第三节　矩、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、小结

一、矩的基本概念

设 X 和 Y 是随机变量,若$E(X^k)$, $k=1,2,\cdots$ 存在,称它为 X 的 k **阶原点矩**,简称 k **阶矩**.

$$\alpha_k = E(X^k), \quad (k=1,2,\cdots)$$

$$k=1, \quad E(X), \quad k=2, \quad E\left(X^2\right)$$

连续型

$$E(X^k)=\int_{-\infty}^{+\infty} x^k f(x)dx$$

若 $\mu_k = E\{[X - E(X)]^k\},\quad k = 2,3,\cdots$

存在,称它为 X 的 k **阶中心矩**.

当 $k=2$时,2阶中心矩 $E(X-EX)^2 = DX$

连续型

$$E\left\{\left[X - E(X)\right]^k\right\} = \int_{-\infty}^{+\infty}\left[X - E(X)\right]^k f\left(x\right)dx$$

若 $E(X^kY^l),\quad k,l=1,2,\cdots$

存在,称它为 X 和 Y 的$k+l$阶混合矩.

若 $E\{[X-E(X)]^k[Y-E(Y)]^l\},\ k,l=1,2,\cdots$

存在,称它为 X 和 Y 的 $k+l$ 阶混合中心矩.

$$k=1,l=1\ ,\ E\{[X-E(X)][Y-E(Y)]\}$$

协方差

以上所有实质为随机变量函数的期望，所以为数

2. 说明

(1) 以上数字特征都是随机变量函数的数学期望;

(2) 随机变量 X 的数学期望 $E(X)$ 是 X 的一阶原点矩,方差为二阶中心矩, 协方差 $\mathrm{Cov}(X,Y)$是 X 与 Y 的二阶混合中心矩;

(3) 在实际应用中,高于 4 阶的矩很少使用.

三阶中心矩$E\{[X-E(X)]^3\}$主要用来衡量随机变量的分布是否有偏.

四阶中心矩 $E\{[X-E(X)]^4\}$主要用来衡量随机变量的分布在均值附近的陡峭程度如何.

偏度系数

假设X为一随机变量，若随机变量X的 k 阶中心距 $\mu_k = E\{[X-E(X)]^k\},\quad k=2,3,\cdots$ 存在，

则称
$$\gamma(X)=\frac{E[X-E(X)]^3}{\left[D(X)\right]^{\frac{3}{2}}}$$

为随机变量X的偏度系数.

峰度系数

假设X为一随机变量，若随机变量X的 k 阶中心距 $\mu_k = E\{[X-E(X)]^k\}, \quad k=2,3,\cdots$ 存在，

则称 $$\kappa(X) = \frac{E[X-E(X)]^4}{\left[D(X)\right]^2}$$

为随机变量X的峰度系数.

峰度系数

$$X \sim N\left(\mu, \sigma^2\right)$$

则

$$\kappa\left(X\right) = \frac{\mu^4}{\sigma^4} = 3$$

X总是跟正态分布比较，故也可写成：

$$\kappa\left(X\right) = \frac{E[X - E(X)]^4}{\left[D\left(X\right)\right]^2} - 3$$

二、协方差与相关系数的概念及性质

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 $$E[(X-EX)(Y-EY)]$$

反映了随机变量X,Y之间的某种关系

对二维随机变量(X,Y)来说，数字特征EX、EY只反映了X与Y各自的平均值，而DX、DY只反映了X与Y各自离开平均值的偏离程度，它们对X与Y之间的相互联系没有提供任何信息.

自然，我们也希望有一个数字特征能够在一定程度上反映这种相互联系.

在证明方差的性质(iii)若X与Y相互独立，则

$D(X+Y)= DX+DY$，$D(X-Y)= DX-DY$

时，我们曾得到

$E(X-EX)(Y-EY)= 0.$

这说明当$E(X-EX)(Y-EY)\neq 0$时，X与Y肯定不独立.

进一步的研究表明$E(X-EX)(Y-EY)$的数值，在一定程度上反映了X与Y之间的相互联系，因而引入如下的定义.

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

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

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

当 $X=Y$:

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

协方差是方差的推广，方差是协方差的特殊情况

考虑到 X 与 Y 的量纲差别有可能太大，故消除量纲影响：

而
$$\rho_{XY}=\frac{\operatorname{Cov}(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$$
称为随机变量 X 与 Y 的相关系数.

