

# Комментарии к теме «Математическое ожидание»

Практические занятия по теории вероятностей

кафедра статистического моделирования <http://statmod.ru>, матмех СПбГУ, 2018 г.

## 1 Математическое ожидание

### 1.1 Определение и основные свойства

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

**Замечание 1.1.** 1. Поскольку мера  $P$  — вероятностная (то есть нормированная на 1), то математическое ожидание — это *среднее значение* случайной величины. Иногда эти термины используют как эквивалентные.

2. Пусть  $\int_{\Omega} |\xi| dP = \infty$ , но  $\int_{\Omega} \xi^- dP < \infty$ .<sup>2</sup> Тогда (допуская некоторую вольность речи) говорят, что математическое ожидание случайной величины  $\xi$  равно  $+\infty$  и пишут  $E\xi = +\infty$ . Аналогичным образом интерпретируется запись  $E\xi = -\infty$ . Если  $\int_{\Omega} \xi^+ dP = \int_{\Omega} \xi^- dP = \infty$ , то математического ожидания случайной величины  $\xi$  *не существует*.

**Простейшие свойства математического ожидания.**

1. Если  $\xi = \text{const} = c$ , то<sup>3</sup>  $E\xi = c$ .
2. Если у  $\xi$  существует математическое ожидание, то для любых постоянных  $a$  и  $b$  имеет место равенство  $E(a\xi + b) = aE\xi + b$ .
3. Если  $\xi \geq \eta$ , то  $E\xi \geq E\eta$ . В частности, математическое ожидание положительной случайной величины положительно. Если  $a \leq \xi \leq b$ , то  $a \leq E\xi \leq b$ .<sup>4</sup>
4. Если  $\xi$  и  $\eta$  имеют конечные математические ожидания, то  $E(\xi + \eta) = E\xi + E\eta$ .<sup>5</sup>
5. Если  $\xi$  и  $\eta$  независимы, то  $E\xi\eta = E\xi E\eta$  (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.<sup>6</sup>

<sup>1</sup>Обозначение математического ожидания буквой  $E$  можно объяснить использованием аббревиатуры английского «Expectation».

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

<sup>3</sup>Среднее значение постоянной равно этой постоянной.

<sup>4</sup>Это свойство может быть полезно при проверке правдоподобности результатов вычислений.

<sup>5</sup>Понятно ведь, что среднее значение суммы должно равняться сумме средних значений!

<sup>6</sup>Ход доказательства этого свойства такой. Если  $\xi$  и  $\eta$  — 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  $E\xi\eta = E\xi E\eta$  не следует, что  $\xi$  и  $\eta$  независимы.<sup>7</sup>

**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  $y$  случайной величины  $\xi$  с распределением  $\mathcal{P}_\xi$  существует математическое ожидание  $E\xi$ . Тогда

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

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

$$Eg(\xi) = \int_{\mathbb{R}} g(x) \mathcal{P}_\xi(dx). \quad (1.2)$$

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

$$Eg(\bar{\xi}) = Eg(\xi_1, \dots, \xi_d) = \int_{\mathbb{R}^d} g(x_1, \dots, x_d) \mathcal{P}_\xi(dx_1 \dots dx_d). \quad (1.3)$$

**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  $\xi$  (или случайного вектора  $\bar{\xi}$ ) является дискретным или абсолютно непрерывным.

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

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

Тогда (1.1) и (1.2) превращаются в

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

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

Если же случайная величина  $\xi$  обладает плотностью распределения  $p_\xi$ , то (1.1) и (1.2) приобретают вид

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

<sup>7</sup>Чтобы убедиться в этом, рассмотрите случайный вектор  $(\xi, \eta)^T$ , равномерно распределенный в круге единичного радиуса с центром в нуле, и примените формулы (1.6) и (1.7) для вычисления  $E\xi$ ,  $E\eta$  и  $E\xi\eta$ .

а конечности этих математических ожиданий соответствует суммируемость функций  $x p_\xi(x)$  и  $g(x) p_\xi(x)$  по мере Лебега.<sup>8</sup>

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

$$Eg(\xi_1, \dots, \xi_d) = \int_{\mathbb{R}^d} g(x_1, \dots, x_d) p_\xi(x_1, \dots, x_d) dx_1 \dots dx_d. \quad (1.7)$$

**Замечание 1.4.** Если распределение  $\mathcal{P}_\xi$  представляется в виде смеси более простых распределений, то это может существенно облегчить вычисление математических ожиданий. Покажем это на примере формулы (1.2) (для многомерного случая все аналогично).

