Комментарии к теме «*Математическое ожидание*»

Практические занятия по теории вероятностей

кафедра статистического моделирования `http://statmod.ru`, матмех СПбГУ, 2014 г.

1 Математическое ожидание

1.1 Определение и основные свойства

Определение 1.1. Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство и $\xi : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B})$ — некоторая случайная величина. Если интеграл $\int_\Omega |\xi|\, d\mathbb{P}$ конечен, то интеграл $\mathbb{E}\xi = \int_\Omega \xi\, d\mathbb{P}$ называется *математическим ожиданием* случайной величины ξ.[1]

Замечание 1.1. 1. Поскольку мера $\mathbb{P}$ — вероятностная (то есть нормированная на 1), то математическое ожидание — это *среднее значение* случайной величины. Иногда эти термины используют как эквивалентные.
2. Пусть $\int_\Omega |\xi|\, d\mathbb{P} = \infty$, но $\int_\Omega \xi^-\, d\mathbb{P} < \infty$.[2] Тогда (допуская некоторую вольность речи) говорят, что математическое ожидание случайной величины ξ равно $+\infty$ и пишут $\mathbb{E}\xi = +\infty$. Аналогичным образом интерпретируется запись $\mathbb{E}\xi = -\infty$. Если $\int_\Omega \xi^+\, d\mathbb{P} = \int_\Omega \xi^-\, d\mathbb{P} = \infty$, то математического ожидания случайной величины ξ *не существует*.

Простейшие свойства математического ожидания.

1. Если $\xi = \mathrm{const} = c$, то[3] $\mathbb{E}\xi = c$.
2. Если у ξ существует математическое ожидание, то для любых постоянных a и b имеет место равенство $\mathbb{E}(a\xi + b) = a\mathbb{E}\xi + b$.
3. Если $\xi \geq \eta$, то $\mathbb{E}\xi \geq \mathbb{E}\eta$. В частности, математическое ожидание положительной случайной величины положительно. Если $a \leq \xi \leq b$, то $a \leq \mathbb{E}\xi \leq b$.[4]
4. Если ξ и η имеют конечные математические ожидания, то $\mathbb{E}(\xi + \eta) = \mathbb{E}\xi + \mathbb{E}\eta$.[5]
5. Если ξ и η независимы, то $\mathbb{E}\xi\eta = \mathbb{E}\xi\, \mathbb{E}\eta$ (если только все математические ожидания, входящие в это равенство, конечны).

Замечание 1.2. 1. Конечно, первые 4 свойства — это просто свойства интегралов по вероятностной мере. Что касается пятого свойства, то оно носит специфически вероятностный характер.[6]

[1] Обозначение математического ожидания буквой $\mathbb{E}$ можно объяснить использованием аббревиатуры английского «Expectation».

[2] Здесь ξ^- — это отрицательная часть функции ξ, а ξ^- — положительная. Они по определению неотрицательны и имеют дизъюнктные носители, так что $\xi = \xi^+ - \xi^-$, а $|\xi| = \xi^+ + \xi^-$.

[3] Среднее значение постоянной равно этой постоянной.

[4] Это свойство может быть полезно при проверке правдоподобности результатов вычислений.

[5] Понятно ведь, что среднее значение суммы должно равняться сумме средних значений!

[6] Ход доказательства этого свойства такой. Если ξ и η — индикаторы независимых событий, то искомое равенство превращается просто в определение независимости двух событий. Далее рассуждения стандартны для математического анализа: от индикаторов мы переходим к их линейным комбинациям (то есть к «простым» функциям), а потом предельный переход завершает все построения.

2. Важно отметить, что свойство, обратное к пятому, вообще говоря, неверно. Иначе говоря, из равенства $\mathbb{E}\xi\eta = \mathbb{E}\xi\,\mathbb{E}\eta$ не следует, что ξ и η независимы.[7]

Представление математических ожиданий через распределения случайных величин. Возникает естественный вопрос — а как можно считать математические ожидания? Соберем нужные нам (и связанные между собой) утверждения в одно предложение.

Предложение 1.1. *1. Пусть у случайной величины ξ с распределением $\mathcal{P}_\xi$ существует математическое ожидание $\mathbb{E}\xi$. Тогда*

$$\mathbb{E}\xi = \int_{\mathbb{R}} x\,\mathcal{P}_\xi(dx). \tag{1.1}$$

2. Пусть случайная величина ξ имеет распределение $\mathcal{P}_\xi$ и g — измеримая функция $(\mathbb{R},\mathcal{B}) \mapsto (\mathbb{R},\mathcal{B})$. Если существует математическое ожидание $\mathbb{E}g(\xi)$, то

$$\mathbb{E}g(\xi) = \int_{\mathbb{R}} g(x)\,\mathcal{P}_\xi(dx). \tag{1.2}$$

3. Пусть случайный вектор $\overline{\xi} = (\xi_1,\ldots,\xi_d)^{\mathrm{T}}$ имеет распределение $\mathcal{P}_{\overline{\xi}}$ и g — измеримая функция $(\mathbb{R}^d,\mathcal{B}_d) \mapsto (\mathbb{R},\mathcal{B})$. Если существует математическое ожидание $\mathbb{E}g(\overline{\xi})$, то

$$\mathbb{E}g(\overline{\xi}) = \mathbb{E}g(\xi_1,\ldots,\xi_d) = \int_{\mathbb{R}^d} g(x_1,\ldots,x_d)\,\mathcal{P}_{\overline{\xi}}(dx_1\ldots dx_d). \tag{1.3}$$

Замечание 1.3. 1. Формулы (1.1) — (1.3) — это ни что другое, как примеры абстрактной замены переменной в интеграле по мере.
2. Конечно, равенство (1.2) следует из (1.3), а (1.1) — из (1.2).
3. На самом деле из существования (в том же смысле, в котором мы говорили о существовании математических ожиданий) правых частей равенств (1.1) — (1.3) следует существование левых частей этих равенств.
4. Из равенств (1.1) — (1.3) легко выводятся простейшие свойства 1 — 5 математических ожиданий.

