

# Конспект по теории 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:

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

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

**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  $n_0$ .

С использованием только что введенного рабочего определения случайной величины и его аналога для непрерывных числовых интервалов (лекция 2) можно получить все асимптотические результаты, доказываемые в первой части курса. Однако для изучения случайных процессов потребуется более абстрактное и мощное определение, которым обычно и пользуются в современной теории вероятностей.

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

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

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  $P(N \in A)$  можно писать  $P(A)$ . Иногда вместо обозначения множества  $A$  будем записывать определяющий его предикат, заменяя «немую» переменную на обозначение случайной величины: например,  $P(N \text{ четное})$ .

**1.5. Пример.** Если  $N$  представляет собой число очков, выпавших на игральной кости, то  $p_N(0) = p_N(7) = p_N(8) = \dots = 0$ . Если, более того, эта кость симметрична, то  $p_N(1) = p_N(2) = \dots = p_N(6) = \frac{1}{6}$ . 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  $\lim_{m,n \rightarrow \infty} |A \cap \{-m+1, -m+2, \dots, n\}|/(m+n)$  (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**  $EN$  случайной величины  $N$  — это сумма ряда

$$EN = \sum_{n \geq 0} n p_N(n),$$

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

$$Ef(N) = \sum_{n \geq 0} f(n) p(n).$$

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

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

$$E[f(N) + g(N)] = Ef(N) + Eg(N), \quad E(\alpha f(N)) = \alpha Ef(N).$$

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

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

Если из контекста ясно, о какой случайной величине идет речь, будем писать  $EN^k = \mu_k$  и  $E(N - EN^k) = \mu_k$ .

**1.10.** Центральный момент второго порядка называется **дисперсией** и обозначается  $DN$ . Дисперсию удобно вычислять по формуле

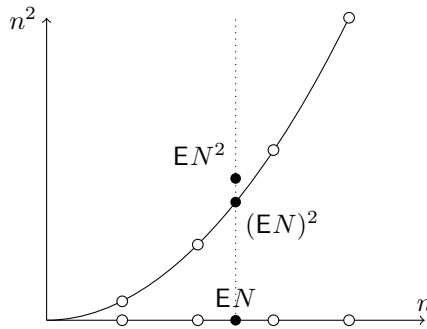
$$DN = EN^2 - (EN)^2,$$

которая получается следующим образом:

$$DN = E[N^2 - 2NEN + (EN)^2] = EN^2 - (EN)^2.$$

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

**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  $EN^2 - (EN)^2 > 0$  имеет следующее геометрическое представление:



**1.13.** Неравенство предыдущего пункта является частным случаем важного **неравенства Иенсена**: если функция  $f$  выпукла, то

$$Ef(N) \geq f(EN).$$

Допустим, что несколько случайных величин надо рассмотреть одновременно. Тогда из них можно образовывать 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.

**1.14.** 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.

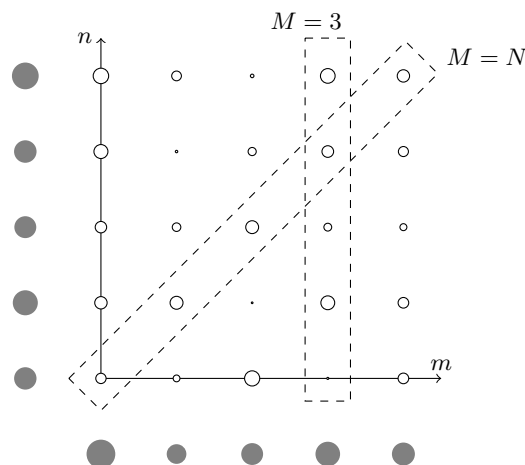
**1.15.** Для совместного распределения  $p_{M,N}$  определяются **маргинальные распределения вероятности**, задаваемые формулами

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

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  $P(M = N) = \sum_{l \geq 0} p_{M,N}(l,l)$ . Следующий рисунок иллюстрирует определения совместного распределения (белые кружки), маргинальных распределений (серые кружки) и событий  $\{M = 3\}$  и  $\{M = N\}$ .



