4.3 协方差和相关系数

4.4 其他数字特征

4.5 综合例题

主讲人：郭峰

办公室：校本部 创新园大厦B1407

电话：84708351-8088

Email：fguo ‘AT’ dlut.edu.cn

4.4 协方差，相关系数

一 协方差

定义 设(X,Y)是二维随机变量，如果

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

存在, 则称它是X与Y的协方差，记为$\mathrm{Cov}(X,Y)$

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

当X与Y是离散型随机变量时，分布律$P(X=x_i,Y=y_j)=p_{ij}$

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

当X与Y是连续型随机变量时，密度函数$f(x,y)$

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

由协方差定义可得，对任意的随机变量X、Y，有

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

——协方差的一个计算公式。

例4.15 设二维随机变量(X,Y)的密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{(b-a)(d-c)} & a\le x\le b, c\le y\le d\\ 0 & \text{其它}\end{cases}$$

求Cov(X,Y)

解 $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_c^d\dfrac{1}{(b-a)(d-c)}dy=\dfrac{1}{b-a}\quad a\le x\le b$

$$E(X)=\int_{-\infty}^{+\infty}xf_X(x)dx=\int_a^b x\cdot\frac{1}{b-a}dx=\frac{b+a}{2}$$

同理 $E(Y)=\dfrac{c+d}{2}$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy=\int_a^b dx\int_c^d xy\frac{1}{(b-a)(d-c)}dy$$

$$=\frac{(a+b)(c+d)}{4}$$

Cov(X,Y)=$E(XY)-E(X)E(Y)$=0

二、协方差的性质

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

(2) $\mathrm{Cov}(X,X)=D(X)$，$\mathrm{Cov}(X,C)=0$；

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

(4) 若 X 与 Y 独立，则$\mathrm{Cov}(X,Y)=0$；

(5) $\mathrm{Cov}(X+Y,Z)=\mathrm{Cov}(X,Z)+\mathrm{Cov}(Y,Z)$；

(6) 和的方差与协方差的关系

$\mathrm{D(aX+bY)=Cov(aX+bY,ax+bY)}$

$\mathrm{=Cov(aX,aX)+2Cov(aX,bY)+Cov(bY,bY)}$

$=a^2\mathrm{D(X)}+b^2\mathrm{D(Y)+2abCov(X,Y)}$

4.4 协方差，相关系数

二、相关系数

定义 设(X,Y)是二维随机变量，当$D(X)>0,D(Y)>0$时称

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

为X与Y的相关系数，或称X与Y的标准协方差。

ρ_{XY}是一个无量纲的量。

例4.16 设二维随机变量(X,Y)的联合分布律为

$X \backslash Y$	0	1
0	q	0
1	0	p

其中$p+q=1$，求相关系数ρ_{XY}。

解 由题意可得X,Y的边缘分布律为

X	0	1
P	q	p

Y	0	1
P	q	p

均为0—1分布，$E(X)=p$，$D(X)=pq$，$E(Y)=p$，$D(Y)=pq$，

所以$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$

$$=0\times0\times q+0\times1\times0+1\times0\times0+1\times1\times p-p\times p$$

$$=p-p^2=pq$$

因此

$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{pq}{\sqrt{pq}\sqrt{pq}}=1$$

三、相关系数的性质

相关系数有如下性质：

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

(2) 当 $\rho_{XY}=1$时，存在常数$a>0$及b，使得
$Y=aX+b$

三、相关系数的性质

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

证明：对于任意实数α

$$D(\alpha X+Y)$$

$$=\alpha^2 D(X)+D(Y)+2Cov(X,Y)\alpha \ge 0$$

所以：α的二次函数$\alpha^2 D(X)+D(Y)+2Cov(X,Y)\alpha$

的判别式≤ 0

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

所以：

$$|\rho_{XY}| = \left|\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}\right| \le 1$$

