第4.3节　协方差及相关系数

一、协方差与相关系数的概念及性质

二、相关系数的意义

三、协方差矩阵

四、小结

一、协方差与相关系数的概念及性质

1. 问题的提出

若随机变量 X 和 Y 相互独立,那么

$$D(X+Y)=D(X)+D(Y).$$

若随机变量 X 和 Y 不相互独立

$$D(X+Y)=?$$

$$D(X+Y)=E(X+Y)^2-[E(X+Y)]^2$$

$$=D(X)+D(Y)+2\boxed{E\{[X-E(X)][Y-E(Y)]\}}.$$

协方差

2. 定义4.5

(X,Y)是二维随机变量,量 $E\{[X-E(X)][Y-E(Y)]\}$ 称为随机变量X 与 Y 的协方差. 记为 $\mathrm{Cov}(X,Y)$, 即

$$\mathrm{Cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\}.$$

而
$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$$

称为随机变量 X 与 Y 的相关系数.

3. 说明

(1) X 和 Y 的相关系数又称为标准 协方差,它是一个无量纲的量.

(2) 若随机变量 X 和 Y 相互独立

$$\Rightarrow \mathrm{Cov}(X,Y) = E\{[X - E(X)][Y - E(Y)]\}$$

$$= E[X - E(X)]E[Y - E(Y)] = 0.$$

(3) 若随机变量 X 和 Y 相互独立

$$\Rightarrow D(X + Y) = D(X) + D(Y)$$

$$+ 2E\{[X - E(X)][Y - E(Y)]\}$$

$$= D(X) + D(Y) + 2\mathrm{Cov}(X,Y) = D(X) + D(Y).$$

4. 协方差的计算公式

(1) $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$;

(2) $D(X+Y) = D(X) + D(Y) + 2\mathrm{Cov}(X,Y)$.

证明 (1) $\mathrm{Cov}(X,Y) = E\{[X - E(X)][Y - E(Y)]\}$

$$= E[XY - YE(X) - XE(Y) + E(X)E(Y)]$$

$$= E(XY) - 2E(X)E(Y) + E(X)E(Y)$$

$$= E(XY) - E(X)E(Y).$$

$$(2) D(X+Y) = E\{[(X+Y) - E(X+Y)]^2\}$$

$$= E\{[(X - E(X)) + (Y - E(Y)]^2\}$$

$$= E\{[X - E(X)]^2\} + E\{[Y - E(Y)]^2\}$$

$$+ 2E\{[X - E(X)][Y - E(Y)]\}$$

$$= D(X) + D(Y) + 2\mathrm{Cov}(X, Y).$$

5. 协方差的性质

(1) $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$;

(2) $\mathrm{Cov}(aX,bY)=ab\,\mathrm{Cov}(X,Y)$ $\quad a,b$为常数;

(3) $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$.

6. 相关系数的性质

(1) $|\rho_{XY}| \le 1.$

(2) $|\rho_{XY}| = 1$ 的充要条件是，存在常数 a, b 使

$$P\{Y = a + bX\} = 1.$$

证明 (1) $\min\limits_{a,b} e = E[(Y - (a + bX))^2]$

$$= (1 - \rho_{XY}^2) D(Y) \ge 0$$

$$\Rightarrow 1 - \rho_{XY}^2 \ge 0$$

$$\Rightarrow |\rho_{XY}| \le 1.$$

例1 设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,试求 X 与 Y 的相关系数.

解 由 $f(x,y)=\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp$

$$\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

$$\Rightarrow f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},\ -\infty<x<\infty,$$

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},\ -\infty<y<\infty.$$

$$\Rightarrow E(X)=\mu_1, E(Y)=\mu_2, D(X)=\sigma_1^2, D(Y)=\sigma_2^2.$$

而

$$\operatorname{Cov}(X,Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_1)(y-\mu_2)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_1)(y-\mu_2)$$

$$\cdot e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right]^2}\mathrm{d}y\,\mathrm{d}x$$

令 $t=\dfrac{1}{\sqrt{1-\rho^2}}\left(\dfrac{y-\mu_2}{\sigma_2}-\rho\dfrac{x-\mu_1}{\sigma_1}\right),\ u=\dfrac{x-\mu_1}{\sigma_1},$