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

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

Конечно, такое определение имеет смысл только при  $p_N(n) > 0$ .

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

$$P(M = m | A) = \frac{P(\{M = m\} \cap A)}{P(A)}.$$

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

**1.17.** 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:

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

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

**1.19.** Несколько случайных величин  $M_1, \dots, M_k$  называются **независимыми в совокупности**, если

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

В последнем определении имеется тонкость: независимость в совокупности набора из трех или большего числа случайных величин — это более сильное свойство, чем независимость каждой пары величин из этого набора. Следующий пример иллюстрирует, как это может быть.

**1.20. Пример.** Каждая из трех случайных величин  $K, M, N$  может принимать значения 0 или 1, причем распределение вероятностей по вершинам булева куба  $\{0, 1\}^3$  устроено таким образом, что в вершине  $(1, 1, 1)$  и в конце каждого координатного вектора (т. е. в вершинах, у которых ровно одна координата равна 1) содержится вероятность  $\frac{1}{4}$ , а в остальных четырех вершинах — нулевая вероятность. Проверьте, что любые две из трех случайных величин  $K, M, N$  независимы, но все три не являются независимыми в совокупности.

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.

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

$$\begin{aligned} E(M + N) &= \sum_{m, n \geq 0} (m + n) p_{M,N}(m, n) = \\ &= \sum_{m, n \geq 0} m p_{M,N}(m, n) + \sum_{m, n \geq 0} n p_{M,N}(m, n) = [E_{M,N} M + E_{M,N} N] \\ &= \sum_{m \geq 0} m \sum_{n \geq 0} p_{M,N}(m, n) + \sum_{n \geq 0} n \sum_{m \geq 0} p_{M,N}(m, n) = \\ &= \sum_{m \geq 0} m p_M(m) + \sum_{n \geq 0} n p_N(n) [E_M M + E_N N]. \end{aligned}$$

**1.22.** Равенство  $EMN = EMEN$  выполнено не всегда. Наиболее важным **достаточным** условием для него является независимость случайных величин:

$$\begin{aligned} EMN &= \sum_{m, n \geq 0} mn p_{MN}(m, n) = [M \perp N] = \sum_{m \geq 0} \sum_{n \geq 0} mn p_M(m) p_N(n) = \\ &= \sum_{m \geq 0} \left( m p_M(m) \sum_{n \geq 0} n p_N(n) \right) = \left( \sum_{n \geq 0} n p_N(n) \right) \sum_{m \geq 0} m p_M(m) = EMEN. \end{aligned}$$

**1.23.** Если случайные величины  $M$  и  $N$  независимы, то  $D(M + N) = DM + DN$ :

$$\begin{aligned} D(M + N) &= E(M + N)^2 - [E(M + N)]^2 = \\ &= E(M^2 + 2MN + N^2) - (EM)^2 - 2EMEN - (EN)^2 = \\ &= EM^2 - (EM)^2 + EN^2 - (EN)^2 + 2(EMN - EMEN) = [M \perp N] = \\ &= EM^2 - (EM)^2 + EN^2 - (EN)^2 = DM + DN. \end{aligned}$$

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.

**1.24. Производящая функция распределения вероятности** случайной величины  $N$ :

$$G_N(z) = \sum_{n \geq 0} z^n p_N(n) = \mathbb{E} z^N.$$

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

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

$$G_N(1) = 1, \quad G'_N(1) = \sum_{n \geq 0} n z^{n-1} p_N(n) \Big|_{z=1} = \sum_{n \geq 0} n p_N(n) = \mathbb{E} N,$$

$$G''_N(1) = \sum_{n \geq 0} n(n-1) z^{n-2} p_N(n) \Big|_{z=1} = \mathbb{E} N(N-1)$$

Вообще,

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

Выражение  $N^{\underline{k}}$  называется  $k$ -й **убывающей факториальной степенью** числа  $N$ , а величина  $\mathbb{E} N^{\underline{k}}$  — **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  $(n+1)^{\underline{k}} - n^{\underline{k}} = k n^{\underline{k-1}}$ ,  $\sum_{k \geq 0} n^{\underline{k}} / k! = 2^n$ ,  $2^{n+1} - 2^n = 2^n$  (ср. в дифференциальном исчислении  $(x+dx)^k - x^k = kx^{k-1} dx$ ,  $\sum_{k \geq 0} x^k / k! = e^x$ ,  $e^{x+dx} - e^x = e^x dx$ ).

