рабочего» определения случайной величины и его аналога для непрерывных числовых интервалов (лекция 2) можно получить бо́льшую часть интересующих нас в этом курсе результатов. Однако для изучения случайных процессов необходимо более сложное определение, которым обычно и пользуются в современной теории вероятностей. Мы увидим, как оно естественно возникнет в последних лекциях курса.

Вариант от 24 октября 2010 г.

1.3. Число $p_N(n)$ интерпретируется как **вероятность** того, что случайная величина N примет значение n. В частности, если $p_N(n) = 0$, то случайная величина N никогда не принимает значение n.

Если $p_N(n_0) = 1$ для некоторого n_0, то $p_N(n) = 0$ при $n \neq n_0$ и случайная величина N является **детерминированной**: N всегда принимает лишь значение n_0.

1.4. Конечные или счетные множества различных значений случайной величины, т. е. несовместных исходов случайного эксперимента, называются **событиями**. Вероятность события $A \subset \mathbf{N}_0$ по определению равна сумме вероятностей составляющих его исходов. Записывается это так:

$$\mathsf{P}(N \in A) = \sum_{n \in A} p_N(n).$$

Если событие A представлено счетным множеством, этот ряд сходится абсолютно в силу условий п. 1.2.

Обычно из контекста ясно, о какой случайной величине речь, и вместо $\mathsf{P}(N \in A)$ можно писать $\mathsf{P}(A)$. Иногда вместо обозначения множества A будем записывать определяющий его предикат, заменяя «немую» переменную на обозначение случайной величины: например, $\mathsf{P}(N$ четное).

1.5. Пример. Если N представляет собой число очков, выпавших на игральной кости, то $p_N(0) = p_N(7) = p_N(8) = \cdots = 0$. Если, более того, эта кость симметрична, то $p_N(1) = p_N(2) = \cdots = p_N(6) = \frac{1}{6}$. Событие «на игральной кости выпало четное число очков N» состоит из исходов 2, 4, 6, каждый из которых может пониматься и как событие $\{N = 2\}$, $\{N = 4\}$, $\{N = 6\}$ (ср. с элементами и одноэлементными множествами в теории множеств).

В сказанном до сих пор существенны два момента: что полная вероятность совокупности всех исходов, а значит и любого события, конечна (ее нормировка на единицу в п. 1.2 — не более, чем естественное и удобное соглашение) и что вероятность счетного множества исходов должна получаться как сумма бесконечного ряда, состоящего из вероятностей отдельных исходов. Приведем пример, в котором эти требования входят в противоречие друг с другом.

1.6. Контрпример. В теории чисел вводят понятие **плотности** $\rho(A)$ множества A целых чисел как предела $\lim_{m,n\to\infty} |A \cap \{-m+1, -m+2, \dots, n\}|/(m+n)$ (если он существует, то обязательно заключен между 0 и 1). Это естественная формализация интуитивного представления о «равномерном распределении вероятности» на множестве целых чисел: например, плотность множества четных чисел равна $\frac{1}{2}$, плотность множества чисел, сравнимых с 5 или 7 по модулю 8, равна $\frac{1}{4}$ и т. п.

Плотность объединения двух (или любого конечного числа) непересекающихся множеств равна сумме плотностей этих множеств. Однако при счетных объединениях плотности не складываются: множество четных чисел, обладающее плотностью $\frac{1}{2}$, есть бесконечное объединение множеств $\{0\}, \{2\}, \{4\}, \dots$, каждое из которых, как нетрудно сообразить, имеет нулевую плотность. Поэтому плотности не соответствуют никакому распределению вероятности в смысле данного выше определения.

Можно поставить вопрос, как охарактеризовать «типичное» значение данной случайной величины N. Наиболее употребительно следующее определение, образованное по аналогии с физическим понятием «центра тяжести».

1.7. Математическое ожидание $\mathsf{E}N$ случайной величины N — это сумма ряда

$$\mathsf{E}N = \sum_{n \geqslant 0} n\, p_N(n),$$

а математическое ожидание функции $f(N)$ случайной величины N — сумма ряда

$$\mathsf{E}f(N) = \sum_{n \geqslant 0} f(n)\, p(n).$$

В этой лекции будем всегда предполагать, что такие ряды сходятся.

1.8. Математическое ожидание **линейно**: для любых функций $f(\cdot)$, $g(\cdot)$ и числа α

$$\mathsf{E}[f(N)+g(N)] = \mathsf{E}f(N) + \mathsf{E}g(N), \quad \mathsf{E}(\alpha N) = \alpha\,\mathsf{E}N.$$

Для обозначения математического ожидания, помимо $\mathsf{E}N$ (англ. Expectation, фр. Espérance) используются и другие обозначения: $\mathsf{M}N$ (англ. Mean, фр. Moyenne), а в физической литературе $\langle N\rangle$ и $\overline{N}$. В данном курсе, однако, последнее обозначение используется в другом смысле (среднее выборки, см. п. 4.3).

1.9. Традиционно рассматривают следующие характеристики случайной величины N: **момент** k-го порядка: $\mathsf{E}N^k$ и **центральный момент** k-го порядка: $\mathsf{E}(N-\mathsf{E}N)^k$.

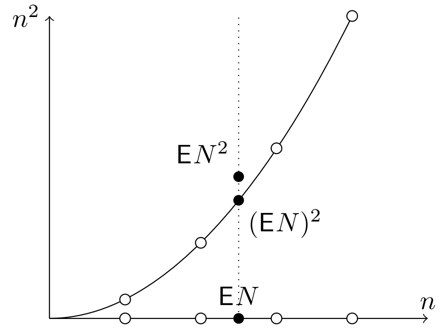
Если из контекста ясно, о какой случайной величине идет речь, будем писать $\mathsf{E}N^k = \mu_k$ и $\mathsf{E}(N - \mathsf{E}N^k) = \mathring{\mu}_k$.

1.10. Центральный момент второго порядка называется **дисперсией** и обозначается $\mathsf{D}N$. Дисперсию удобно вычислять по формуле

$$\mathsf{D}N = \mathsf{E}[N^2 - 2N\,\mathsf{E}N + (\mathsf{E}N)^2] = \mathsf{E}N^2 - (\mathsf{E}N)^2.$$

1.11. При масштабном преобразовании случайной величины дисперсия ведет себя квадратично: $\mathsf{D}(\alpha N) = \alpha^2\mathsf{D}N$.

1.12. Дисперсия всегда неотрицательна и равна нулю только для детерминированной величины. В частности, для распределения $p(1) = p(2) = p(3) = p(4) = \frac{1}{4}$ неравенство $\mathsf{E}N^2 - (\mathsf{E}N)^2 > 0$ имеет следующее геометрическое представление:



Это — частный случай **неравенства Иенсена** для выпуклой функции f: $\mathsf{E}f(N) \geqslant f(\mathsf{E}N)$.

Допустим, что несколько случайных величин надо рассмотреть одновременно. Тогда из них можно образовать вектор, принимающий значения в прямом произведении множеств значений отдельных случайных величин, и рассматривать исходы и события в этом множестве (которое в предположениях настоящей лекции по-прежнему является счетным). Почти все соответствующие определения и свойства можно сформулировать уже в простейшей ситуации, когда имеется всего одна пара случайных величин.

1.13. Множеством значений пары целочисленных случайных величин M и N является прямое произведение их множеств значений $\mathbf{N}_0 \times \mathbf{N}_0$. **Совместное распределение вероятности** по этому множеству обозначается $p_{M,N}(m,n)$ или $p(m,n)$, где $(m,n) \in \mathbf{N}_0 \times \mathbf{N}_0$. При этом предполагаются выполненными аналоги условий п. 1.2.

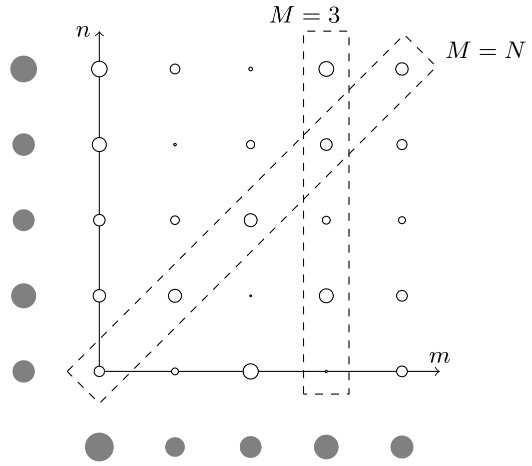
1.14. Для совместного распределения $p_{M,N}$ определяются **маргинальные распределения вероятности**, задаваемые формулами

$$p_M(m) = \sum_{n\geqslant 0} p_{M,N}(m,n), \quad p_N(n) = \sum_{m\geqslant 0} p_{M,N}(m,n).$$

Легко проверить, что это — корректно определенные распределения вероятности, т. е. что условия п. 1.2 выполнены (ср. п. 1.4).

Маргинальные распределения характеризуют одну из совместно рассматриваемых случайных величин, если другая может принимать любые значения, и могут быть определены как вероятности соответствующих событий $\{M = m\}$ или $\{N = n\}$.

Аналогично вероятность события $M = N$ определяется как $\mathsf{P}(M = N) = \sum_{l \geqslant 0} p_{M,N}(l, l)$. Следующий рисунок иллюстрирует определения совместного распределения (белые кружки), маргинальных распределений (серые кружки) и событий $\{M = 3\}$ и $\{M = N\}$.



1.15. Условное распределение вероятности характеризует случайную величину M при фиксированном значении другой: если $N = n$, то

$$\mathsf{P}(\, M = m \mid N = n\,) = \frac{p_{M,N}(m, n)}{p_N(n)}.$$

Конечно, такое определение имеет смысл только при $\mathsf{P}(N = n) > 0$.

Аналогично определяется условное распределение вероятности относительно события A, если $\mathsf{P}(A) > 0$:

$$\mathsf{P}(\, M = m \mid A\,) = \frac{\mathsf{P}(\{M = m\} \cap A)}{\mathsf{P}(A)}.$$

Легко проверить, что условное распределение вероятности в обоих вариантах удовлетворяет требованиям п. 1.2.

1.16. Две случайные величины M, N **независимы** (обозначение $M \perp N$), если условное распределение одной величины остается одним и тем же (и совпадает с соответствующим маргинальным), какое бы значение ни принимала другая:

$$\mathsf{P}(\, M = m \mid N = n\,) = p_M(m) \quad \text{для любого } n.$$

1.17. $M \perp N$ тогда и только тогда, когда $p_{M,N}(m, n) = p_M(m)\, p_N(n)$.

1.18. Несколько случайных величин $M_1, \ldots, M_k$ называются **независимыми в совокупности**, если

$$p_{M_1,\ldots,M_k}(m_1, \ldots, m_k) = p_{M_1}(m_1) \cdot \cdots \cdot p_{M_k}(m_k).$$

В последнем определении имеется тонкость, проиллюстрированная в упр. У1.3: независимость в совокупности набора из трех или большего числа случайных величин — это более сильное свойство, чем независимость каждой пары величин из этого набора.

Заметим, что отношение независимости **симметрично**. Зависимость, т. е. отсутствие независимости — это также симметричное отношение, которое нельзя смешивать с более тонким (и несимметричным) понятием причинной связи между случайными величинами.

1.19. Непосредственно из определений выводится **аддитивность** математического ожидания: $\mathsf{E}(M+N)=\mathsf{E}M+\mathsf{E}N$. Здесь последние два математических ожидания могут быть вычислены как относительно совместного распределения M и N, так и относительно маргинальных распределений:

$$\begin{aligned}\mathsf{E}(M+N) &= \sum_{m,n\geqslant 0}(m+n)\,p_{M,N}(m,n) = \\ &= \sum_{m,n\geqslant 0} m\,p_{M,N}(m,n) + \sum_{m,n\geqslant 0} n\,p_{M,N}(m,n) = && [\mathsf{E}_{M,N}M+\mathsf{E}_{M,N}N] \\ &= \sum_{m\geqslant 0} m\sum_{n\geqslant 0}p_{M,N}(m,n) + \sum_{n\geqslant 0} n\sum_{m\geqslant 0}p_{M,N}(m,n) = \\ &= \sum_{m\geqslant 0} m\,p_M(m) + \sum_{n\geqslant 0} n\,p_N(n) && [\mathsf{E}_M M+\mathsf{E}_N N].\end{aligned}$$

1.20. Равенство $\mathsf{E}MN=\mathsf{E}M\,\mathsf{E}N$ выполнено не всегда. Наиболее важным **достаточным** условием для него является независимость случайных величин:

$$\begin{aligned}\mathsf{E}MN &= \sum_{m,n\geqslant 0} mn\,p_{MN}(m,n) = [M\perp N] = \sum_{m\geqslant 0}\sum_{n\geqslant 0} mn\,p_M(m)\,p_N(n) = \\ &= \sum_{m\geqslant 0}\Big(m\,p_M(m)\sum_{n\geqslant 0} n\,p_N(n)\Big) = \Big(\sum_{n\geqslant 0} n\,p_N(n)\Big)\sum_{m\geqslant 0} m\,p_M(m) = \mathsf{E}M\,\mathsf{E}N.\end{aligned}$$

1.21. Если случайные величины M и N независимы, то $\mathsf{D}(M+N)=\mathsf{D}M+\mathsf{D}N$:

$$\begin{aligned}\mathsf{D}(M+N) &= \mathsf{E}(M+N)^2-[\mathsf{E}(M+N)]^2 = \\ &= \mathsf{E}(M^2+2MN+N^2)-(\mathsf{E}M)^2-2\,\mathsf{E}M\,\mathsf{E}N-(\mathsf{E}N)^2 = \\ &= \mathsf{E}M^2-(\mathsf{E}M)^2+\mathsf{E}N^2-(\mathsf{E}N)^2+2(\mathsf{E}MN-\mathsf{E}M\,\mathsf{E}N) = [M\perp N] = \\ &= \mathsf{E}M^2-(\mathsf{E}M)^2+\mathsf{E}N^2-(\mathsf{E}N)^2 = \mathsf{D}M+\mathsf{D}N.\end{aligned}$$

Таким образом, хотя дисперсия и является квадратичным функционалом (п. 1.11), при наличии независимости она ведет себя **аддитивно**! В дальнейшем нам встретятся многочисленные последствия этого специфически «вероятностного» факта.

1.22. **Производящая функция распределения вероятности** случайной величины N:

$$G_N(z)=\sum_{n\geqslant 0} z^n\,p_N(n)=\mathsf{E}z^N.$$

По образному выражению Д. Пойа, производящая функция позволяет разом охватить вероятности всех исходов в одном объекте, «как камни в мешке».

1.23. Производные функции G_N при $z=1$ имеют вероятностный смысл:

$$G_N(1)=1,\quad G_N'(1)=\sum_{n\geqslant 0} nz^{n-1}\,p_N(n)\Big|_{z=1}=\sum_{n\geqslant 0} n\,p_N(n)=\mathsf{E}N,$$

$$G_N''(1)=\sum_{n\geqslant 0} n(n-1)z^{n-2}\,p_N(n)\Big|_{z=1}=\mathsf{E}N(N-1)$$

Вообще,

$$\frac{\mathrm{d}^k G_N}{\mathrm{d}z^k}\bigg|_{z=1}=\mathsf{E}N^{\underline{k}},\quad \text{где } N^{\underline{k}}=N(N-1)\ldots(N-k+1).$$

Выражение $N^{\underline{k}}$ называется k-й **убывающей факториальной степенью** числа N, а величина $\mathsf{E}N^{\underline{k}}$ — **факториальным моментом** случайной величины N порядка k.

Отвлекаясь от вероятностной темы, заметим, что факториальные степени играют в исчислении конечных разностей роль, аналогичную роли степенных функций в дифференциальном исчислении. Например, имеют место тождества $(n+1)^{\underline{k}} - n^{\underline{k}} = kn^{\underline{k-1}}$, $\sum_{k\geqslant 0} n^{\underline{k}}/k! = 2^n$, $2^{n+1} - 2^n = 2^n$ (ср. в дифференциальном исчислении $(x+\mathrm{d}x)^k - x^k = kx^{k-1}\,\mathrm{d}x$, $\sum_{k\geqslant 0} x^k/k! = \mathrm{e}^x$, $\mathrm{e}^{x+\mathrm{d}x} - \mathrm{e}^x = \mathrm{e}^x\,\mathrm{d}x$).

1.24. Производящая функция моментов случайной величины N:

$$\Psi_N(s) = G_N(\mathrm{e}^s) = \mathsf{E}\mathrm{e}^{sN} = \sum_{n\geqslant 0} \mathrm{e}^{sn}\,\mathsf{P}(N=n).$$

1.25. Производные производящей функции моментов — это обычные, а не факториальные моменты: $\left.\dfrac{\mathrm{d}^k\Psi_N}{\mathrm{d}s^k}\right|_{s=0} = \left.\sum\limits_{n\geqslant 0} n^k \mathrm{e}^{sn}\,p_N(n)\right|_{s=0} = \mathsf{E}N^k$.

1.26. При сложении независимых случайных величин M и N их производящие функции перемножаются:

$$G_{M+N}(z) = \mathsf{E}z^{M+N} = \mathsf{E}(z^M z^N) = \mathsf{E}z^M\,\mathsf{E}z^N = G_M(z)\,G_N(z),$$
$$\Psi_{M+N}(s) = \mathsf{E}\mathrm{e}^{s(M+N)} = \mathsf{E}(\mathrm{e}^{sM}\,\mathrm{e}^{sN}) = \mathsf{E}\mathrm{e}^{sM}\,\mathsf{E}\mathrm{e}^{sN} = \Psi_M(s)\,\Psi_N(s)$$

1.27. Распределение вероятности суммы независимых случайных величин:

$$p_{M+N}(k) = \sum_{0\leqslant n\leqslant k} p_M(k-n)\,p_N(n).$$

Вероятностная интерпретация последней формулы такова: вероятность того, что сумма случайных величин примет значение k, равна сумме по всем разбиениям $k = (k-n)+n$ на два слагаемых вероятностей того, что первая случайная примет значение $k-n$, а вторая — значение n.

Сравнение формул двух последних пунктов показывает, почему при изучении сумм независимых случайных величин производящие функции удобнее прямых вычислений с соответствующими вероятностями.

2. Непрерывные распределения вероятности

Перейдем к изучению случайных величин, принимающих значения в континуальных множествах, т. е. на всей числовой прямой или ее интервале (a,b). Из-за более сложного устройства континуума по сравнению со счетным множеством целых чисел определение основных вероятностных понятий в данном случае связано с рядом тонкостей и требует напоминания некоторых понятий теории меры.

2.1. Множеством значений **скалярной вещественной случайной величины** по определению будем считать числовую прямую $\mathbf{R} = \{\,x\colon\ -\infty < x < \infty\,\}$. Элементы этого множества, т. е. отдельные точки, будем, как и раньше, называть **исходами**. Скалярные вещественные случайные величины обозначаются прописными латинскими буквами $X, Y, Z, \ldots$, а исходы — строчными буквами $x, y, z, \ldots$

Поскольку теперь множество всех возможных исходов континуально, распределение вероятности нельзя определить непосредственно, как в прошлой лекции, явно сопоставляя каждому исходу некоторую положительную вероятность. Вместо этого используется следующая конструкция.

2.2. Распределение вероятности скалярной случайной величины X задается **кумулятивной функцией распределения** $F_X(x)$, которая должна обладать следующими свойствами:

$$F(x) \text{ не убывает;}$$
$$F(-\infty) = 0, \quad F(\infty) = 1;$$
$$F(x) \text{ полунепрерывна справа.}$$

2.3. Значение кумулятивной функции распределения интерпретируется как вероятность того, что случайная величина X не превысит заданного числа x:

$$F_X(x) = \mathsf{P}(X \leqslant x),$$

Первое свойство п. 2.2 соответствует неотрицательности вероятностей отдельных исходов (ср. п. 1.2), а второе свойство — нормировке полной вероятности на единицу. Что касается третьего свойства, то оно условно: можно было бы исходить из соглашения $F_X(x) = \mathsf{P}(X < x)$ и предполагать, что вместо полунепрерывности справа имеет место полунепрерывность слева.

2.4. Любой полуоткрытый интервал $(a, b]$ числовой оси представляет собой **событие**, вероятность которого выражается по формуле

$$\mathsf{P}(X \in (a, b]) = \mathsf{P}(a < X \leqslant b) = F_X(b) - F_X(a).$$

Так же определяются вероятности событий, представленных интервалами других типов:

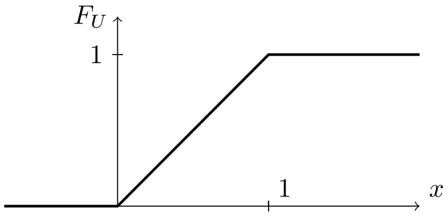
$$\mathsf{P}(X \in [a, b]) = \mathsf{P}(a \leqslant X \leqslant b) = F_X(b) - F_X(a - 0),$$
$$\mathsf{P}(X \in (a, b)) = \mathsf{P}(a < X < b) = F_X(b - 0) - F_X(a) \quad \text{и т. п.,}$$

а также событий, представленных объединением конечного или счетного множества непересекающихся интервалов.

Если интервалы $(a, b]$ и $(c, d]$ пересекаются (например, $a < c < b < d$), то их пересечение и объединение представляют события, вероятности которых определяются по формулам $\mathsf{P}((a, b] \cap (c, d]) = \mathsf{P}((c, b]) = F_X(b) - F_X(c)$, $\mathsf{P}((a, b] \cup (c, d]) = \mathsf{P}((a, d]) = F_X(d) - F_X(a)$ и т. п.

Корректность данного определения вероятности для счетного объединения непересекающихся интервалов обеспечивается абсолютной сходимостью соответствующего ряда в силу того, что вероятность положительна и нормирована на единицу.

2.5. Пример. Пусть случайная величина U равномерно распределена на отрезке $[0, 1]$, т. е. $\mathsf{P}(a \leqslant U \leqslant b) = b - a$ для всех $0 \leqslant a < b \leqslant 1$. Тогда $F_U(x) = x$ на $[0, 1]$ и равна 1 справа и 0 — слева от этого отрезка. В частности, $\mathsf{P}(U = x) = x - x = 0$ для всех $0 \leqslant x \leqslant 1$.



Иначе говоря, вероятность **каждого** отдельного значения непрерывной случайной величины может обращаться в нуль без того, чтобы эти значения становились «запрещенными» (ср. п. 1.3). Но если кумулятивная функция распределения постоянна на некотором интервале, то значения из этого интервала запрещены.

Каков самый широкий класс событий, вероятности которых можно корректно определить, развивая подход п. 2.4? Этот класс включает все события, которые можно представить множествами, получаемыми из интервалов при помощи операций пересечения и объединения, повторенных в любом порядке счетное множество раз. В совокупности эти множества образуют **борелевскую алгебру** относительно операций пересечения и объединения. Класс таких множеств очень широк, но следующий контрпример показывает, что все-таки он включает в себя не все подмножества континуума. (Заметим, однако, что в этом контрпримере существенно используется аксиома выбора.)

2.6. Контрпример. Отнесем две точки отрезка $[0, 1]$ к одному классу эквивалентности, если их разность рациональна. По аксиоме выбора можно образовать множество A, содержащее по одной точке из каждого класса эквивалентности. Введем на отрезке периодические краевые условия, отождествляя 0 и 1; тогда всевозможные сдвиги множества A на рациональные расстояния не пересекаются друг с другом и покрывают весь отрезок.

Очевидно, вероятность множества A относительно равномерного распределения не может быть ни нулевой, ни положительной, поскольку тогда полная вероятность всего отрезка $[0,1]$, состоящего из объединения счетного множества конгруэнтных копий A, должна была бы быть равной соответственно либо 0, либо бесконечности. Следовательно, множество A вообще не имеет вероятности.

Те распределения вероятности, с которыми мы в основном будем иметь дело, устроены достаточно просто: как правило, они будут относиться к классам, описываемым в пп. 2.7 и 2.9, или представлять собой смеси таких распределений (определение смеси см. в упр. У1.11).

2.7. Если существует такая функция p_X, что для всех x

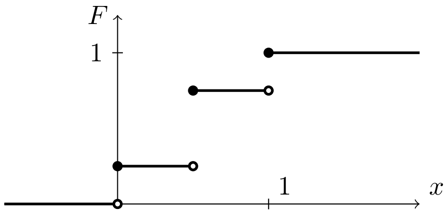
$$F_X(x) = \int_{-\infty}^{x} p_X(\xi)\,\mathrm{d}\xi,$$

то говорят, что распределение вероятности случайной величины X **абсолютно непрерывно**, а функцию p_X называют **функцией плотности вероятности** случайной величины X; она с необходимостью является неотрицательной.

2.8. Пример. Равномерное распределение по отрезку $(0,1)$ (п. 2.5) абсолютно непрерывно и имеет кусочно-постоянную плотность: $p_X(x) = 1$ внутри этого отрезка и $p_X(x) = 0$ вне его. Выбор значений p_X в точках $x = 0$, $x = 1$ может быть произвольным и не влияет на распределение вероятности (почему?).

2.9. Если F_X разрывна в точке x, то $F_X(x) = F_X(x+0) > F_X(x-0)$ и $\mathsf{P}(X = x) = F_X(x) - F_X(x-0) > 0$. Говорят, что в точке x расположен **атом вероятности** массой $\mathsf{P}(X = x)$. Если сумма вероятностей всех атомов равна единице, говорят, что распределение вероятности является **чисто точечным**.

2.10. Пример. Чисто точечное распределение вероятности с атомами в точках $x = 0$, $x = \frac{1}{2}$ и $x = 1$:

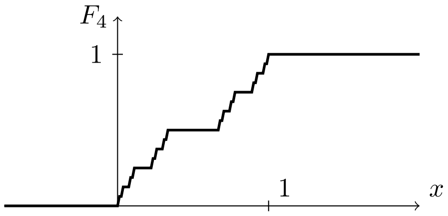


Распределения целочисленных случайных величин являются чисто точечными.

