§ 5.4 协方差和相关系数

问题 对于二维随机变量(X,Y):

$\boxed{\text{已知联合分布}}$ $\underset{\not\longleftarrow}{\longrightarrow}$ $\boxed{\text{边缘分布}}$

这说明对于二维随机变量，除了每个随机变量各自的概率特性以外，相互之间可能还有某种联系. 问题是用一个什么样的数去反映这种联系.

数 $E\big((X-E(X))(Y-E(Y))\big)$

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

反映了随机变量X,Y之间的某种关系

协方差和相关系数的定义

定义 称 $E\big((X-E(X))(Y-E(Y))\big)$ 为X,Y的协方差. 记为

$$\boxed{\operatorname{cov}(X,Y)=E\big((X-E(X))(Y-E(Y))\big)}$$

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

称 $\begin{pmatrix} D(X) & \operatorname{cov}(X,Y) \\ \operatorname{cov}(X,Y) & D(Y) \end{pmatrix}$

为（X, Y ）的协方差矩阵

若$D(X)>0, D(Y)>0$,称

$$E\left(\frac{(X-E(X))(Y-E(Y))}{\sqrt{D(X)}\sqrt{D(Y)}}\right)=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为X,Y的 相关系数，记为

$$\boxed{\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}}$$

事实上， $\rho_{XY}=\operatorname{cov}(X^*,Y^*)$

$$\operatorname{cov}(X^*, Y^*) = E[(X^* - EX^*)(Y^* - EY^*)]$$

$$= E\left[\left(\frac{X-EX}{\sqrt{DX}} - E\left(\frac{X-EX}{\sqrt{DX}}\right)\right)\left(\frac{Y-EY}{\sqrt{DY}} - E\left(\frac{Y-EY}{\sqrt{DY}}\right)\right)\right]$$

$$= E\left[\left(\frac{X-EX}{\sqrt{DX}}\right)\left(\frac{Y-EY}{\sqrt{DY}}\right)\right]$$

$$= \frac{E\left[\left(X-EX\right)\left(Y-EY\right)\right]}{\sqrt{DX}\sqrt{DY}} = \rho_{XY}$$

若 $\boldsymbol{\rho}_{XY} = 0$, 称 X, Y 不相关.

协方差和相关系数的计算

——利用函数的期望或方差计算协方差

- $\boxed{\operatorname{cov}(X,Y)=E\big((X-E(X))(Y-E(Y))\big)}$

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

$$=\frac{1}{2}\big(D(X+Y)-D(X)-D(Y)\big)$$

$$=-\frac{1}{2}\big(D(X-Y)-D(X)-D(Y)\big)$$

❑ 若 (X,Y) 为离散型，

$$\boxed{\operatorname{cov}(X,Y)=E\big((X-E(X))(Y-E(Y))\big)}$$

$$=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(x_i-E(X))(y_j-E(Y))p_{ij}$$

❑ 若 (X,Y) 为连续型，

$$\boxed{\operatorname{cov}(X,Y)=E\big((X-E(X))(Y-E(Y))\big)}$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-E(X))(y-E(Y))f(x,y)dxdy$$

例1 已知 X,Y 的联合分布为

$Y \backslash X$ (p_{ij})	1	0
1	p	0
0	0	q

$0 < p < 1$

$p + q = 1$

求 $\operatorname{cov}(X,Y)$, ρ_{XY}

解

X	1	0
P	p	q

Y	1	0
P	p	q

XY	1	0
P	p	q

X	1	0
P	p	q

Y	1	0
P	p	q

XY	1	0
P	p	q

$$\left.\begin{array}{ll} E(X)=p, & D(X)=pq, \\ E(Y)=p, & D(Y)=pq, \\ E(XY)=p, & D(XY)=pq, \end{array}\right\} \Rightarrow$$

$$\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)$$
$$=p-p\cdot p \quad =pq$$

$$\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \quad =\frac{pq}{\sqrt{pq}\sqrt{pq}}=1$$

$$X^*=\frac{X-p}{\sqrt{pq}},Y^*=\frac{Y-p}{\sqrt{pq}}, \qquad P(X^*=Y^*)=1$$

例2 设 $(X,Y) \sim N(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$，求 ρ_{XY}