Формулы (1.1) — (1.3) дают возможность явно вычислять математические ожидания в том случае, когда распределение случайной величины ξ (или случайного вектора $\overline{\xi}$) является дискретным или абсолютно непрерывным.

Действительно, пусть распределение $\mathcal{P}_\xi$ случайной величины ξ задано таблицей распределения

$$\mathcal{P}_\xi: \quad \begin{pmatrix} x_1 & x_2 & \ldots & x_n & \ldots \\ p_1 & p_2 & \ldots & p_n & \ldots \end{pmatrix}. \tag{1.4}$$

Тогда (1.1) и (1.2) превращаются в

$$\mathbb{E}\xi = \sum_{n\ge 1} x_n p_n \quad \text{и} \quad \mathbb{E}g(\xi) = \sum_{n\ge 1} g(x_n) p_n, \tag{1.5}$$

причем эти математические ожидания конечны, когда соответствующие ряды сходятся абсолютно.

Если же случайная величина ξ обладает плотностью распределения p_ξ, то (1.1) и (1.2) приобретают вид

$$\mathbb{E}\xi = \int_{\mathbb{R}} x p_\xi(x)\,dx \quad \text{и} \quad \mathbb{E}g(\xi) = \int_{\mathbb{R}} g(x) p_\xi(x)\,dx, \tag{1.6}$$

а конечности этих математических ожиданий соответствует суммируемость функций $xp_\xi(x)$ и $g(x)p_\xi(x)$ по мере Лебега.

Аналогично, если случайный вектор $\overline{\xi} = (\xi_1,\ldots,\xi_d)^{\mathrm{T}}$ обладает плотностью распределения $p_{\overline{\xi}}$, то (1.3) переходит в

$$\mathbb{E}g(\xi_1,\ldots,\xi_d) = \int_{\mathbb{R}^d} g(x_1,\ldots,x_d) p_{\overline{\xi}}(x_1,\ldots,x_d)\,dx_1\ldots dx_d\,. \tag{1.7}$$

[7] Чтобы убедиться в этом, рассмотрите случайный вектор $(\xi,\eta)^{\mathrm{T}}$, равномерно распределенный в круге единичного радиуса с центром в нуле, и примените формулы (1.6) и (1.7) для вычисления $\mathbb{E}\xi$, $\mathbb{E}\eta$ и $\mathbb{E}\xi\eta$.

Как все это используется при решении задач? Приведем по этому поводу некоторые соображения, выбрав для краткости $d = 2$.

Пусть случайный вектор $(\xi_1, \xi_2)^{\mathrm{T}}$ имеет плотность распределения $p_{\xi_1\xi_2}(x, y)$. Задача состоит в вычислении математического ожидания $\mathbb{E}g(\xi_1, \xi_2)$ (предполагается, что это математическое ожидание конечно).

Такую задачу можно решать двумя путями. Во-первых, можно найти распределение $\mathcal{P}_\eta$ случайной величины $\eta = g(\xi_1, \xi_2)$. Предположим для простоты, что это распределение абсолютно непрерывно с плотностью p_η. Тогда ответ можно представить в виде

$$\mathbb{E}g(\xi_1, \xi_2) = \mathbb{E}\eta = \int_{\mathbb{R}} x p_\eta(x) dx. \tag{1.8}$$

Вычислив (если нужно) этот одномерный интеграл, получим ответ.

Другой вариант — непосредственно использовать формулу (1.7), то есть записать

$$\mathbb{E}g(\xi_1, \xi_2) = \int_{\mathbb{R}^d} g(x_1, x_2) p_\xi(x_1, x_2)\, dx_1 dx_2\,, \tag{1.9}$$

с последующим вычислением этого двумерного интеграла.

Какой из этих способов предпочтительней? Ответ, конечно, зависит и от совместного распределения (ξ_1, ξ_2) и от функции g. Тем не менее можно высказать несколько общих соображений.

Когда мы находим распределение случайной величины $g(\xi_1, \xi_2)$, мы делаем в каком-то смысле лишнюю работу. Например, найдя плотность p_η, мы вместе с (1.8) можем написать, что

$$\mathbb{E}\,|g(\xi_1, \xi_2)| = \mathbb{E}|\eta| = \int_{\mathbb{R}} |x| p_\eta(x) dx \quad \text{или} \quad \mathbb{E}g^2(\xi_1, \xi_2) = \mathbb{E}\eta^2 = \int_{\mathbb{R}} x^2 p_\eta(x) dx$$

(здесь мы воспользовались формулой (1.2)).

Иначе говоря, в распределении случайной величины содержится гораздо больше информации, чем в ее математическом ожидании. Если она нам вся не нужна, то зачем мы трудимся над ее получением?

