

2. СЛУЧАЙНЫЕ ПЕРЕМЕННЫЕ

2.1. Основные определения

Рассмотрим эксперимент, состоящий в n бросаниях правильной монеты. Построим пространство элементарных событий при $n = 3$:

$$\Omega = \{,,,,,,\}.$$

Это пространство содержит 8 элементарных событий. Правильность монеты дает основание считать элементарные события равновероятными. В таблице 2.1 представлены элементарные события, их вероятности и соответствующие этим событиям количества выпадений герба (X).

Таблица 1:

ω	PPP	PPГ	РГР	РГГ	ГРР	ГРГ	ГГР	ГГГ
$P(\omega)$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
X	0	1	1	2	1	2	2	3

Видно, что X может принимать 4 различных значения: 0, 1, 2, 3. Соответственно определены четыре части Ω , на каждой из которых X сохраняет постоянное значение. Поэтому можно рассмотреть события $X = 0$, $X = 1$, $X = 2$, $X = 3$ и вычислить их вероятности. Результаты сведены в таблицу 2.2.

Таблица 2:

x	0	1	2	3
$P(x)$	1/8	3/8	3/8	1/8

Если целью эксперимента было определение количества выпадений герба, то естественно рассмотреть новое пространство элементарных событий – множество значений X : $\Omega' = \{0, 1, 2, 3\}$ – и заданные на нем таблицей 2.2 вероятности.

Ситуация, когда исходом эксперимента можно считать число, часто встречается в прикладных задачах.

Определение. *Дискретной случайной переменной* называется конечное или счетное *числовое* множество (пространство элементарных событий) с заданными на нем вероятностями.

Замечания. 1. Введенное определение не исключает вырожденной ситуации, когда пространство элементарных событий состоит из одного элемента, вероятность которого равна единице. Такая случайная переменная уже будет "не случайной", и мы исключим ее из рассмотрения.

2. Мы будем обозначать случайные переменные большими буквами латинского алфавита, а их значения – *числа* – соответствующими малыми буквами.

Вспомним понятие *переменная*: переменная – это буква и множество, элементы которого можно подставлять вместо этой буквы. Переменную можно интерпретировать как мешок, содержащий все возможные ее значения. Каждый имеет право выбрать в этом мешке понравившееся ему значение переменной. Аналогично, *случайной переменной* назовем букву и (числовое) множество ее значений, которое мы будем рассматривать как пространство элементарных событий. Дискретную случайную переменную также можно интерпретировать как мешок, содержащий все возможные ее значения. Однако *выбирать эти значения нельзя*. Можно запустить в этот мешок руку и вынуть из него то значение случайной переменной, которое *случайно* попадетсся. Если повторять этот эксперимент *множественно*, то относительные частоты появления различных значений случайной переменной будут, *как правило*, мало отличаться от их вероятностей.

Терминологическое замечание. В книгах по теории вероятностей часто вместо слов "случайная переменная" используется термин "случайная величина".

Итак, дискретная случайная переменная задается множеством своих значений и вероятностями этих значений, т.е. таблицей, которую называют *законом распределения* случайной переменной.

x	x_1	x_2	\dots
$P(x)$	p_1	p_2	\dots

$$p_k > 0; \quad \sum_k p_k = 1.$$

Полезна следующая физическая интерпретация дискретной с.п. (в дальнейшем мы будем писать с.п. вместо "случайная переменная"): значения с.п. рассматриваются как абсциссы точек на оси, а соответствующие вероятности – как массы, сосредоточенные в этих точках. Таким образом, с.п. интерпретируется как система точечных масс на оси. Существенно, что масса всей системы всегда равна единице. Можно сказать, что в нашем распоряжении есть единичная "вероятностная масса" и мы должны "распределить" ее между заданными точками оси – задать *закон распределения вероятностей*.

Эта физическая аналогия помогает найти способ описания с.п. с несчетным множеством значений. Известно, что физика оперирует в таком случае понятием *плотности массы*, считая, что масса в каждой точке равна нулю, но масса отрезка равна интегралу от плотности по этому отрезку. По аналогии мы будем рассматривать так называемую *абсолютно непрерывную* с.п. X , множество значений которой $\Omega = \mathbb{R}$. Мы будем считать, что задана кусочно непрерывная функция (*плотность распределения* с.п. X) $f_X : \mathbb{R} \rightarrow \mathbb{R}$, обладающая двумя свойствами:

$$1) f_X \geq 0; \quad 2) \int_{-\infty}^{+\infty} f_X = 1. \quad (2.1.1)$$

В этой модели вероятность попадания с.п. в промежуток $[a, b]$ будет вычисляться по формуле

$$\mathbf{P}(X \in [a, b]) = \int_a^b f_X.$$

Замечания. 1. Мы записали $X \in [a, b]$, считая промежуток *сегментом*. Но если вспомнить, что значение интеграла не зависит от значений подынтегральной функции в конечном числе точек, то будет ясно, что тип промежутка (интервал, полуинтервал, сегмент) не играет роли.

2. Очевидно, $\mathbf{P}(X = a) = \int_a^a f_X = 0$, т.е. вероятность каждого отдельного значения для абсолютно непрерывной с.п. равна нулю.

3. В приложениях встречаются так называемые дискретно-непрерывные с.п., но их рассмотрение выходит за рамки нашего курса.

Одну из первообразных плотности распределения с.п., а именно,

$$F_X(x) = \int_{-\infty}^x f_X = \mathbf{P}(X < x), \quad (2.1.2)$$

называют *функцией распределения* с.п. X . Согласно (2.1.2), значение функции распределения в точке x – это вероятность события $(X < x)$.

Из свойств первообразной следует, что F_X – непрерывная, неубывающая на \mathbb{R} функция, $F_X(-\infty) = 0$, $F_X(+\infty) = 1$. В точках, где непрерывна плотность f_X , функция распределения имеет производную, причем $F'_X = f_X$.

2.2. Числовые характеристики случайных переменных

В механике для компактного описания точечной системы масс используются так называемые *моменты*. Статический момент системы масс относительно начала координат равен, как известно,

$$\sum_k x_k \cdot m_k.$$

Здесь x_k – абсцисса точки, а m_k – масса, сосредоточенная в этой точке.

По аналогии, *первым начальным моментом* или *математическим ожиданием* дискретной с.п. X называют число

$$M(X) = \sum_k x_k \cdot p_k,$$

где x_k – значение с.п. X , а p_k – вероятность этого значения.

Для абсолютно непрерывной с.п. X с плотностью f_X понятие математического ожидания вводится так:

$$M(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx.$$

Замечания. 1. Если множество значений дискретной с.п. счетно, то первый момент представляет собой *сумму ряда*. В этом случае выдвигается требование *абсолютной* сходимости ряда, так как если ряд сходится не абсолютно, то его сумма зависит от порядка слагаемых, а это невозможно интерпретировать в содержательных задачах. Для абсолютно непрерывной с.п. требуется абсолютная сходимость интеграла, определяющего первый момент. В случае, если ряд (интеграл) не является абсолютно сходящимся, говорят, что с.п. *не имеет математического ожидания*.

2. Учитывая, что $\sum_k p_k = 1$, математическое ожидание дискретной с.п. можно записать так:

$$M(X) = \frac{\sum_k x_k \cdot p_k}{\sum_k p_k}.$$

Аналогично, для абсолютно непрерывной с.п.

$$M(X) = \frac{\int_{-\infty}^{+\infty} x \cdot f_X(x) dx}{\int_{-\infty}^{+\infty} f_X(x) dx}$$

(сравните с физическим понятием центра масс).

В физике разброс массы системы относительно ее центра масс характеризуется моментом инерции. По аналогии вводится понятие *второго центрального момента*, или *дисперсии* с.п. X :

для дискретной с.п. X

$$\mu_2(X) = D(X) = \sum_k (x_k - M(X))^2 \cdot p_k, \quad (2.2.1)$$

для абсолютно непрерывной –

$$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 \cdot f_X(x) dx. \quad (2.2.1')$$

Очевидно, есть смысл говорить о дисперсии с.п. только при наличии у нее математического ожидания. Но даже в этом случае дисперсия не обязана существовать (ряд или интеграл может расходиться).

Теорема. Дисперсия с.п. положительна (если она существует).

Доказательство. Интеграл в формуле (2.2.1'), очевидно, неотрицателен. Более того, если он равен нулю, то подынтегральная функция равна нулю почти всюду, чего не может быть ввиду условия (2.1.1).

Аналогично, сумма в (2.2.1) может равняться нулю только в том случае, когда с.п. X принимает единственное значение $M(X)$ с вероятностью 1. Но такую вырожденную ситуацию мы, как уже говорилось, не рассматриваем.