**1.26. Производящая функция моментов** случайной величины  $N$ :

$$\Psi_N(s) = G_N(e^s) = \mathbb{E} e^{sN} = \sum_{n \geq 0} e^{sn} P(N = n).$$

**1.27. Производные производящей функции моментов** — это обычные, а не факториальные моменты:  $\frac{d^k \Psi_N}{ds^k} \Big|_{s=0} = \sum_{n \geq 0} n^k e^{sn} p_N(n) \Big|_{s=0} = \mathbb{E} N^k$ .

**1.28. При сложении независимых случайных величин**  $M$  и  $N$  их производящие функции перемножаются:

$$G_{M+N}(z) = \mathbb{E} z^{M+N} = \mathbb{E}(z^M z^N) = \mathbb{E} z^M \mathbb{E} z^N = G_M(z) G_N(z),$$

$$\Psi_{M+N}(s) = \mathbb{E} e^{s(M+N)} = \mathbb{E}(e^{sM} e^{sN}) = \mathbb{E} e^{sM} \mathbb{E} e^{sN} = \Psi_M(s) \Psi_N(s)$$

**1.29. Распределение вероятности суммы независимых случайных величин:**

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

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  $\mathbf{R} = \{x: -\infty < x < \infty\}$ . Элементы этого множества, т. е. отдельные точки, будем, как и раньше, называть **исходами**. Скалярные вещественные случайные величины обозначаются прописными латинскими буквами  $X, Y, Z, \dots$ , а исходы — строчными буквами  $x, y, z, \dots$ .

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:

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

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

$$F_X(x) = P(X \leq x),$$

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  $F_X(x) = P(X < x)$  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

$$P(X \in (a, b]) = P(a < X \leq b) = F_X(b) - F_X(a).$$

Так же определяются вероятности событий, представленных интервалами других типов:

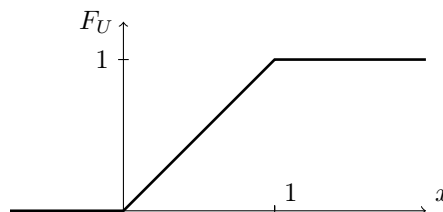
$$\begin{aligned} P(X \in [a, b]) &= P(a \leq X \leq b) = F_X(b) - F_X(a - 0), \\ P(X \in (a, b)) &= P(a < X < b) = F_X(b - 0) - F_X(a) \quad \text{и т. п.,} \end{aligned}$$

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  $P((a, b] \cap (c, d]) = P((c, b]) = F_X(b) - F_X(c)$ ,  $P((a, b] \cup (c, d]) = P((a, d]) = F_X(d) - F_X(a)$  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.  $P(a \leq U \leq b) = b - a$  для всех  $0 \leq a < b \leq 1$ . Тогда  $F_U(x) = x$  на  $[0, 1]$  и равна 1 справа и 0 — слева от этого отрезка. В частности,  $P(U = x) = x - x = 0$  для всех  $0 \leq x \leq 1$ .



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.

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

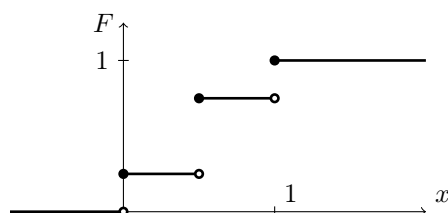
$$F_X(x) = \int_{-\infty}^x p_X(\xi) d\xi,$$

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  $P(X = x) = F_X(x) - F_X(x-0) > 0$ . Говорят, что в точке  $x$  расположен **атом вероятности** массой  $P(X = x)$ . 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  $x = 1$ :

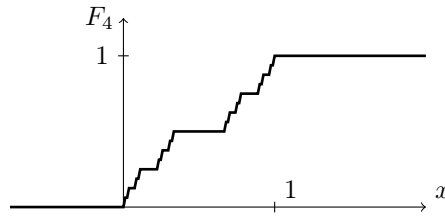


Распределения целочисленных случайных величин являются чисто точечными.