В случае использовании формулы (1.9) такие проблемы, конечно, отсутствуют. Кроме того,[8] равенство (1.9) верно вне зависимости от того, имеет ли случайная величина η абсолютно непрерывное распределение или нет, то есть формула (1.9) гораздо более универсальна.

С другой стороны, если распределение случайной величины η уже известно или легко считается, то вычисление одномерного интеграла (1.8) может оказаться более простой задачей, чем работа с двумерным (а в общем случае — многомерным) интегралом (1.9).

Приведем соответствующий пример.

Пример 1.1. Пусть случайные величины $\xi_1, \xi_2 \in \mathrm{U}(0, 1)$ и являются независимыми. Нужно найти математическое ожидание $\mathbb{E}|\xi_1 - \xi_2|$.
Решение. Обозначим $\beta = \xi_1 - \xi_2$. Известно,[9] что плотность распределения случайной величины β имеет вид

$$p_\beta(x) = \begin{cases} 1 - |x| & \text{при} \quad -1 \le x \le 1, \\ 0 & \text{иначе.} \end{cases}$$

Поэтому[10] случайная величина $\eta = |\beta|$ имеет плотность распределения

$$p_\eta(x) = \begin{cases} 2(1 - x) & \text{при} \ \ 0 \le x \le 1, \\ 0 & \text{иначе.} \end{cases}$$

[8] Это существенно!
[9] Да?
[10] А почему?

Следовательно,

$$\mathbb{E}|\xi_1 - \xi_2| = \int_{\mathbb{R}} x p_\eta(x) dx = 2\int_0^1 x(1-x)dx = 1/3.$$

С другой стороны, можно сразу написать, что

$$\mathbb{E}|\xi_1 - \xi_2| = \int_0^1\int_0^1 |x-y| dx dy,$$

что после несложных преобразований даст тот же результат.

Если же в условиях этого же примера необходимо найти $\mathbb{E}|\xi_1 - 3\xi_2|$, то, вполне возможно, лучше сразу же воспользоваться формулой (1.9), не тратя время на вычисление[11] распределения случайной величины $|\xi_1 - 3\xi_2|$.

2 Дисперсия, ковариация и коэффициент корреляции

2.1 Дисперсия и ковариация.

Дисперсия. Определение. Еще одной стандартной характеристикой распределения случайной величины ξ является ее *дисперсия* $\mathbb{D}\xi$. По определению, если математическое ожидание $\mathbb{E}\xi$ конечно, то

$$\mathbb{D}\xi = \mathbb{E}(\xi - \mathbb{E}\xi)^2. \tag{2.1}$$

Поскольку в определении (2.1) под знаком математического ожидания стоит неотрицательная случайная величина, то принято говорить, что дисперсия бывает конечной или бесконечной. Дисперсия $\mathbb{D}\xi$ конечна тогда и только тогда, когда $\mathbb{E}\xi^2 < \infty$, то есть когда ξ *обладает конечным вторым моментом.*[12]

Само определение (2.1) показывает смысл понятия дисперсии. Рассмотрим гильбертово пространство $\mathrm{L}^2(d\mathbb{P})$ функций, определенных на измеримом пространстве $(\Omega, \mathcal{F})$. Если $\xi \in \mathrm{L}^2(d\mathbb{P})$,[13] то $\mathbb{D}\xi$ является не чем иным, как квадратом расстояния в $\mathrm{L}^2(d\mathbb{P})$ между ξ и ее математическим ожиданием.

Таким образом, дисперсия (точнее, квадратный корень из дисперсии) является мерой разброса случайной величины вокруг ее математического ожидания. Довольно часто дисперсию обозначают σ^2, где $\sigma = \sqrt{\mathbb{D}\xi}$. Число σ называют *стандартным отклонением*[14] случайной величины ξ.

Раскрывая скобки в (2.1) и пользуясь простейшими свойствами математического ожидания, получаем, что

$$\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2. \tag{2.2}$$

Как правило, формула (2.1) оказывается удобной в теоретических построениях, а (2.2) — при вычислениях.

Пусть $\mathcal{P}_\xi$ — распределение случайной величины ξ. Тогда

$$\mathbb{D}\xi = \int_{\mathbb{R}} \big(x - \mathbb{E}\xi\big)^2 \mathcal{P}_\xi(dx) = \int_{\mathbb{R}} x^2 \mathcal{P}_\xi(dx) - \big(\mathbb{E}\xi\big)^2,$$

где $\mathbb{E}\xi$ имеет вид (1.1).

[11] (или поиск в интернете.)

[12] Дальнейшие свойства приводятся при этом предположении.

[13] Точнее, «если случайная величина ξ принадлежит соответствующему классу эквивалентности», но мы будем здесь игнорировать эти тонкости.

[14] Используют и другие термины, например, *среднеквадратическое отклонение.* Часто говорят просто «стандарт».

Соответственно, если распределение случайной величины ξ дискретно и задано таблицей (1.4), то дисперсия $\mathbb{D}\xi$ может быть представлена как

$$\mathbb{D}\xi = \sum_n (x_n - \mathbb{E}\xi)^2 p_n = \sum_n x_n^2 p_n - (\mathbb{E}\xi)^2, \tag{2.3}$$

где $\mathbb{E}\xi$ определяется первой из формул (1.5).

Аналогично, если распределение $\mathcal{P}_\xi$ обладает плотностью распределения p_ξ, то

$$\mathbb{D}\xi = \int_{\mathbb{R}} \big(x - \mathbb{E}\xi\big)^2 p_\xi(x)\, dx = \int_{\mathbb{R}} x^2 p_\xi(x) dx - \big(\mathbb{E}\xi\big)^2,$$

где $\mathbb{E}\xi$ определяется первой из формул (1.6).