Если у с.п. есть дисперсия, то *среднеквадратическим уклонением* (иногда говорят *отклонением*) этой с.п. называют число

$$\sigma(X) = \sqrt{D(X)}.$$

Среднеквадратическое уклонение есть мера разброса значений с.п. относительно ее математического ожидания.

Докажем теперь важное неравенство. Пусть X – с.п., имеющая дисперсию. Тогда для любого $\varepsilon > 0$

$$\mathbf{P}(|X - M(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$$

(при малых ε ($\varepsilon \leq \sigma(X)$) это неравенство становится тривиальным).

Мы ограничимся случаем абсолютно непрерывной с.п. Неравенство

$$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 \cdot f_X(x) dx \geq$$

$$\geq \int_{-\infty}^{M(X)-\varepsilon} (x - M(X))^2 \cdot f_X(x) dx + \int_{M(X)+\varepsilon}^{+\infty} (x - M(X))^2 \cdot f_X(x) dx$$

следует из неотрицательности подынтегральной функции. Учитывая, что в подынтегральных выражениях $|x - M(X)| \geq \varepsilon$, можно записать

$$\begin{aligned} D(X) &\geq \varepsilon^2 \cdot \left(\int_{-\infty}^{M(X)-\varepsilon} f_X(x) dx + \int_{M(X)+\varepsilon}^{+\infty} f_X(x) dx \right) = \\ &= \varepsilon^2 \cdot \int_{|x - M(X)| \geq \varepsilon} f_X(x) dx = \varepsilon^2 \cdot \mathbf{P}(|X - M(X)| \geq \varepsilon). \end{aligned}$$

Доказанное неравенство именуется *неравенством Чебышева*. Его часто интерпретируют так: *большие отклонения с.п. от ее математического ожидания маловероятны*.

Математическое ожидание и дисперсия в какой-то мере характеризуют закон распределения с.п. Однако следует понимать, что существует сколько угодно законов распределения с данными математическим ожиданием и дисперсией. В то же время есть задачи, для решения которых достаточно знать только эти числовые характеристики.

Укажем еще одну часто встречающуюся числовую характеристику. *Модой* распределения абсолютно непрерывной с.п. называют абсциссу локального максимума ее плотности распределения. Если распределение имеет не одну моду, то его называют *многомодальным*.

2.3. Некоторые законы распределения, встречающиеся в приложениях

1. Биномиальное распределение Бернулли¹. Так называется двухпараметрическое семейство дискретных распределений с параметрами $n \in \mathbb{N}$ и $p \in]0, 1[$:

$$\Omega = \{0, 1, \dots, n\}; \quad \mathbf{P}(X = m) = \binom{n}{m} \cdot p^m \cdot (1 - p)^{n-m}; \quad m = 0, 1, \dots, n.$$

Здесь $\binom{n}{m} = \frac{n!}{m! \cdot (n-m)!}$ – *число сочетаний*² из n элементов по m .

Содержательная интерпретация биномиального распределения: эксперимент может заканчиваться одним из двух исходов: успех или неудача. Вероятность успеха равна p . Эксперимент повторяют n раз. С.п. – количество успехов.

Название "биномиальное" объясняется тем, что вероятности элементарных событий совпадают с членами бинома Ньютона $(a + b)^n$ при $a = p$, $b = 1 - p$. Отсюда следует, что

$$\sum_{m=0}^n \binom{n}{m} \cdot p^m \cdot (1 - p)^{n-m} = 1.$$

Для вычисления математического ожидания и дисперсии биномиального распределения нужна некоторая техника суммирования, которая не имеет отношения к теории вероятностей. Поэтому здесь (как и в дальнейшем) мы лишь приводим результаты:

$$M(X) = \sum_{m=0}^n m \cdot \binom{n}{m} \cdot p^m \cdot (1 - p)^{n-m} = n \cdot p,$$

$$D(X) = \sum_{m=0}^n (m - n \cdot p)^2 \cdot \binom{n}{m} \cdot p^m \cdot (1 - p)^{n-m} = n \cdot p \cdot (1 - p).$$

Заметим, что поскольку $p \cdot (1 - p) = 0.25 - (p - 0.5)^2 \leq 0.25$, то $D(X) \leq 0.25 \cdot n$.

¹БЕРНУЛЛИ – семья швейцарских ученых. Распределение носит имя Якоба I БЕРНУЛЛИ (1654-1705), благодаря работам которого теория вероятностей стала играть важную роль в приложениях.

²В отечественной литературе вместо $\binom{n}{m}$ часто пишут C_n^m (обратите внимание на расположение букв n и m в этих символах!).

2. Геометрическое распределение – однопараметрическое семейство дискретных распределений с параметром $p \in]0, 1[$:

$$\Omega = \mathbb{N}; \quad \mathbf{P}(X = n) = p \cdot (1 - p)^{n-1}.$$

Содержательная интерпретация геометрического распределения: вероятность успеха в одном эксперименте равна p ; эксперимент повторяют до первого успеха. С.п. – количество повторений эксперимента³.

Легко видеть, что

$$\sum_{n=1}^{+\infty} p \cdot (1 - p)^{n-1} = \frac{p}{1 - (1 - p)} = 1.$$

Далее,

$$M(X) = \sum_{n=1}^{+\infty} n \cdot p \cdot (1 - p)^{n-1} = \frac{1}{p},$$

$$D(X) = \sum_{n=1}^{+\infty} (n - 1/p)^2 \cdot p \cdot (1 - p)^{n-1} = \frac{1 - p}{p^2}.$$

3. Распределение Пуассона – однопараметрическое семейство дискретных распределений с параметром $\lambda > 0$:

$$\Omega = \mathbb{N} \cup \{0\}; \quad \mathbf{P}(X = n) = \exp(-\lambda) \cdot \frac{\lambda^n}{n!}.$$

Очевидно, что

$$\sum_{n=0}^{+\infty} \exp(-\lambda) \cdot \frac{\lambda^n}{n!} = \exp(-\lambda) \cdot \exp(\lambda) = 1.$$

Далее, вычисление показывает, что

$$M(X) = \sum_{n=0}^{+\infty} n \cdot \exp(-\lambda) \cdot \frac{\lambda^n}{n!} = \lambda;$$

$$D(X) = \sum_{n=0}^{+\infty} (n - \lambda)^2 \cdot \exp(-\lambda) \cdot \frac{\lambda^n}{n!} = \lambda.$$

4. Равномерное распределение – двухпараметрическое семейство абсолютно непрерывных распределений с параметрами $x_0 \in \mathbb{R}$ и $\Delta > 0$:

$$f_X(x) = \begin{cases} \text{const}, & \text{если } |x - x_0| \leq \Delta, \\ 0, & \text{если } |x - x_0| > \Delta. \end{cases}$$

Таким образом, "вероятностная масса" (равная 1) *равномерно* распределена по промежутку длиной 2Δ (с центром в точке x_0). Поэтому значение константы равно $\frac{1}{2\Delta}$.

Одна из возможных интерпретаций: равномерно распределенной считают (по-видимому, не без основания) погрешность округления в цифровых измерительно-вычислительных устройствах.

Очевидно,

$$M(X) = \int_{x_0 - \Delta}^{x_0 + \Delta} x \cdot \frac{1}{2\Delta} dx = x_0;$$

$$D(X) = \int_{x_0 - \Delta}^{x_0 + \Delta} (x - x_0)^2 \cdot \frac{1}{2\Delta} dx = \frac{\Delta^2}{3}.$$

5. Нормальное распределение – двухпараметрическое семейство абсолютно непрерывных распределений с параметрами $m \in \mathbb{R}$ и $\sigma > 0$ (рис.2.1):

$$f_X(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right).$$

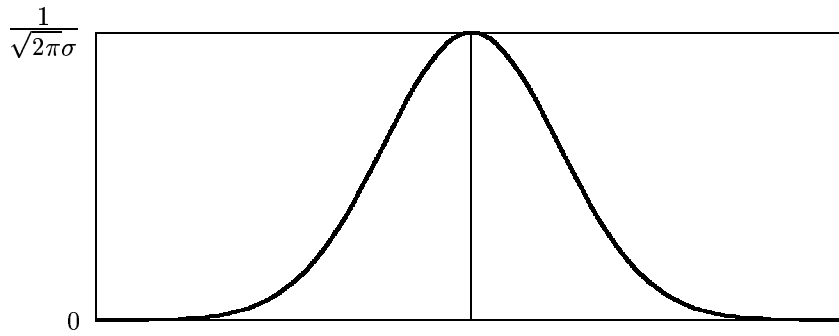


Рис. 1: Плотность нормального распределения с параметрами m и σ

Убедитесь в том, что $\int_{-\infty}^{+\infty} f_X = 1$, используя известный из курса математического анализа интеграл Пуассона

$$\int_{-\infty}^{+\infty} \exp(-y^2) dy = \sqrt{\pi}.$$

Можно показать, что

$$M(X) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^{+\infty} x \cdot \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = m;$$

$$D(X) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^{+\infty} (x-m)^2 \cdot \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = \sigma^2.$$