**2.11.** Полезно выделять еще один тип локального поведения распределения вероятности: если для некоторого  $0 < \alpha < 1$  имеет место асимптотика  $P(|X - x| < \Delta) \sim \Delta^\alpha$  при  $\Delta \rightarrow 0$ , будем говорить, что распределение имеет в точке  $x$  **сингулярность** порядка  $\alpha$ . 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  $x \geq 0$  задается формулой  $F(x) = \min(x^\alpha, 1)$ . Следующий пример менее тривиален (но при этом более типичен).

**2.12. Пример.** Пусть  $M_0$  — отрезок  $[0, 1]$ ,  $M_1$  — пара отрезков  $[0, \frac{1}{3}]$  и  $[\frac{2}{3}, 1]$ , получаемая выбрасыванием из  $M_0$  его средней трети, и вообще  $M_{i+1}$  — совокупность отрезков, получаемых выбрасыванием средней трети из каждого отрезка, входящего в  $M_i$ . Пусть далее  $F_i$  — кумулятивная функция равномерного распределения вероятности на  $M_i$ :



В пределе при  $i \rightarrow \infty$  возникает функция  $F_\infty$ , которая непрерывна, но не может быть представлена на интегралом вида п. 2.7 (проверьте!). Множество  $M_\infty$  точек ее роста называется **Канторовым множеством** средних третей, а соответствующее распределение вероятности — **Канторовой пылью**. Поскольку длина каждого из отрезков, составляющих  $M_i$ , 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); К. 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 Стильтьеса (ср. п. 1.7):

$$EX = \int_{-\infty}^{\infty} x dF_X(x).$$

Для произвольной функции случайной величины  $X$  математическое ожидание определяется интегралом

$$Ef(X) = \int_{-\infty}^{\infty} f(x) dF_X(x)$$

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

$$EX = \int_{-\infty}^{\infty} x p_X(x) dx \quad \text{или} \quad EX = \sum_n x_n P(X = x_n),$$

$$EX = \int_{-\infty}^{\infty} f(x) p_X(x) dx \quad \text{или} \quad EX = \sum_n f(x_n) P(X = x_n),$$

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  $dF_X$  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  $\max_i (x_{i+1} - x_i) \rightarrow 0$  (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, см. п. 2.6). 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):

$$\mu_k = EX^k = \int x^k dF_X(x), \quad \dot{\mu}_k = E(X - EX)^k = \int (x - \mu_1)^k dF_X(x).$$

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

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

**2.16. Медиана** случайной величины  $X$  определяется как такой исход  $x_{\text{med}}$ , для которого события  $X > x_{\text{med}}$  и  $X < x_{\text{med}}$  равновероятны. Точнее,  $x_{\text{med}} = \sup\{x: F(x) \leq \frac{1}{2}\}$ : такой вариант определения работает и в том случае, когда в  $x_{\text{med}}$  находится атом вероятности, так что  $F(x_{\text{med}}) > \frac{1}{2}$  и  $F(x_{\text{med}} - 0) \leq \frac{1}{2}$ . 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:  $p(x_{\text{max}}) = \max_x p(x)$ .

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 сходимостъ соответствующего интеграла, медиана существует у всех случайных величин.

**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  $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \dots$ , а их значения (**исходы**) — соответствующими полужирными строчными буквами  $\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$ .

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

$$\text{если } x_1^- < x_1^+, \dots, x_n^- < x_n^+, \text{ то } \sum_{(\varepsilon_1, \dots, \varepsilon_n)} (\varepsilon_1 1) \dots (\varepsilon_n 1) F(x_1^{\varepsilon_1}, \dots, x_n^{\varepsilon_n}) \geq 0,$$

