§ 4.4协方差和相关系数

问题 对于二维随机变量(X,Y):

已知联合分布之边缘分布



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
$$E((X-E(X))(Y-E(Y)))$$

反映了随机变量X,Y之间的某种关

● 协方差和相关系数的定义

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

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

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

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

为(X,Y)的协方差矩

高维协方差矩阵

若D(X) > 0, D(Y) > 0, 称

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

为X,Y的相关系数,记

为

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

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

若 $\rho_{XY} = 0$, 称 X, Y 不相

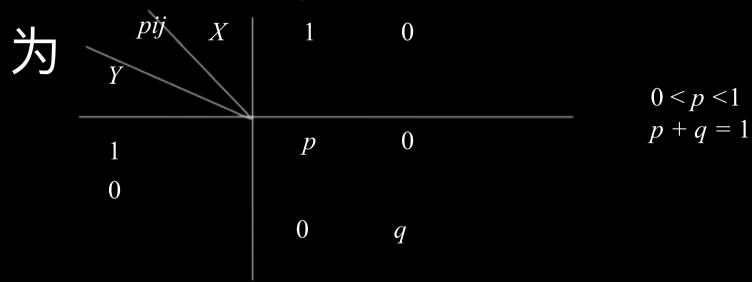
•若(X,Y)为离散型,

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

•若(X,Y)为连续型,

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

例1 已知 X, Y 的联合分布



求 cov(X,Y), $\Box XY$

解

X	1 0	Y	1 0	XY	1 0
P	p q	P	p q	P	p q

$$E(X) = p, E(Y) = p,$$

$$D(X) = pq, D(Y) = pq,$$

$$E(XY) = p, D(XY) = pq,$$

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

设
$$(X,Y) \sim N([]1,[]12,[]2,[]22,[])$$
,求 $[]XY$

$$\text{ for } cov(X,Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_1)(y - \mu_2) f(x,y) dx dy$$

$$\frac{\frac{x-\mu_{1}}{\sigma_{1}}=s}{\frac{y-\mu_{2}}{\sigma_{2}}=t} = \frac{\sigma_{1}\sigma_{2}}{2\pi\sqrt{1-\rho^{2}}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ste^{-\frac{1}{2(1-\rho^{2})}(s-\rho t)^{2}-\frac{1}{2}t^{2}} dsdt$$

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

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

$$=\sigma_1\sigma_2\rho$$

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

若
$$(X,Y) \sim N(\Box 1, \Box 12, \Box 2, \Box 22, \Box 22,$$

例3设 $\square \sim U(0,2\square), X = \cos \square, Y = \cos(\square + \square),$ \(\alpha\) 是给定的常数,求 $\square XY$

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

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

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

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

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

$$D(X) = \frac{1}{2},$$

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

$$D(Y)=\frac{1}{2},$$

$$\rightarrow \rho_{XY} = \cos \alpha$$

$$|\rho_{XY}|=1$$

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

$$\rightarrow$$
 $Y = X$



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

X,Y 有线性关系

若
$$\alpha = \frac{\pi}{2}, \frac{3\pi}{2}, \rho_{XY} = 0$$
 X, Y 不相关,但 X, Y 不独立,

此时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

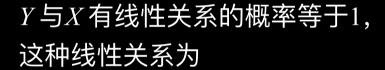
- cov(X,Y) = cov(Y,X) = E(XY) E(X)E(Y)
- cov(aX,bY) = ab cov(X,Y)
- cov(X + Y, Z) = cov(X, Z) + cov(Y, Z)
- $\operatorname{cov}(X, X) = D(X)$

相关系数的性质

 $|\rho_{XY}| \leq 1$

简证

• $|\rho_{XY}|=1$



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

简证

$$\leftarrow$$
 $cov(X,Y) = 0$

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

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

X,Y相互独立



X, Y不相关



若X,Y服从二维正态分布,

X,Y相互独立



X,Y不相关

$$E(X) = E(Y) = 1, D(X) = D(Y) = 4,$$

$$\rho_{XY} = \frac{1}{2}, \quad \text{cov}(X, Y) = 2$$

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

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

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

$$\rho_{XZ} = \frac{6}{2\sqrt{12}} = \frac{\sqrt{3}}{2}$$

作业习题四

A组: 20, 21