6. Распределение Коши – двухпараметрическое семейство абсолютно непрерывных распределений с параметрами $x_0 \in \mathbb{R}$ и $a > 0$ (рис.2.2):

$$f_X(x) = \frac{a}{\pi} \cdot \frac{1}{a^2 + (x - x_0)^2}.$$

Легко видеть, что

$$\int_{-\infty}^{+\infty} f_X = \frac{1}{\pi} \cdot \arctg\left(\frac{x - x_0}{a}\right) \Big|_{x=-\infty}^{x=+\infty} = 1.$$

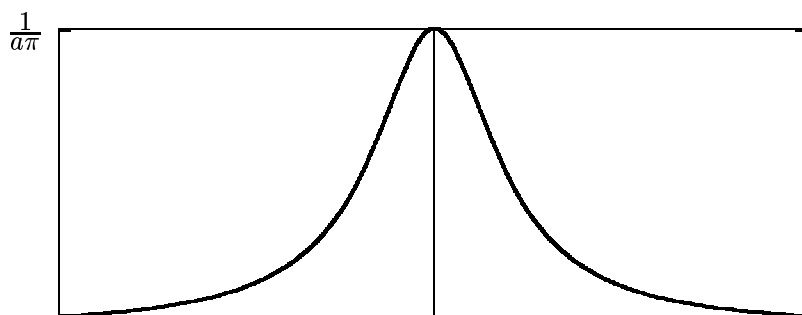


Рис. 2: Плотность распределения Коши с параметрами x_0 и a

Сравнивая рис.2.1 и рис.2.2, можно подумать, что распределение Коши очень похоже на нормальное. Однако интеграл

$$\int_{-\infty}^{+\infty} x \cdot \frac{a}{\pi} \cdot \frac{1}{a^2 + (x - x_0)^2} dx$$

расходится, и следовательно, у распределения Коши нет ни математического ожидания, ни дисперсии. В то же время распределение Коши имеет единственную моду (x_0).

2.4. Случайный вектор

³При $p = 1/2$ это распределение уже встречалось в гл.1.

Случайная переменная может использоваться как математическая модель эксперимента, в котором измеряется значение физической величины. Для описания эксперимента, в котором одновременно фиксируются значения *нескольких* величин, служит понятие *случайного вектора*⁴.

Дискретный случайный вектор задается таблицей, в которой перечислены все возможные значения вектора, и каждому значению приписана вероятность.

Абсолютно непрерывный случайный вектор X задается плотностью его распределения – неотрицательной кусочно непрерывной функцией $f_X : \mathbb{R}^n \rightarrow \mathbb{R}$, такой, что

$$\int_{\mathbb{R}^n} f_X = 1. \quad (2.4.1)$$

При этом каждой области $G \subset \mathbb{R}^n$ с кусочно гладкой границей соответствует вероятность

$$\mathbf{P}(X \in G) = \int_G f_X.$$

Примеры. 1. Двухмерный дискретный случайный вектор задан таблицей:

$x_2 \backslash x_1$	0	1
2	1/2	1/8
4	1/8	1/16
6	1/16	1/8

Множество значений этого случайного вектора (пространство элементарных событий) составляют шесть числовых векторов:

$$\begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \end{bmatrix}.$$

Их вероятности записаны в соответствующих клетках таблицы. Так, например, $\mathbf{P}(X = [1, 6]^T) = 1/8$. Убедитесь в том, что сумма всех вероятностей равна единице.

2. Двухмерный случайный вектор, равномерно распределенный в ограниченной области G с кусочно гладкой границей, задается плотностью распределения:

$$f_X : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad f_X(x) = \begin{cases} \text{const}, & \text{если } x \in G, \\ 0, & \text{если } x \notin G. \end{cases}$$

Таким образом, единичная „вероятностная масса“ *равномерно* распределена по области G . Поэтому значение константы равно $\frac{1}{S(G)}$.

Аналогично определяется n -мерный случайный вектор, равномерно распределенный в n -мерной области.

3. n -мерный *нормально* распределенный случайный вектор задается плотностью распределения

$$f_X(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(B)}} \cdot \exp\left(-\frac{1}{2} \langle B^{-1}(x - m), (x - m) \rangle\right). \quad (2.4.2)$$

Здесь $m = [m_1, \dots, m_n]^T \in \mathbb{R}^n$, B – симметричная, положительно определенная $(n \times n)$ -матрица.

Проверим, что $\int_{\mathbb{R}^n} f_X = 1$. В курсе линейной алгебры было показано, что положительно определенная матрица допускает разложение Холецкого:

$$B = H^* H,$$

где H – верхняя треугольная матрица. Сделаем в интеграле "подстановку" $x = m + H^* y$. Получим

$$\begin{aligned} \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2} \sqrt{\det(B)}} \cdot \exp\left(-\frac{1}{2} \langle B^{-1}(x - m), (x - m) \rangle\right) &= \\ = \int_{\mathbb{R}^n} \frac{|\det(H^*)|}{(2\pi)^{n/2} \sqrt{\det(B)}} \cdot \exp\left(-\frac{1}{2} \langle B^{-1} H^* y, H^* y \rangle\right) &= \end{aligned}$$

⁴в дальнейшем мы будем иногда использовать сокращение "с.в."

Замечая, что $|\det(H^*)| = |\det(H)| = \sqrt{\det(B)}$, и

$$\langle B^{-1}H^*y, H^*y \rangle = \langle H(H^*H)^{-1}H^*y, y \rangle = \langle y, y \rangle,$$

имеем

$$\begin{aligned} \int_{\mathbb{R}^n} f_X &= \frac{1}{(2\pi)^{n/2}} \cdot \int_{\mathbb{R}^n} \exp\left(-\frac{y_1^2 + \dots + y_n^2}{2}\right) dy_1 \dots dy_n = \\ &= \prod_{k=1}^n \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{y_k^2}{2}\right) dy_k \right) = 1 \end{aligned}$$

(все одномерные интегралы равны 1, как следует из примера 5 п.2.3).

Если задан случайный вектор X со значениями в \mathbb{R}^n , то можно построить случайные переменные X_1, \dots, X_n – координаты X . Множество значений каждой такой с.п. известно. Требуется лишь указать распределение вероятностей на нем.

Рассмотрим вначале случай *дискретного* случайного вектора (пример 1). Множество значений с.п. X_1 – это $\Omega = \{0, 1\}$. Событие $X_1 = 0$ есть объединение трех элементарных событий:

$$(X_1 = 0, X_2 = 2), (X_1 = 0, X_2 = 4), (X_1 = 0, X_2 = 6).$$

Поэтому

$$\mathbf{P}(X_1 = 0) = 1/2 + 1/8 + 1/16 = 11/16.$$

Аналогично,

$$\mathbf{P}(X_1 = 1) = 1/8 + 1/16 + 1/8 = 5/16.$$

Мы нашли распределение с.п. X_1 – первой координаты случайного вектора X .

Обобщая этот результат, можно сформулировать алгоритм построения закона распределения координаты *дискретного* случайного вектора:

$$\mathbf{P}(X_k = x_k) = \sum_{x_1} \dots \sum_{x_{k-1}} \sum_{x_{k+1}} \dots \sum_{x_n} \mathbf{P}(X = x).$$

Для каждого возможного значения координаты суммируются вероятности всех элементарных событий, на которых это значение достигается.

В случае абсолютно непрерывного случайного вектора суммирование заменяется интегрированием:

$$f_{X_k}(x_k) = \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_{k-1} \int_{-\infty}^{+\infty} dx_{k+1} \dots \int_{-\infty}^{+\infty} f_X(x) dx_n.$$

Примеры. 4. Пусть случайный вектор Y равномерно распределен на прямоугольнике $|y_1| \leq a$ и $|y_2| \leq b$; $a > 0$, $b > 0$. Тогда

$$f_Y : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad f_Y(y) = \begin{cases} \frac{1}{4ab}, & \text{если } |y_1| \leq a \text{ и } |y_2| \leq b, \\ 0, & \text{если } |y_1| > a \text{ или } |y_2| > b; \end{cases}$$

$$f_{Y_1}(y_1) = \int_{-\infty}^{+\infty} f_Y(y_1, y_2) dy_2 = \begin{cases} \frac{1}{2a}, & \text{если } |y_1| \leq a, \\ 0, & \text{если } |y_1| > a; \end{cases}$$

$$f_{Y_2}(y_2) = \int_{-\infty}^{+\infty} f_Y(y_1, y_2) dy_1 = \begin{cases} \frac{1}{2b}, & \text{если } |y_2| \leq b, \\ 0, & \text{если } |y_2| > b. \end{cases}$$

Координаты этого случайного вектора распределены равномерно на сегментах $|y_1| \leq a$ и $|y_2| \leq b$ соответственно.