2.11. Полезно выделять еще один тип локального поведения распределения вероятности: если для некоторого $0 < \alpha < 1$ имеет место асимптотика $\mathsf{P}(|X - x| < \Delta) \sim \Delta^{\alpha}$ при $\Delta \to 0$, будем говорить, что распределение имеет в точке x **сингулярность** порядка α. Значения $\alpha = 0$ и $\alpha = 1$ соответствуют атомарному и абсолютно непрерывному случаям.

Тривиальный пример сингулярности при $x = 0$ доставляет кумулятивная функция распределения, которая равна нулю при $x < 0$, а при $x \geqslant 0$ задается формулой $F(x) = \min(x^{\alpha}, 1)$. Следующий пример менее тривиален (но при этом более типичен).

2.12. 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 далее F_i — кумулятивная функция равномерного распределения вероятности на $\mathcal{M}_i$:



В пределе при $i \to \infty$ возникает функция F_∞, которая непрерывна, но не может быть представлена интегралом вида п. 2.7 (проверьте!). Множество $\mathcal{M}_\infty$ точек ее роста называется **Канторовым множеством** средних третей, а соответствующее распределение вероятности — **Канторовой пылью**. Поскольку длина каждого из отрезков, составляющих $\mathcal{M}_i$, равна 3^{-i}, а содержащаяся в нем вероятность равна $2^{-i} = (3^{-i})^{\ln 2/\ln 3}$, показатель сингулярности во всех точках Канторовой пыли равен $\alpha = \ln 2/\ln 3$.

Сингулярные распределения вероятности возникают в эргодической теории и математической статистической физике как инвариантные меры диссипативных динамических систем, обладающих т. н. «странными аттракторами». Количественное изучение таких мер относится к геометрической теории меры и известно под названием «фрактальной геометрии». Основные импульсы развития этой дисциплины исходили из работ К. Каратеодори, Ф. Хаусдорфа, А. Безиковича 1920-х годов, а позднее — Б. Мандельброта и многочисленных физиков, которые занимались «динамическим хаосом» в 1980-х годах (П. Грассбергер, И. Прокачча, Дж. Паризи, У. Фриш).

Подробнее о мультифрактальных мерах см., например, книги: Е. Федер, Фракталы, М.: Мир, 1991 («физический» уровень строгости); K. Falconer. Fractal Geometry: Mathematical Foundations and Applications, John Wiley & Sons, 1990 (популярное, но аккуратное изложение для физиков, написанное математиком); Я. Б. Песин, Теория размерности и динамические системы, М.-Ижевск: Ин-т компьютерных исследований, 2002 (математически строгая монография).

2.13. Математическое ожидание случайной величины X определяется интегралом Стилтьеса (ср. п. 1.7):

$$\mathsf{E}X = \int_{-\infty}^{\infty} x \,\mathrm{d}F_X(x).$$

Для произвольной функции случайной величины X математическое ожидание определяется интегралом

$$\mathsf{E}f(X) = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}F_X(x)$$

Для абсолютно непрерывного или чисто точечного распределений эти формулы принимают соответственно вид

$$\mathsf{E}X = \int_{-\infty}^{\infty} x\, p_X(x)\,\mathrm{d}x \quad \text{или} \quad \mathsf{E}X = \sum_n x_n\, \mathsf{P}(X = x_n),$$
$$\mathsf{E}X = \int_{-\infty}^{\infty} f(x)\, p_X(x)\,\mathrm{d}x \quad \text{или} \quad \mathsf{E}X = \sum_n f(x_n)\, \mathsf{P}(X = x_n),$$

а для смеси распределений складываются вклады от абсолютно непрерывной и чисто точечной частей.

Как и в предыдущей лекции, будем пока предполагать, что все такие интегралы сходятся.

В формулах предыдущего пункта интегралы по $\mathrm{d}F_X$ надо понимать как интегралы Римана–Стилтьеса, т. е. как пределы интегральных сумм

$$\sum_i f(\xi_i)\,[F_X(x_{i+1}) - F_X(x_i)],$$

где $x_i < \xi_i < x_{i+1}$, при $\max_i(x_{i+1} - x_i) \to 0$ (для обычного интеграла Римана выражение в квадратных скобках имело бы вид $x_{i+1} - x_i$, т. е. это интеграл Римана–Стилтьеса относительно функции

$F(x) \equiv x$; в остальном обе теории практически совпадают). Чтобы такой интеграл была корректно определен, подынтегральная функция f должна быть непрерывной во всех точках, где находятся атомы вероятности (почему?).

Подводя итог этой части лекции, заметим, что распределения вероятности можно понимать двумя способами: как функции множеств (п. 2.4), т. е. как **меры**, а также как **функционалы**, сопоставляющие непрерывным функциям математические ожидания (п. 2.13). В функциональном анализе показывается, что эти подходы приводят к одним и тем же результатам. Хотя второй подход кажется более абстрактным, он несколько проще идейно и технически, так как свободен от трудностей, связанных с теоретико-множественными тонкостями устройства континуума вещественных чисел (такими, как существование неизмеримых множеств). Особенно эффективным он становится после введения понятия характеристической функции распределения вероятности (см. лекцию 3).

2.14. **Моменты** и **центральные моменты** (в частности, **дисперсию**) непрерывной случайной величины определяют аналогично дискретному случаю (ср. п. 1.9):

$$\mu_k = \mathsf{E}X^k = \int x^k \, \mathrm{d}F_X(x), \quad \mathring{\mu}_k = \mathsf{E}(X - \mathsf{E}X)^k = \int (x - \mu_1)^k \, \mathrm{d}F_X(x).$$

2.15. Квадратный корень из дисперсии называется **стандартным отклонением**:

$$\sigma_X = \sqrt{\mathsf{E}(X - \mathsf{E}X)^2}.$$

2.16. **Медиана** случайной величины X определяется как такой исход x_{med}, для которого события $X > x_{\mathrm{med}}$ и $X < x_{\mathrm{med}}$ равновероятны. Точнее, $x_{\mathrm{med}} = \sup\{\, x\colon F(x) \leqslant \frac{1}{2}\,\}$: такой вариант определения работает и в том случае, когда в x_{med} находится атом вероятности, так что $F(x_{\mathrm{med}}) > \frac{1}{2}$ и $F(x_{\mathrm{med}} - 0) \leqslant \frac{1}{2}$. Если имеется целый интервал, на котором $F(x) = \frac{1}{2}$, то данное определение фиксирует в качестве медианы его левую границу.

2.17. Если у случайной величины X существует непрерывная функция плотности вероятности $p(x)$, то ее **мода** определяется как точка максимума плотности: $p(x_{\max}) = \max_x p(x)$.

Наряду с математическим ожиданием, медиана и мода являются еще двумя употребительными способами придать смысл идее «типичного значения» случайной величины. В отличие от математического ожидания, для которого требуется сходимость соответствующего интеграла, медиана существует у всех случайных величин. Из задачи У2.1 видно, что медиану и математическое ожидание можно понимать как различные частные случаи одного общего понятия.

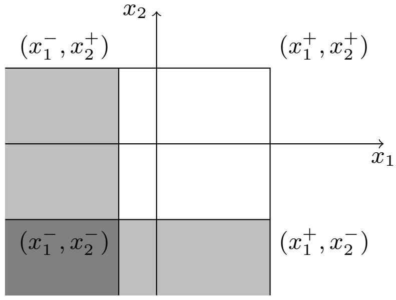
2.18. Конечный набор случайных величин $X_1, X_2, \ldots, X_n$, взятых в совокупности, образует **случайный вектор**. Векторные случайные величины обозначаются прописными полужирными латинскими буквами $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \ldots$, а их значения (**исходы**) — соответствующими полужирными строчными буквами $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots$

2.19. Распределение вероятности случайного n-мерного вектора $\boldsymbol{X}$, или **совместное распределение вероятности** его компонент $X_1, \ldots, X_n$, задается кумулятивной функцией распределения $F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathsf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n)$, которая обладает следующими свойствами:

$$\text{если } x_1^- < x_1^+, \ldots, x_n^- < x_n^+, \text{ то } \textstyle\sum_{(\varepsilon_1,\ldots,\varepsilon_n)} (\varepsilon_1 1) \ldots (\varepsilon_n 1) F(x_1^{\varepsilon_1}, \ldots, x_n^{\varepsilon_n}) \geqslant 0,$$
$$\text{где суммирование проводится по всем } 2^n \text{ наборам знаков } \varepsilon_i = \pm;$$
$$F(x_1, \ldots, x_{i-1}, -\infty, x_{i+1}, \ldots, x_n) = 0 \text{ при любом } 1 \leqslant i \leqslant n;$$
$$F(\infty, \infty, \ldots, \infty) = 1;$$
$$\textstyle\lim_{y_1 \downarrow x_1, y_2 \downarrow x_2, \ldots, y_n \downarrow x_n} F(y_1, y_2, \ldots, y_n) = F(x_1, x_2, \ldots, x_n).$$

Если не возникает неясности, нижний индекс $\boldsymbol{X}$ можно опускать.

Первое из этих свойств выражает положительность вероятности: $\mathsf{P}(x_1^- < X_1 \leqslant x_1^+, \ldots, x_n^- < X_n \leqslant x_n^+) \geqslant 0$ и выражает **принцип включения-исключения** для подсчета вероятностей. Его легче понять на примере $n = 2$, где для любых $x_1^- < x_1^+$, $x_2^- < x_2^+$ должно выполняться неравенство $F(x_1^+, x_2^+) - F(x_1^+, x_2^-) - F(x_1^-, x_2^+) + F(x_1^-, x_2^-) \geqslant 0$:



Действительно, при вычислении вероятности того, что $x_1^- < X \leqslant x_1^+$, $x_2^- < X_2 < x_2^+$ из вероятности «белого» квадранта с вершиной в (x_1^+, x_2^+) вычитаются вероятности «светлосерых» квадрантов и затем прибавляется вероятность «темносерого» квадранта.

Помимо теории вероятностей первое свойство п. 2.19 встречается еще у функций многих переменных в исследовании операций, где оно называется **супермодулярностью** или **свойством Монжа**.

2.20. Аналогично п. 2.4 определение вероятности может быть распространено с событий, представленных множествами-«брусами» вида $\{\boldsymbol{x} \colon x_i^- \leqslant x_i \leqslant x_i^+, 1 \leqslant i \leqslant n\}$, которые играют роль многомерных интервалов, на их всевозможные конечные и объединения и пересечения в счетном числе, образующие борелевскую алгебру **измеримых множеств**.

2.21. Распределение вероятности случайного вектора $\boldsymbol{X}$ называется **абсолютно непрерывным**, если существует такая **функция плотности вероятности** $p_{\boldsymbol{X}}$, что

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{y_1 \leqslant x_1, \ldots, y_n \leqslant x_n} p_{\boldsymbol{X}}(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}.$$

Функция $p_{\boldsymbol{X}}$ с необходимостью неотрицательна.

2.22. Моменты случайного вектора определяются как математические ожидания произведений его компонент: $\mathsf{E} X_{i_1} \ldots X_{i_k}$, где среди индексов $i_1, \ldots, i_k$ могут быть повторяющиеся. Число k сомножителей в этом произведении называется **порядком** момента. Аналогично определяются **центральные моменты**: $\mathsf{E}[(X_{i_1} - \mathsf{E} X_{i_1}) \ldots (X_{i_k} - \mathsf{E} X_{i_k})]$.

Интегралы, появляющиеся в этом определении, будем в абсолютно непрерывном случае понимать как кратные интегралы Римана, а в чисто точечном — как суммы.

В общем случае математическое ожидание определяется как многомерный интеграл Стилтьеса (см., например, Г. Е. Шилов и Б. Л. Гуревич, Интеграл, мера и производная, М.: Наука, 1967).

2.23. Маргинальные распределения для отдельных компонент случайного вектора имеют кумулятивные функции распределения $F_{X_i}(x_i) = F(\infty, \ldots, x_i, \ldots, \infty)$, где $1 \leqslant i \leqslant n$. Для абсолютно непрерывных распределений маргинальные распределения задаются функциями плотности вероятности

$$p_{X_i}(x_i) = \int p(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \ldots \mathrm{d}x_n = \int p(\boldsymbol{x})\, \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}x_i}, \quad 1 \leqslant i \leqslant n.$$

Вообще маргинальные распределения можно определять для любых совокупностей компонент случайного вектора; их также называют **проекциями** распределения случайного вектора $\boldsymbol{X}$ на соответствующие координатные подпространства.

2.24. Пусть $(\boldsymbol{X}, \boldsymbol{Y})$ — случайный вектор, компоненты которого разбиты на два подвектора $\boldsymbol{X}$ и $\boldsymbol{Y}$. **Условное распределение вероятности** вектора $\boldsymbol{X}$ при условии, что

$\boldsymbol{Y} \in A$, где A — измеримое множество положительной вероятности, задается кумулятивной функцией (ср. п. 1.15)

$$F_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{Y} \in A) = \frac{\mathsf{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n, \boldsymbol{Y} \in A)}{\mathsf{P}(\boldsymbol{Y} \in A)}.$$

Для достаточно регулярных абсолютно непрерывных распределений (например, выражаемых кусочно-непрерывными функциями плотности вероятности) можно написать выражение для **функции условной плотности вероятности**: если X, Y — две компоненты случайного вектора и $\mathsf{P}(Y = y) = 0$, то

$$p_X(x \mid Y = y) = \lim_{\delta \to 0} p_X(x \mid y - \delta < Y < y + \delta) = \lim_{\delta \to 0} \frac{\int_{y-\delta}^{y+\delta} p_{X,Y}(x, y')\, \mathrm{d}y'}{\int_{-\infty}^{+\infty} \mathrm{d}x \int_{y-\delta}^{y+\delta} p_{X,Y}(x, y')\, \mathrm{d}y'} = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

Вообще, если $\boldsymbol{X}$, $\boldsymbol{Y}$ — для случайных вектора, рассматриваемых вместе, то функция условной плотности вероятности случайного вектора $\boldsymbol{X}$ при условии, что $\boldsymbol{Y} = \boldsymbol{y}$, задается выражением $p_{\boldsymbol{X}|\boldsymbol{Y}}(\, \boldsymbol{x} \mid \boldsymbol{y}\,) = p(\boldsymbol{x}, \boldsymbol{y})/p_{\boldsymbol{Y}}(\boldsymbol{y})$.

По сравнению с п. 1.15 определение условной плотности в абсолютно непрерывном случае учитывает тот факт, что $\mathsf{P}(Y = y) = 0$. Говоря нестрого, речь идет об условной вероятности по отношению к «событию» $y \leqslant Y \leqslant y + \mathrm{d}y$, имеющему «инфинитезимально малую, но положительную» вероятность. Следующий пример иллюстрирует возникающую здесь тонкость.

2.25. Контрпример. Пусть плотность $p_{X,Y}(x, y)$ задает равномерное распределение на единичном квадрате $(0, 1)^2$. Тогда $p_X(x \mid X = Y)$, понимаемая как предел при $\delta \to 0$ условной плотности $p_X(x \mid -\delta < X - Y < \delta)$, постоянна и равна 1, в то время как аналогичный предел для $p_X(x \mid 1 - \delta < X/Y < t + \delta)$ равен $2x$ (проверьте!).

2.26. Компоненты $X_1, \ldots, X_n$ случайного вектора $\boldsymbol{X}$ называются **независимыми в совокупности**, если их совместное распределение вероятности распадается в произведение индивидуальных распределений отдельных компонент: $F(\boldsymbol{x}) = F_{X_1}(x_1) \ldots F_{X_n}(x_n)$ или $p(\boldsymbol{x}) = p_{X_1}(x_1) \ldots p_{X_n}(x_n)$.

2.27. Свойства пп. 1.19 и 1.21 (аддитивность математического ожидания и аддитивность дисперсии, если случайные величины независимы) выполнены и в непрерывном случае.

3. Характеристические функции распределений вероятности. Гауссовы случайные величины

Введенные в предыдущей лекции конструкции находятся в прямой аналогии с дискретным случаем и позволяют в принципе полностью описать распределение вероятности любой заданной случайной величины или случайного вектора. Тем не менее работать с ними технически неудобно: формулы громоздки, при вычислениях приходится различать абсолютно непрерывные и чисто точечные компоненты распределений и т. п. В этой лекции вводится аппарат характеристических функций, основанный на аналогии с производящими функциями целочисленных случайных величин. Этот аппарат значительно облегчает вероятностные вычисления.

3.1. Характеристической функцией распределения вероятности случайной величины X называется функция

$$\varphi_X(s) = \mathsf{E}\mathrm{e}^{\mathrm{i}sX} = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}sx}\, \mathrm{d}F_X(x)$$

(преобразование Фурье–Стилтьеса распределения F_X).

Для любого распределения вероятности данный интеграл сходится абсолютно, поскольку $|\mathrm{e}^{\mathrm{i}sx}| = 1$. Для абсолютно непрерывного распределения вероятности характеристическая функция совпадает с преобразованием Фурье функции плотности вероятности $p_X(x)$.

Прямым аналогом производящей функции дискретного распределения вероятности является более экзотическое преобразование Меллина–Стилтьеса $\mathsf{E} z^X = \int z^x \mathrm{d}F_X(x)$ (ср. п. 1.22), но на практике им не пользуются из-за аналитических трудностей, связанных с его определением.

Можно было бы также рассматривать прямой аналог производящей функции моментов (п. 1.24) — преобразование Лапласа–Стилтьеса $\mathsf{E}\mathrm{e}^{sX} = \int \mathrm{e}^{sx}\, \mathrm{d}F_X(x)$ распределения вероятности, однако условие сходимости этого интеграла гораздо ограничительнее, чем для характеристической функции.

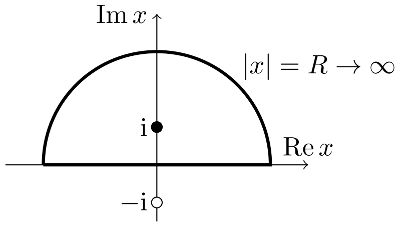
3.2. Производные характеристической функции имеют вероятностный смысл:

$$\left.\frac{\mathrm{d}^k \varphi_X}{\mathrm{d}(\mathrm{i}s)^k}\right|_{s=0} = \mathsf{E}X^k.$$

Следовательно, $\varphi_X(s) = \sum_{k\geqslant 0} \mathsf{E}X^k (\mathrm{i}s)^k/k!$, если данный ряд сходится.

Интегралы, выражающие моменты достаточно высоких порядков, могут расходиться, если распределение вероятности слишком медленно убывает на бесконечности. В таких случаях характеристическая функция не имеет производных соответствующих порядков в нуле. Это проявление общего свойства: поведение «на бесконечности» одной из функций, сопряженных преобразованием Фурье, отображается в поведение другой функции в окрестности нуля.

3.3. Пример. Найдем характеристическую функцию распределения распределения Коши $p(x) = (\pi(1+x^2))^{-1}$. При $s > 0$ методом вычетов получаем



$$\varphi(s) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{\mathrm{i}sx}}{(x-\mathrm{i})(x+\mathrm{i})}\,\mathrm{d}x = 2\pi\mathrm{i}\,\frac{1}{\pi}\left.\frac{\mathrm{e}^{\mathrm{i}sx}}{x+\mathrm{i}}\right|_{x=\mathrm{i}} = \mathrm{e}^{-s}.$$

При $s < 0$ контур интегрирования замыкается через нижнюю полуплоскость, а его ориентация противоположна. Ответ определяется полюсом при $x = -\mathrm{i}$: $\varphi(s) = \mathrm{e}^{s}$. Комбинируя оба ответа, получаем $\varphi(s) = \mathrm{e}^{-|s|}$: в нуле φ не имеет ни одной производной, а все моменты $\mathsf{E}X^k$ расходятся.

Трудность другого типа возникает, если все моменты существуют, но растут слишком быстро, так что ряд Тейлора характеристической функции расходится. В этом случае распределение вероятности не всегда может быть однозначно восстановлено по совокупности своих моментов.

3.4. Пример. Моменты логнормального распределения, которое имеет на полуоси $x > 0$ плотность $p(x) = \mathrm{e}^{-(\ln x)^2/2}/\sqrt{2\pi\sigma^2 x^2}$, конечны и равны $\mathsf{E}X^k = \mathrm{e}^{k^2/2}$, так что ряд Тейлора его характеристической функции расходится. Такие же моменты имеет распределение $\tilde{p}(x) = p(x)(1+\sin(2\pi\ln x))$ (упражнение У2.15).

В лекции 4 будет показано, что по характеристической функции распределения само распределение восстанавливается однозначно. Выяснение точных условий, при которых его можно восстановить по последовательности моментов (т. е. условий аналитичности характеристической функции в терминах моментов) — это так называемая **проблема моментов**. Для этого требуется некоторое ограничение на рост моментов, например такое, как в следующей теореме.

3.5. Теорема Карлемана (без доказательства). Если ряд $\sum_{k\geqslant 0}(\mathsf{E}X^{2k})^{-1/2k}$ расходится, то характеристическая функция аналитична при $s = 0$ и может быть восстановлена по своему ряду Тейлора.

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

◀ Следует из того факта, что $|\mathrm{e}^{\mathrm{i}sx}| = 1$. ▶

3.7. $\operatorname{Re}\varphi_X(-s) = \operatorname{Re}\varphi_X(s)$, $\operatorname{Im}\varphi_X(-s) = -\operatorname{Im}\varphi_X(s)$, $\varphi_X(-s) = \varphi^*(s)$; если распределение случайной величины X симметрично, то $\operatorname{Im}\varphi_X(s) \equiv 0$.

3.8. Характеристическая функция равномерно непрерывна при $-\infty < s < \infty$.

◀ Действительно, выберем сколь угодно малое $\varepsilon > 0$ и оценим разность значений характеристической функции в точках s и $s+h$ следующим образом:

$$|\varphi(s+h) - \varphi(s)| = \left|\int \left(\mathrm{e}^{\mathrm{i}(s+h)x} - \mathrm{e}^{\mathrm{i}s}\right) \mathrm{d}F(x)\right| = \left|\int \mathrm{e}^{\mathrm{i}sx}\left(\mathrm{e}^{\mathrm{i}hx} - 1\right) \mathrm{d}F(x)\right| \leqslant$$
$$\leqslant \int \left|\mathrm{e}^{\mathrm{i}hx} - 1\right| \mathrm{d}F(x) = \int\limits_{|x|\leqslant R} \left|\mathrm{e}^{\mathrm{i}hx} - 1\right| \mathrm{d}F(x) + \int\limits_{|x|>R} \left|\mathrm{e}^{\mathrm{i}hx} - 1\right| \mathrm{d}F(x).$$

Теперь выберем R настолько большим, чтобы $\mathsf{P}(|X| > R) < \varepsilon/4$. Поскольку $|\mathrm{e}^{\mathrm{i}hx} - 1| \leqslant 2$, второй интеграл при этом не превосходит по величине $\varepsilon/2$. После этого выберем h столь малым, чтобы $|\mathrm{e}^{\mathrm{i}hx} - 1| < \varepsilon/2$ при всех $|x| \leqslant R$. Тогда и первый интеграл не превосходит $\varepsilon/2$ и, таким образом, по заданному $\varepsilon > 0$ подобрано столь малое $h > 0$, что $|\varphi(s+h) - \varphi(s)| < \varepsilon$ при любом s. ▶

Иногда возникает необходимость проверить, является ли заданная функция характеристической функцией некоторой случайной величины. Результаты предыдущих пунктов дают необходимые, но не достаточные условия этого. Чтобы получить необходимый и достаточный критерий, заметим, что множество характеристических функций состоит из всевозможных выпуклых комбинаций экспонент $\mathrm{e}^{\mathrm{i}sx}$. Следующая теорема дает двойственное описание этого выпуклого множества как решения некоторой системы линейных неравенств.

3.9. Теорема Бохнера (без доказательства). Функция $\varphi(s)$ является характеристической функцией некоторой случайной величины тогда и только тогда, когда она непрерывна и **положительно определена**: для любых наборов (s_i) и комлексных чисел (ξ_i), $1 \leqslant i \leqslant k$, $k = 1, 2, \ldots$, выполнено неравенство $\sum_{i,j} \varphi(s_i - s_j)\xi_i\xi_j^* \geqslant 0$, где ξ^* обозначает комплексное сопряжение.

3.10. Закон преобразования характеристической функции при аффинном преобразовании скалярной случайной величины:

$$\varphi_{aX+b}(s) = \mathsf{E}\mathrm{e}^{\mathrm{i}s(aX+b)} = \mathsf{E}\mathrm{e}^{\mathrm{i}asX}\,\mathrm{e}^{\mathrm{i}sb} = \mathrm{e}^{\mathrm{i}sb}\,\varphi_X(as).$$

Аналогичная формула для кумулятивной функции распределения:

$$F_{aX+b}(x) = \mathsf{P}(aX + b \leqslant x) = \mathsf{P}(X \leqslant \tfrac{x-b}{a}) = F_X(\tfrac{x-b}{a}), \quad a > 0.$$

3.11. В силу пп. 3.8 и 3.6 в окрестности точки $s = 0$, где $\varphi_X(s) > 0$, может быть определен **характеристический показатель** $\eta_X(s) = \ln \varphi_X(s)$ (это определение однозначно, если положить $\eta_X(0) = 0$ и потребовать непрерывности η_X, т. е. оставаться на главной ветви комплексного логарифма).