Пусть

$$\mathcal{P}_\xi = \sum_{i=1}^n q_i \mathcal{P}_i,$$

тогда

$$Eg(\xi) = \int_{\mathbb{R}} g(x) \mathcal{P}_\xi(dx) = \sum_{i=1}^n q_i \int_{\mathbb{R}} g(x) \mathcal{P}_i(dx).$$

Если интегралы  $\int g(x) \mathcal{P}_i(dx)$  считаются легко, то и вычисление искомого математического ожидания  $Eg(\xi)$  не представляет сложности.

Например, это может произойти, если среди распределений  $\mathcal{P}_i$  есть как (простые) дискретные распределения, так и (простые) абсолютно непрерывные распределения (а других распределений нет). В тоже время непосредственное использование формулы (1.2) может вызвать затруднения, так как распределение  $\mathcal{P}_\xi$  не будет ни дискретным, ни абсолютно непрерывным.

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

Пусть случайный вектор  $(\xi_1, \xi_2)^T$  имеет плотность распределения  $p_{\xi_1 \xi_2}(x, y)$ . Задача состоит в вычислении математического ожидания  $Eg(\xi_1, \xi_2)$  (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  $p_\eta$ . Тогда ответ можно представить в виде

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

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

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

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

с последующим вычислением этого двумерного интеграла.

Какой из этих способов предпочтительней? Ответ, конечно, зависит и от совместного распределения  $(\xi_1, \xi_2)$  и от функции  $g$ . Тем не менее можно высказать несколько общих соображений.

<sup>8</sup>Из формулы (1.6) для математических ожиданий следует, что  $E\xi = 0$ , если плотность  $p_\xi$  является четной функцией и  $\int_0^\infty x p_\eta(x) dx < \infty$ .

Когда, решая эту задачу, мы сначала находим распределение случайной величины  $g(\xi_1, \xi_2)$ , а потом используем формулу (1.8), мы делаем в каком-то смысле лишнюю работу. Например, найдя плотность  $p_\eta$ , мы вместе с (1.8) можем написать, что

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

(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,<sup>9</sup> равенство (1.9) верно вне зависимости от того, имеет ли случайная величина  $\eta$  абсолютно непрерывное распределение или нет, то есть формула (1.9) гораздо более универсальна.

С другой стороны, если распределение случайной величины  $\eta$  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  $\xi_1, \xi_2 \in U(0, 1)$  и являются независимыми. Нужно найти математическое ожидание  $E|\xi_1 - \xi_2|$ .

*Решение.* Обозначим  $\beta = \xi_1 - \xi_2$ . Известно,<sup>10</sup> что плотность распределения случайной величины  $\beta$  имеет вид

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

Поэтому<sup>11</sup> случайная величина  $\eta = |\beta|$  имеет плотность распределения

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

Следовательно,

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

С другой стороны, можно сразу написать, что

$$E|\xi_1 - \xi_2| = \int_0^1 \int_0^1 |x - y|dxdy,$$

что после несложных преобразований даст тот же результат.

Если же в условиях этого же примера необходимо найти  $E|\xi_1 - 3\xi_2|$ , то, вполне возможно, лучше сразу же воспользоваться формулой (1.9), не тратя время на вычисление<sup>12</sup> 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  $\xi$  является ее *дисперсия*  $D\xi$ . По определению, если математическое ожидание  $E\xi$  конечно,

<sup>9</sup>Это существенно!

<sup>10</sup>Да?

<sup>11</sup>А почему?

<sup>12</sup>(или поиск в интернете.)

то

$$D\xi = E(\xi - E\xi)^2. \quad (2.1)$$

Поскольку в определении (2.1) под знаком математического ожидания стоит неотрицательная случайная величина, то принято говорить, что дисперсия бывает конечной или бесконечной. Дисперсия  $D\xi$  конечна тогда и только тогда, когда  $E\xi^2 < \infty$ , то есть когда  $\xi$  обладает конечным вторым моментом.<sup>13</sup>

Само определение (2.1) намекает на смысл понятия дисперсии. Рассмотрим гильбертово пространство  $L^2(dP)$  функций, определенных на измеримом пространстве  $(\Omega, \mathcal{F})$ . Если  $\xi \in L^2(dP)$ ,<sup>14</sup> то  $D\xi$  является не чем иным, как *квадратом расстояния в  $L^2(dP)$  между  $\xi$  и ее математическим ожиданием  $E\xi$* .

