§ 4.4 协方差和相关系数

问题 对于二维随机变量(X,Y):

已知联合分布 $\rightarrow$ 边缘分布；边缘分布 $\not\rightarrow$ 已知联合分布

对二维随机变量,除每个随机变量各自的概率特性外, 相互之间可能还有某种联系问题是用一个怎样的数去反映这种联系.

数 $E\big([X-E(X)][Y-E(Y)]\big)$

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

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

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

可以证明 协方差矩阵 为 半正定矩阵

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

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

→ 无量纲的量

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

若 $\rho_{XY}=0$，称 X,Y 不相关.

协方差和相关系数的计算

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

$$=\pm\frac{1}{2}\left(D(X\pm Y)-D(X)-D(Y)\right)$$

若(X,Y)为离散型，

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

若(X,Y)为连续型，

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

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 $\text{cov}(X, Y)$, ρ_{XY}

解

X	1	0
P	p	q

Y	1	0
P	p	q

XY	1	0
P	p	q

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

$$\text{cov}(X, Y) = pq,\quad \rho_{XY} = 1$$

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

解 $\operatorname{cov}(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)f(x,y)dxdy$

$$\overset{\frac{x-\mu_1}{\sigma_1}=s}{\underset{\frac{y-\mu_2}{\sigma_2}=t}{=\!=\!=}}\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}ste^{-\frac{1}{2(1-\rho^2)}(s-\rho t)^2-\frac{1}{2}t^2}dsdt$$

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

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

$$=\sigma_1\sigma_2\rho$$

$$\rho_{XY}=\rho$$

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$\Theta \sim U(0,2\pi)$, $X=\cos\Theta$, $Y=\cos(\Theta+\alpha)$, α 是给定的常数，求 ρ_{XY}

解

$$f_\Theta(t)=\begin{cases}\dfrac{1}{2\pi}, & 0<t<2\pi,\\ & \text{其他}\end{cases}$$

$$E(X)=\int_0^{2\pi}\cos t\cdot\frac{1}{2\pi}dt=0,$$

$$E(Y)=\int_0^{2\pi}\cos(t+\alpha)\cdot\frac{1}{2\pi}dt=0,$$

$$E(XY)=\int_0^{2\pi}\cos(t)\cos(t+\alpha)\cdot\frac{1}{2\pi}dt=\frac{1}{2}\cos\alpha$$

→ $$\operatorname{cov}(X,Y)=\frac{1}{2}\cos\alpha$$

$$E(X^2)=\int_0^{2\pi}\cos^2 t\cdot\frac{1}{2\pi}dt=\frac{1}{2},\qquad D(X)=\frac{1}{2},$$

$$E(Y^2)=\int_0^{2\pi}\cos^2(t+\alpha)\cdot\frac{1}{2\pi}dt=\frac{1}{2},\qquad D(Y)=\frac{1}{2},$$

→ $$\rho_{XY}=\cos\alpha$$

若 $\alpha=0,\ \rho_{XY}=1 \longrightarrow Y=X$

若 $\alpha=\pi,\ \rho_{XY}=-1 \longrightarrow Y=-X$

$\left.\begin{matrix} Y=X \\ Y=-X \end{matrix}\right\} \longrightarrow$

$|\rho_{XY}|=1 \longrightarrow$ X,Y 有线性关系

若 $\alpha=\dfrac{\pi}{2},\dfrac{3\pi}{2},\ \rho_{XY}=0$ X,Y 不相关，

但 X,Y 不独立，

X,Y 没有线性关系，但有函数关系

$$X^2+Y^2=1$$

协方差和相关系数的性质

协方差的性质

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

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

- $\operatorname{cov}(aX,bY)=ab\operatorname{cov}(X,Y)$

- $\operatorname{cov}(X+Y,Z)=\operatorname{cov}(X,Z)+\operatorname{cov}(Y,Z)$

- $\operatorname{cov}(X,X)=D(X)$

❑ $|\operatorname{cov}(X,Y)|^2 \le D(X)D(Y)$

当$D(X)>0, D(Y)>0$时，当且仅当

$$P\{Y-E(Y)=t_0[X-E(X)]\}=1$$

时，等式成立 —Cauchy-Schwarz不等式

证 令$g(t)=E\{[Y-E(Y)]-t[X-E(X)]\}^2$

$$=D(Y)-2t\operatorname{cov}(X,Y)+t^2D(X)$$

对任何实数 t，$g(t)\ge 0$ ⟶

$$4\operatorname{cov}^2(X,Y)-4D(X)D(Y)\le 0$$

即 $|\operatorname{cov}(X,Y)|^2 \le D(X)D(Y)$

等号成立 ⟷ $g(t)=0$ 有两个相等的实零点

$$t_0=\frac{\operatorname{cov}(X,Y)}{D(X)}=\left(\pm\sqrt{\frac{D(Y)}{D(X)}}\right)$$

$g(t_0)=0$ 即

$$\left.\begin{array}{l} E[(Y-E(Y))-t_0(X-E(X))]^2=0 \\ \text{显然}\quad E[(Y-E(Y))-t_0(X-E(X))]=0 \end{array}\right\}$$

$\Longleftrightarrow \quad D[(Y-E(Y))-t_0(X-E(X))]=0$

$\Longleftrightarrow \quad P[(Y-E(Y))-t_0(X-E(X))=0]=1$

$$P[(Y-E(Y))-t_0(X-E(X))=0]=1$$

即 $P[(Y-E(Y))=t_0(X-E(X))]=1$

即 Y 与 X 有线性关系的概率等于1, 这种线性关系为

$$P\left(\frac{Y-E(Y)}{\sqrt{D(Y)}}=\pm\frac{X-E(X)}{\sqrt{D(X)}}\right)=1$$

完全类似地可以证明

$$E^2(XY)\le E(X^2)E(Y^2)$$

当 $E(X^2)>0, E(Y^2)>0$ 时,当且仅当

$$P(Y=t_0X)=1$$

时, 等式成立.