3.12. При сложении независимых случайных величин их характеристические функции перемножаются:

$$\varphi_{X+Y}(s) = \mathsf{E}\mathrm{e}^{\mathrm{i}sX+Y} = \mathsf{E}(\mathrm{e}^{\mathrm{i}sX}\,\mathrm{e}^{\mathrm{i}sY}) = [X \perp Y] = \mathsf{E}\mathrm{e}^{\mathrm{i}sX}\,\mathsf{E}\mathrm{e}^{\mathrm{i}sY} = \varphi_X(s)\,\varphi_Y(s) \quad (\text{ср. п. } 1.26)$$

и поэтому функции плотности вероятности слагаемых подвергаются свертке (ср. п. 1.27):

$$p_{X+Y}(u) = \int_{-\infty}^{\infty} p_Y(u-x)p_X(x)\,\mathrm{d}x,$$

а характеристические показатели складываются:

$$\eta_{X+Y}(s) = \eta_X(s) + \eta_Y(s).$$

3.13. Коэффициенты разложения характеристического показателя случайной величины X в ряд Тейлора с центром в нуле называются **кумулянтами** случайной величины X (ср. п. 3.2):

$$\eta_X(s) = \sum_{k\geqslant 0} \varkappa_k \frac{(\mathrm{i}s)^k}{k!}, \quad \varkappa_k = \left.\frac{\mathrm{d}\eta_X(s)}{\mathrm{d}s}\right|_{s=0}.$$

3.14. Последовательно дифференцируя по $\mathrm{i}s$ равенство $\varphi(s) = \mathrm{e}^{\eta(s)}$ в окрестности нуля, где характеристический показатель однозначно определен, можно получить выражения для моментов через кумулянты:

$$\begin{aligned}
\varphi' &= \mathrm{e}^\eta \eta' : & \mu_1 &= \varkappa_1;\\
\varphi'' &= \mathrm{e}^\eta (\eta')^2 + \mathrm{e}^\eta \eta'' : & \mu_2 &= \varkappa_1^2 + \varkappa_2;\\
\varphi''' &= \mathrm{e}^\eta (\eta')^3 + 3\mathrm{e}^\eta \eta'\eta'' + \mathrm{e}^\eta \eta''' : & \mu_3 &= \varkappa_1^3 + 3\varkappa_1\varkappa_2 + \varkappa_3;\\
\varphi^{\mathrm{IV}} &= \mathrm{e}^\eta (\eta')^4 + 6\mathrm{e}^\eta (\eta')^2\eta'' + 3\mathrm{e}^\eta (\eta'')^2 + 4\mathrm{e}^\eta \eta'\eta''' + \mathrm{e}^\eta \eta^{\mathrm{IV}} : & \mu_4 &= \varkappa_1^4 + 6\varkappa_1^2\varkappa_2 + 3\varkappa_2^2 + 4\varkappa_1\varkappa_3 + \varkappa_4
\end{aligned}$$

и т. д. Обратные формулы имеют вид

$$\begin{aligned}
\varkappa_1 &= \mu_1; & &\\
\varkappa_2 &= \mu_2 - \mu_1^2; & &[= \mathring{\mu}_2 = \mathsf{D}X]\\
\varkappa_3 &= \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3; & &[= \mathring{\mu}_3]\\
\varkappa_4 &= \mu_4 - 4\mu_1\mu_3 + 12\mu_1^2\mu_2 - 3\mu_2^2 - 6\mu_1^4. & &[= \mathring{\mu}_4 - 3\mathring{\mu}_2^2]
\end{aligned}$$

Несмотря на то, что кумулянты второго и более высоких порядков задаются нелинейными выражениями, при сложении независимых случайных величин они ведут себя **аддитивно**, как и сам характеристический показатель. Этим замечательным свойством и обусловлена особая роль кумулянтов в аналитическом аппарате теории вероятностей. Ранее мы уже проверили его для дисперсии, которая по существу является квадратичным кумулянтом (ее совпадение со вторым центральным моментом — случайность).

3.15. Вероятности и характеристические функции безразмерны. Если случайная величина X размерна, то ее стандартное отклонение σ_X имеет такую же размерность, а функция плотности вероятности p_X и аргумент s ее характеристической функции — обратную размерность.

3.16. «Обезразмеренные» кумулянты 3 и 4 порядков называют **асимметрией** (англ. skewness) $\mathsf{S} = \varkappa_3/\sigma^3$ и **эксцессом** (англ. kurtosis) $\mathsf{K} = \varkappa_4/\sigma^4$. Величину $\mathsf{F} = \mathsf{K} + 3 = \mathring{\mu}_4/\sigma^4$ называют **пологостью** (англ. flatness).

Построенная теория легко обобщается на случайные векторы.

3.17. Случайному вектору $\boldsymbol{X}$ соответствует **характеристическая функция**

$$\varphi_{\boldsymbol{X}}(\boldsymbol{s}) = \mathsf{E}\mathrm{e}^{\mathrm{i}(\boldsymbol{s}\cdot\boldsymbol{X})} = \int \mathrm{e}^{\mathrm{i}\sum_i s_i X_i}\, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}).$$

3.18. $\left.\partial^k \varphi_{\boldsymbol{X}} / \partial(\mathrm{i}s_{i_1}) \ldots \partial(\mathrm{i}s_{i_k})\right|_{\boldsymbol{s}=\boldsymbol{0}} = \mathsf{E}X_{i_1} \ldots X_{i_k}$ (ср. п. 3.2).

3.19. Закон преобразования характеристической функции при аффинном преобразовании случайного вектора (ср. п. 3.10):

$$\varphi_{A\boldsymbol{X}+\boldsymbol{b}}(\boldsymbol{s}) = \mathsf{E}\mathrm{e}^{\mathrm{i}(\boldsymbol{s}\cdot(A\boldsymbol{X}+\boldsymbol{b}))} = \mathrm{e}^{\mathrm{i}(\boldsymbol{s}\cdot\boldsymbol{b})}\varphi_{\boldsymbol{X}}(A^{\mathsf{T}}\boldsymbol{s}).$$

Заметим, что матрица в аргументе характеристической функции транспонирована.

3.20. Сложение независимых случайных векторов соответствует умножению их характеристических функций, свертке плотностей вероятности и сложению характеристических показателей (ср. п. 3.12).

3.21. Кумулянты случайного вектора первых двух порядков — это его математическое ожидание $\boldsymbol{\mu}$ и **матрица ковариации** Γ:

$$\boldsymbol{\mu} := \frac{\partial \varphi_{\boldsymbol{X}}(\boldsymbol{s})}{\partial(\mathrm{i}\boldsymbol{s})}\bigg|_{\boldsymbol{s}=\boldsymbol{0}} = \frac{\partial \ln \varphi_{\boldsymbol{X}}(\boldsymbol{s})}{\partial(\mathrm{i}\boldsymbol{s})}\bigg|_{\boldsymbol{s}=\boldsymbol{0}},$$

$$\Gamma_{ij} := \frac{\partial^2 \ln \varphi_{\boldsymbol{X}}(\boldsymbol{s})}{\partial s_i \partial s_j}\bigg|_{\boldsymbol{s}=\boldsymbol{0}} = \mathsf{E}(X_i X_j) - \mathsf{E}X_i\,\mathsf{E}X_j = \mathsf{E}(X_i - \mathsf{E}X_i)(X_j - \mathsf{E}X_j).$$

Особенно важными в приложениях являются определяемые ниже гауссовы случайные величины и случайные векторы.

3.22. Случайная величина, для которой функция плотности вероятности и характеристическая функция имеют вид

$$p(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}, \qquad \varphi(s) = \mathrm{e}^{-\frac{1}{2}s^2},$$

называется распределенной **нормально**. Произвольное аффинное преобразование нормальной случайной величины называется **гауссовой** случайной величиной. Функция плотности вероятности и характеристическая функция общей гауссовой случайной величины с математическим ожиданием μ и стандартным отклонением σ имеют вид (упражнение У4.1)

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{1}{2\sigma^2}(y-\mu)^2}, \qquad \varphi(s) = \mathrm{e}^{\mathrm{i}\mu s - \frac{\sigma^2}{2}s^2}.$$

3.23. Кумулянты $\varkappa_k$ гауссовой случайной величины при $k \geqslant 3$ равны 0; в частности, она обладает нулевым эксцессом.

3.24. Теорема Марцинкевича (без доказательства). Не существует случайных величин, характеристические показатели которых были бы полиномами выше второго порядка.

В статистике оценка эксцесса используется как критерий того, насколько хорошо некоторая случайная величина может быть моделирована гауссовым распределением. Распределение вероятности с положительным эксцессом (т. е. пологостью, превышающей 3) характеризуется более острым пиком и более тяжелыми «хвостами», чем нормальное, а с отрицательным эксцессом — более пологим пиком и легкими «хвостами».

3.25. Случайный вектор называется **гауссовым**, если его проекция на любое направление является гауссовой случайной величиной.

3.26. Образ гауссова случайного вектора под действием любого линейного преобразования остается гауссовым.

3.27. Если $\boldsymbol{X}$ — гауссов случайный вектор с матрицей ковариации Γ и мат. ожиданием $\mathsf{E}\boldsymbol{X} = \boldsymbol{m}$, то $\varphi_{\boldsymbol{X}}(\boldsymbol{s}) = \mathrm{e}^{\mathrm{i}(\boldsymbol{s}\cdot\boldsymbol{m}) - \frac{1}{2}\sum_{i,j}\Gamma_{ij}s_i s_j}$ (упражнение У4.9).

3.28. Матрица ковариации n-мерного гауссова случайного 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 случайный вектор называется **вырожденным**.

3.29. Если $\det\Gamma > 0$, то функция плотности вероятности n-мерного гауссова случайного вектора $\boldsymbol{X}$ из п. 3.27 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 (упражнение У4.9).

4. Закон больших чисел и слабая сходимость распределений вероятности

С этой лекции начинается новая глава курса, посвященная асимптотическим теоремам теории вероятностей. При изучении асимптотических теорем мы будем пользоваться следующей вероятностной моделью.

4.1. Фиксируем некоторое распределение вероятности и рассмотрим набор из n **независимых одинаково распределенных** по этому закону случайных величин $X_1, X_2, \ldots, X_n$.

4.2. Иногда бывает удобно пользоваться языком математической статистики, на котором фиксированное распределение вероятности называют **генеральной совокупностью**, а набор n независимых одинаково распределенных случайных величин $X_1, \ldots, X_n$ — **выборкой объема** n из генеральной совокупности.

4.3. Закон больших чисел — это асимптотическая теорема о поведении **среднего выборки**

$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n).$$

Если объем выборки n однозначно определен контекстом или не важен, будем сокращать это обозначение до $\overline{X}$.

4.4. Для вывода закона больших числе в форме Чебышёва (п. 4.6) предположим, что $\mathsf{E}X_i = \mu$, $\mathsf{D}X_i = \sigma^2$ для всех X_i. Тогда

$$\mathsf{E}\overline{X}_n = \frac{1}{n}\,n\mu = \mu, \qquad \mathsf{D}\overline{X}_n = \frac{1}{n^2}\,n\sigma^2 = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0.$$

В отличие от математического ожидания $\mathsf{E}X = \mu$, среднее выборки $\overline{X}$ само является случайной величиной. Однако в силу результатов предыдущего пункта можно ожидать, что при больших n случайная величина $\overline{X}$ «стремится» к неслучайной величине μ. Чтобы придать этому точный смысл, воспользуемся следующей стандартной леммой.

4.5. Неравенство Бьенэме–Чебышёва. Если $\mathsf{E}Y = \mu$, $\mathsf{D}Y = \sigma^2$, а ξ — произвольный положительный параметр, то

$$\mathsf{P}(|Y - \mu| > \xi\sigma) \leqslant \frac{1}{\xi^2}.$$

◀ Оценим дисперсию случайной величины Y:

$$\sigma^2 = \int (y-\mu)^2\,\mathrm{d}F_Y(y) = \int_{|y-\mu|\leqslant\xi\sigma} (y-\mu)^2\,\mathrm{d}F_Y(y) + \int_{|y-\mu|>\xi\sigma} (y-\mu)^2\,\mathrm{d}F_Y(y) \geqslant$$
$$\geqslant \int_{|y-\mu|>\xi\sigma} (y-\mu)^2\,\mathrm{d}F_Y(y) \geqslant \xi^2\sigma^2 \int_{|y-\mu|>\xi\sigma} \mathrm{d}F_Y(y) = \xi^2\sigma^2\,\mathsf{P}(|Y-\mu| > \xi\sigma).$$

После деления на $\sigma^2\xi^2$ получаем требуемое неравенство. ▶

4.6. Закон больших чисел в форме Чебышёва. Если $X_1, \ldots, X_n$ — выборка случайных величин с математическим ожиданием μ и дисперсией σ^2 (п. 4.4), то

$$\mathsf{P}(|\overline{X}_n - \mu| > \varepsilon) \leqslant \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

◀ Это следует из неравенства Бьенэме–Чебышёва при $\varepsilon > 0$ и $\xi = \varepsilon/\sqrt{\sigma^2/n}$. ▶

Закон больших чисел можно установить и без предположения о конечности стандартного отклонения (это будет сделано в п. 4.20), но следующий пример показывает, что чрезмерный разброс значений случайных слагаемых слишком велик, усреднение выборки не приводит к возникновению в пределе детерминированной величины.

4.7. Контрпример. Пусть случайные величины $X_1, \ldots, X_n$ распределены по Коши (см. п. 3.3), а значит не имеют ни математического ожидания, ни тем более дисперсии. Тогда

$$\varphi_{\overline{X}_n}(s) = \varphi_{\frac{1}{n}(X_1+\cdots+X_n)}(s) = \varphi_{X_1+\cdots+X_n}\left(\frac{s}{n}\right) = \left(\varphi\left(\frac{s}{n}\right)\right)^n = \mathrm{e}^{-n|s/n|} = \mathrm{e}^{-|s|} = \varphi_X(s),$$

т. е. среднее выборки распределено так же, как и любое из случайных слагаемых.

Таким образом, операция усреднения почему-то оказывается не в состоянии сократить разброс случайных слагаемых. В следующей лекции мы включим это неожиданное явление в общую теорию, а еще через лекцию выясним его причину.

Итак, закон больших чисел утверждает, что сумма большого числа аналогичных малых слагаемых, *если подходящая мера разброса каждого слагаемого обратно пропорциональна их числу*, стремится к некоторой детерминированной величине. В статистической физике в подобных ситуациях говорят, что имеет место «самоусреднение» (self-averaging). Различные обобщения закона больших чисел, возникающие в теории вероятностей, сохраняют две основных особенности: детерминированность результата и характер происходящего перемасштабирования, обратно пропорционального числу слагаемых («скейлинг закона больших чисел»).

Сравнение с упр. У1.7 показывает, что для выполнения закона больших чисел существенна малость не просто ожидаемого вклада каждого слагаемого в сумму, но именно той или иной меры разброса его значений — например, стандартного отклонения. А именно, с помощью «прореживания» (упр. У1.12) можно обеспечить малость вклада каждого слагаемого, обратно пропорциональную их числу, но при этом не изменять масштаб разброса значений слагаемых. Тогда вместо детерминированного предела возникает другой тип предельного распределения (ср. упр. У1.7 и последнее выражение в упр. У2.19) и тем самым другой тип предельной теоремы — предельная теорема Пуассона или «закон малых чисел». Как выяснится в следующей лекции, возможны и другие варианты собственно перемасштабирования, приводящие к совершенно иным результатам и другим классам обобщений («скейлинг центральной предельной теоремы»).

4.8. Сравним утверждение п. 4.6 с обычным определением сходимости последовательности $\overline{x}_n \to \mu$ в математическом анализе, но записанным на вероятностном языке: для сколь угодно малого $\varepsilon > 0$ найдется такое $n_0(\varepsilon)$, что $\mathsf{P}(|\overline{x}_n - \mu| > \varepsilon) = 0$ как только $n > n_0(\varepsilon)$. Явная аналогия между этими утверждениями показывает, что закон больших чисел содержит вероятностное обобщение понятия сходимости.

Поскольку, как мы уже знаем, возможны и другие типы предельных теорем, в открывшемся направлении направлении необходимо пройти на шаг дальше и допустить, что качестве предела такой «новой» сходимости может выступать не только детерминированная, но и случайная величина, точнее, некоторое распределение вероятности.

4.9. Слабая сходимость распределений вероятности. Говорят, что последовательность распределений случайных величин $X_1, X_2, \ldots$ **слабо сходится** к распределению случайной величины X, если

$$\mathsf{P}(a < X_n < b) \xrightarrow[n\to\infty]{} \mathsf{P}(a < X < b).$$

для любых точек $a < b$, не являющихся атомами предельного распределения.

Последняя оговорка важна даже для «обычной» сходимости, когда распределения как раз атомарны: если $x_n \to x$ снизу, то $\mathsf{P}(x_n < x) \equiv 1 \not\to \mathsf{P}(x < x) = 0$, а если сходимость x_n к x немонотонна, то последовательность $\mathsf{P}(x_n < x)$ вообще не сходится.

4.10. Слабая сходимость распределений имеет место тогда и только тогда, когда соответствующая последовательность кумулятивных функций распределения F_{X_n} сходится к монотонной непрерывной справа функции F_X во всех точках непрерывности последней и $F_X(+\infty) = 1$.

◀ $F_{X_n}(x) = \mathsf{P}(X_n \leqslant x) \xrightarrow[n\to\infty]{} \mathsf{P}(X \leqslant x)$ (восстановите детали!). ▶

4.11. Теорема Хелли. Из любой последовательности кумулятивных функций распределения скалярных случайных величин $F_{X_n} \equiv F_n$ можно выбрать подпоследовательность, сходящуюся к некоторой монотонной непрерывной справа функции F в каждой точке непрерывности последней.

◀ Доказательство проводится диагональным методом Вейерштрасса.

Выберем последовательность x_k, предельные точки которой покрывают всю вещественную ось (например, занумерованные тем или иным способом рациональные числа), и рассмотрим числовую последовательность $F_n(x_1)$. Поскольку $0 \leqslant F_n(x_1) \leqslant 1$ для всех n, существует подпоследовательность n_l такая, что $F_{n_l}(x_1)$ сходится к некоторому пределу при $n_l \to \infty$. Переобозначим последовательность F_{n_l} как $F_n^{(1)}$ с новым индексом n и рассмотрим $F_n^{(1)}(x_2)$. Аналогично можно выбрать подпоследовательность $F_{n_l}^{(2)}$ такую, что $F_{n_l}^{(2)}(x_2)$ сходится к некоторому пределу, и т. д. Продолжая эту процедуру для всех $k \geqslant 1$, получим последовательность вложенных подпоследовательностей кумулятивных функций $F_n^{(k)}$ такую, что $F_n^{(k)}(x_k)$ сходится при $n \to \infty$ для любого k.

Рассмотрим теперь «диагональную» последовательность $F_k^{(k)}$, составленную из функций $F_n^{(k)}$ с $n = k$ и занумерованную индексом k, и заметим, что она сходится в каждой из точек x_k к некоторому пределу, который обозначим $\tilde{F}(x_k)$ (проверьте!). Более того, из монотонности функций F_n следует, что $\tilde{F}(x'') \geqslant \tilde{F}(x')$, если $x'' > x'$. Это позволяет по монотонности продолжить определение функции $\tilde{F}$ на такие предельные точки x множества $\{x_k\}$, в которых $\tilde{F}(x+0) = \tilde{F}(x-0)$. Доопределяя $\tilde{F}$ в точках разрыва, например, так, чтобы имела место непрерывность справа, получим монотонную функцию, для которой $F_k^{(k)}(x) \xrightarrow[n\to\infty]{} \tilde{F}(x)$ во всех точках непрерывности функции $\tilde{F}$. ▶

4.12. Контрпример. Если $X_n = n$, то $F_{X_n}(x) \to 0$ при любом x и $F_X(+\infty) = 0$.

В такой ситуации можно сказать, что в пределе вероятность «уходит на бесконечность». Чтобы исключить эту опасность, Ю. В. Прохоров предложил накладывать на последовательность распределений дополнительное условие, которое по интуитивно понятным причинам по-русски называется **плотностью**, а по-английски — «герметичностью» (tightness). Оказывается, что условие плотности необходимо и достаточно для компактности семейства распределений относительно слабой сходимости.

4.13. Условие Прохорова. Дополнительно потребуем от последовательности распределений случайных величин X_n в теореме Хелли, чтобы для любого $\varepsilon > 0$ можно было выбрать $R_\varepsilon > 0$ так, что $\mathsf{P}(|X_n| > R_\varepsilon) < \varepsilon$ равномерно по n. Тогда предельная функция F задает распределение вероятности, т. е. $F(+\infty) = 1$. Обратно, если последовательность распределений слабо сходится, то она удовлетворяет данному условию.

Теперь выясним, как ведут себя в ситуации слабой сходимости распределений их характеристические функции. Оказывается, что они сходятся поточечно, и это условие является не только необходимым, но и почти достаточным для слабой сходимости: надо лишь добавить некоторое естественное условие в нуле, соответствующее «плотности» последовательности распределений на бесконечности. Остаток лекции посвящен доказательству этого факта.

4.14. Пусть последовательность распределений случайных величин X_n, обладающих характеристическими функциями φ_n, слабо сходится к распределению случайной величины X, обладающей характеристической функцией φ. Тогда φ_n сходится к φ поточечно.

◀ Сходимость $\varphi_n(s)$ к $\varphi(s)$ при любом s — это общий факт, не связанный с конкретным определением характеристической функции как преобразования Фурье–Стилтьеса.

Действительно, пусть f — произвольная равномерно ограниченная непрерывная функция и $\varepsilon > 0$. Запишем

$$\int f(x)\,\mathrm{d}F_n(x) = \int_{|x|>R_\varepsilon} f(x)\,\mathrm{d}F_n(x) + \int_{|x|\leqslant R_\varepsilon} f(x)\,\mathrm{d}F_n(x)$$

и выберем R_ε так, чтобы первое слагаемое при любом n не превосходило по модулю $\varepsilon/2$ (это возможно благодаря условию Прохорова). Теперь приближенно заменим f на кусочно-постоянную функцию f_ε, которая принимает конечное число значений на отрезке $|x| \leqslant R_\varepsilon$ и обращается в нуль вне него, причем $|f(x) - f_\varepsilon(x)| \leqslant \varepsilon/2$ равномерно при $|x| \leqslant R_\varepsilon$. Заметим, что

$$(\star) \qquad \int f_\varepsilon(x)\,\mathrm{d}F_n(x) \xrightarrow[n\to\infty]{} \int f_\varepsilon(x)\,\mathrm{d}F(x),$$

поскольку каждый из интегралов в левой части может быть записан как конечная сумма вида

$$\sum_i f_\varepsilon^{(i)} \int_{a_i}^{b_i} \mathrm{d}F_n(x) = \sum_i f_\varepsilon^{(i)}\,\mathsf{P}(a_i \leqslant X_n \leqslant b_i),$$

где (a_i, b_i) — отрезок, на котором функция f_ε принимает постоянное значение $f_\varepsilon^{(i)}$. Но по предположению о слабой сходимости каждое из таких слагаемых сходится к пределу, где в правой части вместо F_n стоит F. Заметим наконец, что благодаря нормировке распределений вероятности на единицу каждый из интегралов слева и справа в $(\star)$ отличается не более чем на $\varepsilon/2$ от соответствующего интеграла для исходной функции f. Поэтому имеем окончательно

$$\int f(x)\,\mathrm{d}F_n(x) \xrightarrow[n\to\infty]{} \int f(x)\,\mathrm{d}F(x),$$

что при $f(x) = \mathrm{e}^{\mathrm{i}sx}$ означает поточечную сходимость характеристических функций. ►

4.15. Утверждение предыдущего пункта можно усилить: на самом деле при тех же предположениях сходимость φ_n к φ равномерно на любом конечном отрезке.

◄ Для доказательства заметим, что при $|s| < S$ семейство функций $\mathrm{e}^{\mathrm{i}sx}$ равномерно ограничено и равностепенно непрерывно по x. В силу равномерной ограниченности в доказательстве из предыдущего пункта значение R_ε можно выбрать общим для всех этих функций, а в силу равностепенной непрерывности кусочно-постоянные аппроксимации для $\mathrm{e}^{\mathrm{i}sx}$ можно построить так, чтобы они были постоянны на одной и той же конечной системе отрезков, образующих разбиение $|x| \leqslant R_\varepsilon$. Отсюда следует, что сходимость в $(\star)$, а следовательно и сходимость φ_n, равномерна по s при $|s| \leqslant S$ (проверьте!). ►

Идея заменить сходимость более сложного объекта — распределений вероятности — на сходимость числовых последовательностей, возникающих при интегрировании этих распределений с всевозможными пробными функциями f, часто встречается в функциональном анализе. Такие пределы называются **слабыми**, откуда происходит и вероятностный термин «слабая сходимость».

4.16. Заметим, что из равномерной сходимости φ_n на конечных отрезках следует, что предельная функция должна быть непрерывной во всех точках, и в частности при $s = 0$.

Покажем теперь, что и наоборот, слабую сходимость распределений можно вывести из сходимости характеристических функций.

4.17. Вспомогательное тождество. Домножим тождество $\mathrm{e}^{-\mathrm{i}sy}\,\varphi(s) = \int \mathrm{e}^{\mathrm{i}s(x-y)}\mathrm{d}F(x)$ на $\mathrm{e}^{-\sigma^2 s^2/2}/2\pi$ и проинтегрируем по s. Преобразовывая правую часть

$$\frac{1}{2\pi}\iint \exp\Bigl(-\frac{\sigma^2 s^2}{2} + \mathrm{i}s(x-y)\Bigr)\,\mathrm{d}F(x)\,\mathrm{d}s = \\ = \frac{1}{2\pi}\int \mathrm{d}F(x)\int \mathrm{d}s\,\exp\Bigl[-\frac{\sigma^2}{2}\Bigl(s-\mathrm{i}\frac{x-y}{\sigma^2}\Bigr)^2\Bigr]\,\mathrm{e}^{-(x-y)^2/2\sigma^2} = \\ = \frac{1}{\sqrt{2\pi\sigma^2}}\int \mathrm{e}^{-(x-y)^2/2\sigma^2}\,\mathrm{d}F(x),$$

получим тождество

$$\frac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}sy}\,\varphi(s)\,\mathrm{e}^{-\frac{1}{2}\sigma^2 s^2}\,\mathrm{d}s = \int \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-(x-y)^2/2\sigma^2}\,\mathrm{d}F(x).$$

Если $\mathrm{d}F(x) = p(x)\,\mathrm{d}x$, то в пределе $\sigma \downarrow 0$ это тождество переходит в формулу обратного преобразования Фурье.