Таким образом, дисперсия (точнее, квадратный корень из дисперсии) является мерой разброса случайной величины вокруг ее математического ожидания. Довольно часто дисперсию обозначают  $\sigma^2$ , где  $\sigma = \sqrt{D\xi}$ . Число  $\sigma$  называют *стандартным отклонением*<sup>15</sup> случайной величины  $\xi$ .

Раскрывая скобки в (2.1) и пользуясь простейшими свойствами математического ожидания,<sup>16</sup> получаем, что

$$D\xi = E\xi^2 - (E\xi)^2. \quad (2.2)$$

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

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

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

где  $E\xi$  имеет вид (1.1).

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

$$D\xi = \sum_n (x_n - E\xi)^2 p_n = \sum_n x_n^2 p_n - (E\xi)^2, \quad (2.3)$$

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

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

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

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

Приведем теперь выражения для математических ожиданий и дисперсий случайных величин, имеющих некоторые стандартные распределения.<sup>17</sup>

а) Дискретные распределения.

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

<sup>13</sup> Дальнейшие свойства приводятся при этом предположении.

<sup>14</sup> Точнее, «если случайная величина  $\xi$  принадлежит соответствующему классу эквивалентности», но мы будем здесь игнорировать эти тонкости.

<sup>15</sup> Используют и другие термины, например, *среднеквадратическое отклонение*. Часто говорят просто «стандарт».

<sup>16</sup> Какими?

<sup>17</sup> Убедитесь в том, что приведенные выражения правильны.

- *Равномерно распределение*  $U_n(X)$  на множестве  $X = \{x_1, \dots, x_n\}$ . Пусть  $P(\xi = x_i) = 1/n$  при  $i = 1, \dots, n$ . Тогда<sup>18</sup>  $E\xi = \bar{x}_n \stackrel{\text{def}}{=} (x_1 + \dots + x_n)/n$  и  $D\xi = \sum_{i=1}^n (x_i - \bar{x}_n)^2/n$ .

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

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

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

Во-вторых, если взять независимые случайные величины  $\varepsilon_1, \dots, \varepsilon_n$ , имеющие распределение  $\text{Ber}(p)$ , и положить  $\xi = \varepsilon_1 + \dots + \varepsilon_n$ , то, как известно,  $\xi \in \text{Bin}(n, p)$ . Поэтому приведенные выражения для  $E\xi$  и  $D\xi$  считаются получаются просто по общим свойствам математического ожидания и дисперсии.<sup>19</sup>

- *Геометрическое распределение*  $\text{Geom}(p)$  с параметром  $p$ . Здесь  $P(\xi = k) = p(1-p)^k$  при  $k \geq 0$ ,  $E\xi = (1-p)/p$  и  $D\xi = (1-p)/p^2$ . Результат получается прямым подсчетом.
- *Распределение Пуассона*  $\Pi(\lambda)$  с параметром  $\lambda > 0$ , то есть  $P(\xi = k) = \lambda^k e^{-\lambda}/k!$  при  $k = 0, 1, \dots$ . Прямой подсчет показывает, что  $E\xi = D\xi = \lambda$ .

С помощью теоремы Пуассона легко убедиться в правдоподобности<sup>20</sup> этого результата. Действительно, рассмотрим случайные величины  $\xi_n \in \text{Bin}(n, p_n)$ , где  $np_n \rightarrow \lambda > 0$  при  $n \rightarrow \infty$ . Тогда  $E\xi_n = np_n \rightarrow \lambda$  и  $D\xi = np_n(1-p_n) \rightarrow \lambda$ . С другой стороны, по Теореме Пуассона  $P(\xi_n = k) \rightarrow \lambda^k e^{-\lambda}/k!$ , то есть при больших  $n$  распределение  $\text{Bin}(n, p_n)$  близко к распределению Пуассона. Поэтому можно ожидать, что и такие их характеристики, как математическое ожидание и дисперсия, тоже близки.

б) Абсолютно непрерывные распределения.