Приведем теперь выражения для математических ожиданий и дисперсий случайных величин, имеющих некоторые стандартные распределения.[15]

а) Дискретные распределения.

- *Распределение Дирака* δ_a, то есть распределение постоянной случайной величины $\xi = a$. В этом случае $\mathbb{E}\xi = a$, $\mathbb{D}\xi = 0$.

- *Равномерно распределение* $\mathrm{U}(X)$ *на множестве* $X = \{x_1, \ldots, x_n\}$. Пусть $\mathbb{P}(\xi = x_i) = 1/n$ при $i = 1, \ldots, n$. Тогда[16] $\mathbb{E}\xi = \overline{x}_n \stackrel{\text{def}}{=} (x_1 + \ldots + x_n)/n$ и $\mathbb{D}\xi = \sum_{i=1}^n (x_i - \overline{x}_n)^2/n$.

- *Распределение Бернулли* $\mathrm{Ber}(p)$ с параметром p. Если $\xi \in \mathrm{Ber}(p)$, то ξ имеет смысл числа успехов в одном испытании Бернулли с вероятностью успеха p. Легко видеть, что $\mathbb{E}\xi = p$ и $\mathbb{D}\xi = p(1 - p)$.

- *Биномиальное распределение* $\mathrm{Bin}(n, p)$ с параметрами (n, p), то есть $\mathbb{P}(\xi = k) = \mathrm{C}_n^k p^k (1-p)^k$ при $k = 0, \ldots, n$. Здесь $\mathbb{E}\xi = np$ и $\mathbb{D}\xi = np(1 - p)$.

  Эти выражения могут быть получены двумя путями. Во-первых, можно впрямую посчитать суммы, используя формулы (1.5) и (2.3).

  Во-вторых, если взять независимые случайные величины $\varepsilon_1, \ldots, \varepsilon_n$. имеющие распределение $\mathrm{Ber}(p)$, и положить $\xi = \varepsilon_1 + \ldots + \varepsilon_n$, то, как известно, $\xi \in \mathrm{Bin}(n, p)$. Поэтому приведенные выражения для $\mathbb{E}\xi$ и $\mathbb{D}\xi$ считаются получаются просто по общим свойствам математического ожидания и дисперсии.[17]

- *Геометрическое распределение* $\mathrm{Geom}(p)$ с параметром p. Здесь $\mathbb{P}(\xi = k) = p(1-p)^k$ при $k \ge 0$, $\mathbb{E}\xi = (1 - p)/p$ и $\mathbb{D}\xi = (1 - p)/p^2$. Результат получается прямым подсчетом.

- *Распределение Пуассона* $\Pi(\lambda)$ с параметром $\lambda > 0$, то есть $\mathbb{P}(\xi = k) = \lambda^k e^{-\lambda}/k!$ при $k = 0, 1, \ldots$ Прямой подсчет показывает, что $\mathbb{E}\xi = \mathbb{D}\xi = \lambda$.

  С помощью теоремы Пуассона легко убедиться в правдоподобности[18] этого результата. Действительно, рассмотрим случайные величины $\xi_n \in \mathrm{Bin}(n, p_n)$, где $np_n \to \lambda > 0$ при $n \to \infty$. Тогда $\mathbb{E}\xi_n = np_n \to \lambda$ и $\mathbb{D}\xi = np_n(1-p_n) \to \lambda$. С другой стороны, по Теореме Пуассона $\mathbb{P}(\xi_n = k) \to \lambda^k e^{-\lambda}/k!$, то есть при больших n распределение $\mathrm{Bin}(n, p_n)$ близко к распределению Пуассона. Поэтому можно ожидать, что и такие их характеристики, как математическое ожидание и дисперсия, тоже близки.

---

[15] Убедитесь в том, что приведенные выражения правильны.

[16] Заметим, что именно по этим формулам вычисляются так называемые выборочное среднее и выборочная дисперсия в математической статистике. Разница состоит в том, что здесь x_i — это вещественные числа, а в статистике — случайные величины.

[17] При этом при подсчете математического ожидания независимость случайных величин ε_i не используется.

[18] Но не более!

b) Абсолютно непрерывные распределения.

- *Равномерное распределение* $\mathrm{U}(a,b)$ на отрезке $[a,b]$ с $p_\xi(x)=1/(b-a)$ при $x\in[a,b]$. В этом случае $\mathbb{E}\xi=(b-a)/2$ и $\mathbb{D}\xi=(b-a)^2/12$.

- *Показательное распределение* $\mathrm{Exp}(\lambda)$ с параметром $\lambda>0$, имеющее плотность $p_\xi(x)=\lambda e^{-\lambda x}$ при $x>0$. Здесь $\mathbb{E}\xi=1/\lambda$ и $\mathbb{D}\xi=1/\lambda^2$.

- *Нормальное распределение* $\mathrm{N}(a,\sigma^2)$ с плотностью $p_\xi(x)=e^{-(x-a)^2/2\sigma^2}/\sqrt{2\pi}\sigma$, $x\in\mathbb{R}$. Тогда $\mathbb{E}\xi=a$ и $\mathbb{D}\xi=\sigma^2$. Таким образом, параметры нормального распределения — это его среднее и дисперсия. В частности, если $\xi\in\mathrm{N}(0,1)$, то $\mathbb{E}\xi=0$ и $\mathbb{D}\xi=1$.