5. Пусть случайный вектор Z равномерно распределен на круге $\|z\| \leq r$. Тогда

$$f_Z : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad f_Z(z) = \begin{cases} \frac{1}{\pi r^2}, & \text{если } \|z\| \leq r, \\ 0, & \text{если } \|z\| > r; \end{cases}$$

$$f_{Z_1}(z_1) = \int_{-\infty}^{+\infty} f_Z(z_1, z_2) dz_2 = \begin{cases} \frac{2\sqrt{r^2 - z_1^2}}{\pi r^2}, & \text{если } |z_1| \leq r, \\ 0, & \text{если } |z_1| > r; \end{cases}$$

$$f_{Z_2}(z_2) = \int_{-\infty}^{+\infty} f_Z(z_1, z_2) dz_1 = \begin{cases} \frac{2\sqrt{r^2 - z_2^2}}{\pi r^2}, & \text{если } |z_2| \leq r, \\ 0, & \text{если } |z_2| > r. \end{cases}$$

Координаты этого случайного вектора распределены одинаково (но неравномерно!).

Отметим, что знание распределений координат не дает, вообще говоря, возможности построить распределение вектора. Исключение составляет случай, для описания которого мы введем важное новое понятие.

Координаты дискретного с.в. называются (*статистически*) *независимыми в совокупности*, если

$$\mathbf{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{k=1}^n \mathbf{P}(X_k = x_k)$$

для всех возможных значений $x = (x_1, \dots, x_n)$ этого вектора.

Возвращаясь к примеру 1, видим, что координаты вектора X статистически зависимы, так как, например, $\mathbf{P}(X_1 = 0, X_2 = 2) = 1/2$, но $\mathbf{P}(X_1 = 0) \cdot \mathbf{P}(X_2 = 2) = 55/128$.

Координаты абсолютно непрерывного с.в. называются независимыми в совокупности, если

$$f_X(x) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \dots f_{X_n}(x_n),$$

т.е. если плотность распределения случайного вектора равна произведению плотностей распределения его координат.

Легко видеть, что в примере 4 это условие выполнено, т.е. координаты случайного вектора статистически независимы. А вот в примере 5 координаты оказываются зависимыми.

Итак, если координаты случайного вектора независимы в совокупности, то изучение распределения с.в. можно заменить изучением распределений его координат. Поэтому эксперимент обычно стараются планировать так, чтобы "из физических соображений" координаты случайного вектора (математической модели эксперимента) можно было считать независимыми в совокупности. Такое предположение существенно упрощает задачу и уде- шевляет ее решение. Подчеркнем, что ответственность за правильность этих "физических соображений" несет экспериментатор.

2.5. Числовые характеристики случайного вектора

Определение. Математическим ожиданием n -мерного с.в. X называется числовой вектор $m \in \mathbb{R}^n$, вычисляемый по правилу:

в дискретном случае

$$m = M(X) = \sum_x x \cdot \mathbf{P}(X = x)$$

(суммирование ведется по всем значениям случайного вектора);

в абсолютно непрерывном случае

$$m = M(X) = \int_{\mathbb{R}^n} x \cdot f_X(x)$$

(напомним, что интегрирование вектора выполняется по координатно).

Вычислим первую координату вектора $M(X)$ для дискретного с.в.:

$$m_1 = \sum_{x_1} x_1 \cdot \sum_{x_2, \dots, x_n} \mathbf{P}(X = x) = \sum_{x_1} x_1 \cdot \mathbf{P}(X_1 = x_1) = M(X_1)$$

(суммирование по x_2, \dots, x_n дает распределение X_1).

В случае абсолютно непрерывного с.в.

$$m_1 = \int_{-\infty}^{+\infty} x_1 dx_1 \int_{\mathbb{R}^{n-1}} f_X(x) dx_2 \dots dx_n = \int_{-\infty}^{+\infty} x_1 \cdot f_{X_1}(x_1) dx_1 = M(X_1)$$

(интегрирование по x_2, \dots, x_n дает плотность распределения X_1).

Аналогично, $m_k = M(X_k)$, $k = 1, \dots, n$.

Итак, координаты математического ожидания случайного вектора равны математическим ожиданиям соответствующих координат этого вектора. При этом, конечно, предполагается *абсолютная* сходимости всех числовых рядов или несобственных интегралов.

Определение. Ковариационной матрицей n -мерного с.в. X называется числовая матрица, вычисляемая по правилу:

$$B(X) = \text{cov}(X) = \sum_x (x - M(X)) \cdot (x - M(X))^T \cdot \mathbf{P}(X = x)$$

в дискретном случае (суммирование ведется по всем значениям с.в.);

$$B(X) = \text{cov}(X) = \int_{\mathbb{R}^n} (x - M(X)) \cdot (x - M(X))^T \cdot f_X(x)$$

в абсолютно непрерывном случае (естественно, интегрирование матрицы выполняется поэлементно).

Поскольку $(x - M(X)) \cdot (x - M(X))^T$ – матрица порядка n , то и $B(X)$ – матрица порядка n . Вычислим ее первый *диагональный* элемент.

$$\begin{aligned} b_{11} &= \sum_{x_1} (x_1 - m_1)^2 \cdot \sum_{x_2, \dots, x_n} \mathbf{P}(X = x) = \\ &= \sum_{x_1} (x_1 - m_1)^2 \cdot \mathbf{P}(X_1 = x_1) = D(X_1). \end{aligned}$$

В случае абсолютно непрерывного с.в.

$$\begin{aligned} b_{11} &= \int_{-\infty}^{+\infty} (x_1 - m_1)^2 dx_1 \int_{\mathbb{R}^{n-1}} f_X(x) dx_2 \dots dx_n = \\ &= \int_{-\infty}^{+\infty} (x_1 - m_1)^2 \cdot f_{X_1}(x_1) dx_1 = D(X_1). \end{aligned}$$

Аналогично, $b_{kk} = D(X_k)$, $k = 1, \dots, n$.

Таким образом, диагональные элементы ковариационной матрицы равны дисперсиям соответствующих координат случайного вектора.

Рассмотрим теперь внедиагональные элементы матрицы $B(X)$:

$$b_{kj} = b_{jk} = \sum_x (x_k - m_k)(x_j - m_j) \cdot \mathbf{P}(X = x)$$

в дискретном случае;

$$b_{kj} = b_{jk} = \int_{\mathbb{R}^n} (x_k - m_k)(x_j - m_j) \cdot f_X(x)$$

в абсолютно непрерывном случае.

Число b_{kj} называется *ковариацией* случайных переменных X_k и X_j или *вторым смешанным центральным моментом* (читателю, знакомому с механикой, рекомендуем вспомнить центробежный момент инерции).

Напомним, что для существования матрицы ковариаций требуется существование вектора математического ожидания и абсолютная сходимости всех определяющих ее элементы рядов (интегралов).

Важное свойство матрицы ковариаций доказывает

Теорема. Матрица $B(X)$ неотрицательно определена, т.е.

$$\langle B(X) \alpha, \alpha \rangle \geq 0 \quad \text{для любого } \alpha \in \mathbb{R}^n.$$

Доказательство. Пусть $\alpha \in \mathbb{R}^n$. Для дискретного случайного вектора X имеем

$$\begin{aligned}\langle B(X) \alpha, \alpha \rangle &= \alpha^T B(X) \alpha = \sum_x \alpha^T (x - m) \cdot (x - m)^T \alpha \cdot \mathbf{P}(X = x) = \\ &= \sum_x |\langle (x - m), \alpha \rangle|^2 \cdot \mathbf{P}(X = x);\end{aligned}\quad (2.5.1)$$

аналогично, для абсолютно непрерывного случайного вектора

$$\langle B(X) \alpha, \alpha \rangle = \int_{\mathbb{R}^n} |\langle (x - m), \alpha \rangle|^2 \cdot f_X(x). \quad (2.5.1')$$

Интеграл в (2.5.1'), очевидно, неотрицателен. Более того, если $\alpha \neq \theta_n$, то он не может равняться нулю, ибо в силу кусочной непрерывности f_X и условия (2.4.1) найдется область, в которой оба сомножителя положительны.

Аналогично, если $\alpha \neq \theta_n$, то сумма в (2.5.1) может равняться нулю только в случае, когда все точки x , для которых $\mathbf{P}(X = x) > 0$, расположены в плоскости $\langle (x - m), \alpha \rangle = 0$. А это равенство, как известно из курса линейной алгебры, означает, что одна из координат вектора $x - m$ есть линейная комбинация остальных:

$$x_k = m_k - \frac{1}{\alpha_k} \sum_{j \neq k} \alpha_j \cdot (x_j - m_j). \quad (2.5.2)$$

Таким образом, матрица ковариации является даже положительно определенной, за исключением вырожденного случая, когда одна из координат случайного вектора – полином первой степени от остальных координат.