4.18. Распределение вероятности однозначно восстанавливается по своей характеристической функции.

◄ Предположим, напротив, что одна и та же характеристическая функция φ возникает как преобразование Фурье–Стилтьеса двух разных распределений, задаваемых кумулятивными функциями F и G. Выберем такой интервал (a, b), на котором разность $F(x) - G(x)$ возрастает (почему это возможно?). Тогда при $a < y < b$ и достаточно малом σ правая часть почленной разности соответствующих тождеств вида, установленного в п. 4.17, будет положительной (почему?), а левая должна быть равна нулю. ►

Теперь перейдем к доказательству основного утверждения последней части лекции.

4.19. Пусть последовательность случайных величин X_n обладает характеристическими функциями φ_n, которые поточечно сходятся к функции φ, непрерывной при $s = 0$. Тогда φ является характеристической функцией распределения вероятности некоторой случайной величины X, к которому слабо сходятся распределения X_n.

◄ Пусть $\varphi_n(s) \to \varphi(s)$ при всех s. Выберем из последовательности F_n подпоследовательность, сходящуюся к некоторой монотонной предельной функции F во всех точках непрерывности последней. Это возможно по теореме Хелли.

Записывая тождество п. 4.17 для φ_n и F_n, замечаем, что обе части сходятся к пределам, соответствующим предельным функциями φ и F. Действительно, к правой части непосредственно применимо доказательство сходимости из п. 4.14. Чтобы доказать сходимость левой части, разобьем ее на слагаемые, содержащие $\varphi_{\mathrm{Re},+}(s) = \max(\operatorname{Re}\varphi(s), 0)$, $\varphi_{\mathrm{Re},-}(s) = \max(-\operatorname{Re}\varphi(s), 0)$, а также аналогичные положительные и отрицательные срезки мнимой части $\varphi(s)$. Теперь можно применить то же доказательство сходимости почленно к каждому из этих слагаемых, которые все имеют вид $\int \mathrm{e}^{-\mathrm{i}sy}\,\mathrm{d}G_n(s)$ с положительной плотностью меры $\mathrm{d}G_n(s) = \frac{1}{2\pi}\varphi_n(s)\mathrm{e}^{-\frac{1}{2}\sigma^2 s^2}\,\mathrm{d}s$, слабо сходящейся к пределу $\mathrm{d}G(s) = \frac{1}{2\pi}\varphi(s)\mathrm{e}^{-\frac{1}{2}\sigma^2 s^2}\,\mathrm{d}s$.

Рассуждение п. 4.18 показывает, что не только распределение вероятности, но и любое положительное распределение, не обязательно нормированное на единицу, восстанавливается по функции φ однозначно. Поэтому любая сходящаяся подпоследовательность F_n имеет один и тот же предел F, а значит имеет место и сходимость всей последовательности $F_n \to F$. Это доказывает, что распределения случайных величин X_n слабо сходятся к некоторому предельному распределению, которое, однако, может и не быть распределением вероятности: необходимо еще гарантировать отсутствие оттока вероятности «на бесконечность».

Чтобы доказать последнее, домножим обе стороны тождества п. 4.17 на $\sqrt{2\pi\sigma^2}$ и устремим теперь σ к бесконечности, а не к нулю. Поскольку функция φ непрерывна в нуле, левая часть тождества п. 4.17 будет стремиться к $\varphi(0) = 1$. Правая же часть при любом σ не превосходит $\int \mathrm{d}F(x)$, поскольку $\mathrm{e}^{-(x-y)^2/2\sigma^2} \leqslant 1$. Отсюда следует $\int \mathrm{d}F(x) \geqslant \varphi(0) = 1$, т. е. масса распределения F не может быть меньше единицы (но, очевидно, не может быть и больше). ►

4.20. Еще раз о законе больших чисел. Пусть $X_1, \ldots, X_n$ — независимые и одинаково распределенные случайные величины, причем их характеристическая функция φ дифференцируема в нуле. Тогда слабым пределом среднего выборки $\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ является $-\mathrm{i}\varphi'(0)$.

◄ Поскольку вещественная часть характеристической функии четна (п. 3.7), ее производная в нуле исчезает. Поэтому

$$\varphi(s) = 1 + \varphi'(0)s + o(s) = 1 + \mathrm{i}\mu s + o(s).$$

Следовательно,

$$\varphi_{\overline{X}_n}(s) = \Big[\varphi_X\Big(\frac{s}{n}\Big)\Big]^n = \Big[1 + \frac{\mathrm{i}\mu s}{n} + o\Big(\frac{s}{n}\Big)\Big]^n \xrightarrow[n\to\infty]{} \mathrm{e}^{\mathrm{i}\mu s},$$

а это — характеристическая функция распределения вероятности, полностью сосредоточенного в точке $x = \mu$. Остается применить результат предыдущего пункта. ►

Если существует $\mathsf{E}X = \mu$, то $\varphi'(0) = \mathrm{i}\mu$. Однако производная может существовать и тогда, когда интеграл, выражающий математическое ожидание, расходится на бесконечности в обычном смысле, но сходится в смысле главного значения. В этом случае для существования производной дополнительно требуется, чтобы «хвосты» распределения убывали достаточно быстро, а именно чтобы $\lim_{R\to\infty} R\,[F_X(-R) + 1 - F_X(R)] = 0$. Точное необходимое и достаточное условие дифференцируемости характеристической функции в нуле см., напр., в § 2 гл. XVII т. 2 книги Феллера.

4.21. В многомерном случае определение слабой сходимости аналогично: последовательность распределений случайных величин $\boldsymbol{X}_n$ слабо сходится к распределению случайной величины $\boldsymbol{X}$, если $\mathsf{P}(\boldsymbol{X} \in A) \to \mathsf{P}(\boldsymbol{X} \in A)$ при $n \to \infty$ для любого такого открытого множества A, что $\mathsf{P}(\boldsymbol{X} \in \partial A) = 0$ (ср. условие п. 4.9). Аналогично переносится на многомерный случай критерий плотности Прохорова: если для любого $\varepsilon > 0$ найдется такой компакт K_ε, что $\mathsf{P}(X_n \in K_\varepsilon) < \varepsilon$, то компактность последовательности распределений случайных векторов $\boldsymbol{X}_n$ относительно слабой сходимости гарантирована. Переносится на многомерный случай и теорема о сходимости характеристических функций к непрерывному пределу.

5. Центральная предельная теорема и устойчивые распределения

5.1. Как и в предыдущей лекции, пусть $X_1, X_2, \ldots, X_n$ — набор n независимых одинаково распределенных случайных величин, обладающих конечной дисперсией $\mathsf{E}X_i = \mu$ и математическим ожиданием $\mathsf{D}X_i = \sigma^2$.

5.2. Перенормируем сумму $S_n = X_1 + \cdots + X_n$ таким образом, чтобы дисперсия стремилась не к нулю, а к конечному положительному пределу (если $\mu \neq 0$, то необходим еще и сдвиг):

$$\mathsf{E}\frac{1}{\sigma\sqrt{n}}(S_n - \mu n) = 0, \qquad \mathsf{D}\frac{1}{\sigma\sqrt{n}}(S_n - \mu n) = 1.$$

Иначе говоря, рассмотрим «под микроскопом» окрестность точки $x = \mu$ размером порядка $1/\sqrt{n}$ и тем самым уточним асимптотику в законе больших чисел.

5.3. Удобно рассматривать случайные величины $\tilde{X}_i = X_i - \mu$, так что $\mathsf{E}\tilde{X}_i = 0$, $\mathsf{D}\tilde{X}_i = \sigma^2$ и $\tilde{S}_n = \tilde{X}_1 + \dots + \tilde{X}_n = S_n - \mu n$. Для характеристической функции φ случайной величины $\tilde{X}_i$ имеет место разложение в нуле

$$\varphi(s) = 1 + \mathrm{i}\,\mathsf{E}\tilde{X}\,s + \frac{1}{2}\mathrm{i}^2\,\mathsf{D}\tilde{X}\,s^2 + o(s^2) = 1 - \frac{1}{2}\sigma^2 s^2 + o(s^2).$$

Поэтому

$$\varphi_{\frac{1}{\sigma\sqrt{n}}\tilde{S}_n}(s) = \varphi_{\tilde{S}_n}\Big(\frac{s}{\sigma\sqrt{n}}\Big) = \Big[\varphi\Big(\frac{s}{\sigma\sqrt{n}}\Big)\Big]^n = \Big[1 - \frac{s^2}{2n} + o\Big(\frac{s^2}{n}\Big)\Big]^n \xrightarrow[n\to\infty]{} \mathrm{e}^{-s^2/2}.$$

Вычисляя обратное преобразование Фурье (для этого достаточно выбрать $\mathrm{d}F(x) = \delta(x)\,\mathrm{d}x$, $\varphi(s) \equiv 1$ в тождестве п. 4.17), получаем, что элемент распределения вероятности, которое задает предельная функция, имеет вид

$$p(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x.$$

Это распределение вероятности называется **нормальным**. Одномерным **гауссовым распределением**, как и выше (п. 3.22), будем называть результат произвольного перемасштабирования и сдвига нормального распределения:

$$p(x \mid \mu, \sigma)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}\,\sigma}\mathrm{e}^{-\frac{1}{2}(x-\mu)^2}.$$

Пользуясь результатом п. 4.19, получаем следующее утверждение.

5.4. Центральная предельная теорема. Если при каждом n случайные величины $X_1, X_2, \dots, X_n$ независимы и одинаково распределены, $\mathsf{E}X_i = \mu$, $\mathsf{D}X_i = \sigma^2$, то последовательность распределений случайных величин $(S_n - \mu n)/(\sigma\sqrt{n})$ слабо сходится к нормальному распределению.

Идея доказательства центральной предельной теоремы методом характеристических функций принадлежит А. А. Ляпунову. Позднее метод характеристических функций был применен П. Леви для получения значительно более тонких результатов, касающихся безгранично делимых распределений (п. 5.12).

Сделаем несколько замечаний о полученном результате.

5.5. Пример. Пусть закон распределения X_i задается плотностью

$$p(x) = \frac{\lambda^\nu}{\Gamma(\nu)}x^{\nu-1}\,\mathrm{e}^{-\lambda x} \quad \text{при } x > 0,$$

где $\lambda > 0$, $\nu > 1$, так что $\mathsf{E}X = \nu/\lambda$, $\mathsf{D}X = \nu/\lambda^2$ (гамма-распределение, см. п. П1.11). Тогда

$$p_{S_n}(x) = \frac{\lambda^{n\nu}}{\Gamma(n\nu)}x^{n\nu-1}\,\mathrm{e}^{-\lambda x} \quad \text{при } x > 0,$$

т. е. распределение суммы сосредоточено на положительной полуоси и экспоненциально убывает при $x \to +\infty$. Однако перемасштабированное распределение $(\lambda S_n - n\nu)/\sqrt{n\nu}$ сходится к нормальному распределению, которое симметрично и на бесконечности убывает сверхэкспоненциально.

5.6. Логнормальное распределение. Пусть случайные величины $Y_1, Y_2, \dots, Y_n$ независимы и одинаково распределены на положительной полуоси и $\Pi_n = Y_1 Y_2 \dots Y_n$. Логарифмируя, получим $\log Y_i = X_n$, $\log \Pi_n = X_1 + \dots + X_n = S_n$; если $\mathsf{E}X_n = \mu$ и $\mathsf{D}X_n = \sigma^2$, то из центральной предельной теоремы следует, что распределение случайной величины

$(Y_1Y_2\dots Y_n\mathrm{e}^{-\mu n})^{1/\sigma\sqrt{n}}$ сходится к пределу, задаваемому функцией плотности вероятности логнормального распределения (см. п. П1.15)

$$p(y)=\frac{1}{\sqrt{2\pi}\,y}\mathrm{e}^{-\frac{1}{2}(\ln y)^2}.$$

5.7. Еще раз о контрпримере п. 4.7. Если распределение вероятности характеризуется медленным (степенным) убыванием «на бесконечности», то, вообще говоря, $\mathsf{D}X=\infty$ и центральная предельная теорема неприменима. Однако для распределения Коши результат, отчасти напоминающий центральную предельную теорему, выполняется при другой нормировке: нетривиальным предельным распределением характеризуется не $S_n/\sqrt{n}$, а $\overline{X}=S_n/n$. С другой стороны, из-за степенного убывания функции плотности вероятности «на бесконечности» характеристическая функция распределения Коши недифференцируема в нуле: $\varphi_C(s)=\exp(-|s|)$.

Заметим, что перемасштабирование суммы умножением на $\frac{1}{n}$ оказывается «согласовано» с этой особенностью распределения Коши: $[\varphi_C(\frac{1}{n}s)]^n=\varphi_C(s)$. Аналогичное соотношение для нормального распределения имеет вид $[\varphi(\frac{1}{\sqrt{n}}s)]^n=\varphi(s)$ и согласуется с перенормировкой сумм в центральной предельной теореме.

Чтобы понять, как в общем случае соотносятся: (а) медленное убывание на бесконечности, (б) характер особенности характеристической функции в нуле и (в) «правильная» перенормировка суммы случайных величин, рассмотрим следующий пример.

5.8. Симметричное распределение со степенным убыванием. Пусть распределение вероятности задано функцией плотности

$$p(x)=\frac{A}{2}\ \text{при}\ |x|\leqslant 1,\quad p(x)=\frac{A}{2|x|^{\alpha+1}}\ \text{при}\ |x|\geqslant 1,$$

где $\alpha>0$. Из условия нормировки следует, что $A=\frac{\alpha}{\alpha+1}$ (проверьте!). Для характеристической функции получаем

$$\varphi(s)=\int p(x)\,\mathrm{e}^{\mathrm{i}sx}\,\mathrm{d}x=\frac{A}{2}\int_{-1}^{1}\mathrm{e}^{\mathrm{i}sx}\,\mathrm{d}x+\frac{A}{2}\int_{|x|\geqslant 1}\frac{1}{|x|^{\alpha+1}}\,\mathrm{e}^{\mathrm{i}sx}\,\mathrm{d}x.$$

Учитывая симметрию и заменяя переменную на $\xi=|s|x$, получим

$$\varphi(s)=A\int_0^1\cos sx\,\mathrm{d}x+A\int_1^\infty\frac{\cos sx}{x^{\alpha+1}}\,\mathrm{d}x=\frac{A}{|s|}\int_0^{|s|}\cos\xi\,\mathrm{d}\xi+A|s|^\alpha\int_{|s|}^\infty\frac{\cos\xi}{\xi^{\alpha+1}}\,\mathrm{d}\xi.$$

Изучим асимптотику этой функции при малых $|s|$. Первое слагаемое равно $A\sin|s|/|s|$ и при $s\to 0$ стремится к A, а во втором слагаемом при $s\to 0$ возникает неопределенность вида $0\cdot\infty$. Чтобы разрешить ее, перепишем интеграл в виде

$$A|s|^\alpha\int_{|s|}^\infty\frac{\cos\xi-1+1}{\xi^{\alpha+1}}\,\mathrm{d}\xi=A|s|^\alpha\int_{|s|}^\infty\frac{\mathrm{d}\xi}{\xi^{\alpha+1}}-A|s|^\alpha\int_{|s|}^\infty\frac{1-\cos\xi}{\xi^{\alpha+1}}\,\mathrm{d}\xi.$$

При $0<\alpha<2$ и малых положительных $|s|$ последний интеграл положителен и отличается от конечной константы $I_1(\alpha)=\int_0^\infty(1-\cos\xi)\,\mathrm{d}\xi/\xi^{\alpha+1}$ на величину порядка $O(|s|^{2-\alpha})$ (проверьте!). Поэтому при $0<\alpha<2$ получаем

$$\varphi(s)=A\frac{\sin|s|}{|s|}+\frac{A}{\alpha}-AI_1(\alpha)\,|s|^\alpha+O(|s|^2)$$

или с учетом того, что $A=\frac{\alpha}{\alpha+1}$ и $\frac{\sin|s|}{|s|}=1+O(|s|^2)$,

$$\varphi(s)=1-AI_1(\alpha)\,|s|^\alpha+O(|s|^2).$$

Пусть теперь $X_1, X_2, \ldots, X_n$ — независимые одинаково распределенные случайные величины, обладающие указанной плотностью вероятности. Для характеристической функции их суммы имеем

$$\varphi_{S_n}(s) = [1 - AI_1(\alpha)\,|s|^\alpha + O(|s|^2)]^n,$$

откуда получаем, что распределение $S_n/n^{1/\alpha}$ при $n \to \infty$ сходится к предельному распределению, задаваемому характеристической функцией

$$\varphi_\alpha(s) = \mathrm{e}^{-\Gamma|s|^\alpha},$$

где $\Gamma = AI_1(\alpha)$.

Теперь разберем ситуацию, в которой распределение несимметрично. Если «хвосты» распределения на положительной и отрицательной полупрямой имеют разные показатели α, α', то асимптотически при больших n будет играть роль только «хвост», убывающий более медленно. Поэтому достаточно рассмотреть «хвосты» разной массы, но с одинаковым показателем убывания.

5.9. Асимметричное распределение со степенным убыванием. Пусть $p(x) = A/2$ при $|x| \leqslant 1$, а вероятностные массы положительной и отрицательной полуосей различны:

$$p(x) = \frac{A(1+\beta)}{2x^{\alpha+1}} \text{ при } x \geqslant 1, \quad p(x) = \frac{A(1-\beta)}{2|x|^{\alpha+1}} \text{ при } x \leqslant -1,$$

где $|\beta| \leqslant 1$. Для характеристической функции получаем

$$\varphi(s) = \frac{A}{2}\int_{-1}^{1} \mathrm{e}^{\mathrm{i}sx}\,\mathrm{d}x + \frac{A}{2}\int_1^\infty \frac{\mathrm{e}^{\mathrm{i}sx} + \mathrm{e}^{-\mathrm{i}sx}}{x^{\alpha+1}}\,\mathrm{d}x + \frac{A}{2}\int_1^\infty \frac{\beta\mathrm{e}^{\mathrm{i}sx} - \beta\mathrm{e}^{-\mathrm{i}sx}}{x^{\alpha+1}}\,\mathrm{d}x$$

или после замены $\xi = |s|x$

$$\varphi(s) = A\frac{\sin|s|}{|s|} + A|s|^\alpha \int_{|s|}^\infty \frac{\cos\xi}{\xi^{\alpha+1}}\,\mathrm{d}\xi + \mathrm{i}A\beta|s|^\alpha \operatorname{sign} s \int_{|s|}^\infty \frac{\sin\xi}{\xi^{\alpha+1}}\,\mathrm{d}\xi.$$

Асимптотика первых двух слагаемых при малых $|s|$ нам уже известна. Для третьего слагаемого надо рассмотреть три случая:

1) $0 < \alpha < 1$: интеграл $I_2(\alpha) = \int_0^\infty \sin\xi\,\mathrm{d}\xi/\xi^{\alpha+1}$ сходится. Тогда (проверьте!)

$$\varphi(s) = 1 - AI_1(\alpha)|s|^\alpha\bigl(1 - \mathrm{i}\beta \operatorname{sign} s \tfrac{I_2(\alpha)}{I_1(\alpha)}\bigr) + O(|s|).$$

Можно показать, что $I_2(\alpha)/I_1(\alpha) = \operatorname{tg} \pi\alpha/2$. При перенормировке $S_n/n^{1/\alpha}$ в пределе при $n \to \infty$ получается распределение с характеристической функцией

$$\varphi(s) = \exp\bigl[-AI_1(\alpha)|s|^\alpha\bigl(1 - \mathrm{i}\beta \operatorname{tg} \tfrac{\pi\alpha}{2} \operatorname{sign} s\bigr)\bigr].$$

2) $1 < \alpha < 2$: третье слагаемое можно переписать в виде

$$\mathrm{i}A\beta|s|^\alpha \operatorname{sign} s \int_{|s|}^\infty \frac{\sin\xi - \xi + \xi}{\xi^{\alpha+1}}\,\mathrm{d}\xi = \frac{\mathrm{i}A\beta|s| \operatorname{sign} s}{\alpha - 1} - \mathrm{i}A\beta|s|^\alpha \operatorname{sign} s \int_{|s|}^\infty \frac{\xi - \sin\xi}{\xi^{\alpha+1}}\,\mathrm{d}\xi.$$

Введем конечную положительную константу $I_3(\alpha) = \int_0^\infty (\xi - \sin\xi)\,\mathrm{d}\xi/\xi^{\alpha+1}$, которая отличается от последнего интеграла на $O(|s|^{3-\alpha})$, и учтем, что $|s| \operatorname{sign} s = s$. Поэтому

$$\varphi(s) = 1 + \frac{\mathrm{i}A\beta s}{\alpha - 1} - AI_1(\alpha)|s|^\alpha\bigl(1 + \mathrm{i}\beta \operatorname{sign} s \tfrac{I_3(\alpha)}{I_1(\alpha)}\bigr) + O(|s|^2).$$

Можно показать, что $I_3(\alpha)/I_1(\alpha) = -\operatorname{tg}\pi\alpha/2$. Чтобы теперь получить нетривиальный предел, необходимо, как при выводе обычной центральной предельной теоремы, сдвинуть слагаемые: $\tilde{X}_i = X_i - A\beta/(\alpha - 1)$. Тогда распределение суммы $\tilde{S}_n = \tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_n$ после деления на $n^{1/\alpha}$ сходится к пределу, выражение для характеристической функции которого совпадает с полученным выше:

$$\varphi(s) = \exp\bigl[-AI_1(\alpha)|s|^\alpha\bigl(1 - \mathrm{i}\beta \operatorname{tg} \tfrac{\pi\alpha}{2} \operatorname{sign} s\bigr)\bigr].$$

3) $\alpha = 1$: расходимость в третьем слагаемом логарифмическая. Записывая его в виде

$$\mathrm{i}A\beta|s|\operatorname{sign}s\int_{|s|}^{\infty}\frac{\sin\xi}{\xi^2}\,\mathrm{d}\xi = \mathrm{i}A\beta|s|\operatorname{sign}s\left(\int_{|s|}^{1}\frac{\mathrm{d}\xi}{\xi} + \int_{|s|}^{1}\frac{\sin\xi-\xi}{\xi^2}\,\mathrm{d}\xi + \int_{1}^{\infty}\frac{\sin\xi}{\xi^2}\,\mathrm{d}\xi\right)$$

и учитывая, что $|s|\operatorname{sign}s = s$, получим

$$-\mathrm{i}A\beta|s|\ln|s|\operatorname{sign}s + \mathrm{const}\,\mathrm{i}s + O(|s|^2).$$

Таким образом, полученное выше выражение для характеристической функции

$$\varphi(s) = A\frac{\sin|s|}{|s|} + A|s|^{\alpha}\int_{|s|}^{\infty}\frac{\cos\xi}{\xi^{\alpha+1}}\,\mathrm{d}\xi + \mathrm{i}A\beta|s|^{\alpha}\operatorname{sign}s\int_{|s|}^{\infty}\frac{\sin\xi}{\xi^{\alpha+1}}\,\mathrm{d}\xi$$

при $\alpha = 1$ может быть записано в виде

$$\varphi(s) = 1 + \mathrm{const}\,\mathrm{i}s - AI_1(1)|s|(1 + \mathrm{i}\beta\tfrac{1}{I_1(1)}\ln|s|\operatorname{sign}s) + O(|s|^2)$$

Можно показать, что $I_1(1) = \pi/2$. Поэтому в пределе при $n \to \infty$, применяя подходящий сдвиг случайных слагаемых X_i и перемасштабирование их суммы делением на n, получим

$$\varphi(s) = \exp[-AI_1(1)|s|(1 + \mathrm{i}\tfrac{2\beta}{\pi}\ln|s|\operatorname{sign}s)].$$

Естественно, что для получения правильного предела при $1 \leqslant \alpha < 2$ потребовались сдвиги слагаемых, поскольку распределение несимметрично. Несколько более неожиданно то, что при $0 < \alpha < 1$, когда математическое ожидание расходится, необходимости в сдвигах не возникает.

5.10. Распределение вероятности, задаваемое характеристической функцией

$$\varphi(s) = \exp\bigl[\mathrm{i}\mu s - \Gamma|s|^{\alpha}(1 - \mathrm{i}\beta\operatorname{tg}\tfrac{\pi\alpha}{2}\operatorname{sign}s)\bigr], \quad 0 < \alpha \leqslant 2,\ \alpha \neq 1,$$
$$\varphi(s) = \exp\bigl[\mathrm{i}\mu s - \Gamma|s|(1 - \mathrm{i}\tfrac{2\beta}{\pi}\ln\tfrac{1}{|s|}\operatorname{sign}s)\bigr], \quad \alpha = 1,$$

где μ — произвольный вещественный параметр, $\Gamma > 0$, $|\beta| \leqslant 1$, называется **распределением Леви–Парето** с показателем α (**показатель Леви**).

Явные выражения для соответствующей функции плотности вероятности p известны в трех случаях: $\alpha = 2$ (гауссово распределение), $\alpha = 1$ и $\beta = 0$ (распределение Коши) и $\alpha = \frac{1}{2}$ и $\beta = \pm 1$ («распределение времени выхода», см. п. П1.20).

Интересно отметить, что О. Коши изучал преобразование Фурье функции $\exp(-|s|^{\alpha})$ и обнаружил для него явное выражение при $\alpha = 1$. Хотя эта работа не была вероятностной, именно благодаря ей позднее возник и закрепился термин «распределение Коши».

5.11. Распределение Леви–Парето обладает свойством **устойчивости**: если $X_1, \dots, X_n$ — совокупность независимых случайных величин, распределенных по закону п. 5.10 с одинаковыми параметрами, то можно найти такие положительные константы a_n, b_n, что случайная величина $(X_1 + \dots + X_n - a_n)/b_n$ распределена по Леви–Парето с теми же значениями параметров. Если аналогичное утверждение выполнено с $a_n \equiv 0$, говорят, что распределение случайных величин X_i обладает свойством **строгой устойчивости**.