- *Равномерное распределение*  $U(a, b)$  на отрезке  $[a, b]$  с  $p_\xi(x) = 1/(b-a)$  при  $x \in [a, b]$ . В этом случае  $E\xi = (b+a)/2$  и  $D\xi = (b-a)^2/12$ .<sup>21</sup>
- *Показательное распределение*  $\text{Exp}(\lambda)$  с параметром  $\lambda > 0$ , имеющее плотность  $p_\xi(x) = \lambda e^{-\lambda x}$  при  $x > 0$ . Здесь  $E\xi = 1/\lambda$  и  $D\xi = 1/\lambda^2$ .
- *Нормальное распределение*  $N(a, \sigma^2)$  с плотностью  $p_\xi(x) = e^{-(x-a)^2/2\sigma^2}/\sqrt{2\pi}\sigma$ ,  $x \in \mathbb{R}$ . Тогда  $E\xi = a$  и  $D\xi = \sigma^2$ . Таким образом, параметры нормального распределения — это его среднее и дисперсия. В частности, если  $\xi \in N(0, 1)$ , то  $E\xi = 0$  и  $D\xi = 1$ .
- А вот у *распределения Коши*, которое имеет плотность

$$p(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R},$$

математического ожидания (и, следовательно, дисперсии) не существует. Действительно, поскольку функция  $|x|/(1+x^2)$  ведет себя как  $|x|^{-1}$  при  $x \rightarrow \infty$ , то интеграл  $\int_{\mathbb{R}} |x|p(x)dx$  расходится.

<sup>18</sup>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.

<sup>19</sup>При этом при подсчете математического ожидания независимость случайных величин  $\varepsilon_i$  не используется.

<sup>20</sup>Но не более!

<sup>21</sup>Не правда ли — формула для математического ожидания здесь “очевидна”? А что можно сказать про формулу для дисперсии?

**Дисперсия и ковариация. Свойства.** Пусть теперь  $\xi$  и  $\eta$  — случайные величины, обладающие конечными вторыми моментами. *Ковариацией*<sup>22</sup> этих случайных величин называется число

$$\text{Cov}(\xi, \eta) = E(\xi - E\xi)(\eta - E\eta). \quad (2.4)$$

Тем самым  $\text{Cov}(\xi, \eta)$  — это ни что иное, как скалярное произведение<sup>23</sup> случайных величин  $\xi - E\xi$  и  $\eta - E\eta$  в  $L^2(dP)$ . Заметим, что по определению  $\text{Cov}(\xi, \xi) = D\xi$ .

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

$$\text{Cov}(\xi, \eta) = E\xi\eta - E\xi E\eta. \quad (2.5)$$

Ковариация легко расписывается через совместное распределение случайных величин  $\xi, \eta$ . Например, равенство (2.5) приобретает вид

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

переход к частным случаям абсолютно непрерывного или дискретного совместного распределения  $(\xi, \eta)$  не представляет труда.<sup>24</sup>

Перечислим основные свойства дисперсий и ковариаций. Удобнее начать со свойств ковариации.<sup>25</sup>

1.  $\text{Cov}^2(\xi, \eta) \leq D\xi D\eta$ .
2.  $\text{Cov}(a\xi + b, c\eta + d) = ac\text{Cov}(\xi, \eta)$ . 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.  $\text{Cov}(\xi_1 + \xi_2, \eta) = \text{Cov}(\xi_1, \eta) + \text{Cov}(\xi_2, \eta)$ .
4. Если  $\xi$  и  $\eta$  независимы, то  $\text{Cov}(\xi, \eta) = 0$ .

Случайные величины, для которых  $\text{Cov}(\xi, \eta) = 0$ , называются *некоррелированными*. Тем самым независимые случайные величины (с конечными вторыми моментами) являются некоррелированными. Обратное, вообще говоря, неверно.<sup>26</sup>

Теперь свойства дисперсии.

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

<sup>22</sup>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.

<sup>23</sup>Напомним, что скалярное произведение элементов  $\beta_1, \beta_2 \in L^2(dP)$  определяется как  $E\beta_1\beta_2$ .

<sup>24</sup>Действительно?

<sup>25</sup>Конечно, свойства 1 – 3 тесно связаны с общими свойствами скалярного произведения.

<sup>26</sup>См. второй пункт Замечания 1.2 и равенство (2.5).