Примеры. Найдем числовые характеристики случайных векторов, рассмотренных в предыдущем пункте.

1. Дискретный случайный вектор из примера 1.

$x_2 \backslash x_1$	0	1
2	1/2	1/8
4	1/8	1/16
6	1/16	1/8

$$m_1 = M(X_1) = 0 \cdot (1/2 + 1/8 + 1/16) + 1 \cdot (1/8 + 1/16 + 1/8) = 5/16.$$

Аналогичные подсчеты дают

$$M(X) = \begin{bmatrix} 5/16 \\ 25/8 \end{bmatrix}; \quad cov(X) = \begin{bmatrix} 55/256 & 35/128 \\ 35/128 & 159/64 \end{bmatrix}.$$

2. Абсолютно непрерывный случайный вектор из примера 4:

$$f_Y : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad f_Y(y) = \begin{cases} \frac{1}{4ab}, & \text{если } |y_1| \leq a \text{ и } |y_2| \leq b, \\ 0, & \text{если } |y_1| > a \text{ или } |y_2| > b. \end{cases}$$

$$m_1 = M(Y_1) = \int_{-\infty}^{+\infty} y_1 \cdot f_{Y_1}(y_1) dy_1 = \frac{1}{2a} \cdot \int_{-a}^a y_1 dy_1 = 0$$

(интеграл от нечетной функции по симметричному относительно нуля промежутку равен нулю).

Аналогично, $m_2 = 0$, т.е. $M(Y) = [0, 0]^T$.

$$\begin{aligned}b_{11} = D(Y_1) &= \int_{-\infty}^{+\infty} (y_1 - m_1)^2 \cdot f_{Y_1}(y_1) dy_1 = \\ &= \frac{1}{2a} \cdot \int_{-a}^a (y_1 - 0)^2 dy_1 = a^2/3.\end{aligned}$$

Аналогично, $b_{22} = D(Y_2) = b^2/3$. Далее,

$$\begin{aligned} b_{12} = b_{21} &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (y_1 - m_1)(y_2 - m_2) \cdot f_Y(y_1, y_2) dy_1 dy_2 = \\ &= \frac{1}{4ab} \cdot \int_{-a}^a y_1 dy_1 \int_{-b}^b y_2 dy_2 = 0 \end{aligned}$$

(интеграл от нечетной функции по симметричному относительно нуля промежутку равен нулю).

Итак,

$$\text{cov}(Y) = \begin{bmatrix} \frac{a^2}{3} & 0 \\ 0 & \frac{b^2}{3} \end{bmatrix}.$$

3. Для абсолютно непрерывного случайного вектора из примера 5

$$f_Z : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad f_Z(z) = \begin{cases} \frac{1}{\pi r^2}, & \text{если } \|z\| \leq r, \\ 0, & \text{если } \|z\| > r \end{cases}$$

аналогичные выкладки дают

$$M(Z) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}^T; \quad \text{cov}(Z) = \begin{bmatrix} \frac{r^2}{4} & 0 \\ 0 & \frac{r^2}{4} \end{bmatrix}.$$

4. Для нормально распределенного случайного вектора с параметрами m и B (пример 3), не приводя достаточно громоздких выкладок, укажем, что $M(X) = m$, $\text{cov}(X) = B$.

Иногда вместо ковариационной матрицы используют *корреляционную матрицу*, элементы которой определяются формулой

$$r_{jk} = \frac{b_{jk}}{\sqrt{b_{jj} \cdot b_{kk}}}.$$

Очевидно, что $r_{ii} = 1$. Покажем, что $|r_{ik}| \leq 1$. Действительно, применив доказанную выше теорему к вектору

$$\alpha = \lambda e^{(j)} - e^{(k)}$$

(здесь $\lambda \in \mathbb{R}$, а $e^{(j)}$ и $e^{(k)}$ – векторы стандартного базиса), получим

$$b_{jj}^2 \lambda^2 - 2b_{jk} \lambda + b_{kk}^2 \geq 0.$$

Известно, что квадратный трехчлен *всюду* неотрицателен, только если его дискриминант неположителен. Поэтому $b_{jk}^2 - (b_{jj} b_{kk})^2 \leq 0$ или $r_{jk}^2 \leq 1$, т.е. $|r_{jk}| \leq 1$.

Замечание. Равенство $|r_{jk}| = 1$ означает, что координаты случайного вектора $X_j - M(X_j)$ и $X_k - M(X_k)$ линейно зависимы⁵ (см. (2.5.2)).

Число r_{jk} называется *коэффициентом корреляции* между случайными переменными X_j и X_k . Если $r_{jk} = 0$, то с.п. X_j и X_k называются *некоррелированными*.

Покажем, что независимые случайные переменные некоррелированы. Пусть $f_X(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$. Тогда

$$\begin{aligned} b_{12} &= \int_{\mathbb{R}^2} (x_1 - m_1)(x_2 - m_2) f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 = \\ &= \left(\int_{-\infty}^{+\infty} (x_1 - m_1) f_{X_1}(x_1) dx_1 \right) \times \left(\int_{-\infty}^{+\infty} (x_2 - m_2) f_{X_2}(x_2) dx_2 \right). \end{aligned}$$

⁵ Не путайте линейную зависимость со статистической!

Легко видеть, что каждый из полученных интегралов равен нулю. Поэтому $b_{12} = 0$.

Серьезное предупреждение. В рассмотренных выше примерах (случайные векторы, равномерно распределенные на прямоугольнике и на круге) координаты случайных векторов оказались *некоррелированными*. В то же время в предыдущем пункте было показано, что координаты первого случайного вектора независимы, а второго – зависимы. Следовательно, равенство нулю коэффициента корреляции между двумя с.п. не несет никакой информации о наличии статистической зависимости между ними. К сожалению, слово "корреляция" иногда употребляется в смысле "статистическая зависимость". Исторические причины этой путаницы будут изложены ниже.

Замечание. Можно показать, что координаты нормального случайного вектора – нормально распределенные с.п.:

$$f_{X_k}(x_k) = \frac{1}{\sqrt{2\pi} \cdot \sigma_k} \cdot \exp\left(-\frac{(x_k - m_k)^2}{2\sigma_k^2}\right) \quad (\sigma_k = \sqrt{b_{kk}}).$$

Если координаты *нормального* случайного вектора попарно некоррелированы, то $B = \text{diag} [\sigma_1^2, \dots, \sigma_n^2]$ и

$$f_X(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n \sigma_1 \dots \sigma_n} \times \\ \times \exp\left(-\frac{(x_1 - m_1)^2}{2\sigma_1^2} - \dots - \frac{(x_n - m_n)^2}{2\sigma_n^2}\right) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \dots f_{X_n}(x_n).$$

Таким образом (в исключение из общего правила), некоррелированность координат *нормального* случайного вектора равносильна их независимости в совокупности.

2.6. Преобразования случайных переменных

Пусть задан случайный вектор X , т.е. задано пространство элементарных событий $\Omega_X \subset \mathbb{R}^n$, и на нем заданы либо вероятности (для дискретного случайного вектора), либо плотность распределения вероятностей (для абсолютно непрерывного случайного вектора).

Зададим теперь функцию $\phi: \Omega_X \rightarrow \mathbb{R}^m$ и обозначим множество ее значений Ω_Y . Тогда на Ω_Y естественным образом задается распределение вероятностей, порожаемое распределением вероятностей на Ω_X : вероятность каждого события $A \subset \Omega_Y$ равна вероятности его полного прообраза, т.е. множества всех точек из Ω_X , которые переводятся в A функцией ϕ . Таким образом определяется новый случайный вектор, обозначаемый $Y = \phi(X)$.

Легко видеть, что распределение вероятностей образа *дискретного* случайного вектора $Y = \phi(X)$ вычисляется по формуле

$$\mathbf{P}(Y = y) = \sum_{\{x \mid \phi(x)=y\}} \mathbf{P}(X = x).$$

Пусть, например, $\Omega_X = \{-3, -2, -1, 0, 1, 2, 3\}$ и эти значения *равновероятны*. Зададим функцию $g: \Omega_X \rightarrow \mathbb{R}$ формулой $g(x) = x^2$. Видно, что множество значений с.п. $Y = X^2$ $\Omega_Y = \{0, 1, 4, 9\}$. При этом у значения 0 один прообраз, а у остальных значений – по два. Теперь легко сосчитать вероятности всех значений. С.п. Y задана таблицей:

y	0	1	4	9
$\mathbf{P}(y)$	1/7	2/7	2/7	2/7

Вопрос о преобразовании плотности распределения абсолютно непрерывного случайного вектора легко решается в ситуации, когда ϕ – гладкая функция, взаимно однозначно отображающая $\Omega_X \subset \mathbb{R}^n$ на $\Omega_Y \subset \mathbb{R}^n$, причем $\det(\phi') \neq 0$. В этом случае $X \in G \iff Y \in \phi(G)$. Поэтому $\mathbf{P}(X \in G) = \mathbf{P}(Y \in \phi(G))$, и по теореме о преобразовании интеграла

$$\int_G f_X = \int_{\phi(G)} f_Y = \int_G (f_Y \circ \phi) \cdot |\det(\phi')|. \quad (2.6.1)$$

Поскольку это равенство должно быть выполнено для любой области G с кусочно гладкой границей, подынтегральные функции слева и справа в (2.6.1) совпадают, откуда

$$f_Y(\phi(x)) = \frac{f_X(x)}{|\det(\phi'(x))|}.$$

Если отображение ϕ не является взаимно однозначным, то вопрос о распределении $Y = \phi(X)$ более сложен. Остановимся подробнее на случае, когда $\phi: \Omega_X \rightarrow \mathbb{R}$ – функционал.