Дисперсия и ковариация. Свойства. Пусть теперь ξ и η — случайные величины, обладающие конечными вторыми моментами. *Ковариацией*[19] этих случайных величин называется число

$$\mathrm{Cov}(\xi,\eta)=\mathbb{E}\big(\xi-\mathbb{E}\xi\big)\big(\eta-\mathbb{E}\eta\big). \tag{2.4}$$

Тем самым $\mathrm{Cov}(\xi,\eta)$ — это ни что иное, как скалярное произведение[20] случайных величин $\xi-\mathbb{E}\xi$ и $\eta-\mathbb{E}\eta$ в $\mathrm{L}^2(d\mathbb{P})$. Заметим, что по определению $\mathrm{Cov}(\xi,\xi)=\mathbb{D}\xi$.

Раскрывая скобки в правой части (2.4), получим, что

$$\mathrm{Cov}(\xi,\eta)=\mathbb{E}\xi\eta-\mathbb{E}\xi\,\mathbb{E}\eta. \tag{2.5}$$

Ковариация легко расписывается через совместное распределение случайных величин ξ,η. Например, равенство (2.5) приобретает вид

$$\mathrm{Cov}(\xi,\eta)=\int_{\mathbb{R}^2}xy\,\mathbb{P}_{\xi\eta}(dxdy)-\left(\int_{\mathbb{R}}x\,\mathbb{P}_\xi(dx)\right)\left(\int_{\mathbb{R}}y\,\mathbb{P}_\eta(dy)\right),$$

переход к частным случая абсолютно непрерывного или дискретного совместного распределения (ξ,η) не представляет труда.[21]

Перечислим основные свойства дисперсий и ковариаций. Удобнее начать со свойств ковариации.[22]

1. $\mathrm{Cov}^2(\xi,\eta)\le\mathbb{D}\xi\,\mathbb{D}\eta$.

2. $\mathrm{Cov}(a\xi+b,c\eta+d)=ac\mathrm{Cov}(\xi,\eta)$. В частности, из этого следует, что при рассмотрении ковариаций можно всегда считать, что математические ожидания случайных величин равны нулю (они все равно вычитаются — см. определение (2.4)).

3. $\mathrm{Cov}(\xi_1+\xi_2,\eta)=\mathrm{Cov}(\xi_1,\eta)+\mathrm{Cov}(\xi_2,\eta)$.

4. Если ξ и η независимы, то $\mathrm{Cov}(\xi,\eta)=0$.
   Случайные величины, для которых $\mathrm{Cov}(\xi,\eta)=0$, называются *некоррелированными*. Тем самым независимые случайные величины (с конечными вторыми моментами) являются некоррелированными. Обратное, вообще говоря, неверно.[23]

Теперь свойства дисперсии.

[19] Термин можно объяснить следующим образом. Вариация — это «изменчивость» (по английски дисперсия — variance), а «ко» означает «совместность» (например, ко-операция, ко-ординация). Так что ковариация — это характеристика совместной изменчивости. Конечно, это все латынь.

[20] Напомним, что скалярное произведение элементов $\beta_1,\beta_2\in\mathrm{L}^2(d\mathbb{P})$ определяется как $\mathbb{E}\beta_1\beta_2$.

[21] Действительно?

[22] Конечно, свойства 1 – 3 тесно связаны с общими свойствами скалярного произведения.

[23] См. второй пункт Замечания 1.2 и равенство (2.5).

1. $\mathbb{D}\xi \geq 0$, причем $\mathbb{D}\xi = 0$ тогда и только тогда, когда $\xi = \mathrm{const}$ ($\mathbb{P}$-п.в.).
2. $\mathbb{D}(a\xi + b) = a^2\mathbb{D}\xi$.
3. $\mathbb{D}(\xi + \eta) = \mathbb{D}\xi + \mathbb{D}\eta + 2\,\mathrm{Cov}(\xi, \eta)$.
4. Если ξ и η независимы, то $\mathbb{D}(\xi + \eta) = \mathbb{D}\xi + \mathbb{D}\eta$. Обратное, вообще говоря, неверно.

Коэффициент корреляции.

Определение 2.1. Пусть ξ и η — случайные величины, обладающие конечными вторыми моментами и ненулевыми дисперсиями. Тогда *коэффициентом корреляции* между ξ и η называется число

$$\rho(\xi, \eta) = \frac{\mathrm{Cov}(\xi, \eta)}{\sqrt{\mathbb{D}\xi}\,\sqrt{\mathbb{D}\eta}}\,. \tag{2.6}$$

Прокомментируем это определение. Если $\beta_1, \beta_2 \in \mathrm{L}^2(d\mathbb{P})$, то отношение $\mathbb{E}\beta_1\beta_2\big/\sqrt{\mathbb{E}\beta_1^2}\sqrt{\mathbb{E}\beta_2^2}$ является косинусом[24] угла между β_1 и β_2. Этот косинус может быть проинтерпретирован как мера пропорциональности β_1 и β_2: если $\beta_1 = a\beta_2$ с $a > 0$, то косинус равен $+1$, это же равенство с $a < 0$ дает значение косинуса -1, а при полном отсутствии пропорциональности (то есть при ортогональности β_1 и β_2) косинус принимает значение ноль.