## Коэффициент корреляции.

**Определение 2.1.** Пусть  $\xi$  и  $\eta$  — случайные величины, обладающие конечными вторыми моментами и ненулевыми дисперсиями. Тогда *коэффициентом корреляции* между  $\xi$  и  $\eta$  называется число

$$\rho(\xi, \eta) = \frac{\text{Cov}(\xi, \eta)}{\sqrt{D\xi} \sqrt{D\eta}}. \quad (2.6)$$

Прокомментируем это определение. Если  $\beta_1, \beta_2 \in L^2(dP)$ , то отношение  $E\beta_1\beta_2 / \sqrt{E\beta_1^2} \sqrt{E\beta_2^2}$  является косинусом<sup>27</sup> угла между  $\beta_1$  и  $\beta_2$ . Этот косинус может быть проинтерпретирован как мера пропорциональности  $\beta_1$  и  $\beta_2$ : 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  $\beta_1$  и  $\beta_2$ ) косинус принимает значение ноль.

В нашем случае  $\beta_1 = \xi - E\xi$  и  $\beta_2 = \eta - E\eta$ . Поэтому пропорциональность  $\beta_1$  и  $\beta_2$  превращается в линейную зависимость между  $\xi$  и  $\eta$ . Следовательно, коэффициент корреляции между  $\xi$  и  $\eta$  можно интерпретировать как *меру линейной зависимости* между этими случайными величинами.<sup>28</sup>

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

$$\rho(\xi_1, \eta_1) = \text{sign}(ac) \rho(\xi, \eta).$$

2. Обозначим<sup>29</sup>  $\xi_1 = (\xi - E\xi) / \sqrt{D\xi}$  и  $\eta_1 = (\eta - E\eta) / \sqrt{D\eta}$ . Тогда

$$\text{Cov}(\xi_1, \eta_1) = \rho(\xi_1, \eta_1) = \rho(\xi, \eta).$$

3.  $|\rho(\xi, \eta)| \leq 1$ .

4.  $\rho(\xi, \eta) = 1$  тогда и только тогда,<sup>30</sup> когда  $\xi = a\eta + b$  и  $a > 0$ . В частности,  $\rho(\xi, \xi) = 1$ .

5.  $\rho(\xi, \eta) = -1$  тогда и только тогда,<sup>31</sup> когда  $\xi = a\eta + b$  и  $a < 0$ . В частности,  $\rho(\xi, -\xi) = 1$ .

6. Если  $\xi$  и  $\eta$  независимы, то  $\rho(\xi, \eta) = 0$ . Обратное, вообще говоря, неверно.<sup>32</sup>

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  $\pm 1$ ) корреляции между  $\xi$  и  $\eta$  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  $+0.67$ .<sup>33</sup> 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.<sup>34</sup>

<sup>27</sup>В общем случае косинус угла между элементами  $\beta_1, \beta_2$  Гильбертова пространства  $\mathbb{H}$  имеет вид  $(\beta_1, \beta_2) / \|\beta_1\| \|\beta_2\|$ .

<sup>28</sup>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.

<sup>29</sup>Так как  $E\xi_1 = E\eta_1 = 0$  и  $D\xi_1 = D\eta_1 = 1$ , то такая операция называется *стандартизацией*, то есть центрированием (вычитанием среднего) и нормировкой (делением на квадратный корень из дисперсии) случайных величин  $\xi$  и  $\eta$ .

<sup>30</sup>Точнее: если  $\rho(\xi, \eta) = 1$ , то существуют такие постоянные  $a > 0$  и  $b$ , что  $\xi = a\eta + b$  с вероятностью 1.

<sup>31</sup>См. предыдущее замечание с заменой условия  $a > 0$  на  $a < 0$ .

<sup>32</sup>Снова см. второй пункт Замечания 1.2.

<sup>33</sup>Нам сейчас не важно, по каким данным и как было сосчитано это число.

<sup>34</sup>Конечно, интерпретаций может быть гораздо больше. Представим себе, например, что данные просто-напросто сфальсифицированы. Тогда сам вопрос нужно ставить совершенно по-другому.

1. *Оптимистичный вариант.* Чем больше растет преступность, тем больше выделяется средств на борьбу с ней.<sup>35</sup>
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-следственные (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 (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.<sup>36</sup>

Такого рода рассуждения можно попытаться формализовать. Докажем следующий факт.