Воспользуемся свойством (2.1.2) функции распределения:

$$F_Y(y) = \mathbf{P}(\phi(X) < y) = \int_{\{x \mid \phi(x) < y\}} f_X. \quad (2.6.2)$$

Если удастся вычислить функцию распределения, то плотность f_Y находится из нее дифференцированием.

Примеры. 1. Мгновенное значение синусоидального напряжения определяется по формуле $u = U_{max} \cdot \sin(\varphi)$. Измерение мгновенного значения производится в *случайный* момент времени, т.е. предполагается, что с.п. Φ (фаза) распределена равномерно на $[-\pi, \pi[$:

$$f_\Phi(\varphi) = \begin{cases} \frac{1}{2\pi} & \text{при } -\pi \leq \varphi < \pi, \\ 0 & \text{при } \varphi < -\pi \text{ или } \varphi \geq \pi. \end{cases}$$

При этих условиях определена с.п. U , значения которой заполняют сегмент $\Omega_U = [-U_{max}, U_{max}]$. Найдем плотность ее распределения, положив для упрощения, что $U_{max} = 1$.

Начнем с построения $F_U(u) = \mathbf{P}(U < u)$.

Очевидно, что $F_U(u) = 0$ при $u \leq -1$ и $F_U(u) = 1$ при $u > 1$.

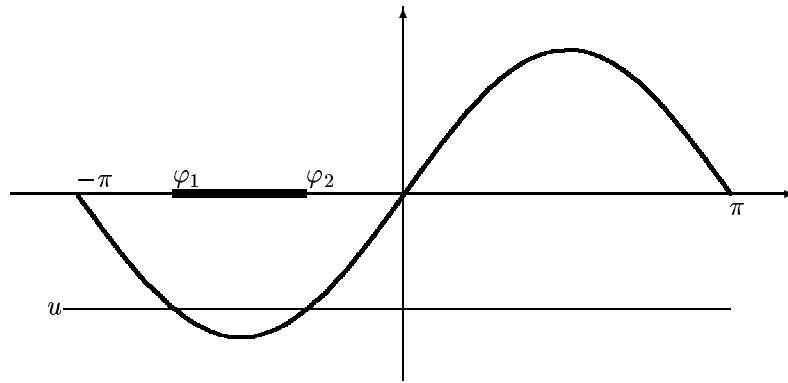


Рис. 3:

На рис.2.3 выделен участок оси абсцисс, на котором $U < u$, для случая $u \in]-1, 0[$, а на рис.2.4 – для случая $u \in [0, 1]$.

Из рис.2.3 видно, что при $-1 < u < 0$

$$F_U(u) = \mathbf{P}(U < u) = \frac{\varphi_2 - \varphi_1}{2\pi} = \frac{\pi + 2\arcsin(u)}{2\pi}.$$

Аналогично, из рис.2.4 видно, что при $0 \leq u \leq 1$

$$F_U(u) = \mathbf{P}(U < u) = \frac{\varphi_1 + 2\pi - \varphi_2}{2\pi} = \frac{\pi + 2\arcsin(u)}{2\pi}.$$

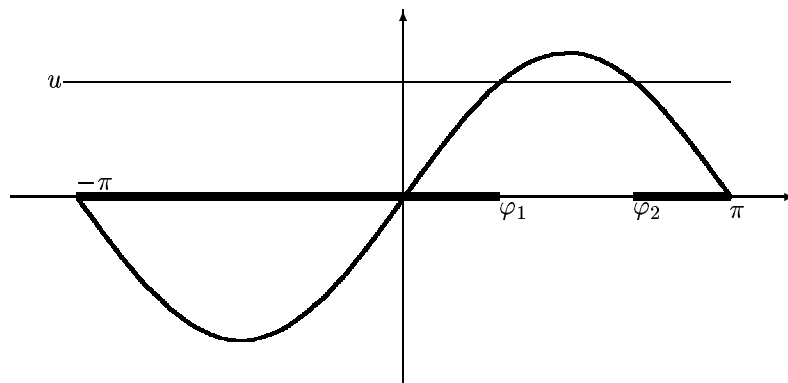


Рис. 4:

Итак,

$$F_U(u) = \begin{cases} 0 & \text{при } u \leq -1, \\ \frac{\pi + 2\arcsin(u)}{2\pi} & \text{при } -1 \leq u \leq 1, \\ 1 & \text{при } u \geq 1. \end{cases}$$

Дифференцируя, найдем плотность распределения

$$f_U(u) = F'_U(u) = \begin{cases} \frac{1}{\pi\sqrt{1-u^2}} & \text{при } |u| < 1, \\ \text{не определена} & \text{при } |u| = 1, \\ 0 & \text{при } |u| > 1. \end{cases}$$

Мы получили так называемый *закон арксинуса*.

2. Задан двухмерный случайный вектор X с плотностью распределения f_X . Найдем распределение с.п. $Y = X_1 + X_2$.

Формула (2.6.2) дает

$$F_Y(y) = \int_{\{x \mid x_1 + x_2 < y\}} f_X(x_1, x_2) dx_1 dx_2.$$

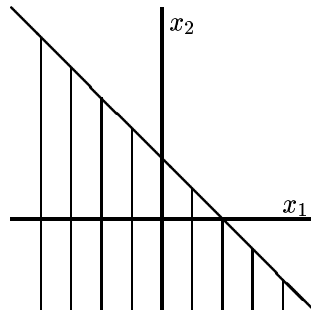


Рис. 5: $x_1 + x_2 < y$

Как известно, множество $\{x \in \mathbb{R}^2 \mid x_1 + x_2 < y\}$ – полуплоскость (рис.2.5).

Преобразуем двойной интеграл в повторный:

$$F_Y(y) = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{y-x_1} f_X(x_1, x_2) dx_2 = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^y f_X(x_1, z - x_1) dz$$

(во внутреннем интеграле мы сделали подстановку $x_2 = z - x_1$).

Теперь преобразуем повторный интеграл обратно в двойной, а затем – в повторный в другом порядке ("поменяем порядок интегрирования"):

$$F_Y(y) = \int_{-\infty}^y dz \int_{-\infty}^{+\infty} f_X(x_1, z - x_1) dx_1.$$

Дифференцируя по y , получим плотность распределения:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_X(x_1, y - x_1) dx_1. \quad (2.6.3)$$

Замечание. Если координаты x_1 и x_2 *независимы*, то формула (2.6.3) перепишется так:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X_1}(x_1) \cdot f_{X_2}(y - x_1) dx_1 = (f_{X_1} * f_{X_2})(y). \quad (2.6.4)$$

Плотность распределения суммы двух независимых случайных переменных равна свертке плотностей распределения слагаемых.

Рассмотрим два примера:

3. Найдем распределение суммы координат двухмерного *нормального* случайного вектора.

Подставляя в формулу (2.4.2)

$$m = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}; \quad B = \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

(r – коэффициент корреляции), получим

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times \\ \times \exp\left(-\frac{\left(\frac{x_1-m_1}{\sigma_1}\right)^2 - 2r \cdot \frac{x_1-m_1}{\sigma_1} \cdot \frac{x_2-m_2}{\sigma_2} + \left(\frac{x_2-m_2}{\sigma_2}\right)^2}{2(1-r^2)}\right).$$

Если подставить f_X в формулу (2.6.3) и вычислить интеграл, то получим

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right),$$

где $m = m_1 + m_2$; $\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2r \cdot \sigma_1 \cdot \sigma_2$.

Можно показать, что и в n -мерном случае сумма любого числа координат нормального случайного вектора имеет нормальное распределение.