В нашем случае $\beta_1 = \xi - \mathbb{E}\xi$ и $\beta_2 = \eta - \mathbb{E}\eta$. Поэтому пропорциональность β_1 и β_2 превращается в линейную зависимость между ξ и η. Следовательно, коэффициент корреляции между ξ и η можно интерпретировать как *меру линейной зависимости* между этими случайными величинами.[25]

Приведем основные свойства коэффициента корреляции.

1. Пусть a, b, c и d — постоянные, причем $a, c \neq 0$. Обозначим $\xi_1 = a\xi + b$ и $\eta_1 = c\eta + d$. Тогда
$$\rho(\xi_1, \eta_1) = \mathrm{sign}(ac)\,\rho(\xi, \eta).$$
2. Обозначим[26] $\xi_1 = (\xi - \mathbb{E}\xi)/\sqrt{\mathbb{D}\xi}$ и $\eta_1 = (\eta - \mathbb{E}\eta)/\sqrt{\mathbb{D}\eta}$. Тогда
$$\mathrm{Cov}(\xi_1, \eta_1) = \rho(\xi_1, \eta_1) = \rho(\xi, \eta).$$
3. $|\rho(\xi, \eta)| \leq 1$.
4. $\rho(\xi, \eta) = 1$ тогда и только тогда,[27] когда $\xi = a\eta + b$ и $a > 0$. В частности, $\rho(\xi, \xi) = 1$.
5. $\rho(\xi, \eta) = -1$ тогда и только тогда,[28] когда $\xi = a\eta + b$ и $a < 0$. В частности, $\rho(\xi, -\xi) = 1$.
6. Если ξ и η независимы, то $\rho(\xi, \eta) = 0$. Обратное, вообще говоря, неверно.[29]

[24] В общем случае косинус угла между элементами β_1, β_2 Гильбертова пространства $\mathbb{H}$ имеет вид $(\beta_1, \beta_2)/\|\beta_1\|\,\|\beta_2\|$.

[25] Тем самым у нас появляется три понятия, связанные с зависимостью случайных величин: функциональная зависимость $\eta = f(\xi)$ для некоторой функции f, статистическая зависимость, то есть отсутствие независимости между этими случайными величинами, и линейная зависимость. Коэффициент корреляции является мерой именно этой последней зависимости.

[26] Так как $\mathbb{E}\xi_1 = \mathbb{E}\eta_1 = 0$ и $\mathbb{D}\xi_1 = \mathbb{D}\eta_1 = 1$, то такая операция называется *стандартизацией*, то есть центрированием (вычитанием среднего) и нормировкой (делением на квадратный корень из дисперсии) случайных величин ξ и η.

[27] Точнее: если $\rho(\xi, \eta) = 1$, то существуют такие постоянные $a > 0$ и b, что $\xi = a\eta + b$ с вероятностью 1.

[28] См. предыдущее замечание с заменой условия $a > 0$ на $a < 0$.

[29] Снова см. второй пункт Замечания 1.2.

Понятие коэффициента корреляции очень часто используется при обработке результатов наблюдений (то есть в прикладной математической статистике), где важна интерпретация результатов. Поэтому имеет смысл подчеркнуть, что наличие большой (то есть достаточно близкой к ± 1) корреляции между ξ и η ничего не говорит о причинно-следственных связях между этими переменными. Приведем реальный пример.

В 60-х годах прошлого века коэффициент корреляции между уровнем преступности и затратами на борьбу с преступностью в США был примерно равен $+0.67$.[30] Это говорит об относительно большой положительной (линейной) связи между этими характеристиками. Какая может быть интерпретация такой связи? Можно предложить три (самых простых) варианта такой интерпретации.[31]

1. *Оптимистичный вариант.* Чем больше растет преступность, тем больше выделяется средств на борьбу с ней.[32]

2. *Пессимистический вариант.* Деньги, выделяемые на борьбу с преступностью, поступают на самом деле в распоряжение преступных группировок. То есть — чем больше выделяется денег, тем больше совершается преступлений. Такое тоже бывает.

3. *Конспирологический вариант.* В обоих предыдущих случаях рассматривались причинно-следственые (в ту и другую сторону) интерпретации большой корреляции между уровнем преступности и затратами на борьбу с преступностью. Но непосредственной причинно-следственной связи здесь вообще может не быть!

   Представим себе, что где-то есть (конечно, глубоко законспирированный) центр, который по неизвестным нам (но, конечно, глубоко аморальным) причинам выделяет деньги одновременно и на развитие преступности и на борьбу с ней. Тогда тоже возникнет положительная связь между этими переменными, наблюдаемая на практике.[33]

Такого рода рассуждения можно попытаться формализовать. Докажем следующий факт.

Предложение 2.1. *Пусть β — некоторая случайная величина, $\eta_1 = f(\beta)$, $\eta_2 = g(\beta)$, где измеримые функции f и g монотонно не убывают.*[34] *Предположим, кроме того, что η_1, η_2 обладают конечными вторыми моментами и ненулевыми дисперсиями.*

Далее, пусть ε_1 и ε_2 — независимые между собой и с β случайные величины, имеющие нулевые средние и дисперсии σ_1^2 и σ_2^2 соответственно. Положим $\xi_1 = \eta_1 + \varepsilon_1$ и $\xi_2 = \eta_2 + \varepsilon_2$. Тогда $\rho(\xi_1, \xi_2) \geq 0$.