解 $\operatorname{cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\boldsymbol{\mu}_1)(y-\boldsymbol{\mu}_2)f(x,y)dxdy$

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

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

令 $\frac{x-\mu_1}{\sigma_1}=s$

$\frac{y-\mu_2}{\sigma_2}=t$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\sigma_1 s\cdot\sigma_2 t\cdot\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$e^{-\frac{1}{2(1-\rho^2)}(s^2-2\rho st+t^2)}(\sigma_1 ds)(\sigma_2 dt)$$

$$=\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} ste^{-\frac{1}{2(1-\rho^2)}[(s-\rho t)^2+(1-\rho^2)t^2]}dsdt$$

$$=\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} ste^{-\frac{1}{2(1-\rho^2)}(s-\rho t)^2-\frac{1}{2}t^2}dsdt$$

$$\overset{令 s-\rho t=u}{=}\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} t(\rho t+u)e^{-\frac{u^2}{2(1-\rho^2)}-\frac{1}{2}t^2}dudt$$

$$= \frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\left[\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\rho\, t^2\cdot e^{-\frac{u^2}{2(1-\rho^2)}-\frac{1}{2}t^2}dudt + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} t\, ue^{-\frac{u^2}{2(1-\rho^2)}-\frac{1}{2}t^2}dudt\right]$$

$$= \frac{\sigma_1\sigma_2\rho}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{u^2}{2(1-\rho^2)}}\cdot t^2 e^{-\frac{1}{2}t^2}dtdu$$

$$=\frac{\sigma_1\sigma_2\rho}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}e^{-\frac{u^2}{2(1-\rho^2)}}du\int_{-\infty}^{+\infty}t^2e^{-\frac{1}{2}t^2}dt$$

$$=\frac{\sigma_1\sigma_2\rho}{2\pi\sqrt{1-\rho^2}}\cdot\sqrt{2\pi\cdot(1-\rho^2)}\cdot\sqrt{2\pi}\quad=\boldsymbol{\sigma_1\sigma_2\rho}$$

$$\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{\sigma_1\sigma_2\rho}{\sqrt{\sigma_1^2}\sqrt{\sigma_2^2}}=\rho$$

若 $(X,Y)\sim N(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2,\rho)$,

则 X,Y 相互独立 $\Longleftrightarrow$ X,Y 不相关

例3 设 X,Y 相互独立，且都服从 $N(0,\sigma^2)$，
$U = aX + bY$，$V = aX - bY$，a,b 为常数，
且都不为零，求 ρ_{UV}

解 $\operatorname{cov}(U,V) = E(UV) - E(U)E(V)$

$$= E(a^2X^2 - b^2Y^2)$$

$$-\left[E(aX+bY)\right]\left[E(aX-bY)\right]$$

$$= a^2E(X^2) - b^2E(Y^2)$$

$$-\left[aE(X)+bE(Y)\right]\left[aE(X)-bE(Y)\right]$$

$$= a^2 E(X^2) - b^2 E(Y^2)$$

$$-\left[a^2 E^2(X) - b^2 E^2(Y) \right]$$

$$= a^2 \left[E(X^2) - E^2(X) \right] - b^2 \left[E(Y^2) - E^2(Y) \right]$$

$$= a^2 D(X) - b^2 D(Y)$$

$$= (a^2 - b^2)\sigma^2$$

而
$$D(U)=D(aX+bY)$$
$$=a^2D(X)+b^2D(Y) \quad =(a^2+b^2)\sigma^2$$
$$D(V)=D(aX-bY)$$
$$=a^2D(X)+b^2D(Y) \quad =(a^2+b^2)\sigma^2$$

故
$$\rho_{UV}=\frac{Cov(U,V)}{\sqrt{D(U)}\cdot\sqrt{D(V)}}$$
$$=\frac{(a^2-b^2)\sigma^2}{\sqrt{(a^2+b^2)\sigma^2}\cdot\sqrt{(a^2+b^2)\sigma^2}}$$
$$=\frac{a^2-b^2}{a^2+b^2}$$