4. Если двухмерный случайный вектор X равномерно распределен на квадрате $[-1, 1] \times [-1, 1]$, то, как было показано в п.2.4, его координаты независимы и равномерно распределены на сегменте $[-1, 1]$:

$$f_{X_1}(t) = f_{X_2}(t) = \begin{cases} \frac{1}{2} & \text{при } |t| \leq 1, \\ 0 & \text{при } |t| > 1. \end{cases}$$

По формуле (2.6.4)

$$f_{X_1+X_2}(y) = \int_{-\infty}^{+\infty} f_{X_1}(t) \cdot f_{X_2}(y-t) dt.$$

Подынтегральная функция здесь отлична от нуля только при выполнении одновременно двух неравенств: $-1 \leq t \leq 1$ и $-1 \leq y-t \leq 1$ (в этом случае она равна $\frac{1}{4}$).

Поэтому $f_{X_1+X_2}(y) = \frac{1}{4} \cdot \Delta$, где Δ – длина общей части промежутков $[-1, 1]$ и $[y-1, y+1]$.

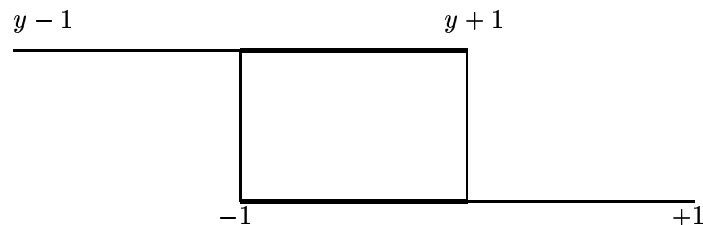


Рис. 6: $-2 \leq y \leq 0$

Из рис.2.6 видно, что при $y \in [-2, 0]$ $\Delta = (y+1) - (-1) = y+2$. Аналогично, при $y \in [0, 2]$ $\Delta = 2 - y$. Наконец, очевидно, что при $|y| \geq 2$ $\Delta = 0$.

График плотности распределения X_1+X_2 (так называемое *треугольное* распределение) изображен на рис.2.7.

Вычислим теперь математическое ожидание и ковариационную матрицу образа случайного вектора.

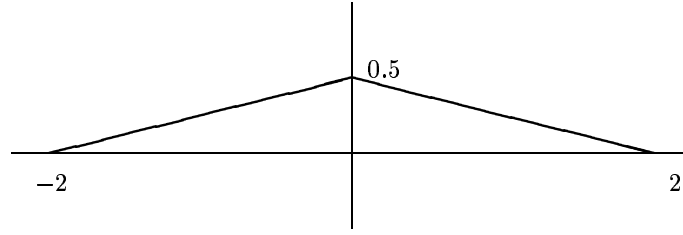


Рис. 7:

Пусть X – n -мерный дискретный случайный вектор, $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $Y = \phi(X)$. Тогда из определения видно, что

$$M(Y) = \sum_x \phi(x) \cdot \mathbf{P}(X = x). \quad (2.6.5)$$

Можно показать, что и для абсолютно непрерывного случайного вектора X справедлива аналогичная формула

$$M(Y) = \int_{\mathbb{R}^n} \phi(x) \cdot f_X(x) \quad (2.6.5')$$

(как всегда, предполагаем, что ряд или интеграл абсолютно сходится).

Замечание. Формулы (2.2.1), (2.2.1') можно теперь переписать так:

$$D(X) = M\left((X - M(X))^2\right).$$

Дисперсия с.п. – это математическое ожидание квадрата отклонения с.п. от ее математического ожидания.

Аналогично, из определения ковариационной матрицы для дискретного случайного вектора видно, что

$$B(Y) = \sum_x (\phi(x) - M(Y)) \cdot (\phi(x) - M(Y))^T \cdot \mathbf{P}(X = x), \quad (2.6.6)$$

и можно показать, что для абсолютно непрерывного случайного вектора

$$B(Y) = \int_{\mathbb{R}^n} (\phi(x) - M(Y)) \cdot (\phi(x) - M(Y))^T \cdot f_X(x). \quad (2.6.6')$$

Пример. В приложениях часто приходится вычислять математическое ожидание и дисперсию линейной комбинации случайных переменных.

Пусть X – n -мерный случайный вектор, $Y = \alpha_1 X_1 + \dots + \alpha_n X_n$. В этом случае $\phi(x) = \langle x, \alpha \rangle$, где $\alpha = [\alpha_1, \dots, \alpha_n]^T$. Поэтому, вынося в формулах (2.6.5), (2.6.5') постоянный множитель α за знак суммы (интеграла), получаем

$$M(Y) = \langle M(X), \alpha \rangle$$

(в предположении, что $M(X)$ существует).

Аналогично, подставив $\phi(x) = \langle x, \alpha \rangle$ и $M(Y) = \langle M(X), \alpha \rangle$ в формулы (2.6.6), (2.6.6'), можно вынести постоянные множители за знак суммы (интеграла) и получить

$$D(Y) = \langle B(X) \alpha, \alpha \rangle$$

(если существует $B(X)$).

В важном частном случае, когда $Y = \sum_{k=1}^n X_k$, имеем $\alpha = [1, \dots, 1]^T$. Поэтому

$$M(Y) = \sum_{k=1}^n M(X_k); \quad D(Y) = \sum_{j=1}^n \sum_{k=1}^n b_{jk}.$$

Если координаты случайного вектора имеют математические ожидания, то математическое ожидание суммы координат равно сумме математических ожиданий слагаемых. Если координаты случайного вектора имеют дисперсии, то дисперсия суммы координат равна сумме всех элементов ковариационной матрицы.

Отметим очень важное обстоятельство: *дисперсия суммы координат, вообще говоря, не равна сумме дисперсий этих координат*. Она может быть и больше, и меньше.

Если считать дисперсию с.п. мерой "неупорядоченности" этой с.п., то при суммировании случайных переменных эта "неупорядоченность" может как увеличиваться, так и уменьшаться.

Если же координаты случайного вектора *попарно некоррелированы*, то дисперсия суммы координат равна сумме дисперсий слагаемых.

2.7. Датчик псевдослучайных чисел

Задача о преобразовании случайной переменной часто встречается в приложениях (например, задача о преобразовании "случайной помехи" при ее прохождении через приемно-усилительное устройство). В силу сложности структуры физических преобразователей "аналитические" методы (некоторые из них рассматривались в предыдущем пункте) нередко оказываются неприменимыми. Хорошие результаты дает в этом случае машинное моделирование: алгоритм преобразования, осуществляемого физическим устройством, реализуется в виде машинной программы, на вход которой подается достаточно длинный набор чисел, имитирующих значения случайной переменной ("помехи"). Вопрос о качестве имитации будет рассмотрен в главе 3. Числа эти именуются *псевдослучайными*.

В библиотеках стандартных подпрограмм и в средах конечного пользователя имеются программные датчики (генераторы) псевдослучайных чисел. В их основе лежат алгоритмы, вырабатывающие псевдослучайные числа, равномерно распределенные на $]0, 1[$. Устройство таких алгоритмов весьма сложно и не может рассматриваться в нашем курсе. Взяв достаточно большую выборку, можно убедиться, что относительные частоты попадания псевдослучайных чисел в части интервала $]0, 1[$ примерно пропорциональны длинам этих частей.

Имея равномерный датчик, можно конструировать датчики с заданными законами распределения. Покажем, как это делается, на примере абсолютно непрерывной с.п.

Пусть X – с.п., равномерно распределенная на $]0, 1[$, и Ψ – заданная функция распределения (ψ – соответствующая плотность распределения). Рассмотрим объединение промежутков, на которых $\psi > 0$. На этом множестве (обозначим его Q) функция Ψ возрастает и, следовательно, имеет обратную. Поскольку $0 \leq \Psi \leq 1$, то Ψ^{-1} определена почти всюду на $[0, 1]$, т.е. на множестве значений с.п. X . Докажем, что с.п. $Y = \Psi^{-1}(X)$ имеет функцию распределения Ψ .

Действительно, из возрастания Ψ следует возрастание Ψ^{-1} , и

$$\Psi^{-1}(X) < y \iff X < \Psi(y).$$

Отсюда

$$\begin{aligned} F_Y(y) &= \mathbf{P}(Y < y) = \mathbf{P}(\Psi^{-1}(X) < y) = \mathbf{P}(X < \Psi(y)) = \\ &= \int_{-\infty}^{\Psi(y)} f_X(x) dx = \int_0^{\Psi(y)} 1 \cdot dx = \Psi(y). \end{aligned}$$

Построим, например, датчик *экспоненциального* распределения с параметром ($\alpha > 0$) :

$$F_Y(y) = \begin{cases} 1 - \exp(-\alpha \cdot y) & \text{при } y > 0, \\ 0 & \text{при } y \leq 0. \end{cases}$$

Здесь $Q =]0, +\infty[$,

$$F_Y^{-1}(x) = -\frac{\ln(1-x)}{\alpha} \quad \text{при } 0 < x < 1.$$

Следовательно, преобразуя числа x_k , полученные от равномерного датчика, по формуле $y_k = -\frac{1}{\alpha} \cdot \ln(1-x_k)$, мы построим набор значений экспоненциальной с.п.

