13. Ковариация и корреляция

13.1. Предварительные сведения

Пусть ξ и η — случайные величины, имеющие конечные вторые моменты: $\mathbb{E}\xi^2 < \infty$, $\mathbb{E}\eta^2 < \infty$.

Ковариацией $\operatorname{Cov}(\xi, \eta)$ называется число $\mathbb{E}(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta) = \mathbb{E}\xi\eta - \mathbb{E}\xi\mathbb{E}\eta$.

Ковариация линейна по каждому своему аргументу (следует из свойств математического ожидания).

Для дискретных случайных величин с совместным распределением $p_{ij} = \mathbb{P}(\xi = x_i, \eta = y_j)$ и одномерными распределениями $p_{i\cdot} = \mathbb{P}(\xi = x_i) = \sum\limits_j p_{ij}$ и $p_{\cdot j} = \mathbb{P}(\eta = y_j) = \sum\limits_i p_{ij}$ ковариация находится по формуле

$$\operatorname{Cov}(\xi, \eta) = \sum_{ij} x_i y_j p_{ij} - \sum_i x_i p_{i\cdot} \sum_j y_j p_{\cdot j}.$$

Для абсолютно непрерывных случайных величин с совместной плотностью $p_{\xi,\eta}(x, y)$ и одномерными плотностями $p_\xi(x) = \int\limits_{-\infty}^{\infty} p_{\xi,\eta}(x, y)\, dy$ и $p_\eta(y) = \int\limits_{-\infty}^{\infty} p_{\xi,\eta}(x, y)\, dx$ ковариация находится по формуле

$$\operatorname{Cov}(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p_{\xi,\eta}(x, y)\, dx\, dy - \int_{-\infty}^{\infty} x p_\xi(x)\, dx \int_{-\infty}^{\infty} y p_\eta(y)\, dy.$$

Коэффициентом корреляции случайных величин ξ и η называется число

$$\rho(\xi, \eta) = \frac{\mathbb{E}\xi\eta - \mathbb{E}\xi\mathbb{E}\eta}{\sqrt{\operatorname{Var}\xi \operatorname{Var}\eta}} = \frac{\operatorname{Cov}(\xi, \eta)}{\sqrt{\operatorname{Var}\xi \operatorname{Var}\eta}}.$$

- ξ и η независимы $\quad\Rightarrow\quad \rho(\xi, \eta) = 0 \quad\Leftrightarrow\quad$ ξ и η некоррелированы;
- $|\rho(\xi, \eta)| \le 1$ и $\rho(\xi, \eta) = \pm 1 \quad\Leftrightarrow\quad$ ξ и η связаны линейно: $\eta = a\xi + b$, $a, b \in \mathbb{R}$, $a \ne 0$.

13.2. Практическое занятие

1. Игральная кость подброшена два раза. Пусть ξ — количество выпавших единиц, а η — количество выпавших шестерок. Выписать совместное распределение (ξ, η) и найти ковариацию $\operatorname{Cov}(\xi, \eta)$ и коэффициент корреляции $\rho(\xi, \eta)$. Согласуется ли знак коэффициента корреляции с вашими представлениями о характере зависимости ξ и η?

Ответ. $\operatorname{Cov}(\xi, \eta) = -1/18$, $\rho(\xi, \eta) = -1/5$.

2. Случайная точка с координатами (ξ, η) равномерно распределена в треугольнике с вершинами в точках с координатами $(0, 0)$, $(1, 0)$ и $(0, 1)$. Найти совместную плотность распределения $p_{\xi,\eta}(x, y)$, одномерные плотности $p_\xi(x)$ и $p_\eta(y)$ и вычислить ковариацию $\operatorname{Cov}(\xi, \eta)$ и коэффициент корреляции $\rho(\xi, \eta)$. Согласуется ли знак $\rho(\xi, \eta)$ с вашими представлениями о характере зависимости ξ и η?

Ответ. $p_{\xi,\eta}(x, y) = \begin{cases} 2, & x \ge 0,\ y \ge 0,\ x + y \le 1, \\ 0, & \text{иначе,} \end{cases}$ $\quad p_\xi(x) = \begin{cases} 0, & x < 0 \text{ или } x > 1, \\ 2 - 2x, & 0 \le x \le 1, \end{cases}$

$p_\eta(y) = \begin{cases} 0, & y < 0 \text{ или } y > 1, \\ 2 - 2y, & 0 \le y \le 1, \end{cases}$ $\quad \operatorname{Cov}(\xi, \eta) = -1/36$, $\rho(\xi, \eta) = -1/2$.