**Предложение 2.1.** Пусть  $\beta$  — некоторая случайная величина,  $\eta_1 = f(\beta)$ ,  $\eta_2 = g(\beta)$ , где измеримые функции  $f$  и  $g$  монотонно не убывают.<sup>37</sup> Предположим, кроме того, что  $\eta_1, \eta_2$  обладают конечными вторыми моментами и ненулевыми дисперсиями.

Далее, пусть  $\varepsilon_1$  и  $\varepsilon_2$  — независимые между собой и с  $\beta$  случайные величины, имеющие нулевые средние и дисперсии  $\sigma_1^2$  и  $\sigma_2^2$  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 это утверждение, дополнительно предполагая, что функция  $f$  является непрерывной.<sup>38</sup> Конечно, нам достаточно показать, что  $\text{Cov}(\xi_1, \xi_2) \geq 0$ . Как нетрудно видеть,  $\text{Cov}(\xi_1, \xi_2) = \text{Cov}(\eta_1, \eta_2)$ .<sup>39</sup>

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

$$(f(x) - f(y))(g(x) - g(y)) \geq 0. \quad (2.7)$$

Действительно, монотонное возрастание функции  $f$  означает, что  $(f(x) - f(y))(x - y) \geq 0$  для любых  $x$  и  $y$ , аналогично этому  $(g(x) - g(y))(x - y) \geq 0$ . 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,<sup>40</sup> так как функция  $f$  непрерывна). Тогда условие монотонности (2.7) превратится в

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

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

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

Осталось заметить, что  $\text{Cov}(\xi, \eta) = \mathbb{E}(\xi - \mathbb{E}\xi)(\eta - a)$ , где  $a$  — произвольная постоянная.<sup>41</sup> Утверждение доказано.  $\square$

**Замечание 2.1.** Обсудим этот результат. Прежде всего заметим, что при (совместном) монотонном убывании функций  $f$  и  $g$  коэффициент корреляции между  $\xi_1$  и  $\xi_2$  остается положительным. Если же одна из этих функций возрастает, а другая — убывает, то это приводит к отрицательной корреляции.

Предположим теперь, что нас интересуют переменные  $\eta_1$  и  $\eta_2$ .<sup>42</sup> При этом  $\varepsilon_1$  и  $\varepsilon_2$  — ошибки наших наблюдений,<sup>43</sup> так что результатом реальных измерений являются переменные  $\xi_1$  и  $\xi_2$ . Функциональные

<sup>35</sup> Или: там, где преступность сходит на нет, на борьбу с ней выделяется меньше средств. Так политкорректнее.

<sup>36</sup> Эта интерпретация выглядит несерьезно, но на самом деле подобные ситуации встречаются часто.

<sup>37</sup> В дальнейшем мы для краткости будем употреблять термин «возрастают».

<sup>38</sup> Попробуйте обойтись без этого предположения.

<sup>39</sup> Проверьте!

<sup>40</sup> Действительно?

<sup>41</sup> Проверьте это.

<sup>42</sup> А переменная  $\beta$ , вообще говоря, недоступна для непосредственного наблюдения.

<sup>43</sup> 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.

зависимости  $\eta_1 = f(\beta)$  и  $\eta_2 = g(\beta)$  трактуются в духе причинно-следственной связи.<sup>44</sup> В Предложении 2.1 утверждается, что, если эти зависимости имеют одинаковую монотонность, то результаты наблюдений будут положительно коррелированы. Эта может служить формальной моделью «конспирологической» версии большой положительной зависимости между уровнем преступности и уровнем затрат на борьбу с ней.

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

Обсудим теперь роль дисперсий  $\sigma_1^2$  и  $\sigma_2^2$  случайных величин  $\varepsilon_1$  и  $\varepsilon_2$ , 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  $D\xi_1 = D\eta_1 + \sigma_1^2$  и  $D\xi_2 = D\eta_2 + \sigma_2^2$ .

Тем самым чем больше разброс ошибок наблюдения, тем ближе (положительный) коэффициент корреляции  $\rho(\xi_1, \xi_2)$  к нулю. При больших  $\sigma_1$  и/или  $\sigma_2$  «зашумление» результатов наблюдений будет настолько велико, что  $\rho(\xi_1, \xi_2)$  фактически перестанет обнаруживать (линейную) связь между  $\eta_1$  и  $\eta_2$ .