Отметим, что у описанного метода есть трудность: необходимо уметь вычислять значения функции, обратной к F_Y .

2.8. Центральная предельная теорема

Рассмотрим n -мерный абсолютно непрерывный случайный вектор $X = [X_1, \dots, X_n]^T$ с одинаково распределенными и независимыми в совокупности координатами. Будем предполагать, что координаты с.в. имеют математическое ожидание m , дисперсию σ^2 и *абсолютный третий центральный момент*, определяемый равенством

$$\mu_3(X_k) = M(|X_k - M(X_k)|^3).$$

Введем новые случайные переменные $Y_k = \frac{X_k - m}{\sigma\sqrt{n}}$. Из формул, полученных в п.2.6, видно, что

$$M(Y_k) = 0, \quad D(Y_k) = \frac{1}{n}, \quad \mu_3(Y_k) = \frac{\mu_3(X_k)}{\sigma^3 n \sqrt{n}}. \quad (2.8.1)$$

Обозначим f_Y общую плотность распределения с.п. Y_k и рассмотрим ее преобразование Фурье:

$$\widetilde{f_Y}(\omega) = \int_{-\infty}^{+\infty} \exp(-i\omega y) \cdot f_Y(y) dy. \quad (2.8.2)$$

Формула Тейлора третьего порядка дает

$$\exp(-i\omega y) = 1 - i\omega y - \frac{(\omega y)^2}{2} + i \frac{(\omega y)^3}{6} \cdot g(\omega y),$$

где $g(\omega y) = \cos(\gamma_1) - i \cdot \sin(\gamma_2)$, γ_1 и γ_2 – некоторые точки между 0 и ωy .

Подставим это выражение в (2.8.2):

$$\begin{aligned} \widetilde{f_Y}(\omega) &= \int_{-\infty}^{+\infty} f_Y(y) dy - i\omega \int_{-\infty}^{+\infty} y f_Y(y) dy - \frac{\omega^2}{2} \int_{-\infty}^{+\infty} y^2 f_Y(y) dy + \\ &+ i \frac{\omega^3}{6} \int_{-\infty}^{+\infty} g(\omega y) y^3 f_Y(y) dy. \end{aligned} \quad (2.8.3)$$

Очевидно, что первый интеграл в этой сумме равен единице. Далее, в силу (2.8.1) второй интеграл равен нулю, а третий – $\frac{1}{n}$.

Оценим модуль последнего слагаемого в (2.8.3):

$$\left| i \frac{\omega^3}{6} \int_{-\infty}^{+\infty} g(\omega y) y^3 f_Y(y) dy \right| \leq \left| \frac{\omega^3}{3} \int_{-\infty}^{+\infty} |y|^3 f_Y(y) dy \right| = \left| \frac{\mu_3(Y_k) \omega^3}{3} \right|.$$

Поэтому (2.8.3) можно переписать так:

$$\widetilde{f_Y}(\omega) = 1 - \frac{\omega^2}{2n} + \frac{\omega^3 h(\omega)}{n \sqrt{n}}, \quad (2.8.4)$$

где $|h(\omega)| \leq \frac{\mu_3(X_k)}{3\sigma^3}$.

Пусть теперь $Z_n = \sum_{k=1}^n Y_k$. Плотность распределения случайной переменной Z_n – суммы (независимых в совокупности) с.п. Y_k – есть, как показано в п.2.6, свертка их плотностей распределения, а образ Фурье свертки, как указано в главе "Приближение функций", есть произведение образов Фурье этих плотностей, т.е.

$$\widetilde{f_{Z_n}}(\omega) = \left(\widetilde{f_Y}(\omega) \right)^n \quad (2.8.5)$$

(образы Фурье всех с.п. Y_k одинаковы).

Подставим (2.8.4) в (2.8.5):

$$\widetilde{f_{Z_n}}(\omega) = \left(1 - \frac{\omega^2}{2n} + \frac{\omega^3 h(\omega)}{n \sqrt{n}} \right)^n.$$

Выясним теперь, что произойдет при неограниченном увеличении n . Логарифмируя $\widetilde{f_{Z_n}}$ и вспоминая степенной ряд для логарифма, получим

$$\ln \left(\widetilde{f_{Z_n}} \right) = n \cdot \ln \left(1 - \frac{\omega^2}{2n} + \dots \right) = n \cdot \left(-\frac{\omega^2}{2n} + \dots \right) \xrightarrow{n \rightarrow +\infty} -\frac{\omega^2}{2}$$

(многоточие обозначает совокупность слагаемых, содержащих степени $1/n$ выше первой).

Отсюда находим предел образов Фурье с.п. Z_n :

$$\lim_{n \rightarrow +\infty} \left(\widetilde{f_{Z_n}} \right) = \exp \left(-\frac{\omega^2}{2} \right).$$

Таким образом, предел $\widetilde{f_{Z_n}}$ не зависит от распределения компонент исходного вектора X .

Терминологическое замечание. Поскольку

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n (X_k - m), \quad M(Z_n) = 0, \quad D(Z_n) = 1,$$

случайную переменную Z_n называют *центрированной и нормированной суммой* координат с.в. X .

Можно показать, что $\exp\left(-\frac{\omega^2}{2}\right)$ – преобразование Фурье плотности *стандартного* нормального распределения, т.е. нормального распределения с нулевым математическим ожиданием и единичной дисперсией.

Итак, если координаты абсолютно непрерывного случайного вектора

- одинаково распределены,
- независимы в совокупности,
- имеют дисперсию и абсолютный третий центральный момент,

то при неограниченном увеличении размерности вектора образ Фурье плотности центрированной и нормированной суммы координат (*при любом, но одинаковом их распределении*) стремится к образу Фурье плотности стандартного нормального распределения.

Этот факт не случаен. Имеет место

Центральная предельная теорема. Пусть (X_n) – последовательность с.п., имеющих дисперсии и абсолютные третьи центральные моменты. Пусть при любом n координаты случайного вектора $[X_1, \dots, X_n]^T$ независимы в совокупности. Обозначим

$$\mathcal{D}_n = \sum_{k=1}^n D(X_k), \quad \mathcal{M}_n = \sum_{k=1}^n \mu_3(X_k)$$

и предположим, что

$$\lim_{n \rightarrow +\infty} \mathcal{M}_n / (\mathcal{D}_n)^{3/2} = 0. \quad (2.8.6)$$

Тогда распределение последовательности центрированных и нормированных сумм

$$\left(\frac{1}{\sqrt{\mathcal{D}_n}} \sum_{k=1}^n (X_k - M(X_k)) \right)_{n=1}^{+\infty}$$

сходится к стандартному нормальному распределению.

Слово "сходится" здесь означает, что для любого сегмента $[a, b]$

$$\mathbf{P} \left(\frac{1}{\sqrt{\mathcal{D}_n}} \sum_{k=1}^n (X_k - M(X_k)) \in [a, b] \right) \xrightarrow{n \rightarrow +\infty} \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{t^2}{2}\right) dt.$$

Эта теорема была доказана А.М. Ляпуновым.

Замечание. В рассмотренном нами случае, когда все X_k имеют одинаковые распределения, имеем $\mathcal{D}_n = nD(X_k)$, $\mathcal{M}_n = n\mu_3(X_k)$, и условие (2.8.6) выполнено автоматически.

Подчеркнем еще раз, что в центральной предельной теореме случайные переменные X_k могут быть любыми (не обязательно абсолютно непрерывными); более того, их распределения *могут быть различными*, лишь бы наборы X_1, \dots, X_n *при всех n* были независимы в совокупности и выполнялось условие (2.8.6).

Из центральной предельной теоремы следует, что если бы всегда можно было иметь дело с суммой *большого* числа таких случайных переменных, то единственным нужным распределением стало бы нормальное. Именно поэтому нормальное распределение играет исключительную роль в теории вероятностей и ее приложениях.

Серьезное предупреждение. Возможно, под влиянием центральной предельной теоремы возникло распространенное суеверие, суть которого сводится к тому, что в прикладных задачах *все* случайные переменные обязаны иметь нормальное распределение. В частности, бытует мнение, что случайные погрешности измерительных приборов всегда имеют нормальное распределение. Эта точка зрения опровергается экспериментом П.В. Новицкого⁶, обработавшего большое количество протоколов поверки приборов и не обнаружившего ожидавшейся нормальности погрешностей.

Упомянутая выше ошибка – отождествление независимости и некоррелированности – также проистекает, по-видимому, из убеждения, что все случайные переменные – нормальные.

⁶Петр Васильевич НОВИЦКИЙ (1922-2000) – профессор Санкт-Петербургского технического университета, основатель и разработчик информационных критериев качества измерительных устройств.