三、相关系数的性质

(2) 当 $\rho_{XY}=1$时，存在常数$a>0$及b，使得$Y=aX+b$

当 $\rho_{XY}=1$ 时，

$$Cov(X,Y)=\sqrt{D(X)}\cdot\sqrt{D(Y)}$$

所以：

$$D(\sqrt{D(X)}\,Y-\sqrt{D(Y)}\,X)$$

$$=D(X)D(Y)-2\sqrt{D(X)D(Y)}Cov(X,Y)+D(Y)D(X)$$

$$=0$$

三、相关系数的性质

(2) 当 $\rho_{XY} = 1$时，存在常数$a > 0$及b，使得$Y = aX + b$

所以：

$$\sqrt{D(X)}\ Y - \sqrt{D(Y)}\ X = b$$

$$Y = \frac{\sqrt{D(Y)}}{\sqrt{D(X)}}\ X + b$$

当 $\rho_{XY}=1$ 时，X,Y 正线性相关

三、相关系数的性质

(2)* $\rho_{XY} = 1$ 当且仅当 存在常数$a > 0$及b，使得 $Y = aX + b$

证明：假设存在常数$a > 0$及b，使得 $Y = aX + b$ 则

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{Cov(X,aX+b)}{\sqrt{D(X)}\sqrt{D(aX+b)}}$$

$$= \frac{aCov(X,X)}{\sqrt{D(X)}|a|\sqrt{D(X)}} = 1$$

(3) $\rho_{XY} = -1$ 当且仅当存在常数$a < 0$及b，使得 $Y = aX + b$ 。（X, Y 负线性相关）

定义 若$\rho_{XY}=0$，则称X与Y不相关。

若X与Y相互独立，则必有$\rho_{XY}=0$，即X与Y不相关。

证明 X与Y相互独立，有$E(XY)=E(X)E(Y)$

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=0$$

所以

$$\rho_{XY}=0$$

即X与Y不相关。

注意：X与Y不相关，X与Y未必相互独立。

所谓不相关只是就线性关系而言，而相互独立是就一般关系而言的。

$\rho_{XY}=0$ 称X与Y不相关

$\rho_{XY}>0$ 称X与Y正相关

特别$\rho_{XY}=1$，X与Y完全正相关，有
$Y=aX+b$（a>0）

$\rho_{XY}<0$ 称X与Y负相关

特别$\rho_{XY}=-1$，X与Y完全负相关，
有$Y=aX+b$（a<0）

注：这里的相关都是指的**线性**相关。

定义: 称 $X^* = \dfrac{X - E(X)}{\sqrt{D(X)}}$ 为 X 的标准化变量

性质: (1) $E(X^*)=0$， $D(X^*)=1$

(2) $\rho_{XY} = \rho_{X^*Y^*} = \mathrm{Cov}(X^*, Y^*)$

$$\rho_{XY} = \mathrm{Cov}(X^*, Y^*) = \mathrm{Cov}\left[\frac{X - E(X)}{\sqrt{D(X)}}, \frac{Y - E(Y)}{\sqrt{D(Y)}}\right] = \frac{\mathrm{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

例4.17 设(X, Y)在$D=\{(x,y)|x^2+y^2\le r^2\}$上服从均匀分布，(1)求$\rho_{XY}$；(2)讨论$X$与$Y$的独立性。

解 (1) $$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy=\int_{-r}^{r}dx\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}}x\cdot\frac{1}{\pi r^2}dy=0$$

$$E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy=\int_{-r}^{r}dy\int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}}y\cdot\frac{1}{\pi r^2}dx=0$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy=\int_{-r}^{r}dy\int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}}xy\cdot\frac{1}{\pi r^2}dx=0$$

$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=0$， 所以$\rho_{XY}=0$，$X$与$Y$不相关。

(2) $$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\begin{cases}\dfrac{2\sqrt{r^2-x^2}}{\pi r^2} & -r\le x\le r\\ 0 & \text{其它}\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\begin{cases}\dfrac{2\sqrt{r^2-y^2}}{\pi r^2} & -r\le y\le r\\ 0 & \text{其它}\end{cases}$$

显然

$$f(x,y)\ne f_X(x)\cdot f_Y(y)$$

X与Y不独立。

例4.18 设二维随机变量 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2;\rho)$

则可求得协方差 $\mathrm{Cov}(X,Y)=\rho\sigma_1\sigma_2$

且相关系数 $\rho_{XY}=\rho$

二维正态变量(X,Y)，X与Y相互独立的充分必要条件是$\rho=0$；

而$\rho_{XY}=\rho=0$表示X与Y不相关,

可见，X与Y独立的充分必要条件是X与Y不相关。

二维正态随机变量(X,Y)，

X与Y独立　　等价于　　X与Y不相关

定理：设二维随机变量 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$，令 $Z=aX+bY, W=cX+dY$, 其中 a,b,c,d 为任意常数，则 (Z,W) 也服从二维正态分布。