Доказательство. Докажем утверждение дополнительно предполагая, что функция f является непрерывной.[35] Конечно, нам достаточно показать, что $\mathrm{Cov}(\xi_1, \xi_2) \geq 0$. Кроме того, как нетрудно видеть, $\mathrm{Cov}(\xi_1, \xi_2) = \mathrm{Cov}(\eta_1, \eta_2)$.[36]

Поскольку функции f и g монотонно возрастают, то для любых x, y имеет место неравенство

$$\big(f(x) - f(y)\big)\big(g(x) - g(y)\big) \geq 0. \tag{2.7}$$

Действительно, монотонное возрастание функции f означает, что $\big(f(x) - f(y)\big)(x - y) \geq 0$ для любых x и y, аналогично этому $\big(g(x) - g(y)\big)(x - y) \geq 0$. Перемножая эти неравенства при $x \neq y$, получаем требуемое.

Теперь положим $x = \beta$, а y выберем так, чтобы $f(y) = \mathbb{E}f(\beta)$ (этот выбор возможен,[37] так как функция f непрерывна). Тогда условие монотонности (2.7) превратится в

$$\big(f(\beta) - \mathbb{E}f(\beta)\big)\big(g(\beta) - g(y)\big) \geq 0. \tag{2.8}$$

Беря от обеих частей (2.8) математическое ожидание, получаем, что

$$\mathbb{E}\big(f(\beta) - \mathbb{E}f(\beta)\big)\big(g(\beta) - g(y)\big) \geq 0.$$

Осталось заметить, что $\mathrm{Cov}(\xi, \eta) = \mathbb{E}(\xi - \mathbb{E}\xi)(\eta - a)$, где a — произвольная постоянная.[38] Утверждение доказано. □

[30] Нам сейчас не важно, по каким данным и как было сосчитано это число.

[31] Конечно, интерпретаций может быть гораздо больше. Представим себе, например, что данные просто-напросто сфальсифицированы. Тогда сам вопрос нужно ставить совершенно по-другому.

[32] Или: там, где преступность сходит на нет, на борьбу с ней выделяется меньше средств. Так политкорректнее.

[33] Эта интерпретация выглядит несерьезно, но на самом деле подобные ситуации встречаются часто.

[34] В дальнейшем мы для краткости будем употреблять термин «возрастают».

[35] Попробуйте обойтись без этого предположения.

[36] Проверьте!

[37] Действительно?

[38] Проверьте это.

Замечание 2.1. Обсудим этот результат. Прежде всего заметим, что при (совместном) монотонном убывании функций f и g коэффициент корреляции между ξ_1 и ξ_2 остается положительным. Если же одна из этих функций возрастает, а другая — убывает, то это приводит к отрицательной корреляции.

Предположим теперь, что нас интересуют переменные η_1 и η_2.[39] При этом ε_1 и ε_1 — ошибки наших наблюдений,[40] так что результатом реальных измерений являются переменные ξ_1 и ξ_2. Функциональные зависимости $\eta_1 = f(\beta)$ и $\eta_2 = g(\beta)$ трактуются в духе причинно-следственной связи.[41] В Предложении 2.1 утверждается, что, если эти зависимости имеют одинаковую монотонность, то результаты наблюдений будут положительно коррелированы. Эта может служить формальной моделью «конспирологичекой» версии большой положительной зависимости между уровнем преступность и уровнем затрат на борьбу с ней.

Если же $f(x) = x$ (и $\mathbb{D}\beta < \infty$), то мы приходим в одному из вариантов непосредственной причинно-следственной интерпретации этой зависимости, то есть к «оптимистической» или «пессимистической» версиям.

Обсудим теперь роль дисперсий σ_1^2 и σ_2^2 случайных величин ε_1 и ε_2, трактуемых как ошибки наблюдений. Как мы уже видели, числитель дроби, определяющей коэффициент корреляции $\rho(\xi_1, \xi_2)$, не зависит от этих дисперсий. Что касается знаменателя, то он равен квадратному корню из произведения дисперсий $\mathbb{D}\xi_1 = \mathbb{D}\eta_1 + \sigma_1^2$ и $\mathbb{D}\xi_2 = \mathbb{D}\eta_2 + \sigma_2^2$.

Тем самым чем больше разброс ошибок наблюдения, тем ближе (положительный) коэффициент корреляции $\rho(\xi_1, \xi_2)$ к нулю. При больших σ_1 и/или σ_2 «зашумление» результатов наблюдений будет настолько велико, что $\rho(\xi_1, \xi_2)$ фактически перестанет обнаруживать (линейную) связь между η_1 и η_2.

3 Ковариационная и корреляционная матрицы

Пусть $\overline{\xi} = (\xi_1, \ldots, \xi_d)^{\mathrm{T}}$ — случайный вектор, причем $\mathbb{E}|\xi_i| < \infty$ при всех i. Математическим ожиданием вектора $\overline{\xi}$ называется вектор $\mathbb{E}\overline{\xi} = \big(\mathbb{E}\xi_1, \ldots, \mathbb{E}\xi_d\big)^{\mathrm{T}}$. Аналогичным (покомпонентным) образом определяется математическое ожидание случайной матрицы.

Математические ожидания случайных векторов и матриц обладают следующими простыми свойствами.[42]

1. Пусть $\overline{\xi} \in \mathbb{R}^d$. Рассмотрим (детерминированные) матрицу $A : \mathbb{R}^d \mapsto \mathbb{R}^k$ и вектор $\overline{b} \in \mathbb{R}^k$ и положим $\overline{\eta} = A\overline{\xi} + \overline{b}$. Тогда $\mathbb{E}\overline{\eta} = A\mathbb{E}\overline{\xi} + \overline{b}$.