5.12. Распределение вероятности с положительной дисперсией называется **безгранично делимым**, если при любом $n > 1$ случайная величина X, распределенная по этому закону, может быть представлена суммой n независимых одинаково распределенных величин: $X = Y_1^{(n)} + Y_2^{(n)} + \dots + Y_n^{(n)}$. При этом распределения $Y_i^{(n)}$ могут и не совмещаться с распределением X никаким сдвигом или перемасштабированием. Примерами безгранично делимых распределений являются устойчивые, а также гамма-распределение и распределение Пуассона (проверьте!). Дискретные распределения, как правило, не являются безгранично делимыми.

Важное замечание: в выкладках пп. 5.8, 5.9 конкретный вид плотности при значениях $|x|$, ограниченных любой наперед заданной константой, не играет большой роли. Можно показать, что поведение характеристической функции в нуле определяется только асимптотикой распределения вероятности при больших $|x|$ (см., например, гл. XVII т. 2 Феллера). Точнее, имеет место следующее утверждение.

5.13. Обобщенная центральная предельная теорема. Пусть $X_1, X_2, \ldots, X_n$ — совокупность независимых одинаково распределенных случайных величин, распределения вероятности которых обладают степенным убыванием на бесконечности:

$$x^\alpha[1 - F(x)] \to C_+, \quad x^\alpha F(-x) \to C_-$$

при $x \to \infty$, где хотя бы она из констант C_+, C_- положительна и $0 < \alpha < 2$. Тогда существует такая константа a (которую можно выбрать равной нулю при $0 < \alpha < 1$), что распределение случайной величины $(X_1 + \cdots + X_n - an)/n^{1/\alpha}$ при $n \to \infty$ сходится к распределению Леви–Парето с $\mu = 0$.

Строго говоря, это утверждение не обобщает центральную предельную теорему п. 5.4, поскольку случай $\alpha = 2$ им не покрывается. Действительно, при $\alpha = 2$ регуляризация слагаемого $\int_{|s|}^\infty \cos\xi \, d\xi/\xi^3$ приводит к логарифмической расходимости. Можно показать (§ 5 гл. XVII т. 2 книги Феллера), что сходимость к $\varphi(s) = \exp(-\frac{1}{2}s^2)$ все же имеет место, если нормирующий множитель a_n определяется из соотношения $a_n^2/(n \log a_n) \to \mathrm{const} > 0$. Заметим, что «естественный» нормирующий множитель $a_n = \sqrt{n}$ растет недостаточно быстро.

Говорят, что при $0 < \alpha < 2$ распределения вероятности распадаются на «классы универсальности» по отношению к суммированию случайных величин, представителями которых являются соответствующие распределения Леви–Парето. Еще один класс универсальности охватывает распределения с $2 < \alpha \leqslant \infty$ (в том числе и распределения, убывающие быстрее любой степени x — например, гауссово или гамма-распределение). У таких распределений дисперсия конечна и ситуация описывается центральной предельной теоремой, а нормирующий множитель есть $1/\sqrt{n}$.

5.14. Вернемся еще раз к примеру п. 4.7 с распределением Коши. Если независимые случайные величины X, $X_1, \ldots, X_n$ распределены не по Коши, а по распределению Леви–Парето с $0 < \alpha < 1$, то из только что развитой теории следует, что сумма $X_1 + \cdots + X_n$ распределена так же, как $n^{1/\alpha}X$, и поэтому выборочное среднее $\overline{X}$ распределено как $n^{(1-\alpha)/\alpha}X$, т. е. при больших n *шире*, чем любое отдельное слагаемое! Причиной этого парадокса служит явление, анализу которого посвящена следующая лекция.

6. Статистика экстремальных значений

6.1. Наблюдение. Пусть случайные величины X, Y независимы, а их распределения вероятности задаются кумулятивными функциями F, G. Тогда кумулятивная функция распределения случайной величины $\max\{X, Y\}$ есть произведение FG:

$$\mathsf{P}(\max\{X, Y\} \leqslant x) = \mathsf{P}(X \leqslant x, Y \leqslant x) = \mathsf{P}(X \leqslant x)\,\mathsf{P}(Y \leqslant x) = F(x)\,G(x).$$

По индукции это наблюдение нетрудно распространить на совокупность любого числа случайных величин.

6.2. И вновь о распределении Коши. Кумулятивная функция распределения Коши имеет вид $F(x) = \frac{1}{\pi}\operatorname{arctg} x + \frac{1}{2}$, так что «хвост» при $x \to \infty$ имеет вид

$$\tilde{F}(x) := 1 - F(x) = \frac{1}{\pi x} + o\Big(\frac{1}{x}\Big).$$

Кумулятивная функция распределения суммы $S_n = X_1 + X_2 + \cdots + X_n$ независимых слагаемых, распределенных по Коши, имеет вид

$$F_{S_n}(x) = F_{nX}(x) = F(\tfrac{x}{n}) = \frac{1}{\pi}\operatorname{arctg}\frac{x}{n} + \frac{1}{2},$$

так что $\tilde{F}_{S_n}(x) = n/\pi x + o(1/x)$ при $x \to \infty$. Наконец, кумулятивная функция распределения максимального значения выборки $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$ имеет вид

$$F_{(n)}(x) = [F_X(x)]^n = \left(\frac{1}{\pi}\operatorname{arctg} x + \frac{1}{2}\right)^n = \left[1 - \frac{1}{\pi x} + o\left(\frac{1}{x}\right)\right]^n = 1 - \frac{n}{\pi x} + o\left(\frac{1}{x}\right),$$

так что «хвост» $\tilde{F}_{(n)}$ этого распределения имеет такую же асимптотику, как $\tilde{F}_{S_n}$.

Этот пример наводит на мысль, что для величин, распределенных по Коши, распределение самых больших значений суммы $S_n = X_1 + X_2 + \cdots + X_n$ в основном определяется одним или несколькими максимальными слагаемыми. Это противоположно поведению сумм случайных величин с конечной дисперсией, в которых все слагаемые вносят сравнимый вклад. Такая структура сумм случайных величин, распределенных с «тяжелыми хвостами», была открыта Б. Мандельбротом в 1960-х годах.

Тем самым для распределения Коши, как и для распределений с показателем Леви $\alpha < 1$, перестает работать механизм подавления случайности при помощи усреднения, на котором основан закон больших чисел, и полученные парадоксальные результаты находят объяснение.

Изучим закон распределения максимума выборки более детально.

6.3. Пусть $F_{(n)}(x) = [F(x)]^n$ — кумулятивная функция распределения максимума $X_{(n)}$ набора независимых одинаково распределенных случайных величин $X_1, X_2, \ldots X_n$. При $n \to \infty$ либо $F_{(n)}(x) = [F(x)]^n \to 0$, если $F(x) < 1$ при всех x, либо распределение $X_{(n)}$ слабо сходится к неслучайному распределению, сосредоточенному в точке $x_0 = \min\{x\colon F(x) = 1\}$.

6.4. Пусть $x_{(n)}$ — решение уравнения

$$F(x) = 1 - \frac{1}{n}.$$

Будем называть $x_{(n)}$ **типичным наибольшим значением** n-элементной выборки из распределения, заданного кумулятивной фукнцией F (ср. определение медианы в п. 2.16). Примем без потери общности, что $x_{(n)} > 0$.

Поскольку $\mathsf{P}(X > x_{(n)}) = 1/n$, в выборке объема n в среднем будет встречаться одно значение, превышающее $x_{(n)}$. Если функция F разрывна и не принимает значения $1 - 1/n$, то можно положить $x_{(n)} = \inf\{x\colon F(x) \geqslant 1 - 1/n\}$, что приведет к незначительным изменениям в дальнейших выкладках.

6.5. Будем искать предел перемасштабированной кумулятивной функции распределения максимума выборки в следующем виде:

$$F_{(n)}(x_{(n)}\xi) = [F(x_{(n)}\xi)]^n = [1 - \tilde{F}(x_{(n)}\xi)]^n \xrightarrow[n\to\infty]{} \mathrm{e}^{-\varphi(\xi)},$$

причем $\varphi(\infty) = 0$, поскольку $F_{(n)}(\infty) = 1$. Логарифмируя и замечая, что $n\tilde{F}(x_{(n)}\xi)$ ограничена, получим

$$-n\log[1 - \tilde{F}(x_{(n)}\xi)] = n\tilde{F}(x_{(n)}\xi) + o(1) \xrightarrow[n\to\infty]{} \varphi(\xi).$$

Поскольку $n\tilde{F}(x_{(n)}) = 1$, имеем

$$\varphi(\xi) = \lim_{n\to\infty} \frac{n\tilde{F}(x_{(n)}\xi)}{n\tilde{F}(x_{(n)})} = \lim_{x_{(n)}\to\infty} \frac{\tilde{F}(x_{(n)}\xi)}{\tilde{F}(x_{(n)})}.$$

6.6. Заметим, что благодаря монотонности функции $\tilde{F}$ при $x_{(n)} \leqslant t \leqslant x_{(n+1)}$ выполнены неравенства

$$\frac{\tilde{F}(x_{(n+1)}\xi)}{\tilde{F}(x_{(n)})} \leqslant \frac{\tilde{F}(t\xi)}{\tilde{F}(t)} \leqslant \frac{\tilde{F}(x_{(n)}\xi)}{\tilde{F}(x_{(n+1)})}.$$

Поскольку $n\tilde{F}(x_{(n)}) = (n+1)\tilde{F}(x_{(n+1)}) = 1$, имеем

$$\frac{n}{n+1}\frac{\tilde{F}(x_{(n+1)}\xi)}{\tilde{F}(x_{(n+1)})} \leqslant \frac{\tilde{F}(t\xi)}{\tilde{F}(t)} \leqslant \frac{n+1}{n}\frac{\tilde{F}(x_{(n)}\xi)}{\tilde{F}(x_{(n)})}.$$

что позволяет переходить к пределу в отношении $\tilde{F}(t\xi)/\tilde{F}(t)$ по параметру $t \to \infty$, принимающему не только целые, а произвольные вещественные значения.

6.7. Поскольку при $\xi_1, \xi_2 > 0$

$$\varphi(\xi_1\xi_2) \xleftarrow[t\to\infty]{} \frac{\tilde{F}(t\xi_1\xi_2)}{\tilde{F}(t)} = \frac{\tilde{F}(t\xi_1\xi_2)}{\tilde{F}(t\xi_2)}\frac{\tilde{F}(t\xi_2)}{\tilde{F}(t)} \xrightarrow[t\to\infty]{} \varphi(\xi_1)\,\varphi(\xi_2),$$

имеем $\varphi(\xi_1\xi_2) = \varphi(\xi_1)\,\varphi(\xi_2)$. Общим решением этого уравнения, удовлетворяющим условиям $\varphi(1) = 1$ и $\varphi(\infty) = 0$, является $\varphi(\xi) = \xi^{-\alpha}$ при $0 < \alpha \leqslant \infty$ (считаем $x^{-\infty} \equiv 0$ при $x > 1$).

Монотонная функция $F(x)$, удовлетворяющая условию $\lim_{t\to+\infty} F(tx)/F(t) = x^{-\alpha}$ при конечном α, по Й. Карамате называется **правильно меняющейся** с показателем α. Понятие правильного изменения позволяет грубо выделить основную «степенную» часть в функциях типа $x^{-\alpha}\log\log x/\log x$; на этом же языке может быть сформулирован и наиболее широкий вариант обобщенной центральной предельной теоремы для устойчивых распределений (см. предыдущую лекцию). Подробнее о теории правильно меняющихся функций см. т. 2 учебника Феллера, §§8–9 гл. VIII.

В пп. 6.5– 6.7 установлено следующее утверждение.

6.8. Пусть $X_1, X_2, \ldots, X_n$ — совокупность независимых одинаково распределенных случайных величин, характеризуемых последовательностью типичных наибольших значений $x_{(n)} > 0$ (п. 6.4). Тогда распределение случайной величины $\max(X_1, X_2, \ldots X_n)/x_{(n)}$ слабо сходится к распределению, сосредоточенному на положительной полуоси $x \geqslant 0$ и заданному там кумулятивной функцией

$$F_\alpha(x) = \mathrm{e}^{-x^{-\alpha}}, \quad 0 < \alpha \leqslant +\infty.$$

При конечных α оно называется **распределением Фреше**.

Заметим, что распределение Фреше характеризуется степенным «хвостом»: $\tilde{F}_\alpha(x) = x^{-\alpha}$. При этом степенное убывание на бесконечности присуще и исходному распределению случайных величин X_i, а $x_{(n)} \sim n^{1/\alpha}$. Если же «хвосты» исходного распределения убывают быстрее, то $\alpha = +\infty$ и нетривиальное предельное распределение для максимума возникает при ином преобразовании $F_{(n)}$.

6.9. Рассмотрим предел

$$F_{(n)}(x_{(n)} + \eta) = [F(x_{(n)} + \eta)]^n = [1 - \tilde{F}(x_{(n)} + \eta)]^n \xrightarrow[n\to+\infty]{} \mathrm{e}^{-\psi(\eta)}.$$

Действуя аналогично пп. 6.5– 6.7, получим $\psi(\eta) = \lim_{t\to+\infty} \tilde{F}(t+\eta)/\tilde{F}(t)$, причем $\psi(0) = 1$. Отсюда следует функциональное уравнение $\psi(\eta_1 + \eta_2) = \psi(\eta_1)\psi(\eta_2)$ с общим решением $\psi(\eta) = \exp(-\eta/\eta_0)$.

6.10. При тех же условиях, как в п. 6.8, распределение $\max(X_1, X_2, \ldots, X_n) - x_{(n)}$ с точностью до фиксированного масштабного преобразования слабо сходится к распределению, определенному на всей вещественной оси кумулятивной функцией $F(x) = \mathrm{e}^{-\exp(-x)}$, которое называется **распределением Гумбеля**.

Логарифм случайной величины, распределенной по Фреше, распределен по Гумбелю.

6.11. Осталось рассмотреть еще один частный случай, когда исходное распределение $F(x) = 1$ при $x \geqslant x_0$, т. е. исходная случайная величина ограничена сверху. Если x_0 — наименьшая такая верхняя грань, то преобразование $Y = (x_0 - X)^{-1}$ переводит данную задачу в задачу о распределении Фреше [соответствующее перемасштабирование имеет вид

$F_{(n)}(\zeta/y_{(n)})]$. Предельное распределение в исходной задаче сосредоточено на полуоси $x \leqslant x_0$, задается там кумулятивной функцией

$$F_\beta(x) = \mathrm{e}^{-(x_0 - x)^\beta},$$

где $0 < \beta \leqslant +\infty$, и называется **распределением Вейбулла**.

Часто распределение Вейбулла пишут для минимума совокупности случайных величины, ограниченных снизу, так что кумулятивная функция принимает вид $1 - F_\beta(-x)$. При $\beta = 1$ это распределение совпадает с показательным.

Можно показать, что результаты пп. 6.8, 6.10 и 6.11 вместе с тривиальным распределением п. 6.3 исчерпывают возможные предельные формы распределения максимума совокупности независимых одинаково распределенных случайных величин (**предельная теорема Фишера–Типпета–Гнеденко**).

6.12. Пусть $X_1, X_2, \ldots, X_n$ — выборка, т. е. набор случайных величин, независимых и одинаково распределенных по некоторому закону. Отсортируем по возрастанию набор значений, которые приняли эти случайные величины:

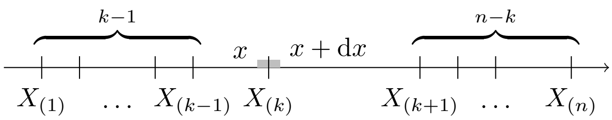
$$\min\{X_1, X_2, \ldots, X_n\} = X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}.$$

Отсортированный набор называется **вариационным рядом** выборки $X_1, X_2, \ldots, X_n$, а его члены — **порядковыми статистиками**. При фиксированном n обозначение $X_{(k)}$ для k-й порядковой статистики (а также $F_{(k)}$ для ее кумулятивной функции распределения) не вызывает путаницы.

6.13. Элемент вероятности распределения k-й порядковой статистики имеет вид

$$\mathrm{d}F_{(k)}(x) = n \cdot C_{n-1}^{k-1} \cdot [F(x)]^{k-1} \cdot [1 - F(x)]^{n-k} \cdot \mathrm{d}F(x).$$

Смысл каждого из сомножителей в правой части ясен из следующей схемы.



Одно из значений $X_1, X_2, \ldots, X_k$ можно n способами выбрать на роль k-й порядковой статистики, лежащей в интервале $(x, x+\mathrm{d}x)$. После этого остается C_{n-1}^{k-1} способов выбрать $k-1$ значение, попадающее левее $X_{(k)} = x$. Вероятность того, что $k-1$ независимых случайных величин все примут значения, не превосходящие x, равна $[F(x)]^{k-1}$, а вероятность того, что остальные $n-k$ независимых случайных величин все примут значения, не меньшие x (точнее, $x+\mathrm{d}x$), равна $[1-F(x)]^{n-k}$. Наконец, элемент вероятности распределения случайной величины, выбранной на роль k-й порядковой статистики и принимающей значения в отрезке $(x, x+\mathrm{d}x)$, есть $\mathrm{d}F(x)$.

6.14. Если распределение случайных величин X_i не имеет атомов, то замена переменных $Y = F(X)$ переводит их в равномерно распределенные случайные величины на отрезке $(0,1)$ а элемент вероятности $Y_{(k)}$ принимает вид

$$\mathrm{d}F_{(k)}(y) = nC_{n-1}^{k-1} y^{k-1}(1-y)^{n-k}\,\mathrm{d}y$$

(бета-распределение с параметрами k, $n-k+1$, см. п. П1.12).

6.15. В частности, элемент вероятности распределения медианы выборки $Y_{(k+1)}$ при $n = 2k+1$ имеет вид

$$\mathrm{d}F_{\mathrm{med}}(y) = (2k+1)C_{2k}^{k} y^k (1-y)^k\,\mathrm{d}y = (2k+1)C_{2k}^{k} \frac{(1-\xi^2/2k)^k}{4^k} \frac{1}{2\sqrt{2k}}\,\mathrm{d}\xi,$$

где $y = (1+\xi/\sqrt{2k})/2$. Нетрудно проверить, что при $n, k \to \infty$ это распределение сходится к нормальному. Поэтому, если $\mathrm{d}F/\mathrm{d}x|_{x_{\mathrm{med}}} = \lambda > 0$, то распределение медианы исходной выборки $X_1, X_2, \ldots, X_n$ также является асимптотически нормальным с дисперсией $(2\lambda\sqrt{n})^{-1}$ и сосредотачивается на истинном значении медианы.

Следовательно, медиана выборки является хорошей оценкой истинной медианы для любого распределения. Ср. это с поведением среднего выборки $\overline{X}_n$, которое может не стремиться ни к какому пределу при $n \to \infty$, если исходное распределение слишком медленно убывает на бесконечности.

6.16. Полезное семейство предельных распределений получается, если от переменных $Y_{(k)}$, распределенных на отрезке $(0, 1)$, перейти к переменным $Z_{(k)} = n(1 - Y_{(k)})$, распределенных на положительной полуоси. Переходя к пределу при $n \to \infty$ в формуле п. 6.14 с учетом замены $y = 1 - z/n$, нетрудно проверить, что предельное распределение $Z_{(k)}$ имеет вид гамма-распределения с параметрами $\nu = k$, $\alpha = 1$. В частности, $Z_{(1)}$ распределена по показательному закону (ср. п. 6.11 при $\beta = 1$).

7. Большие уклонения I: энтропия и теорема Санова

В лекции 5 мы видели, что центральная предельная теорема и ее обобщения уточняют закон больших чисел для сумм независимых числовых случайных величин: они описывают асимптотическое поведение среднего выборки или других нормировок суммы в малой окрестности предельного значения, около которого происходит концентрация вероятности. В этой лекции нас будет интересовать, напротив, относительное распределение вероятности по тем реализациям, которые находятся «вдали» от области концентрации вероятности. Говорят, что такие реализации характеризуются **большими уклонениями** от наиболее вероятных.

7.1. Пример. Набор случайных величин $X_1, X_2, \ldots, X_n$, каждая из которых принимает значение 1 с вероятностью p и -1 с вероятностью $1 - p$, можно интерпретировать как **случайное блуждание** по целым точкам числовой прямой, начинающееся из начала координат: смещение за k шагов по определению равно $X_1 + X_2 + \cdots + X_k$, $1 \leqslant k \leqslant n$. Согласно закону больших чисел (лекция 4), распределение вероятности среднего смещения $\overline{X}_n$ за n шагов концентрируется на неслучайном значении $\mathsf{E}X = 1 \cdot p + (-1) \cdot (1 - p) = 2p - 1$.

Рассмотрим большое уклонение случайного блуждания, т. е. событие вида $\overline{X}_n = \alpha$, где $\alpha \neq 2p - 1$ (или, точнее, вида $\{|\overline{X}_n - \alpha| \leqslant \varepsilon\}$ при малом $\varepsilon > 0$, поскольку случайная величина $\overline{X}_n$ может принимать лишь рациональные значения, а параметр α можно выбрать любым вещественным). О таком событии закон больших чисел утверждает лишь то, что его вероятность стремится к нулю. Однако, интересно (а для некоторых приложений и важно) выяснить скорость этой сходимости.

Кроме того, закон больших чисел утверждает, что относительная доля шагов вправо в реализации случайного блуждания при больших n стремится к p. Аналогично обстоит дело и для т. н. «ленивого» случайного блуждания, когда X_i может принимать с положительными вероятностями три значения ± 1 и 0: доли шагов вправо, влево и «на месте» стремятся к соответствующим вероятностям.

Если произошло большое уклонение указанного вида, то нетрудно сообразить, что для обычного случайного блуждания доля шагов вправо должна составить $(1 + \alpha)/2$. Однако для «ленивого» блуждания фиксации одного параметра α уже недостаточно для определения предельных значений долей шагов вправо, влево и «на месте», и для их нахождения необходимо привлекать дополнительные идеи.

Оказывается, что и «невероятные» события происходят определенным наиболее вероятным образом, т. е. при подходящей перенормировке концентрация вероятности наблюдается и вне области наиболее вероятных реализаций. Результаты такого типа называются **принципами больших уклонений**.

7.2. Пусть $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ — совокупность независимых одинаково распределенных случайных величин.

В этой лекции будем предполагать, что каждая из них может принимать конечное множество различных значений, природа которых несущественна. Чтобы подчеркнуть это (и выявить связь последующих рассмотрений с теорией информации), примем, что значениями случайных величин X_i являются не числа, а произвольные символы $a, b, \ldots, z$, образующие конечный **алфавит** $\mathcal{A}$. Вероятности символов обозначим p_a, $a \in \mathcal{A}$, а их общее число в алфавите — $|\mathcal{A}|$ и будем предполагать, что $|\mathcal{A}| > 1$.

Кроме этого, в данной лекции, в отступление от принятой в курсе общей системы обозначений, некоторые случайные величины будут обозначаться строчными буквами.

7.3. Каждая реализация совокупности случайных величин $\boldsymbol{X}$ образует случайную n-буквенную **строку** $s = a_1 a_2 \ldots a_n$ в алфавите $\mathcal{A}$. Поскольку величины X_i независимы, вероятность этой строки $\mathsf{P}(s) = \mathsf{P}(a_1 a_2 \ldots a_n)$ равна произведению $p_{a_1} p_{a_2} \ldots p_{a_n}$. В совокупности вероятности строк образуют распределение на множестве $\mathcal{A}^n$, число элементов которого с ростом n растет экспоненциальным образом, как $|\mathcal{A}|^n$.

7.4. Обозначим $\nu_a(s)$ относительную частоту буквы a в строке s, т. е. число вхождений буквы a в строку s, деленное на n. Тогда $\nu_a(s)$ есть случайная величина, принимающая значения от 0 до 1. По закону больших чисел (лекция 4) ее распределение при $n \to \infty$ сосредотачивается на неслучайном значении p_a.

7.5. Чтобы описать структуру распределения вероятности по множеству $\mathcal{A}^n$, определенного в п. 7.3, назовем **типом** строки s набор $\boldsymbol{\nu}(s) = (\nu_a(s), \nu_b(s), \ldots, \nu_z(s))$. Разумеется, для любой строки числа, составляющие ее тип, образуют распределение вероятности по символам алфавита $\mathcal{A}$. Однако в отличие от исходного распределения $\boldsymbol{p} = (p_a, p_b, \ldots, p_z)$, в котором вероятности p_a, $a \in \mathcal{A}$, могут быть произвольными вещественными числами из отрезка $(0, 1)$, вероятности $\nu_a(s)$ необходимо являются рациональными числами.

7.6. Число различных возможных типов равно $C_{n+|\mathcal{A}|-1}^{|\mathcal{A}|-1}$ и представляет собой многочлен степени $|\mathcal{A}| - 1$ от n.

◀ Доказательство проводится стандартным комбинаторным подсчетом «распределений шаров по ящикам»: каждый тип можно представить одной и только одной строкой вида

$$\bullet\bullet \mid \bullet \mid\mid \bullet\bullet\bullet\bullet \mid \bullet\bullet\bullet,$$

где $|\mathcal{A}| - 1$ вертикальных черт разделяют «ящики», соответствующие различным буквам алфавита, перечисляемым слева направо, а n точек изображают отдельные символы строки без учета их порядка (на схеме $n = 10$ и $|\mathcal{A}| = 5$, причем относительная частота первой буквы равна $\frac{2}{10}$, второй — $\frac{1}{10}$, третьей — нулю и т. д.). Каждая из таких строк состоит из $n + |\mathcal{A}| - 1$ символов, ровно $|\mathcal{A}| - 1$ из которых должны быть вертикальными чертами. ▶

Поскольку число всех возможных строк $|\mathcal{A}|^n$ растет экспоненциально по n, а число возможных типов растет лишь как конечная степень n, следует ожидать, что и число строк внутри каждого типа растет по n также экспоненциально.

