конкретная теория вероятностей

А. Н. Соболевский

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 бóльшую 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.

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 $n \neq n_0$ (почему?) 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 $\lim_{m,n\to\infty} |A \cap \{-m+1, -m+2, \ldots, n\}|/(m+n)$ (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 $\{0\}, \{2\}, \{4\}, \ldots$, 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

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

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, обозначение $\overline{N}$ 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

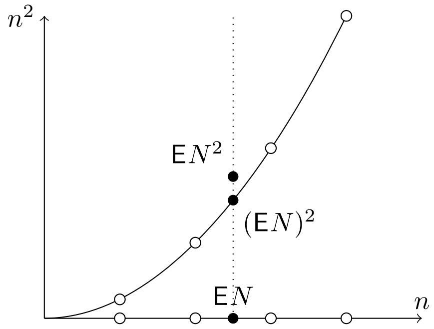
$$\mathsf{D}N = \mathsf{E}N^2 - (\mathsf{E}N)^2,$$

которая получается следующим образом:

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

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.

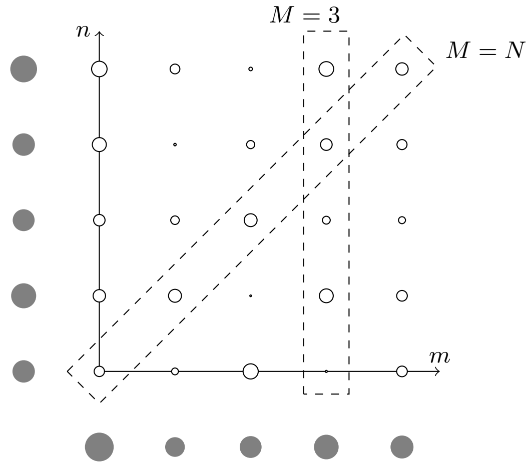
1.14. Для совместного распределения $p_{M,N}$ определяются **маргинальные распределения вероятности**, задаваемые формулами

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

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 $p_N(n) > 0$.

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. **??**: 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,

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

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: $\left.\dfrac{\mathrm{d}^k\Psi_N}{\mathrm{d}s^k}\right|_{s=0} = \left.\sum_{n\geqslant 0} n^k \mathrm{e}^{sn}\, p_N(n)\right|_{s=0} = \mathsf{E}N^k$.

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 использования некоторых простых 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:

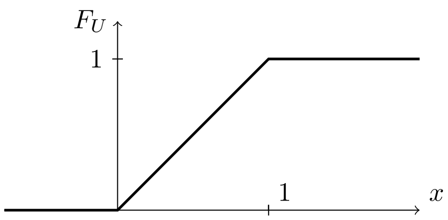
$$\mathsf{P}(X \in [a, b]) = \mathsf{P}(a \leqslant X \leqslant b) = F_X(b) - F_X(a - 0),$$
$$\mathsf{P}(X \in (a, b)) = \mathsf{P}(a < X < b) = F_X(b - 0) - F_X(a) \quad \text{и т. п.},$$

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. **??**).

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

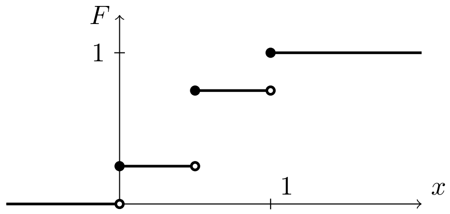
$$F_X(x) = \int_{-\infty}^{x} p_X(\xi)\,\mathrm{d}\xi,$$

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 $x = 1$:

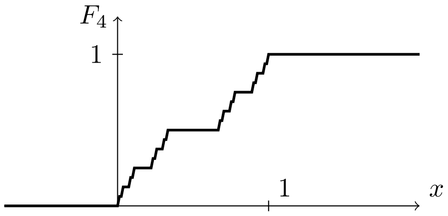


Распределения целочисленных случайных величин являются чисто точечными.

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 был 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 **??** видно, что медиану и математическое ожидание можно понимать как различные частные случаи одного общего понятия.

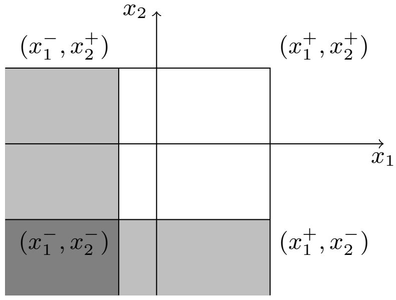
2.18. Конечный набор случайных величин $X_1, X_2, \dots, X_n$, взятых в совокупности, образует **случайный вектор**. Векторные случайные величины обозначаются прописными полужирными латинскими буквами $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \dots$, а их значения (**исходы**) — соответствующими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

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

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

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

Первое из этих свойств выражает положительность вероятности: $\mathsf{P}(x_1^- < X_1 \leqslant x_1^+, \dots, x_n^- < X_n \leqslant x_n^+) \geqslant 0$ 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.

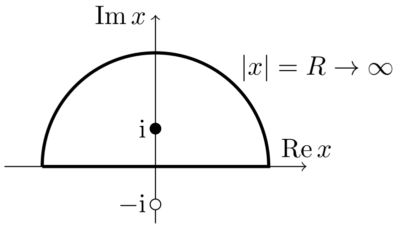
3.2. Производные характеристической функции имеют вероятностный смысл:

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

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

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

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:

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

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

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 **??**)

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

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 **??**).

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 **??**).

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 когда 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, усреднение выборки случайных величин, распределенных по Коши, 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. **??** 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. **??**) 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. **??** и последнее выражение в упр. **??**) 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

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

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,

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

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.

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 У**??**. 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 $p_1, \ldots, p_k$ $(0 < p_i < 1,\ p_1 + \cdots + p_k = 1)$ и $n \geqslant 1$:

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

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

П1.3. Распределение Пуассона (У**??**). 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 (У**??**). 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 (У**??**). 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.