3. Случайная величина ξ равномерно распределена на отрезке $[-\pi, \pi]$. Докажите, что случайные величины $\eta_1 = \sin\xi$ и $\eta_2 = \cos\xi$ некоррелированы, хотя и являются зависимыми.

4. Некоторая величина отклоняется от своего среднего значения под воздействием двух случайных факторов A и B. Стандартное отклонение, вызванное фактором A, равно 1.2, а фактором B — 1.1. Коэффициент

корреляции между этими уклонениями равен $1/3$. Найти стандартное уклонение, вызываемое совместным действием обоих факторов.
Указание. Вычислить дисперсию уклонения, вызванного совместным действием обоих факторов.
Ответ. $\sqrt{1.2^2 + 1.1^2 + 2 \cdot \frac{1}{3} \cdot 1.2 \cdot 1.1} \approx 1.88.$

13.3. Домашнее задание

5. В одной из групп 19 человек получили оценку за контрольную работу по теории вероятностей шкале от 2 до 5 баллов и экзаменационную оценку на зимней сессии. Пусть (ξ, η) — случайный вектор, представляющий собой пару оценок, полученных наудачу выбранным студентом. Распределение этого случайного вектора имеет вид:

$\xi \setminus \eta$	2	3	4	5
2	0	1/19	0	0
3	1/19	1/19	2/19	3/19
4	0	0	1/19	3/19
5	0	0	0	7/19

Найти $\mathbb{E}\xi$, $\mathbb{E}\eta$, $\operatorname{Var}\xi$, $\operatorname{Var}\eta$, $\operatorname{Cov}(\xi, \eta)$ и $\rho(\xi, \eta)$. Согласуется ли знак $\rho(\xi, \eta)$ с вашими представлениями о характере зависимости ξ и η?
Ответ. $\mathbb{E}\xi = \frac{74}{19} \approx 3.89$, $\mathbb{E}\eta = \frac{85}{19} \approx 4.47$, $\operatorname{Var}\xi = \frac{338}{361} \approx 0.94$, $\operatorname{Var}\eta = \frac{280}{361} \approx 0.78$, $\operatorname{Cov}(\xi, \eta) = \frac{189}{361} \approx 0.52$, $\rho(\xi, \eta) = \frac{189}{\sqrt{338 \cdot 280}} \approx 0.61$. Положительная корреляция должна быть ожидаема.
6. Случайные величины ξ и η независимы и имеют стандартное нормальное распределение. Найти коэффициент корреляции величин $\alpha\xi + \beta\eta$ и $\alpha\xi - \beta\eta$, $\alpha, \beta \in \mathbb{R}$.
Ответ. $\frac{\alpha^2 - \beta^2}{\alpha^2 + \beta^2}$.
7. На отрезок $[0, 1]$ наудачу брошены две точки. Пусть ζ — координата левой точки, η — координата правой точки. Найти плотность совместного распределения ζ и η и вычислить коэффициент корреляции $\rho(\zeta, \eta)$.
Ответ. $p_{\zeta,\eta}(x, y) = \begin{cases} 2, & 0 < x < y < 1, \\ 0, & \text{иначе}, \end{cases}$ $\rho(\zeta, \eta) = 1/2$.
8. Случайные величины ξ и η (возможно, зависимые) обладают конечными дисперсиями: $\operatorname{Var}\xi = \sigma_\xi^2$, $\operatorname{Var}\eta = \sigma_\eta^2$. Указать верхнюю и нижнюю границы, в которых может находиться дисперсия $\operatorname{Var}(\xi + \eta)$. Приведите примеры, когда указанные границы достигаются.
Указание. Использовать ограниченность коэффициента корреляции.
Ответ. $(\sigma_\xi - \sigma_\eta)^2 \leq \operatorname{Var}(\xi + \eta) \leq (\sigma_\xi + \sigma_\eta)^2$.
9. В продукции завода брак вследствие дефекта A составляет 3%, а вследствие дефекта B — 4.5%. Годная продукция составляет 95%. Найти коэффициент корреляции дефектов A и B.
Указание. Пусть $\xi = 1$, если изделие имеет дефект A, и $\xi = 0$, если изделие не имеет дефекта A. Аналогично ввести величину η, указывающую на присутствие дефекта B. По данным из условия задачи составить таблицу совместного распределения (ξ, η).
Ответ. ≈ 0.669.