7.7. Зафиксируем тип $\boldsymbol{\mu} = (\mu_a)$, $a \in \mathcal{A}$, и положим $\boldsymbol{m} = (m_a) = (n\mu_a) = n\boldsymbol{\mu}$. Тогда, обобщая обычные рассуждения при выводе биномиального коэффициента, получаем формулу для числа строк, относящихся к типу $\boldsymbol{\mu}$:

$$|\{s \colon \boldsymbol{\nu}(s) = \boldsymbol{\mu}\}| = \frac{n!}{m_a!\, m_b! \ldots m_z!}.$$

7.8. По формуле Стирлинга $n! = n^n \mathrm{e}^{-n}\sqrt{2\pi n}\,[1+O(\frac{1}{n})]$. Учитывая только экспоненциальную асимптотику по n, получим

$$|\{s\colon \boldsymbol{\nu}(s)=\boldsymbol{\mu}\}| = \frac{n!}{m_a!\,m_b!\,\ldots m_z!} \approx \frac{n^n\cdot \mathrm{e}^{-n}}{m_a^{m_a}m_b^{m_b}\ldots m_z^{m_z}\cdot \mathrm{e}^{-m_a-m_b-\cdots-m_z}} =$$

$$= \frac{1}{\mu_a^{m_a}\mu_b^{m_b}\ldots\mu_z^{m_z}} = (\mathrm{e}^{-\mu_a\log\mu_a}\,\mathrm{e}^{-\mu_b\log\mu_b}\,\ldots \mathrm{e}^{-\mu_z\log\mu_z})^n = \mathrm{e}^{n\mathsf{H}(\boldsymbol{\mu})},$$

где

$$\mathsf{H}(\boldsymbol{\mu}) = \mathsf{H}(\mu_a,\mu_b,\ldots,\mu_z) := -\sum_{a\in\mathcal{A}}\mu_a\log\mu_a > 0.$$

Функция $\mathsf{H}(\cdot)$ может быть определена не только для типа $\boldsymbol{\mu}$, но и для произвольного распределения вероятности. Она называется **энтропией по Шеннону** данного распределения вероятности.

7.9. Более точная оценка $|\{s\colon \boldsymbol{\nu}(s)=\boldsymbol{\mu}\}|$ при *конечных* n дается следующими неравенствами, которые включают степенной предэкспоненциальный множитель:

$$(C_{n+|\mathcal{A}|-1}^{|\mathcal{A}|-1})^{-1}\mathrm{e}^{n\mathsf{H}(\boldsymbol{\mu})} < |\{s\colon \boldsymbol{\nu}(s)=\boldsymbol{\mu}\}| < \mathrm{e}^{n\mathsf{H}(\boldsymbol{\mu})}.$$

◀ Заметим, что

$$n^n = (m_a+m_b+\cdots+m_z)^n = \sum_{k_a+k_b+\cdots+k_z=n}\frac{n!}{k_a!\,k_b!\,\ldots k_z!}m_a^{k_a}m_b^{k_b}\ldots m_z^{k_z},$$

причем наибольшим слагаемым этой суммы является то, в котором $\boldsymbol{k}=\boldsymbol{m}$. Действительно, если для некоторого слагаемого $k_r \geqslant m_r+1$ и $k_s \leqslant m_s-1$, то при уменьшении k_r и увеличении k_s на единицу это слагаемое умножится на коэффициент

$$k_r\frac{1}{k_s+1}\frac{1}{m_r}m_s \geqslant \frac{k_r}{m_r} > 1.$$

С другой стороны, слагаемые в этой сумме находятся во взаимно однозначном соответствии с возможными типами, и поэтому их общее число равно $C_{n+|\mathcal{A}|-1}^{|\mathcal{A}|-1}$. Поэтому

$$\frac{n!}{m_a!\,m_b!\,\ldots m_z!}m_a^{m_a}m_b^{m_b}\ldots m_z^{m_z} < n^n < C_{n+|\mathcal{A}|-1}^{|\mathcal{A}|-1}\frac{n!}{m_a!\,m_b!\,\ldots m_z!}m_a^{m_a}m_b^{m_b}\ldots m_z^{m_z}.$$

Отсюда следуют неравенства

$$(C_{n+|\mathcal{A}|-1}^{|\mathcal{A}|-1})^{-1}\frac{n^n}{m_a^{m_a}m_b^{m_b}\ldots m_z^{m_z}} < \frac{n!}{m_a!\,m_b!\,\ldots m_z!} < \frac{n^n}{m_a^{m_a}m_b^{m_b}\ldots m_z^{m_z}},$$

а из них вытекают искомые оценки. ▶

Теперь можно вычислить распределение вероятности по различным типам.

7.10. Все строки, относящиеся к типу $\boldsymbol{\mu}$, в соответствии с определением п. 7.3 имеют одну и ту же вероятность $p_a^{m_a}p_b^{m_b}\ldots p_z^{m_z}$. Поэтому вероятность того, что случайная строка принадлежит типу $\boldsymbol{\mu}$, удовлетворяет оценкам (проверьте!)

$$(C_{n+|\mathcal{A}|-1}^{|\mathcal{A}-1})^{-1}\mathrm{e}^{-n\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p})} < \mathsf{P}(\boldsymbol{\nu}(s)=\boldsymbol{\mu}) < \mathrm{e}^{-n\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p})},$$

где введена **относительная энтропия** распределения $\boldsymbol{\mu}$ по отношению к распределению $\boldsymbol{p}$

$$\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p}) := \sum_{a\in\mathcal{A}}\mu_a\log\frac{\mu_a}{p_a}.$$

Для лучшего понимания полученных результатов необходимо разобраться в свойствах энтропии Шеннона и относительной энтропии.

7.11. Пусть алфавит $\mathcal{A}$ содержит всего две буквы a, b, так что $\mu_b = 1 - \mu_a$ и $H(\boldsymbol{\mu}) = H(\mu_a) = -\mu_a \ln \mu_a - (1-\mu_a)\ln(1-\mu_a)$. Полагая $0 \ln 0 = 0$ по непрерывности, получаем, что функция $H(\mu_a)$ обращается в нуль при $\mu_a = 0$ или $\mu_a = 1$. Ее производные имеют вид $H'(\mu_a) = -\ln \mu_a - 1 + \ln(1-\mu_a) + 1 = \ln(1-\mu_a) - \ln \mu_a$, $H''(\mu_a) = -\mu_a^{-1} - (1-\mu_a)^{-1} < 0$, т. е. функция $H(\mu_a)$ выпукла вверх на отрезке $0 \leqslant \mu_a \leqslant 1$ и достигает максимального значения $\ln 2$ при $\mu_a = 1 - \mu_a = \frac{1}{2}$. Заметим, что график этой функции симметричен относительно замены $\mu_a \mapsto 1 - \mu_a$, переставляющей концы отрезка, и имеет в этих концах вертикальные касательные.

7.12. Множество распределений вероятности

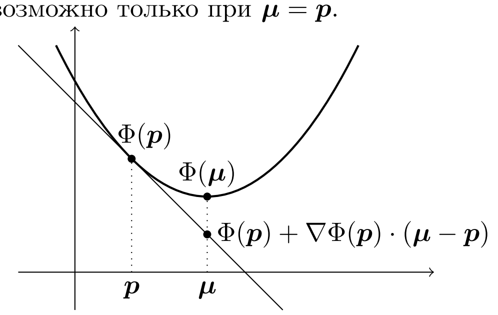
$$\Big\{\boldsymbol{p} = (p_a, p_b, \ldots, p_z) \colon p_a \geqslant 0 \text{ для всех } a \in \mathcal{A},\ \sum_{a\in\mathcal{A}} p_a = 1\Big\}$$

образует **вероятностный симплекс** размерности $|\mathcal{A}| - 1$ (см. рис. к п. 7.16), грани которого в свою очередь также являются вероятностными симплексами меньших размерностей (проверьте!). В частности, при $|\mathcal{A}| = 2$ вероятностный симплекс представляет собой отрезок (ср. п. 7.11). Энтропия по Шеннону $H(\boldsymbol{\mu})$ — гладкая, строго выпуклая вверх функция на вероятностном симплексе, достигающая максимального значения $\ln|\mathcal{A}|$ на равномерном распределении и симметричная относительно действия группы перестановок вершин симплекса.

7.13. Вспомогательный факт из выпуклого анализа. Предположим, что $\Phi(\boldsymbol{p})$ — непрерывно дифференцируемая выпуклая функция векторного аргумента. Тогда для всех $\boldsymbol{\mu}$ имеет место неравенство

$$\Phi(\boldsymbol{\mu}) \geqslant \Phi(\boldsymbol{p}) + \nabla\Phi(\boldsymbol{p}) \cdot (\boldsymbol{\mu} - \boldsymbol{p}),$$

причем если функция Φ строго выпукла (т. е. ее график не содержит линейных участков, ср. У1.1), то равенство возможно только при $\boldsymbol{\mu} = \boldsymbol{p}$.



◀ См. рисунок. ▶

7.14. Функция $\Phi(\boldsymbol{p}) = \sum_{a\in\mathcal{A}} p_a \log p_a$ строго выпукла. Если векторы $\boldsymbol{\mu}$ и $\boldsymbol{p}$ — распределения вероятности, то

$$\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p}) = \Phi(\boldsymbol{\mu}) - \Phi(\boldsymbol{p}) - \nabla\Phi(\boldsymbol{p}) \cdot (\boldsymbol{\mu} - \boldsymbol{p})$$

и поэтому $\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p}) \geqslant 0$, причем равенство достигается лишь при $\boldsymbol{\mu} = \boldsymbol{p}$.

◀ Очевидно, $\mathrm{d}^2(p\log p)/\mathrm{d}p^2 = 1/p > 0$ при $p > 0$. Функция $\Phi(\boldsymbol{p})$ является суммой выпуклых слагаемых и поэтому сама выпукла. Заметим теперь, что $\partial\Phi(\boldsymbol{p})/\partial p_a = \log p_a + 1$, и вычислим правую часть доказываемого равенства:

$$\Phi(\boldsymbol{\mu}) - \Phi(\boldsymbol{p}) - \nabla\Phi(\boldsymbol{p}) \cdot (\boldsymbol{\mu} - \boldsymbol{p}) = \sum_{a\in\mathcal{A}} [\mu_a \log \mu_a - p_a \log p_a - (\log p_a + 1)(\mu_a - p_a)] =$$

$$= \sum_{a\in\mathcal{A}} [\mu_a \log \mu_a - p_a \log p_a - \mu_a \log p_a + p_a \log p_a] = \sum_{a\in\mathcal{A}} \mu_a \log \frac{\mu_a}{p_a} = \mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p}),$$

поскольку $\sum_{a\in\mathcal{A}}\mu_a = \sum_{a\in\mathcal{A}} p_a = 1$. ►

Таким образом, относительная энтропия $\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p})$ есть мера несовпадения распределений $\boldsymbol{\mu}$ и $\boldsymbol{p}$, обращающаяся в нуль тогда и только тогда, когда ее аргументы совпадают (хотя ее нельзя считать «расстоянием», потому что при $\boldsymbol{\mu}\to\boldsymbol{p}$ она стремится к нулю не линейно, а квадратично — это как бы «деформированный» квадрат расстояния).

В частности, энтропия Шеннона $\mathsf{H}(\boldsymbol{p})$ с точностью до знака и постоянного слагаемого $-\ln n$ совпадает с относительной энтропией $\mathsf{D}(\boldsymbol{p}\|\boldsymbol{u})$ по отношению к равномерному распределению $\boldsymbol{u} = (\frac{1}{|\mathcal{A}|}, \frac{1}{|\mathcal{A}|}, \dots, \frac{1}{|\mathcal{A}|})$. Некоторую непоследовательность в выборе знаков в определениях $\mathsf{H}(\boldsymbol{p})$ (п. 7.8) и $\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p})$ (п. 7.10) допускают для того, чтобы обе величины были положительными.

Следствием п. 7.10 является следующий аналог закона больших чисел, в котором рассматривается сходимость не сумм случайных величин, а типов строк, вместо расстояния фигурирует относительная энтропия, а сходимость происходит с экспоненциальной скоростью.

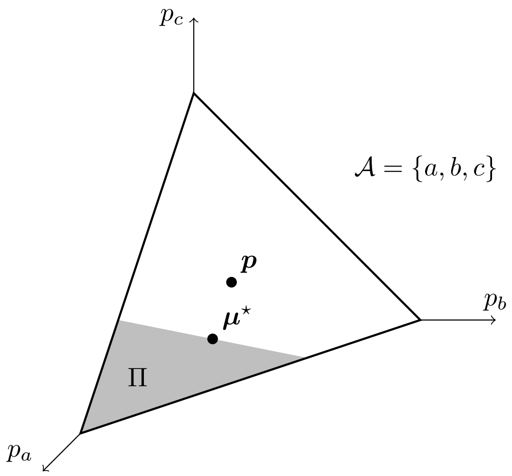
7.15. Для сколь угодно малого $\delta > 0$ полная вероятность множества строк, типы которых отличаются от $\boldsymbol{p}$ по относительной энтропии не менее, чем на δ, стремится к нулю:

$$\mathsf{P}(\mathsf{D}(\boldsymbol{\nu}(s)\|\boldsymbol{p}) \geqslant \delta) \leqslant C_{n+|\mathcal{A}|-1}^{|\mathcal{A}|-1} \mathrm{e}^{-n\delta} \xrightarrow[n\to\infty]{} 0.$$

◄ Действительно, для любого такого типа $\boldsymbol{\mu}$, что $\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p}) > \delta$, имеем $\mathsf{P}(\mu) < \mathrm{e}^{-n\delta}$ согласно п. 7.10, а полное число таких типов не превышает $C_{n+|\mathcal{A}|-1}^{|\mathcal{A}|-1}$. ►

7.16. Теорема Санова. Пусть Π — непустое замкнутое подмножество вероятностного симплекса, совпадающее с замыканием своей внутренности, и $\mathsf{D}(\Pi\|\boldsymbol{p}) := \min_{\boldsymbol{\mu}\in\Pi}\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p})$. Тогда

$$-\frac{1}{n}\log\mathsf{P}(\boldsymbol{\nu}(s)\in\Pi) \xrightarrow[n\to\infty]{} \mathsf{D}(\Pi\|\boldsymbol{p}).$$



◄ Пусть Π_n — множество тех типов n-буквенных слов, которые лежат в Π. Смысл топологического условия на Π в том, чтобы множество Π_n было непустым и величина $\mathsf{D}(\Pi_n\|\boldsymbol{p})$ стремилась к $\mathsf{D}(\Pi\|\boldsymbol{p})$ при $n\to\infty$; это условие при желании нетрудно ослабить. Если оно выполнено, то из п. 7.15 следует оценка сверху

$$\mathsf{P}(\boldsymbol{\nu}(s)\in\Pi_n) \leqslant C_{n+|\mathcal{A}|-1}^{|\mathcal{A}|-1}\mathrm{e}^{-n\mathsf{D}(\Pi_n\|\boldsymbol{p})}.$$

Для оценки снизу можно использовать тот класс $\boldsymbol{\mu}_n^\star\in\Pi_n$, на котором $\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p})$ достигает минимума на Π_n. Тогда

$$\mathsf{P}(\boldsymbol{\nu}(s)\in\Pi_n) \geqslant (C_{n+|\mathcal{A}|-1}^{|\mathcal{A}|-1})^{-1}\exp(-n\mathsf{D}(\Pi_n\|\boldsymbol{p}))$$

согласно п. 7.10. ▶

Можно получить и результат о концентрации вероятности внутри П, аналогичный п. 7.15.

7.17. Если область П выпукла, то минимум $\mathsf{D}(\boldsymbol{\mu}\|\boldsymbol{p})$ достигается на ней в единственной точке $\boldsymbol{\mu}^\star$. Для любого $\delta > 0$ условная вероятность множества строк внутри П, типы которых отличаются от $\boldsymbol{\mu}^\star$ по относительной энтропии не менее, чем на δ, стремится к нулю:

$$\frac{\mathsf{P}(\mathsf{D}(\boldsymbol{\nu}(s)\|\boldsymbol{p}) \geqslant \delta + \mathsf{D}(\Pi\|\boldsymbol{p}))}{\mathsf{P}(\boldsymbol{\nu}(s) \in \Pi)} \leqslant P(n)\,\mathrm{e}^{-n\delta} \xrightarrow[n\to\infty]{} 0,$$

где $P(n)$ — функция степенного роста по n.

7.18. Еще раз случайное блуждание. В примере «ленивого случайного блуждания» положим $p_a = \mathsf{P}(X = -1)$, $p_b = \mathsf{P}(X = 0)$, $p_c = \mathsf{P}(X = 1)$ и введем частоты ν_a, ν_b, ν_c, имеющие аналогичный смысл. Тогда среднее смещение $\overline{X}_n = \nu_c - \nu_a$ является аффинной функцией на вероятностном симплексе, а множество П, выделяемое, например, условием $\nu_c - \nu_a \leqslant \alpha < p_c - p_a$, очевидно удовлетворяет условиям теоремы Санова, причем в точке $\boldsymbol{\mu}^\star$ выполнено равенство $\mu_c^\star - \mu_a^\star = \alpha$ (почему?). Значения координат $\boldsymbol{\mu}^\star$ находятся из уравнений

$$\begin{gathered}\ln(\mu_a/p_a) + 1 - \xi + \eta = 0,\\ \ln(\mu_b/p_b) + 1 - \xi = 0,\\ \ln(\mu_c/p_c) + 1 - \xi - \eta = 0,\end{gathered}$$

которые возникают как условия оптимальности в задаче минимизации относительной энтропии с ограничениями $\mu_a+\mu_b+\mu_c = 1$ (множитель Лагранжа ξ) и $\mu_c-\mu_a = \alpha$ (множитель Лагранжа η). Для определения значений ξ и η получаем

$$\begin{gathered}p_a\mathrm{e}^{-1+\xi-\eta} + p_b\mathrm{e}^{-1+\xi} + p_c\mathrm{e}^{-1+\xi+\eta} = 1,\\ p_c\mathrm{e}^{-1+\xi+\eta} - p_a\mathrm{e}^{-1+\xi-\eta} = \alpha.\end{gathered}$$

8. Большие уклонения II: теорема Крамера

В этой лекции устананавливается принцип больших уклонений для суммы большого числа *непрерывных* случайных величин (как всегда, независимых и одинаково распределенных). Это — теорема Краме́ра, один из самых ранних результатов в теории больших уклонений (Гаральд Крамер, 1938). Поскольку в непрерывном случае неясно, как построить подходящий аналог теории типов, здесь используется более прямой подход, а полученный результат ближе к нашему «источнику вдохновения» — закону больших чисел (лекция 4).

8.1. Пусть $X_1, X_2, \ldots, X_n$ — набор n независимых, одинаково распределенных скалярных случайных величин.

Начнем с простого рассуждения, позволяющего почти «из ничего» получить знакомую экспоненциальную асимптотику вероятности большого уклонения: $\mathsf{P}(a \leqslant \overline{X}_n \leqslant b) \sim \mathrm{e}^{-ns(a,b)}$. Точный смысл употребленного здесь знака «грубой асимптотики» $\sim$ определяется следующим образом:

8.2. $-\lim\limits_{n\to\infty} \dfrac{1}{n} \ln \mathsf{P}(a \leqslant \overline{X}_n \leqslant b) = s(a,b)$ для любых $a < b$, где $s(a,b) \geqslant 0$.

Доказательство сформулированного утверждения проводится в два этапа (пп. 8.3, 8.4), второй из которых — хорошо известная и полезная лемма о субаддитивных числовых последовательностях.

8.3. Последовательность неотрицательных чисел $-\ln \mathsf{P}(a \leqslant \overline{X}_n \leqslant b)$ **субаддитивна**: для любых n, $n' \geqslant 1$

$$-\ln \mathsf{P}(a \leqslant \overline{X}_{n+n'} \leqslant b) \leqslant -\ln \mathsf{P}(a \leqslant \overline{X}_n \leqslant b) - \ln \mathsf{P}(a \leqslant \overline{X}_{n'} \leqslant b).$$

◀ Неотрицательность очевидна, поскольку вероятности всегда заключены между 0 и 1. Заметим, что неравенства

$$a \leqslant \overline{X}_{n+n'} = \frac{1}{n+n'}(X_1 + X_2 + \cdots + X_n + X_{n+1} + X_{n+2} + \ldots X_{n+n'}) \leqslant b$$

заведомо выполнены, если имеют место неравенства

$$a \leqslant \overline{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \leqslant b, \quad a \leqslant \frac{1}{n'}(X_{n+1} + X_{n+2} + \cdots + X_{n+n'}) \leqslant b$$

(действительно, домножим первое из них на $n/(n+n')$, второе на $n'/(n+n')$ и сложим). Поэтому событие $\{a \leqslant \overline{X}_{n+n'} \leqslant b\}$ включает в себя событие, заданное двумя последними неравенствами, т. е. его вероятность не меньше, чем вероятность этого последнего события. Она в свою очередь, благодаря независимости случайных величин $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+n'}$, может быть представлена как произведение:

$$\begin{aligned}\mathsf{P}(a \leqslant \overline{X}_{n+n'} \leqslant b) \geqslant \mathsf{P}(a \leqslant \overline{X}_n \leqslant b,\ a \leqslant \frac{1}{n'}(X_{n+1} + \cdots + X_{n+n'}) \leqslant b) = \\ = \mathsf{P}(a \leqslant \overline{X}_n \leqslant b)\,\mathsf{P}(a \leqslant \frac{1}{n'}(X_{n+1} + \cdots + X_{n+n'}) \leqslant b).\end{aligned}$$

Логарифмируя это произведение и учитывая, что величина $(X_{n+1} + \cdots + X_{n+n'})/n'$ распределена так же, как $\overline{X}_{n'}$, после перемены знака получим требуемое неравенство. ▶

8.4. Пусть $h_n \geqslant 0$, $n = 1, 2, \ldots$ — субаддитивная последовательность, т. е. $h_{n+n'} \leqslant h_n + h_{n'}$ для любых $n, n' \geqslant 1$. Тогда последовательность h_n/n имеет конечный предел.

◀ Зафиксируем некоторое натуральное число $n_0 \geqslant 1$ и представим произвольное $n \geqslant 1$ в виде $n = qn_0 + r$, где q — частное, а r — остаток от деления n на n_0: в частности, $0 \leqslant r < n_0$. Из субаддитивности следует, что $h_n = h_{qn_0+r} \leqslant h_{qn_0} + h_r \leqslant h_{(q-1)n_0} + h_{n_0} + h_r \leqslant \ldots \leqslant qh_{n_0} + h_r$ (при $r = 0$ считаем $h_0 = 0$). Имеем

$$\frac{1}{n}h_n \leqslant \frac{q}{qn_0 + r}h_{n_0} + \frac{1}{n}h_r \leqslant \frac{1}{n_0}h_{n_0} + \frac{1}{n}h_r.$$

Поскольку $\max h_r = \max\{h_0, h_1, \ldots, h_{n_0-1}\}$ конечен и не зависит от n, второе слагаемое в правой части при переходе $n \to \infty$ исчезает, и для произвольного $n_0 \geqslant 1$ получаем

$$\limsup_{n\to\infty} \frac{1}{n}h_n \leqslant \frac{1}{n_0}h_{n_0}.$$

Данный верхний предел неотрицателен, поскольку $h_n \geqslant 0$, и потому конечен. Обозначим его s и заметим, что $s \leqslant h_{n_0}/n_0$ для любого n_0, и потому

$$s \leqslant \liminf_{n_0\to\infty} \frac{1}{n_0}h_{n_0} \leqslant \limsup_{n\to\infty} \frac{1}{n}h_n \leqslant s,$$

так что в действительности $\lim_{n\to\infty} h_n/n = s$. ▶

При выводе результата п. 8.2 не было сделано никаких предположений о характере распределения X_i, не требовалось даже, чтобы для этих случайных величин был выполнен закон больших чисел. В частности, если X_i распределены по Коши, то $s(a,b) \equiv 0$ (почему?), и в этом случае полученный результат, конечно, несет мало информации. Наша очередная задача — выразить $s(a,b)$ через распределение вероятности случайных величин X_i и при этом выделить условия, при которых $s(a,b) > 0$, т. е. результат п. 8.2 нетривиален.

Следующее эвристическое рассуждение показывает, чего нужно ожидать. Допустим, что случайные величины X_i имеют функцию плотности вероятности $p(x)$, и обозначим функцию плотности вероятности величины $\overline{X}_n$ через $p_n(x)$. Исходя из только что полученного результата, естественно предположить, что $p_n(x) \sim \mathrm{e}^{-n\varkappa^*(x)}$. Поэтому

$$\mathsf{P}(a \leqslant \overline{X}_n \leqslant b) \sim \int_a^b \mathrm{e}^{-n\varkappa^*(x)}\,\mathrm{d}x \sim \exp(-n \min_{a\leqslant x\leqslant b} \varkappa^*(x)),$$

так что $s(a,b) = -\lim_{n\to\infty} \ln \mathsf{P}(a \leqslant \overline{X}_n \leqslant b)/n = \min_{a\leqslant x\leqslant b} \varkappa^*(x)$. Задача, таким образом, заключается в вычислении функции $\varkappa^*$.

8.5. Пусть распределение вероятности случайных величин X_i обладает функцией плотности вероятности $p(x)$ и характеристической функцией $\varphi(s) = \mathrm{e}^{\eta(s)}$, где η — характеристический показатель (п. 3.11). Повторяя вычисления п. 4.7, получаем для величины $\overline{X}_n = (X_1 + X_2 + \cdots + X_n)/n$ характеристическую функцию $\varphi_n(s) = (\varphi(s/n))^n = \mathrm{e}^{n\eta(s/n)}$ и функцию плотности вероятности, выраженную через обратное преобразование Фурье:

$$p_n(x) = \frac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}sx}\varphi_n(s)\,\mathrm{d}s = \frac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}sx+n\eta(s/n)}\,\mathrm{d}s = \frac{n}{2\pi}\int \mathrm{e}^{-n[\mathrm{i}ux-\eta(u)]}\,\mathrm{d}u,$$

где в интеграле произведена замена переменной $s = nu$.