2. Пусть Θ — случайная $d\times k$-матрица. Рассмотрим (детерминированные) матрицы $A : \mathbb{R}^m \mapsto \mathbb{R}^d$ и $B : \mathbb{R}^k \mapsto \mathbb{R}^n$. Обозначим $\Xi = A\Theta B$. Тогда $\mathbb{E}\,\Xi = A(\mathbb{E}\Theta)B$.

Определение 3.1. Пусть $\overline{\xi} = (\xi_1, \ldots, \xi_d)^{\mathrm{T}}$ — случайный вектор, причем все его координаты обладают конечными вторыми моментами. Тогда матрица $\Sigma_\xi = \big\{\mathrm{Cov}(\xi_i, \xi_j)\big\}_{i,j=1}^d$ называется *ковариационной матрицей вектора* $\overline{\xi}$.

Если дополнительно $\sigma_i^2 \stackrel{\mathrm{def}}{=} \mathbb{D}\xi_i > 0$ при всех i, то тогда существует матрица $\Delta_\xi = \big\{\rho(\xi_i, \xi_j)\big\}_{i,j=1}^d$, которую называют *корреляционной матрицей вектора* $\overline{\xi}$.

Отметим простейшие свойства ковариационных и корреляционных матриц.[43]

1. Обозначим $\overline{\eta} = \overline{\xi} - \mathbb{E}\overline{\xi}$. Тогда $\Sigma_\xi = \mathbb{E}\overline{\eta}\,\overline{\eta}^{\mathrm{T}}$.

2. Если $\overline{b} \in \mathbb{R}^d$ и $\overline{\eta} = \overline{\xi} + \overline{b}$, то $\Sigma_\eta = \Sigma_\xi$ и $\Delta_\eta = \Delta_\xi$.

[39] А переменная β, вообще говоря, недоступна для непосредственного наблюдения.

[40] Этим и объясняются условия $\mathbb{E}\varepsilon_1 = \mathbb{E}\varepsilon_2 = 0$, которые формально не используются при доказательстве предложения: нам ничего не мешает иметь дело с функциями $f_1 = f + \mathbb{E}\varepsilon_1$ и $g_1 = g + \mathbb{E}\varepsilon_2$ вместо f и g. Если среднее значение ошибок измерений равно нулю, то это означает, что измерения не имеют систематической погрешности.

[41] «Из β следуют η_1 и η_2».

[42] Для их доказательства достаточно расписать в явном виде выражения для компонент вектора $\overline{\eta}$ (и матрицы Ξ) через компоненты вектора $\overline{\xi}$ (и матрицы Θ), применить свойства математических ожиданий для соответствующих случайных величин и записать полученные результаты в векторно-матричной форме. Естественно, здесь предполагается, что все математические ожидания конечны.

[43] Проверьте их!

3. Пусть $a_i \neq 0$, $\eta_i = \xi_i/a_i$ и $\overline{\eta} = (\eta_1, \ldots, \eta_d)$. Тогда $\Delta_\eta = \Delta_\xi$.

4. Корреляционная матрица Δ_ξ является ковариационной матрицей случайного вектора $\overline{\eta}$ с координатам $\eta_i = (\xi_i - \mathbb{E}\xi_i)/\sigma_i$, $1 \leq i \leq d$.

Перечислим еще некоторые важные свойства ковариационных и корреляционных матриц.

Предложение 3.1. *1. Ковариационная Σ_ξ и корреляционная Δ_ξ матрицы любого случайного вектора $\overline{\xi} \in \mathbb{R}^d$ являются неотрицательно определенными.*
2. $\operatorname{rank} \Sigma_\xi = \operatorname{rank} \Delta_\xi$.
3. Если $\operatorname{rank} \Sigma_\xi = k$, то существует такое k-мерное линейное многообразие $\mathcal{L}_k \subset \mathbb{R}^d$, что $\mathbb{P}\big(\overline{\xi} \in \mathcal{L}_k\big) = 1$.

Замечание 3.1. 1. Неотрицательную определенность корреляционной матрицы специально доказывать не нужно, так как она является ковариационной матрицей, только другого случайного вектора.
2. Если $\operatorname{rank} \Sigma_\xi = k < d$, то случайный вектор $\overline{\xi}$ не является абсолютно непрерывным.

Предложение 3.2. *Пусть случайный вектор $\overline{\xi} \in \mathbb{R}^d$ обладает ковариационной матрицей Σ_ξ. Обозначим $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \geq 0$ собственные (с учетом кратности) числа матрицы Σ_ξ и пусть $U_1, \ldots, U_d$ — соответствующая ортонормальная система собственных векторов.*[44]

Обозначим $\mathrm{U} = [U_1 : \ldots : U_d]$ и введем диагональную матрицу $\Lambda = \{\lambda_{ij}\}_{i,j=1}^d$ с $\lambda_{ii} = \lambda_i$. Тогда случайный вектор

$$\overline{\eta} = (\eta_1, \ldots, \eta_d)^{\mathrm{T}} = \mathrm{U}^{\mathrm{T}}\big(\overline{\xi} - \mathbb{E}\overline{\xi}\big)$$

имеет нулевое среднее и диагональную ковариационную матрицу $\Sigma_\eta = \Lambda$.

[44]То есть $U_i^{\mathrm{T}} U_j = \delta_{ij}$, где δ_{ij} — символы Кронекера.