**Замечание 2.2.** Это замечание — о размерности рассматриваемых характеристик случайных величин. В математике случайная величина  $\xi$  принимает числовые значения, ни о какой размерности здесь речи нет. В практических интерпретациях может быть совершенно по-другому. Значения случайных величин могут быть просто числами (например, число шаров в ящике или число гербов при подбрасывании монеты), а могут иметь размерность (если, скажем, это результаты измерений длины, промежутка времени, объема и пр.)

Пусть случайная величина  $\xi$  измеряется в метрах, а случайная величина  $\eta$  — в секундах. Какие размерности имеют а) математические ожидания  $\xi$  и  $\eta$ , б) дисперсии  $\xi$  и  $\eta$ , в) среднеквадратические отклонения  $\xi$  и  $\eta$ , г) ковариация между  $\xi$  и  $\eta$ , е) коэффициент корреляции между  $\xi$  и  $\eta$ ?

### 3 Ковариационная и корреляционная матрицы

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

Математические ожидания случайных векторов и матриц обладают следующими простыми свойствами.<sup>45</sup>

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

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

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

Отметим простейшие свойства ковариационных и корреляционных матриц.<sup>46</sup>

<sup>44</sup> «Из  $\beta$  следуют  $\eta_1$  и  $\eta_2$ ».

<sup>45</sup> Для их доказательства достаточно расписать в явном виде выражения для компонент вектора  $\bar{\eta}$  (и матрицы  $\Xi$ ) через компоненты вектора  $\bar{\xi}$  (и матрицы  $\Theta$ ), применить свойства математических ожиданий для соответствующих случайных величин и записать полученные результаты в векторно-матричной форме. Естественно, здесь предполагается, что все математические ожидания конечны.

<sup>46</sup> Проверьте их!

1. Обозначим  $\bar{\eta} = \bar{\xi} - E\bar{\xi}$ . Тогда  $\Sigma_{\xi} = E\bar{\eta}\bar{\eta}^T$ .
2. Если  $\bar{b} \in \mathbb{R}^d$  и  $\bar{\eta} = \bar{\xi} + \bar{b}$ , то  $\Sigma_{\eta} = \Sigma_{\xi}$  и  $\Delta_{\eta} = \Delta_{\xi}$ .
3. Пусть  $a_i \neq 0$ ,  $\eta_i = \xi_i/a_i$  и  $\bar{\eta} = (\eta_1, \dots, \eta_d)$ . Тогда  $\Delta_{\eta} = \Delta_{\xi}$ .
4. Корреляционная матрица  $\Delta_{\xi}$  является ковариационной матрицей случайного вектора  $\bar{\eta}$  с координатам  $\eta_i = (\xi_i - E\xi_i)/\sigma_i$ ,  $1 \leq i \leq d$ .

Перечислим еще некоторые важные свойства ковариационных и корреляционных матриц.

**Предложение 3.1.** 1. Ковариационная  $\Sigma_{\xi}$  и корреляционная  $\Delta_{\xi}$  матрицы любого случайного вектора  $\bar{\xi} \in \mathbb{R}^d$  являются неотрицательно определенными.

2.  $\text{rank } \Sigma_{\xi} = \text{rank } \Delta_{\xi}$ .

3. Если  $\text{rank } \Sigma_{\xi} = k$ , то существует такое  $k$ -мерное линейное многообразие  $\mathcal{L}_k \subset \mathbb{R}^d$ , что  $P(\bar{\xi} \in \mathcal{L}_k) = 1$ .

**Замечание 3.1.** 1. Неотрицательную определенность корреляционной матрицы специально доказывать не нужно, так как она является ковариационной матрицей, только другого случайного вектора.

2. Если  $\text{rank } \Sigma_{\xi} = k < d$ , то случайный вектор  $\bar{\xi}$  не является абсолютно непрерывным.

**Предложение 3.2.** Пусть случайный вектор  $\bar{\xi} \in \mathbb{R}^d$  обладает ковариационной матрицей  $\Sigma_{\xi}$ . Обозначим  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d \geq 0$  собственные (с учетом кратности) числа матрицы  $\Sigma_{\xi}$  и пусть  $U_1, \dots, U_d$  — соответствующая ортонормальная система собственных векторов.<sup>47</sup>

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

$$\bar{\eta} = (\eta_1, \dots, \eta_d)^T = U^T(\bar{\xi} - E\bar{\xi})$$

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

<sup>47</sup>То есть  $U_i^T U_j = \delta_{ij}$ , где  $\delta_{ij}$  — символы Кронекера.