Нас интересует асимптотика последнего интеграла при $n \to \infty$. Для ее получения допустим, что переменная u может изменяться на всей комплексной плоскости, и продеформируем контур интегрирования так, чтобы облегчить оценку интеграла. Соответствующее построение описано в курсах комплексного анализа и называется «методом перевала». Напомним основные идеи этого метода.

Рассмотрим при $n \gg 1$ интеграл $\int \mathrm{e}^{n\Phi(u)}\,\mathrm{d}u$, где Φ аналитическая функция. Вклад в этот интеграл от участка контура интегрирования, проходящего через точку u, определяется двумя факторами: абсолютной величиной подынтегрального выражения $\mathrm{e}^{n\Phi(u)}$ в этой точке и его осцилляциями вдоль контура интегрирования. При больших n быстрые осцилляции из-за изменения аргумента функции $\mathrm{e}^{n\Phi(u)}$, т. е. $\operatorname{Im}\Phi(u)$, вдоль контура, приводят к взаимному сокращению вкладов от разных точек и уменьшению абсолютной величины интеграла. Поэтому контур желательно провести так, чтобы вдоль него исчезала производная $\operatorname{Im}\Phi$. При этом условии наибольший вклад в интеграл внесут те точки, в которых абсолютная величина подынтегрального выражения, а значит и $\operatorname{Re}\Phi(u)$, достигает вдоль контура своего максимума. Поскольку в точке максимума производная $\operatorname{Re}\Phi$ также исчезает, из аналитичности функции Φ (т. е. независимости ее производной от направления дифференцирования) следует, что в такой точке $\Phi'(u) = 0$.

Поскольку функция $\operatorname{Re}\Phi$ гармоническая, такие точки u^* являются для нее седловыми, или «точками перевала»: при движении через седловую точку в одних направлениях $\operatorname{Re}\Phi$ проходит локальный максимум, а в других — локальный минимум. Если $\Phi''(u^*) \neq 0$, условие $\operatorname{Im}\Phi = \mathrm{const}$ выделяет два направления наискорейшего изменения $\operatorname{Re}\Phi$; контур надо проводить в том из них, в котором $\operatorname{Re}\Phi$ имеет локальный максимум. В остальной части контура осцилляции можно специально не подавлять, поскольку при больших n вклад этой части по абсолютному значению и так пренебрежимо мал по сравнению с вкладом от точек перевала, в которых $\operatorname{Re}\Phi$ достигает наибольших значений. Разумеется, деформации контура возможны лишь внутри области аналитичности функции $\Phi(u)$, без пересечения ее особых точек.

В окрестности каждой точки перевала u^* функция Φ может быть разложена в ряд Тейлора:

$$\Phi(u) = \Phi(u^*) + \frac{1}{2}\Phi(u^*)''\,(u-u^*)^2 + \ldots,$$

где квадратичное слагаемое вдоль контура вещественно и отрицательно, поскольку $\operatorname{Re}\Phi$ достигает максимума в точке u^*, а $\operatorname{Im}\Phi$ не изменяется. Поэтому вклад в интеграл от точки перевала u^* можно приближенно заменить на соответствующий нормировочный интеграл от распределения Гаусса:

$$\int \mathrm{e}^{-n\Phi(u)}\,\mathrm{d}u \sim \sum_{u^*}\int \mathrm{e}^{n\Phi(u^*)+\frac{n}{2}\Phi(u^*)''\,(u-u^*)^2+\ldots}\,\mathrm{d}u \sim \sum_{u^*}\mathrm{e}^{n\Phi(u^*)}\sqrt{\frac{2\pi}{n|\Phi(u^*)''|}},$$

где суммирование распространено на все точки перевала u^*, отвечающие глобально максимальному значению $\operatorname{Re}\Phi(u)$.

8.6. Потребуем, чтобы преобразование Лапласа функции плотности вероятности p

$$\mathrm{e}^{\varkappa(s)} = \int_{-\infty}^{\infty} \mathrm{e}^{sx}\,p(x)\,\mathrm{d}x$$

существовало как аналитическая функция переменной s при $a < s < b$ (а значит, и в полосе $a < \operatorname{Re} s < b$), т. е. чтобы p убывало на бесконечности по меньшей мере экспоненциально. Без потери общности можно выбрать максимальный такой интервал (a, b); тогда $a < 0 < b$, поскольку при $s = 0$ интеграл сходится.

8.7. В предположенях п. 8.6 характеристический показатель $\eta(u) = \varkappa(\mathrm{i}u)$ аналитичен в полосе $-b < \operatorname{Im} u < -a$, причем максимальное значение $\operatorname{Re}\eta(u)$ для любого фиксированного $\operatorname{Im} u$ достигается на мнимой оси.

◀ Доказательства требует только последнее утверждение. Действительная часть характеристического показателя $\operatorname{Re}\eta$ достигает максимума вместе с абсолютным значением самой характеристической функции, а оно для любого u удовлетворяет соотношениям

$$|\mathrm{e}^{\eta(u)}| = \left|\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}x \operatorname{Re} u}\, \mathrm{e}^{-x \operatorname{Im} u}\, p(x)\, \mathrm{d}x\right| \leqslant \int_{-\infty}^{\infty} |\mathrm{e}^{\mathrm{i}x \operatorname{Re} u}|\, \mathrm{e}^{-x \operatorname{Im} u}\, p(x)\, \mathrm{d}x = \mathrm{e}^{\eta(\mathrm{i} \operatorname{Im} u)} = \mathrm{e}^{\varkappa(-\operatorname{Im} u)},$$

т. е. при фиксированном $\operatorname{Im} u$ абсолютное значение $|\mathrm{e}^{\eta(u)}|$ достигает максимума при $\operatorname{Re} u = 0$.

Кроме того, заметим, что при $\operatorname{Re} u \neq 0$ комплексное значение интеграла под знаком модуля есть выпуклая комбинация точек границы единичного круга с непрерывно распределенным весом. Поскольку единичный круг является строго выпуклым множеством, это значение лежит в его внутренности, т. е. при $\operatorname{Re} u \neq 0$ неравенство является строгим. ▶

8.8. Функция $\varkappa(s)$ строго выпукла на интервале $a < s < b$ вещественной оси, причем $\varkappa(a) = \varkappa(b) = \infty$.

◀ Достаточно проверить, что $\varkappa''(s) > 0$ для вещественных значений $a < s < b$. Имеем

$$\frac{\mathrm{d}\mathrm{e}^{\varkappa(s)}}{\mathrm{d}s} = \mathrm{e}^{\varkappa(s)}\, \varkappa'(s) = \int \mathrm{e}^{sx}\, xp(x)\, \mathrm{d}x.$$

Заметим, что $\exp[sx - \varkappa(s)]\, p(x)$ есть функция плотности вероятности некоторого распределения (неотрицательность очевидна, нормировка на единицу следует из формулы п. 8.6). Обозначим математическое ожидание относительно этого распределения через E_s, тогда $\varkappa'(s) = \mathsf{E}_s X$. Вычисляя вторую производную $\exp\varkappa(s)$, получим

$$\frac{\mathrm{d}^2\mathrm{e}^{\varkappa(s)}}{\mathrm{d}s^2} = \mathrm{e}^{\varkappa(s)}[\varkappa'(s)]^2 + \mathrm{e}^{\varkappa(s)}\varkappa''(s) = \int \mathrm{e}^{sx}\, x^2 p(x)\, \mathrm{d}x,$$

откуда $\varkappa''(s) = \mathsf{E}_s X^2 - (\mathsf{E}_s X)^2 = \mathsf{D}_s X > 0$.

Поскольку $\exp\varkappa(s) > 0$, расходимость преобразования Лапласа при $s = a, b$ означает, что $\exp\varkappa(a) = \exp\varkappa(b) = \infty$, откуда и $\varkappa(a) = \varkappa(b) = \infty$. ▶

8.9. Выпуклая функция $\varkappa(s)$ может быть представлена в виде

$$\varkappa(s) = \sup_y [sy - \varkappa^*(y)],$$

где функция $\varkappa^*(y)$ также выпукла. Она называется **преобразованием Лежандра** $\varkappa(s)$, причем

$$\varkappa^*(y) = \sup_{a<s<b} [sy - \varkappa(s)].$$

Если функция $\varkappa$ непрерывно дифференцируема и строго выпукла, то такими же свойствами обладает и $\varkappa^*$, причем производные $\varkappa'$ и $\varkappa^{*\prime}$ взаимно обратны.

◀ Для непрерывной выпуклой функции $\varkappa$ ее «надграфик», т. е. множество $\{(s, t)\colon t \geqslant \varkappa(s)\}$ есть замкнутое выпуклое множество. Первая формула п. 8.9 следует из геометрического представления этого выпуклого множества как пересечения полуплоскостей вида $\{(s, t)\colon t \geqslant sy - \varkappa^*(y)\}$. Ее можно записать в более симметричном виде (**неравенство Юнга**)

$$\varkappa(s) + \varkappa^*(y) \geqslant sy \quad \text{для всех } s,\ y,$$

откуда следует, что $\varkappa^*(y) \geqslant \sup_{a<s<b}[sy - \varkappa(s)]$. Для определенности будем считать, что $\varkappa^*$ наименьшая из всех функций, удовлетворяющих неравенству Юнга при заданной функции $\varkappa$; тогда она выражается через $\varkappa$ по второй формуле п. 8.9 и, следовательно, тоже является выпуклой.

Заметим, что для дифференцируемой функции $\varkappa$ равенство в неравенстве Юнга достигается при таком y, что $\varkappa'(s) = y$. Если $\varkappa$ является строго выпуклой, то ее производная строго возрастает и обратима, откуда следуют строгая монотонность обратной функции $\varkappa^{*\prime}$ и строгая выпуклость $\varkappa^*$. Поэтому условие достижения равенства можно записать и в виде $s = \varkappa^{*\prime}(y)$. ►

Теперь можно завершить вычисления по методу перевала.

8.10. В рассматриваемой ситуации $\Phi(u) = -\mathrm{i}ux + \eta(u)$. Результаты пп. 8.7 и 8.9 показывают, что точкой перевала является точка $u^* = -\mathrm{i}s^*$ мнимой оси, где в s^* выражение $\Phi(-\mathrm{i}s) = \eta(-\mathrm{i}s) - sx = \varkappa(s) - sx$ достигает минимума, а контур интегрирования должен проходить через нее в направлении, параллельном вещественной оси. Из второй формулы п. 8.9 следует, что $\Phi(u^*) = \Phi(-\mathrm{i}s^*) = \varkappa(s^*) - s^*x = -\varkappa^*(x)$, причем в точке u^* вторая производная $\Phi''(u^*) = \eta''(u^*) = -\varkappa''(s^*) < 0$, поскольку $\eta(u) = \varkappa(\mathrm{i}u)$. Следовательно,

$$p_n(x) \sim \mathrm{e}^{-n\varkappa^*(x)}\sqrt{\frac{n}{2\pi|\varkappa''(s^*)|}},$$

причем степенной множитель в грубой асимптотике несущественен:

$$-\lim_{n\to\infty}\frac{1}{n}\ln p_n(x) = \varkappa^*(x).$$

Тем самым обосновано предположение, сделанное перед п. 8.5.

Собирая результаты предыдущих пунктов, получаем следующее утверждение.

8.11. Теорема Краме́ра. Если распределение вероятности, заданное функцией плотности $p_X(x)$, удовлетворяет условиям п. 8.6, то логарифмическая асимптотика вероятности события $a \leqslant \overline{X}_n \leqslant b$ имеет вид

$$-\lim_{n\to\infty}\frac{1}{n}\ln \mathsf{P}(a \leqslant \overline{X}_n \leqslant b) = \min_{a\leqslant x\leqslant b}\varkappa^*(x).$$

8.12. Функция $\varkappa^*(x)$ (преобразование Лежандра $\varkappa(s)$ — логарифма преобразования Лапласа исходной функции плотности вероятности) называется **функцией Краме́ра** или **функционалом действия**.

Распространенным синонимичным английским термином является rate function.

8.13. Пример (тривиальный). Если $p_X(x) = \exp(-x^2/2)/\sqrt{2\pi}$, то $\eta(s) = \frac{1}{2}s^2$, $\varkappa^*(y) = \frac{1}{2}y^2$, и асимптотическое выражение для $p_{\overline{X}_n}(x)$ нетрудно проверить, сравнив с точной формулой.

Очевидно, центральная предельная теорема следует из квадратичной аппроксимации для $\varkappa^*(x)$ в окрестности нуля. Однако теорема Крамера характеризует поведение $p_{\overline{X}_n}$ на всей оси и в этом смысле является далеко идущим уточнением центральной предельной теоремы.

Следующий пример показывает, что предполагать существование функции плотности вероятности $p(x)$ не обязательно.

8.14. Пример. Если $\mathsf{P}(X_i = 1) = p$, $\mathsf{P}(X_i = 0) = 1 - p$, то $e^{\varkappa(s)} = 1 - p + p\mathrm{e}^s$ и

$$\varkappa^*(x) = x\ln\frac{x}{p} + (1-x)\ln\frac{1-x}{1-p} = \mathsf{D}(x, 1-x\|p, 1-p)$$

У1. Упражнения к лекции 1

У1.1. Покажите, что если выпуклая вниз функция f строго выпукла (т. е. ее график не содержит прямолинейных отрезков), то из $\mathsf{E} f(N) = f(\mathsf{E} N)$ следует, что случайная величина N детерминирована. Имеет ли место этот факт без условия строгой выпуклости?

У1.2. Выведите формулу п. 1.27.

У1.3. Пусть случайные величины L, M, N, обладающие совместным распределением $p_{L,M,N}(l,m,n)$, распределены таким образом, что любая пара из них независима относительно соответствующего маргинального распределения:

$$p_{L,M}(l,m) = p_L(l) \cdot p_M(m),$$
$$p_{M,N}(m,n) = p_M(m) \cdot p_N(n),$$
$$p_{L,N}(l,n) = p_L(l) \cdot p_N(n).$$

Следует ли отсюда, что

$$p_{L,M,N}(l,m,n) = p_L(l) \cdot p_M(m) \cdot p_N(n),$$

т. е. что случайные величины L, M, N независимы в совокупности?

У1.4. Биномиальное распределение. Пусть k раз производятся независимые случайные испытания, каждое из которых может завершиться успехом с вероятностью p и неудачей с вероятностью $\bar{p} = 1 - p$. Найти выражение для распределения вероятности числа N успехов в k испытаниях $\mathsf{P}(N = n)$ и получить выражения для G_N, $\mathsf{E}N$ и $\mathsf{D}N$. При каком n вероятность $\mathsf{P}(N = n)$ достигает максимума при заданных k, p?

Вероятностная модель последовательности независимых испытаний, в каждом из которых с одним и тем же распределением вероятности может достигаться один из конечного числа исходов, называется **последовательностью испытаний Бернулли**.

У1.5. Какова вероятность, что в группе из k человек ни у кого не совпадают дни рождения? Для решения принять, что год состоит из 365 дней и все даты рождения равновероятны. При каком k эта вероятность впервые становится меньше $\frac{1}{2}$?

У1.6. Задача Банаха. Как утверждал польский математик Гуго Штейнгауз, другой польский математик, великий Стефан Банах, всегда носил с собой два коробка спичек. Когда ему хотелось закурить, он случайно, с вероятностью $\frac{1}{2}$, выбирал любой из них, доставал спичку и прятал коробок обратно в карман, даже если спичек в нем больше не было. Пусть вначале в каждом из коробков было по n спичек. Когда Банах впервые наткнется на пустой коробок, число спичек в другом коробке будет случайной величиной; найдите ее распределение вероятности. Какова асимптотика математического ожидания этой величины при больших n?

Указание. Заметим, что условие нормировки на единицу искомого распределения вероятности эквивалентно комбинаторному тождеству $\sum_{0 \leqslant l \leqslant m} C_{l+m}^{l} 2^{-l} = 2^m$. Приведите формулу для математического ожидания числа спичек к такой форме, чтобы можно было воспользоваться этим же тождеством, и учтите формулу Стирлинга $n! \simeq n^n \mathrm{e}^{-n} \sqrt{2\pi n}$.

У1.7. Распределение Пуассона. Найти предел распределения вероятности в У1.4, если $p = \lambda/k$ и $k \to \infty$. Получить выражения для его производящей функции, математического ожидания и дисперсии. При каком n вероятность $\mathsf{P}(N = n)$ достигает максимума при заданном λ?

У1.8. Геометрическое и отрицательное биномиальное распределения. Пусть N_1 — число неудач до первого успеха в последовательности независимых случайных испытаний с вероятностью успеха p. Найти распределение вероятностей случайной величины N_1 и получить выражения для $G_{N_1}(z)$, $\mathsf{E}N_1$ и $\mathsf{D}N_1$. Проделать те же вычисления для числа N_k неудач до k-го успеха.

У1.9. Задача о сборщике купонов. На каждой упаковке овсянки печатается купон одного из k различных цветов. Считая, что при отдельной покупке купон каждого цвета может встретиться с равной вероятностью и различные покупки независимы, найти производящую функцию распределения вероятности, математическое ожидание и дисперсию числа упаковок, которые придется купить для того, чтобы собрать купоны всех k цветов.

Можно показать, что $p(n) = k!\left\{{n-1 \atop k-1}\right\}k^{-n}$, где $\left\{{n \atop k}\right\}$ — число способов разбить n-элементное множество на k непустых подмножеств, или число Стирлинга второго рода. Числа Стирлинга возникают также при выражении обычных степеней через факториальные: $x^n = \sum_{0\leqslant k\leqslant n} \left\{{n \atop k}\right\} x^{\underline{k}}$. Подробнее см.: Грэхем Р., Кнут Д., Паташник О. Конкретная математика. Основание информатики. М.: Мир, 1998, гл. 6, 8.

У1.10. Пусть случайная величина M имеет производящую функцию $G(z)$. Выразите производящую функцию величины $N = 2M$.

У1.11. Пусть случайные величины M и N имеют производящие функции распределения вероятности $F(z)$ и $G(z)$ и $0 < p < 1$. Является ли $pF(z) + (1-p)G(z)$ производящей функцией распределения вероятности какой-либо случайной величины? Как можно описать словами соответствующее случайное испытание? Ответьте на те же вопросы для функции $F(G(z))$.

У1.12. «Прореживание» (thinning) по Реньи. Пусть N — целочисленная случайная величина и $0 \leqslant \alpha \leqslant 1$. Образуем случайную величину $T_\alpha N$, сложением N независимых случайных слагаемых, каждое из которых равно 1 с вероятностью α и 0 с вероятностью $1-\alpha$. Выразите производящую функцию распределения вероятности $T_\alpha N$ через производящую функцию $G(z)$ случайной величины N. Как прореживание действует на факториальные моменты? Как изменяются при прореживании биномиальное, отрицательное биномиальное, геометрическое распределения, распределение Пуассона?

При доказательстве асимптотических теорем для непрерывных распределений вероятности большую роль играет операция масштабного преобразования. «Прореживание» — это ее аналог, сохраняющий дискретный характер случайной величины.

У1.13. Вырождение ветвящегося процесса. Пусть число потомков в одном поколении ветвящегося процесса — случайная величина с производящей функцией распределения вероятности $G(z)$. Найти предел вероятности, что в k-м поколении число потомков будет равно нулю, при $k \to \infty$. Каково необходимое и достаточное условие того, что этот предел равен единице (т.е. в пределе наступает вырождение)?

У1.14. Красавица и чудовище. В лесах дремучих стоит дом не дом, чертог не чертог, а дворец зверя лесного, чуда морского, весь в огне, в серебре и золоте и каменьях самоцветных. Красная девица входит на широкий двор, в ворота растворенные, и находит там три двери, а за ними три горницы красоты несказанной, а в каждой из тех горниц еще по три двери, ведущие в горницы краше прежних.

Походив по горницам, красная девица начинает догадываться, что дворец построен ярусами: двери со двора ведут в горницы первого яруса, из тех — в горницы второго яруса, и так далее. В каждой горнице есть вход и три выхода, ведущие в три горницы следующего яруса. В горницах последнего, n-го яруса растворены окна широкие во сады диковинные, плодовитые, а в садах птицы поют и цветы растут.

Вернувшись на широкий двор и отдохнув, красная девица видит, что произошла перемена: ворота, через которые она вошла, и часть дверей внутри дворца сами собой затворились, да не просто так, а каждая дверь с вероятностью $\frac{1}{3}$ независимо от других. Немного обеспокоенная, красная девица начинает метаться из горницы в горницу сквозь оставшиеся незатворенными двери в поисках выхода. Покажите, что при больших n вероятность, что она сможет добраться до окон, растворенных в сады, близка к $(9-\sqrt{27})/4 \approx 95\%$.

У1.15. Случайная перестановка. На корабле служит k матросов. После увольнительной они возвращаются навеселе и занимают койки в кубрике как придется. Принимая, что все перестановки матросов по койкам равновероятны, найти вероятность того, что наутро ни один из них не проснется в своей койке. Какова асимптотика этой вероятности при $k \to \infty$? Показать, что такое же предельное значение имеет и вероятность отсутствия неподвижных точек в произвольном отображении k-элементного множества в себя (без условия взаимной однозначности).

У2. Упражнения к лекции 2

У2.1. Пусть X — скалярная случайная величина. При каких значениях m и μ обращаются в минимум $\mathsf{E}|X-m|$ и $\mathsf{E}(|X-\mu|^2)$? Считается, что необходимые интегралы сходятся, т. е. математические ожидания $\mathsf{E}|X|$ и (во второй части упражнения) $\mathsf{E}X^2$ конечны.

У2.2. Пусть X — случайная величина, распределение которой сосредоточено на полуоси $x > 0$. Предполагая, что момент μ_k порядка $k \geqslant 1$ существует, найти интегральную формулу, выражающую его через кумулятивную функцию распределения $F_X(x)$, не прибегая к дифференцированию.

У2.3. Получить формулу для кумулятивной функции распределения условной вероятности $F_X(x \mid Y = y)$, если в точке y имеется атом вероятности маргинального распределения F_Y: $\mathsf{P}(Y = y) = p_0 > 0$ (ср. п. 2.24).

У2.4. Проверить формулы п. 2.25. Можно ли получить для $p_X(x \mid X = Y)$ еще какие-либо другие ответы?

У2.5. Пусть X, Y — две независимые случайные величины, распределенные по одному закону с кумулятивной функцией F с конечным математическим ожиданием. Покажите, что средняя абсолютная разность

$$\Delta_1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x-y| \, \mathrm{d}F(x)\mathrm{d}F(y)$$

может быть выражена как

$$\Delta_1 = 2\int_{-\infty}^{+\infty} F(x)\,[1-F(x)]\,\mathrm{d}x.$$

У2.6. Пусть в условиях предыдущего упражнения распределение F обладает дисперсией σ^2. Найдите величину отношения

$$\frac{1}{\sigma^2}\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x-y|^2 \, \mathrm{d}F(x)\mathrm{d}F(y).$$

У2.7. Найдите математическое ожидание и дисперсию для гамма-распределения, плотность которого имеет вид

$$p(x \mid \alpha, \nu) = \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} \mathrm{e}^{-\alpha x} \text{ при } x > 0.$$

Интегральное представление гамма-функции Эйлера имеет вид $\Gamma(\nu) = \int_0^{+\infty} \xi^{\nu-1}\mathrm{e}^{-\xi}\,\mathrm{d}\xi$, откуда следует известное тождество $\Gamma(\nu+1) = \nu\Gamma(\nu)$.

У2.8. Покажите, что замену переменной $y = f(x)$ в распределении вероятности с плотностью $p_X(x)$ можно сделать по формуле

$$p_Y(y) = \int \delta(y - f(x)) p_X(x)\,\mathrm{d}x,$$

где δ — дельта-функция Дирака.

У2.9. Пусть случайная величина X распределена нормально:

$$p_X(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2}.$$

Найдите функцию плотности вероятности для случайной величины $Y = X^2$.

У2.10. Проделайте то же вычисление для распределения Коши $p_X(x) = \frac{1}{\pi}(1+x^2)^{-1}$, если $Y = 1/X$. Каково может быть простое геометрическое объяснение полученного результата?

У2.11. Случайные величины X, Y обладают совместным распределением с плотностью $p_{X,Y}(x, y)$. Выразите $p_{X+Y}(t)$ и $p_{X/Y}(t)$.

У2.12. Случайные величины X, Y независимы и одинаково распределены с плотностью $p(x)$. Найдите плотность вероятности: (a) суммы $X + Y$, если $p(x)$ — плотность равномерного распределения на $(-\frac{1}{2}, \frac{1}{2})$; (b) отношения X/Y, если $p(x)$ — плотность нормального распределения У2.9.

У2.13. Найдите функцию плотности вероятности суммы двух независимых случайных величин, обладающих гамма-распределением У2.7 с одинаковым масштабным параметром α, но различными ν', ν''.

Напомним известное тождество для бета-функции Эйлера: $\int_0^1 \xi^{\mu-1}(1-\xi)^{\nu-1}\,\mathrm{d}\xi = \frac{\Gamma(\mu)\Gamma(\nu)}{\Gamma(\mu+\nu)}$.

У2.14. Пусть X, Y_1, …, Y_k — независимые случайные величины, распределенные по нормальному распределению У2.9. Найдите функции плотности распределения: (a) суммы $V = Y_1^2 + \ldots + Y_k^2$ (распределение χ^2 с k степенями свободы); (b) отношения $X/\sqrt{V}$ (распределение Стьюдента с k степенями свободы).