相关系数的性质

- $|\rho_{XY}| \le 1$
- $|\rho_{XY}| = 1$ ⟷ Cauchy-Schwarz不等式的等号成立

⟷ 即Y与X有线性关系的概率等于1，这种线性关系为

$$P\left(Y^* = \pm X^*\right) = 1$$

$$X^* = (X - EX)/\sqrt{D(X)},\quad Y^* = (Y - EY)/\sqrt{D(Y)}.$$

$\rho_{XY} = 1$ → $\operatorname{cov}(X, Y) > 0$

$$P\left(Y^* = X^*\right) = 1$$

$\rho_{XY} = -1$ → $\operatorname{cov}(X, Y) < 0$

$$P\left(Y^* = -X^*\right) = 1$$

如例1中 X, Y 的联合分布为

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

$0 < p < 1$
$p + q = 1$

已求得 $\rho_{XY} = 1$，则必有

$$P(X^* = Y^*) = 1$$

其中 $X^* = (X - p)/\sqrt{pq}$, $Y^* = (Y - p)/\sqrt{pq}$.

❑ $\rho_{XY} = 0 \longleftrightarrow X, Y$ 不相关

$\longleftrightarrow \operatorname{cov}(X, Y) = 0$

$\longleftrightarrow E(XY) = E(X)E(Y)$

$\longleftrightarrow D(X \pm Y) = D(X) + D(Y)$

X, Y 相互独立 $\underset{\not\Leftarrow}{\Rightarrow}$ X, Y 不相关

若 (X, Y) 服从二维正态分布，

X, Y 相互独立 $\longleftrightarrow$ X, Y 不相关

例4 设 $(X, Y) \sim N(1,4;\ 1,4;\ 0.5)$, $Z = X + Y$, 求 ρ_{XZ}

解 $E(X) = E(Y) = 1, D(X) = D(Y) = 4,$

$$\rho_{XY} = 1/2,\ \operatorname{cov}(X,Y) = 2$$

$$\operatorname{cov}(X,Z) = \operatorname{cov}(X,X) + \operatorname{cov}(X,Y) = 6$$

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

$$= D(X) + D(Y) + 2\operatorname{cov}(X,Y) = 12$$

$$\therefore\quad \rho_{XZ} = 3/\sqrt{12} = \sqrt{3/2}.$$

作业 P.173 习题四

23 25

26 28

30 ~ 32

附录 矩在线性回归中 的应用

若 X, Y 是两个r.v., 用 X 的线性函数去逼近 Y 所产生的平均平方误差为

$$E[Y-(aX+b)]^2$$

当取 $\hat{a}=\dfrac{\mathrm{cov}(X,Y)}{D(X)}$,

$$\hat{b}=E(Y)-\hat{a}E(X)=E(Y)-\rho_{XY}\sqrt{\frac{D(Y)}{D(X)}}E(X)$$

平均平方误差最小.

附例 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 $\mathrm{cov}(U,V)=E(UV)-E(U)E(V)$

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

$$-[aE(X)+bE(Y)][aE(X)-bE(Y)]$$

由 $\left.\begin{aligned}&E(X)=E(Y)=0,\\&D(X)=D(Y)=\sigma^2\end{aligned}\right\}\longrightarrow\begin{aligned}&E(X^2)=\sigma^2\\&E(Y^2)=\sigma^2\end{aligned}$

$\longrightarrow$ $\mathrm{cov}(U,V)=(a^2-b^2)\sigma^2$

而 $D(U)=a^2D(X)+b^2D(Y)=(a^2+b^2)\sigma^2$

$D(V)=a^2D(X)+b^2D(Y)=(a^2+b^2)\sigma^2$

故 $\rho_{UV}=\dfrac{a^2-b^2}{a^2+b^2}$

继续讨论

a,b 取何值时, U与V不相关?

此时, U与V是否独立?

若 $a=b$, $\rho_{UV}=0$, 则 U,V 不相关.

但 $U\sim N(0,\ 2a^2\sigma^2)$, $V\sim N(0,\ 2a^2\sigma^2)$,

$$\begin{cases} U=a(X+Y) \\ V=a(X-Y) \end{cases} \longrightarrow \begin{cases} X=\dfrac{1}{2a}(U+V) \\ Y=\dfrac{1}{2a}(U-V) \end{cases}$$

$$f_{UV}(u,v)=\left|\begin{vmatrix} \dfrac{1}{2a} & \dfrac{1}{2a} \\ \dfrac{1}{2a} & -\dfrac{1}{2a} \end{vmatrix}\right| f_{XY}\left(\frac{1}{2a}(u+v),\frac{1}{2a}(u-v)\right)$$

$$f_{UV}(u,v)=\frac{1}{2a^2}f_X\left(\frac{1}{2a}(u+v)\right)f_Y\left(\frac{1}{2a}(u-v)\right)$$

$$=\frac{1}{2a^2}\frac{1}{2\pi\sigma^2}e^{-\frac{\left(\frac{u+v}{2a}\right)^2+\left(\frac{u-v}{2a}\right)^2}{2\sigma^2}}$$

$$=\frac{1}{2\pi(\sqrt{2}a\sigma)^2}e^{-\frac{u^2+v^2}{4a^2\sigma^2}}$$

$(U,V)\sim N(0,2a^2\sigma^2;0,2a^2\sigma^2;0)$,

且U,V相互独立