17. Коэффициент корреляции и его свойства.

Status	Completed

Ковариация

Ковариацией $cov(X,Y)$ случайных величин X,Y называется число

$$cov(X,Y) = E((X - EX)(Y - EY))$$

Свойства ковариации

1. $cov(X,Y) = E(XY) - EXEY$
2. $cov(X,X) = DX$
3. $cov(X,Y) = cov(Y,X)$
4. $cov(cX,Y) = c \cdot cov(X,Y)$

Коэффициент корреляции

Коэффициентом корреляции $p(X,Y)$ случайных величин X,Y дисперсии которых существуют и отличны от нуля, называется число

$$p(X,Y) = \frac{cov(X,Y)}{\sqrt{DX}\sqrt{DY}}$$

Свойства коэффициента корреляции

1. Если X,Y - независимы, то $p(X,Y) = 0$
2. $|p(X,Y)| \leq 1$

Докажем свойство (2). Рассмотрим преобразование $\widehat{\xi} = \dfrac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}}$ случайной величины, называемое *стандартизацией*. Случайная величина $\widehat{\xi}$ имеет нулевое математическое ожидание и единичную дисперсию:

$$\mathsf{E}\,\widehat{\xi} = \mathsf{E}\,\frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}} = \frac{\mathsf{E}\,\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}} = 0; \quad \mathsf{E}\,\widehat{\xi}^2 = \mathsf{D}\,\widehat{\xi} = \mathsf{D}\,\frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}} = \frac{\mathsf{D}\,(\xi - \mathsf{E}\,\xi)}{\mathsf{D}\,\xi} = 1.$$

Коэффициент корреляции теперь запишется проще: $\rho(\xi, \eta) = \mathsf{E}\left(\widehat{\xi} \cdot \widehat{\eta}\right)$.

Далее, неравенство $(x \pm y)^2 \geqslant 0$ равносильно неравенству

$$-\frac{1}{2}(x^2 + y^2) \leqslant xy \leqslant \frac{1}{2}(x^2 + y^2).$$

Подставив в него $\widehat{\xi}$ вместо x, $\widehat{\eta}$ вместо y и взяв математические ожидания всех частей неравенства, получим свойство (2):

$$-1 = -\frac{1}{2}\mathsf{E}\left(\widehat{\xi}^2 + \widehat{\eta}^2\right) \leqslant \rho(\xi, \eta) = \mathsf{E}\left(\widehat{\xi} \cdot \widehat{\eta}\right) \leqslant \frac{1}{2}\mathsf{E}\left(\widehat{\xi}^2 + \widehat{\eta}^2\right) = 1. \quad (20)$$

3. $|p(X,Y)| = 1 \Leftrightarrow X, Y$ - линейно связаны ($P(X = aY + b) = 1$)

Докажем свойство (3). В одну сторону утверждение проверяется непосредственно: если $\eta = a\xi + b$, то

$$\rho(\xi,\, a\xi + b) = \frac{\mathsf{E}\,(\xi(a\xi + b)) - \mathsf{E}\,\xi \cdot \mathsf{E}\,(a\xi + b)}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,(a\xi + b)}} = \frac{a\mathsf{D}\,\xi}{\sqrt{\mathsf{D}\,\xi}\sqrt{a^2\mathsf{D}\,\xi}} = \begin{cases} 1, & a > 0, \\ -1, & a < 0. \end{cases}$$

Докажем вторую часть свойства (3): если $|\rho(\xi, \eta)| = 1$, то существуют числа $a \neq 0$ и b такие, что $\mathsf{P}(\eta = a\xi + b) = 1$.

Рассмотрим сначала случай $\rho(\xi, \eta) = \rho\left(\widehat{\xi}, \widehat{\eta}\right) = 1$. Тогда второе неравенство в формуле (20) превращается в равенство:

$$\mathsf{E}\left(\widehat{\xi} \cdot \widehat{\eta}\right) = \frac{1}{2}\mathsf{E}\left(\widehat{\xi}^2 + \widehat{\eta}^2\right), \quad \text{т. е.} \quad \mathsf{E}\left(\widehat{\xi} - \widehat{\eta}\right)^2 = 0.$$

Если математическое ожидание неотрицательной случайной величины $\left(\widehat{\xi} - \widehat{\eta}\right)^2$ равно нулю, то $\left(\widehat{\xi} - \widehat{\eta}\right)^2 = 0$, п. н. Поэтому с единичной вероятностью

$$\frac{\eta - \mathsf{E}\,\eta}{\sqrt{\mathsf{D}\,\eta}} = \frac{\xi - \mathsf{E}\,\xi}{\sqrt{\mathsf{D}\,\xi}}, \qquad \eta = \frac{\sqrt{\mathsf{D}\,\eta}}{\sqrt{\mathsf{D}\,\xi}}\,\xi + \mathsf{E}\,\eta - \frac{\sqrt{\mathsf{D}\,\eta}}{\sqrt{\mathsf{D}\,\xi}}\,\mathsf{E}\,\xi = a\xi + b.$$

В случае $\rho(\xi, \eta) = -1$ нужно рассмотреть первое неравенство в формуле (20) и повторить рассуждения. Тем самым теорема 33 доказана. □