У2.15. Проблема моментов. Найдите моменты всех порядков случайной величины, заданной на положительной полуоси функцией плотности вероятности

$$p(x \mid a) = \mathrm{const}\, \frac{\mathrm{e}^{-\frac{1}{2}(\ln x)^2}}{x} \bigl[1 + a \sin(2\pi \ln x)\bigr],$$

где $|a| < 1$. Зависят ли они от параметра a?

У2.16. Оценки Фреше–Хёфдинга. Покажите, что для любого совместного распределения случайных величин X, Y выполнены неравенства

$$F_{X,Y}(x, y) \geqslant \max[0, F_X(x) + F_Y(y) - 1],$$
$$F_{X,Y}(x, y) \leqslant \min[F_X(x), F_Y(y)],$$

где $F_X(x)$, $F_Y(y)$ — кумулятивные функции маргинальных распределений.

У2.17. Пусть $F_X(x)$, $F_Y(y)$ — кумулятивные функции распределения и

$$C_\theta(u, v) = uv[1 - \theta(1-u)(1-v)].$$

При каких значениях θ функция $C_\theta(F_X(x), F_Y(y))$ есть кумулятивная функция распределения? Каким общим условиям должна удовлетворять функция $C(u, v)$, чтобы при аналогичной подстановке получалась корректно определенная кумулятивная функция распределения?

Такая функция C называется **копулой** (copula). Копула выражает ту информацию о распределении случайных величин X, Y, которая не связана с конкретным видом маргинальных распределений, т. е. информацию об их взаимной зависимости.

У2.18. При $\nu = 1$ гамма-распределение переходит в показательное распределение $p(x) = \alpha \mathrm{e}^{-\alpha x}$. Найдите соответствующую кумулятивную функцию распределения $F(x)$ и покажите, что

$$1 - \frac{1 - F(x)}{1 - F(x_0)}$$

при $x \geqslant x_0 > 0$ можно понимать как кумулятивную функцию распределения некоторой случайной величины. Как можно описать словами соответствующее случайное испытание?

У2.19. Пусть F, G — кумулятивные функции распределения скалярных случайных величин, $0 < p < 1$, $\lambda > 0$. Проверьте, что выражения

$$pF(x) + (1-p)G(x), \quad F(x) \cdot G(x),$$
$$\frac{pF(x)}{1-(1-p)F(x)}, \quad F(x)\,\mathrm{e}^{\lambda(F(x)-1)}$$

задают кумулятивные функции распределений некоторых случайных величин. Как можно описать словами соответствующие случайные испытания?

У2.20. Случайное испытание совершается следующим образом: сначала разыгрывается случайная величина Λ, распределенная по закону У2.7 с параметрами α, m (где m — положительное целое число), а затем разыгрывается случайная величина, распределенная по Пуассону с параметром Λ. Найдите распределение вероятности получаемой целочисленной случайной величины.

У2.21. Случайное испытание совершается следующим образом: сначала разыгрывается случайная величина D, распределенная с плотностью

$$p_D(x) = \frac{\lambda}{\sqrt{2\pi x^3}}\mathrm{e}^{-\frac{\lambda^2}{2x}}, \quad x > 0,$$

а затем — гауссова случайная величина (У2.9) с дисперсией D. Найдите функцию плотности вероятности получаемой случайной величины.

У3. Упражнения к лекции 3

У3.1. Что можно сказать о распределении вероятности, характеристическая функция которого периодична с периодом S?

У3.2. Пусть φ — характеристическая функция. Будут ли φ^*, $|\varphi|^2$, φ^2, $\operatorname{Re}\varphi$ характеристическими функциями каких-либо распределений вероятности?

У3.3. Пусть φ — характеристическая функция некоторой случайной величины. Являются ли характеристическими функциями

$$\int_0^{+\infty} \varphi(su)\mathrm{e}^{-u}\,\mathrm{d}u, \quad \frac{1}{2-\varphi(s)}\ ?$$

Как можно описать соответствующие случайные испытания?

У3.4. Составное распределение Пуассона. Случайное испытание совершается следующим образом: сначала разыгрывается целочисленная случайная величина N, распределенная по Пуассону с параметром λ, а затем находят сумму N независимых одинаково распределенных случайных величин с характеристической функцией φ. Найти характеристическую функцию полученной суммы.

У3.5. Пусть независимым случайным величинам X, Y соответствуют кумулятивные функции распределения F и G и характеристические функции φ и γ. Покажите, что

$$\varphi_{XY}(s) = \int \varphi(sy)\,\mathrm{d}G(y) = \int \gamma(sx)\,\mathrm{d}F(x)$$

и получите аналогичную формулу для $\varphi_{X/Y}$.

У3.6. Найдите характеристические функции: (a) равномерного распределения на отрезке $(-\frac{1}{2}, \frac{1}{2})$ и на отрезке (a, b), $a < b$; (b) распределения Коши $p(x) = \frac{1}{\pi(1+x^2)}$; (c) показательного распределения $p(x) = \alpha\mathrm{e}^{-\alpha x}$; (d) гамма-распределения

$$p(x) = \frac{\alpha^\nu}{\Gamma(\nu)}x^{\nu-1}\mathrm{e}^{-\alpha x};$$

(e) нормального распределения

$$p(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2};$$

(f) распределения Стьюдента (см. У2.14).

У3.7. Являются ли следующие функции характеристическими функциями каких-либо распределений вероятности?

$$\varphi_1(s) = \max(1 - |s|, 0);$$
$$\varphi_2(s) = \max(1 - |s|, \tfrac{1}{2} - \tfrac{1}{4}|s|, 0);$$
$$\varphi_3(s) = \mathrm{e}^{-|s|^{1/2}}; \quad \varphi_4(s) = \mathrm{e}^{-s^4}.$$

У3.8. Распределение вероятности случайной величины X называется **безгранично делимым**, если для любого $n \geqslant 1$ найдутся такие независимые одинаково распределенные случайные величины $Y_1^{(n)}, \dots, Y_n^{(n)}$, что $X = Y_1^{(n)} + \dots + Y_n^{(n)}$. Покажите, что гамма-распределение, нормальное распределение, распределение Пуассона безгранично делимы. Безгранично делимое распределение вероятности называется **устойчивым**, если случайные величины $Y_i^{(n)}$ получаются из X линейным преобразованием. Какие из перечисленных распределений устойчивы?

У3.9. Пусть X, Y — две одинаково распределенных независимых случайных величины с нулевым математическим ожиданием, единичной дисперсией и характеристической функцией φ. Проверьте, что если случайные величины $X + Y$ и $X - Y$ независимы, то $\varphi(2s) = \varphi^3(s) \cdot \varphi(-s)$. Выведите отсюда, что $\varphi(s) = \mathrm{e}^{-\frac{1}{2}s^2}$.

Это характеристическое свойство нормального распределения применяют для того, чтобы определить аналог нормального распределения на произвольной коммутативной группе — т. н. **нормальное распределение в смысле Бернштейна**.

У3.10. Найдите распределение вероятности суммы двух независимых величин, обладающих плотностями вида

$$p(x \mid \lambda) = \mathrm{e}^{-\lambda^2/2x} \lambda / \sqrt{2\pi x^3}, \qquad x \geqslant 0$$

с различными значениями λ', λ''.

У4. Многомерное нормальное распределение

В упражнениях этого параграфа **выборка** $X_1, X_2, \dots, X_n$ понимается как совокупность независимых случайных величин, распределенных по одному и тому же закону.

У4.1. Характеристическая функция нормальной случайной величины Y есть $\varphi(s) = \mathrm{e}^{-\frac{1}{2}s^2}$. Общая гауссова скалярная случайная величина с математическим ожиданием μ и стандартным отклонением σ определяется как $X = \mu + \sigma Y$. Найти выражения для характеристической функции $\varphi_X(s)$ и плотности вероятности $p_X(x)$.

У4.2. Получить выражение для функций плотности вероятности длины вектора, состоящего из n независимых нормальных случайных величин, и квадрата его длины. Как нормировочная константа в этих формулах связана с площадью единичной $(n-1)$-мерной сферы? Как будет распределен квадрат длины вектора, если на его компоненты дополнительно наложено p однородных линейных условий $\sum_i a_i X_i = 0$?

Закон распределения квадрата длины свободного вектора с независимыми нормальными компонентами называется **χ^2-распределением с n степенями свободы**. Каждое дополнительное ограничение уменьшает число степеней свободы на единицу.

У4.3. Предположим, что абсолютно непрерывное распределение вероятности скалярной случайной величины X обладает следующим свойством: плотность совместного распределения набора n независимых случайных величин, распределенных по такому же закону, зависит только от радиальной координаты $\sqrt{\sum_i x_i^2}$. Показать, что распределение X нормально (точнее, гауссово с нулевым математическим ожиданием).

У4.4. Предполагая все необходимые моменты конечными, найти математические ожидания **среднего выборки** и **дисперсии выборки**

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n), \qquad S^2 = \frac{1}{n} \sum_{1 \leqslant i \leqslant n} (X_i - \overline{X})^2$$

Выразить дисперсии этих величин через центральные моменты или кумулянты случайных величин X_i и показать, что с ростом n эти дисперсии убывают как $1/n$.

У4.5. Пусть выборка состоит из гауссовых величин. Показать, что среднее выборки $\overline{X}$ и дисперсия выборки S^2 являются независимыми случайными величинами, и найти законы их распределения.

У4.6. Пусть математическое ожидание и дисперсия каждой из величин X_i равны μ и σ^2 соответственно. Показать, что если $\overline{X}$ и S^2 независимы, то распределение случайных величин X_i с необходимостью является гауссовым.

Указание: характеристическая функция совместного распределения имеет вид

$$\varphi(u, v) = \int \mathrm{e}^{\mathrm{i}u\overline{x} + \mathrm{i}v\, s^2} \mathrm{d}F(x_1) \ldots \mathrm{d}F(x_n),$$

где $\overline{x} = \frac{1}{n}(x_1 + \cdots + x_n)$, $s^2 = \frac{1}{n} \sum_{1 \leqslant i \leqslant n} (x_i - \overline{x})^2$. Вычислите $\partial\varphi(u, v)/\partial v|_{v=0}$ двумя способами: дифференцируя под интегралом и пользуясь тем, что по условию $\varphi(u, v) = \varphi_1(u)\, \varphi_2(v)$, где $\varphi_1(u) = (\varphi(u/n))^n$.

У4.7. В условиях упражнения У4.5 найти функцию плотности вероятности для отношения $Z = \overline{X}/\sqrt{S^2}$ (ср. п. (b) упражнения У2.14).

Напомним известные соотношения для эйлеровых интегралов:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} \mathrm{d}x = \left[x = \frac{y}{1+y}\right] = \int_0^{+\infty} \frac{y^{\alpha-1} \mathrm{d}y}{(1+y)^{\alpha+\beta}} = \frac{\Gamma(\alpha)\, \Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Случайная величина $T = Z\sqrt{n-1}$ называется **t-статистикой Стьюдента** (У. Госсет, 1908). Ее закон распределения при $n = 1$ совпадает с распределением Коши, а при $n \to \infty$ стремится к нормальному.

У4.8. Пусть $X'_1, X'_2, \ldots, X'_{n'}$ и $X''_1, X''_2, \ldots, X''_{n''}$ — две независимых выборки из нормального распределения. Получить функцию плотности вероятности величины (**F-статистика Фишера**)

$$F = \frac{\frac{1}{n'-1} \sum_{i'} (X'_{i'} - \overline{X'})^2}{\frac{1}{n''-1} \sum_{i''} (X''_{i''} - \overline{X''})^2}.$$

У4.9. Общий гауссов случайный вектор $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ — это векторная случайная величина, все скалярные проекции которой $\boldsymbol{Y} \cdot \boldsymbol{s}$ суть гауссовы скалярные случайные величины (п. 3.27). Принимая, что $\mathsf{E}\boldsymbol{Y} = \boldsymbol{\mu}$ и $\mathsf{E}Y_k Y_l - \mathsf{E}Y_l\, \mathsf{E}Y_l = \Gamma_{kl}$, найти выражения для характеристической функции $\varphi_{\boldsymbol{Y}}(\boldsymbol{s})$ и плотности вероятности $p_{\boldsymbol{Y}}(\boldsymbol{x})$ гауссова случайного вектора, если **матрица ковариации** Γ не имеет нулевых собственных значений. Может ли матрица ковариации иметь отрицательные собственные значения?

У4.10. **Коэффициентом корреляции** компонент Y_k и Y_l случайного вектора, обладающего матрицей ковариации Γ, называется $\rho_{kl} = \Gamma_{kl}/\sqrt{\Gamma_{kk}\, \Gamma_{ll}}$. Показать, что $|\rho_{kl}| \leqslant 1$. Какой вид имеет распределение гауссова случайного вектора, если для некоторой пары его компонент $\rho_{kl} = \pm 1$?

У4.11. Записать функцию плотности вероятности двухкомпонентного гауссова случайного вектора, если $\mu = \mathbf{0}$ и заданы дисперсии $\mathsf{D}Y_1 = \sigma_1^2$, $\mathsf{D}Y_2 = \sigma_2^2$ и коэффициент корреляции ρ. Получить выражение для условной плотности вероятности $p_{Y_2}(x_2 \mid Y_1 = x_1)$.

У4.12. В условиях предыдущего упражнения найти условное математическое ожидание $\mathsf{E}_{y_1} Y_2$ и вычислить его дисперсию, подставляя вместо x_1 случайную величину Y_1. Найти условную дисперсию $\mathsf{D}_{y_1} Y_2$ и вычислить ее математическое ожидание как функции случайной величины Y_1. Проверить, что $\mathsf{E}(\mathsf{D}_{Y_1} Y_2) + \mathsf{D}(\mathsf{E}_{Y_1} Y_2) = \mathsf{D}Y_2$. Выполнено ли это тождество в общем случае?

Конструкция случайной величины как условного математического ожидания, предложенная в этом упражнении, играет важную роль в теории случайных процессов.

У4.13. Пусть $p_{X,Y}(x,y)$ — дважды дифференцируемая функция плотности вероятности совместного распределения. Покажите, что линейное преобразование, переводящее X, Y в пару независимых случайных величин, существует тогда и только тогда, когда для некоторых констант A, B, C

$$\Big(A\frac{\partial^2}{\partial x^2} + 2B\frac{\partial^2}{\partial x\,\partial y} + C\frac{\partial^2}{\partial y^2}\Big)\log p_{X,Y} = 0.$$

У4.14. Под **k-й порядковой статистикой** понимается случайная величина $X_{(k)}$, которая равна k-му по счету значению в векторе, получаемом сортировкой $(X_1, X_2, \ldots, X_n)$ в порядке возрастания (в частности, $X_{(1)} = \min_i X_i$, $X_{(n)} = \max_i X_i$). Получить выражение для функций плотности вероятности: (a) распределения k-й порядковой статистики; (b) совместного распределения k-й и l-й порядковых статистик ($k < l$).

У4.15. Медианой выборки $X_1, X_2, \ldots, X_n$ называется $X_{(\frac{n+1}{2})}$ при нечетном n или среднее арифметическое $X_{(\frac{n}{2})}$ и $X_{(\frac{n}{2}+1)}$ при четном n. Каким условиям должно удовлетворять распределение случайных величин X_i, чтобы дисперсия медианы выборки с ростом n убывала как $1/n$? Пусть $\mathsf{D}X_i = \sigma^2$; что больше — дисперсия среднего или медианы выборки?

У4.16. Квартилями выборки называются порядковые статистики, номера которых равны $n/4$ и $3n/4$, округленным до ближайшего целого. Показать, что дисперсия разности квартилей выборки при больших n убывает как $1/n$.

П1. Приложение 1. «Зоопарк» распределений вероятности

В каждом из пунктов данного раздела символы M, M_i или X, X_i обозначают случайную величину или набор величин, определяемых в данном пункте, m, m_i, x x_i — их значения. Значения целочисленных величин везде неотрицательны. Производящая функция моментов: $\Phi(u) = G(\mathrm{e}^u)$. Символы Γ_{ij}, S и K обозначают коэффициенты ковариации, асимметрию и эксцесс.

П1.1. Биномиальное распределение. Параметры $0 < p, \bar{p} < 1$, $p + \bar{p} = 1$, $n \geqslant 1$:

$$\mathsf{P}(m) = C_n^m p^m \bar{p}^{n-m}, \quad G(z) = (\bar{p} + pz)^n;$$
$$\mathsf{E}M = np, \quad \mathsf{D}M = np\bar{p},$$
$$\mathsf{S} = (\bar{p} - p)/\sqrt{np\bar{p}}, \quad \mathsf{K} = (1 - p\bar{p})/np\bar{p}.$$

Вывод см. в УУ1.4. Здесь $C_n^m = \frac{n!}{m!\,(n-m)!}$ при $0 \leqslant m \leqslant n$, иначе C_n^m=0.

П1.2. Мультиномиальное распределение. Параметры $p_1, \dots, p_k$ ($0 < p_i < 1$, $p_1 + \dots + p_k = 1$) и $n \geqslant 1$:

$$\mathsf{P}(m_1, \dots, m_k) = \frac{n!}{m_1! \dots m_k!} \, p_1^{m_1} \dots p_k^{m_k},$$
$$G(z_1, \dots, z_k) = \Big(\sum_{1 \leqslant i \leqslant k} z_i p_i \Big)^n;$$
$$\mathsf{E} M_i = np_i, \quad \mathsf{E} M_i M_j = (n^2 - n) p_i p_j,$$
$$\mathsf{D} M_i = \Gamma_{ii} = np_i(1 - p_i),$$
$$\Gamma_{ij} = -np_i p_j \text{ при } i \neq j.$$

Мультиномиальное распределение обобщает биномиальное на случай более двух исходов. Отрицательность коэффициентов ковариации объясняется тем, что сумма случайных величин $M_1, \dots, M_k$ фиксирована и равна n. Мультиномиальное распределение возникает, например, при выводе критерия χ^2.

П1.3. Распределение Пуассона (УУ1.7). Параметр $\lambda > 0$:

$$\mathsf{P}(m) = \frac{\lambda^m}{m!} \mathrm{e}^{-\lambda}, \quad G(z) = \mathrm{e}^{\lambda(z-1)};$$
$$\mathsf{E} M = \mathsf{D} M = \lambda, \quad \mathsf{S} = 1/\sqrt{\lambda}, \quad \mathsf{K} = 1/\lambda.$$

П1.4. Геометрическое распределение (УУ1.8). Параметр $0 < p < 1$:

$$\mathsf{P}(m) = p\bar{p}^{m-1} \ (m \geqslant 1), \quad G(z) = \frac{pz}{1 - \bar{p}z};$$
$$\mathsf{E} M = 1/p, \quad \mathsf{D} M = \bar{p}/p^2,$$
$$\mathsf{S} = (1 + \bar{p})/\sqrt{\bar{p}}, \quad \mathsf{K} = (1 + 4\bar{p} + \bar{p}^2)/\bar{p}.$$

Название связано с тем, что вероятности образуют геометрическую прогрессию.

П1.5. Отрицательное биномиальное распределение (УУ1.8). Параметры $0 < p < 1$, $k > 0$:

$$\mathsf{P}(m) = C_{m-1}^{k-1} p^k \bar{p}^{m-k} \text{ при } m \geqslant k,$$
$$\text{иначе } \mathsf{P}(m) = 0, \quad G(z) = \Big(\frac{pz}{1 - \bar{p}z} \Big)^k;$$
$$\mathsf{E} M = m/p, \mathsf{D} M = m\bar{p}/p^2,$$
$$\mathsf{S} = (1 + \bar{p})/\sqrt{m\bar{p}}, \quad \mathsf{K} = (1 + 4\bar{p} + \bar{p}^2)/(m\bar{p}).$$

Геометрическое распределение п. П1.4 является частным случаем отрицательного биномиального при $k = 1$.

П1.6. Равномерное распределение на отрезке $(\mu - \Gamma, \mu + \Gamma)$, $\Gamma > 0$ (см. п. 2.5):

$$p(x) = 1/2\Gamma \text{ при } |x - \mu| < \Gamma,$$
$$p(x) = 0, \quad \varphi(s) = \mathrm{e}^{\mathrm{i}\mu s} \frac{\sin \Gamma s}{\Gamma s};$$
$$\mathsf{E} X = \mu, \quad \mathsf{D} X = \Gamma^2/3,$$
$$\mathsf{S} = 0, \quad \mathsf{K} = -1\frac{1}{5}.$$

П1.7. Треугольное распределение. Параметры μ, $\Gamma > 0$:

$$p(x) = \max\Big(\frac{1}{\Gamma} - \frac{|x-\mu|}{\Gamma^2}, 0\Big), \quad \varphi(s) = \frac{\sin^2(\frac{\Gamma s}{2})}{(\frac{\Gamma s}{2})^2};$$
$$\mathsf{E}X = \mu, \quad \mathsf{D}X = \Gamma^2/6, \quad \mathsf{S} = 0, \quad \mathsf{K} = -\frac{3}{5}.$$

Характеристическая функция треугольного распределения неотрицательна. Поэтому $\Gamma \cdot p(x)$ является характеристической функцией по отношению к $\Gamma \cdot \varphi(s)$, рассматриваемой как функция плотности вероятности.

П1.8. Распределение Коши (Брейта–Вигнера, Лоренца). Параметры μ и $\Gamma > 0$:

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

Моменты всех порядков этого распределения выражаются расходящимися интегралами: в частности, его математическое ожидание не определено, а дисперсия бесконечна. Параметр μ является медианой и модой распределения Коши.

П1.9. Показательное (экспоненциальное) распределение. Параметр $\lambda > 0$:

$$p(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x > 0, \\ 0, & x < 0, \end{cases} \quad \varphi(s) = \frac{1}{1 - \mathrm{i}s/\lambda};$$
$$\mathsf{E}X = 1/\lambda, \quad \mathsf{D}X = 1/\lambda^2, \quad \mathsf{S} = 2, \quad \mathsf{K} = 6.$$

Показательное распределение можно рассматривать как «непрерывный аналог» геометрического распределения.

П1.10. Распределение $\boldsymbol{\chi}^2$. Параметр $n \geqslant 1$ (число степеней свободы):

$$p(x) = \frac{(x/2)^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}}{2\Gamma(n/2)}, \quad \varphi(s) = \frac{1}{(1-2\mathrm{i}s)^{\frac{n}{2}}},$$
$$\mathsf{E}X = n, \ \mathsf{D}X = 2n, \ \mathsf{S} = 2\sqrt{2/n}, \ \mathsf{K} = 12/n.$$

П1.11. Гамма-распределение. Параметры $\lambda > 0$, $\nu > 0$, распределение сосредоточено на положительной полуоси:

$$p(x) = \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1}\mathrm{e}^{-\lambda x}, \quad \varphi(s) = \frac{1}{(1 - \mathrm{i}s/\lambda)^\nu};$$
$$\mathsf{E}X = \nu/\lambda, \ \mathsf{D}X = \nu/\lambda^2, \ \mathsf{S} = 2/\sqrt{\nu}, \ \mathsf{K} = 6/\nu.$$

Гамма-распределение можно рассматривать как «непрерывный аналог» отрицательного биномиального распределения. Распределение χ^2 с n степенями свободы является частным случаем гамма-распределения ($\lambda = \frac{1}{2}$, $\nu = \frac{n}{2}$).

П1.12. Бета-распределение. Параметры $\alpha > 0$, $\beta > 0$, распределение сосредоточено на отрезке $(0, 1)$:

$$p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\beta(\alpha,\beta)};$$
$$\mathsf{E}X = \frac{\alpha}{\alpha+\beta}, \ \mathsf{D}X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Выражения для пологости и эксцесса имеют довольно громоздкий вид. В частности, при $\alpha = k+1$, $\beta = n+1$ коэффициент имеет вид $(n+k+1)!/k!\,n! = (n+k+1)C_n^k$.

П1.13. Нормальное распределение (распределение Гаусса). Параметры μ и σ^2:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad \varphi(s) = \mathrm{e}^{\mathrm{i}\mu s - \frac{\sigma^2 s^2}{2}};$$
$$\mathsf{E}X = \mu, \quad \mathsf{D}X = \sigma^2, \quad \mathsf{S} = 0, \quad \mathsf{K} = 0.$$

П1.14. Многомерное нормальное распределение. Параметры: n компонентный вектор $\boldsymbol{m} = (m_i)$ и положительно определенная матрица Γ размера $n \times n$:

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

Здесь $\hat{\Gamma}$ — матрица, обратная к матрице Γ. Кумулянты порядков 3 и выше 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.

П1.15. Логнормальное распределение. Параметры μ, σ:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma\, x} \mathrm{e}^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} \text{ при } x > 0;$$
$$\mathsf{E}X = \mathrm{e}^{\mu + \frac{1}{2}\sigma^2}, \quad \mathsf{D}X = \mathrm{e}^{2\mu}\xi(\xi - 1),$$
$$\mathsf{S} = \sqrt{\xi - 1}(\xi + 2),$$
$$\mathsf{K} = (\xi - 1)(\xi^3 + 2\xi^2 + 6\xi + 6), \text{ где } \xi = \mathrm{e}^{\sigma^2}.$$

П1.16. Распределение Фреше. Параметры $\alpha > 0, \mu, \sigma > 0$:

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

П1.17. Распределение Вейбулла. Параметры $\alpha > 0, \mu, \sigma > 0$:

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

П1.18. Распределение Гумбеля. Параметры $\mu, \sigma > 0$:

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

П1.19. Распределение Стьюдента. Параметр $n \geqslant 1$ (число степеней свободы):

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

$$\mathsf{E}X = 0 \text{ при } n > 1, \quad \mathsf{D}X = \frac{n}{n-2} \text{ при } n > 2,$$

$$\mathsf{S} = 0 \text{при } n > 3, \quad \mathsf{K} = \frac{6}{n-4} \text{ при } n > 4.$$

П1.20. Распределение времени выхода. Параметр $\Gamma > 0$:

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

Моменты всех порядков расходятся.