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

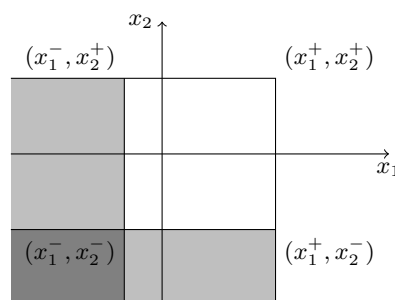
$$F(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = 0 \text{ при любом } 1 \leq i \leq n;$$

$$F(\infty, \infty, \dots, \infty) = 1;$$

$$\lim_{y_1 \downarrow x_1, y_2 \downarrow x_2, \dots, y_n \downarrow x_n} F(y_1, y_2, \dots, y_n) = F(x_1, x_2, \dots, x_n).$$

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

Первое из этих свойств выражает положительность вероятности:  $P(x_1^- < X_1 \leq x_1^+, \dots, x_n^- < X_n \leq x_n^+) \geq 0$  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  $F(x_1^+, x_2^+) - F(x_1^+, x_2^-) - F(x_1^-, x_2^+) + F(x_1^-, x_2^-) \geq 0$ :



Действительно, при вычислении вероятности того, что  $x_1^- < X \leq x_1^+, x_2^- < X_2 \leq x_2^+$  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  $\{\mathbf{x}: x_i^- \leq x_i \leq x_i^+, 1 \leq i \leq n\}$ , 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  $\mathbf{X}$  называется **абсолютно непрерывным**, если существует такая **функция плотности вероятности**  $p_{\mathbf{X}}$ , что

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{y_1 \leq x_1, \dots, y_n \leq x_n} p_{\mathbf{X}}(\mathbf{y}) d\mathbf{y}.$$

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

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

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 Стильеса (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  $F_{X_i}(x_i) = F(\infty, \dots, x_i, \dots, \infty)$ , где  $1 \leq i \leq n$ . Для абсолютно непрерывных распределений маргинальные распределения задаются функциями плотности вероятности

$$p_{X_i}(x_i) = \int p(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n = \int p(\mathbf{x}) \frac{d\mathbf{x}}{dx_i}, \quad 1 \leq i \leq n.$$

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

**2.24.** Пусть  $(\mathbf{X}, \mathbf{Y})$  — случайный вектор, компоненты которого разбиты на два подвектора  $\mathbf{X}$  и  $\mathbf{Y}$ . **Условное распределение вероятности** вектора  $\mathbf{X}$  при условии, что  $\mathbf{Y} \in A$ , где  $A$  — измеримое множество положительной вероятности, задается кумулятивной функцией (ср. п. 1.16)

$$F_{\mathbf{X}}(\mathbf{x} | \mathbf{Y} \in A) = \frac{P(X_1 \leq x_1, \dots, X_n \leq x_n, \mathbf{Y} \in A)}{P(\mathbf{Y} \in A)}.$$

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  $P(Y = y) = 0$ , то

$$p_{\mathbf{X}}(x | Y = y) = \lim_{\delta \rightarrow 0} p_{\mathbf{X}}(x | y - \delta < Y < y + \delta) = \lim_{\delta \rightarrow 0} \frac{\int_{y-\delta}^{y+\delta} p_{\mathbf{X},Y}(x, y') dy'}{\int_{-\infty}^{+\infty} dx \int_{y-\delta}^{y+\delta} p_{\mathbf{X},Y}(x, y') dy'} = \frac{p_{\mathbf{X},Y}(x, y)}{p_Y(y)}.$$

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

По сравнению с п. 1.16 определение условной плотности в абсолютно непрерывном случае учитывает тот факт, что  $P(Y = y) = 0$ . Говоря нестрого, речь идет об условной вероятности по отношению к «событию»  $y \leq Y \leq y + dy$ , 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  $p_{\mathbf{X}}(x | X = Y)$ , понимаемая как предел при  $\delta \rightarrow 0$  условной плотности  $p_{\mathbf{X}}(x | -\delta < X - Y < \delta)$ , постоянна и равна 1, в то время как аналогичный предел для  $p_{\mathbf{X}}(x | 1 - \delta < X/Y < t + \delta)$  равен  $2x$  (проверьте!).

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

**2.27.** Свойства пп. 1.21 и 1.23 (аддитивность математического ожидания и аддитивность дисперсии, если случайные величины независимы) выполнены и в непрерывном случае.