$\mathrm{Cov}(X,Y)=$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_1\sigma_2\sqrt{1-\rho^2}tu+\rho\sigma_1\sigma_2u^2)e^{-\frac{u^2}{2}-\frac{t^2}{2}}\,\mathrm{d}t\,\mathrm{d}u$$

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}\left(\int_{-\infty}^{\infty}u^2e^{-\frac{u^2}{2}}\,\mathrm{d}u\right)\left(\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}\,\mathrm{d}t\right)$$

$$+\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{2\pi}\left(\int_{-\infty}^{\infty}ue^{-\frac{u^2}{2}}\,\mathrm{d}u\right)\left(\int_{-\infty}^{\infty}te^{-\frac{t^2}{2}}\,\mathrm{d}t\right)$$

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}\sqrt{2\pi}\cdot\sqrt{2\pi},$$

故有$\mathrm{Cov}(X,Y)=\rho\sigma_1\sigma_2$.

于是 $$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \rho.$$

结论

(1)二维正态分布密度函数中,参数 ρ 代表了X与 Y 的相关系数;

(2)二维正态随机变量 X 与 Y 相关系数为零等价于 X 与 Y 相互独立.

例2 已知随机变量 X,Y分别服从 $N(1,3^2),N(0,4^2)$, $\rho_{XY}=-1/2$,设 $Z=X/3+Y/2$.

(1) 求 Z 的数学期望和方差.

(2) 求 X 与 Z 的相关系数.

解 (1)由$E(X)=1, D(X)=9, E(Y)=0, D(Y)=16.$

得
$$E(Z)=E(\frac{X}{3}+\frac{Y}{2})=\frac{1}{3}E(X)+\frac{1}{2}E(Y)$$
$$=\frac{1}{3}.$$

$$D(Z) = D(\frac{X}{3}) + D(\frac{Y}{2}) + 2\mathrm{Cov}(\frac{X}{3}, \frac{Y}{2})$$

$$= \frac{1}{9}D(X) + \frac{1}{4}D(Y) + \frac{1}{3}\mathrm{Cov}(X, Y)$$

$$= \frac{1}{9}D(X) + \frac{1}{4}D(Y) + \frac{1}{3}\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}$$

$$= 1 + 4 - 2 = 3.$$

$$(2)\ \ \mathrm{Cov}(X, Z) = \mathrm{Cov}(X, \frac{X}{3} + \frac{Y}{2})$$

$$= \frac{1}{3}\mathrm{Cov}(X, X) + \frac{1}{2}\mathrm{Cov}(X, Y)$$

$$= \frac{1}{3}D(X) + \frac{1}{2}\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)} = 3 - 3 = 0.$$

故 $\rho_{XY} = \mathrm{Cov}(X, Z) \big/ (\sqrt{D(X)}\sqrt{D(Z)}) = 0.$

二、相关系数的意义

1. 相关系数的意义

当 $|\rho_{XY}|$ 较大时，表明 X, Y 的线性关系较密切.

当 $|\rho_{XY}|$ 较小时, X,Y 线性相关的程度较差.

定义：当 $\rho_{XY}=0$ 时, 称 X 和 Y 不相关.

2. 注意

(1) 不相关与相互独立的关系

$$\text{相互独立} \ \underset{\not\Longleftarrow}{\Longrightarrow}\ \text{不相关}$$

(2) 不相关的充要条件

1° X, Y 不相关 $\Leftrightarrow \rho_{XY} = 0;$

2° X, Y 不相关 $\Leftrightarrow \mathrm{Cov}(X, Y) = 0;$

3° X, Y 不相关 $\Leftrightarrow E(XY) = E(X)E(Y).$

四、小结

协方差与相关系数的定义

量 $E\{[X-E(X)][Y-E(Y)]\}$ 称为随机变量 X 与 Y 的协方差，记为 $\mathrm{Cov}(X,Y)$,

$$\mathrm{Cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$$

称 $\rho_{XY}=\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$ 为随机变量 X 与 Y 的相关系数.

协方差的性质

1. $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$.

2. $\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$. ($a,b$ 为常数)

3. $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$.

相关系数的意义

当 $|\rho_{XY}|$ 较大时，表明 X, Y 的线性关系较密切.

当 $|\rho_{XY}|$ 较小时, X,Y 线性相关的程度较差.

当 $\rho_{XY}=0$ 时,称 X 和 Y 不相关.