3. 协方差的计算公式

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

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

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

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

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

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

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

4. 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

$$\Rightarrow \operatorname{Cov}(X,Y) = E\{[X-E(X)][Y-E(Y)]\}$$
$$= E[X-E(X)]E[Y-E(Y)]$$
$$= 0.$$

(3) 若随机变量 X 和 Y 相互独立

$$\Rightarrow D(X+Y) = D(X)+D(Y)$$
$$+2E\{[X-E(X)][Y-E(Y)]\}$$
$$= D(X)+D(Y)+2\operatorname{Cov}(X,Y) = D(X)+D(Y).$$

5. 性质

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

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

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

$$=Cov(Y,X)$$

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

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

$$\begin{aligned}\mathrm{Cov}(X,C)&=E\left\{\left[X-E\left(X\right)\right]\left[C-E\left(C\right)\right]\right\}\\&=E\left\{\left[X-E\left(X\right)\right]\times 0\right\}\\&=E\left(0\right)=0\end{aligned}$$

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

$$\mathrm{Cov}(aX,bY)=E\left\{\left[aX-E\left(aX\right)\right]\left[bY-E\left(bY\right)\right]\right\}$$

$$=E\left\{a\left[X-E\left(X\right)\right]b\left[Y-E\left(Y\right)\right]\right\}$$

$$=abCov(X,Y)$$

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

$$\mathrm{Cov}(X_1+X_2,Y)=E\left\{\left[X_1+X_2-E\left(X_1+X_2\right)\right]\left[Y-E\left(Y\right)\right]\right\}$$

$$=E\left\{\left[X_1+X_2-\left[E\left(X_1\right)+E\left(X_2\right)\right]\right]\left[Y-E\left(Y\right)\right]\right\}$$

$$=E\left\{\left[X_1-E\left(X_1\right)+X_2-E\left(X_2\right)\right]\left[Y-E\left(Y\right)\right]\right\}$$

$$=E\left\{\left[X_1-E\left(X_1\right)\right]\left[Y-E\left(Y\right)\right]+\left[X_2-E\left(X_2\right)\right]\left[Y-E\left(Y\right)\right]\right\}$$

$$=E\left\{\left[X_1-E\left(X_1\right)\right]\left[Y-E\left(Y\right)\right]\right\}+E\left\{\left[X_2-E\left(X_2\right)\right]\left[Y-E\left(Y\right)\right]\right\}$$

$$=Cov(X_1,Y)+Cov(X_2,Y)$$

(4) 若随机变量 X 和 Y 相互独立

$$\Rightarrow \mathrm{Cov}(X,Y)=0$$

$$\not\Leftarrow$$

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

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

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

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

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

而

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

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

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

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

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

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

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

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

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

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

$$\rho=\frac{\mathrm{Cov}(X,Y)}{\sigma_1\sigma_2}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)D(Y)}}=\rho_{XY}$$

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

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.

协方差的大小在一定程度上反映了 X 和 Y 相互间的关系，但它还受X与Y本身度量单位的影响.例如：

$$Cov(kX, kY)=k^2Cov(X,Y)$$

为了克服这一缺点，对协方差进行标准化，这就引入了相关系数 .

二、相关系数的意义

相关系数刻划了X和Y间“线性相关”的程度.

考虑以X的线性函数$a+bX$来近似表示Y,

1. 问题的提出

问 a,b 应如何选择,可使 $aX+b$ 最接近 Y? 接近的程度又应如何来 衡量?

设 $e=E[(Y-(a+bX))^2]$

则 e 可用来衡量 $a+bX$ 近似表达 Y 的好坏程度.

当 e 的值越小,表示 $a+bX$ 与 Y 的近似程度越好.

确定 a,b 的值,使 e 达到最小.

$$e = E[(Y-(a+bX))^2]$$

$$= E(Y^2) + b^2E(X^2) + a^2 - 2bE(XY) + 2abE(X) - 2aE(Y).$$

将 e 分别关于 a,b 求偏导数,并令它们等于零,得

$$\begin{cases} \dfrac{\partial e}{\partial a} = 2a + 2bE(X) - 2E(Y) = 0, \\ \dfrac{\partial e}{\partial b} = 2bE(X^2) - 2E(XY) + 2aE(X) = 0. \end{cases}$$

解得 $b_0 = \dfrac{\mathrm{Cov}(X,Y)}{D(X)}, a_0 = E(Y) - E(X)\dfrac{\mathrm{Cov}(X,Y)}{D(X)}.$

将 a_0,b_0 代入 $e=E[(Y-(a+bX))^2]$中,得

$$\min_{a,b} e = E[(Y-(a+bX))^2]$$
$$= E[(Y-(a_0+b_0X))^2]$$

$$= DY - \frac{Cov^2(X,Y)}{DX} = \left[1 - \frac{Cov^2(X,Y)}{DXDY}\right] DY = (1-\rho_{XY}^2)D(Y).$$

二、相关系数

定义：设$D(X)>0$, $D(Y)>0$, 称

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

为随机变量 X 和 Y 的相关系数 .

在不致引起混淆时，记 ρ_{XY} 为 ρ .

若 $\rho=0$则称X与Y不相关

2. 相关系数的意义

当 $|\rho_{XY}|$ 较大时 e 较小, 表明 X,Y 的线性关系联系较紧密.

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.

3. 注意

(1) 不相关与相互独立的关系

相互独立 $\underset{\not\longleftarrow}{\longrightarrow}$ 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 $\Leftrightarrow E(XY) = E(X)E(Y).$

4. 相关系数的性质

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

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

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

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

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

(2) $|\rho_{XY}|=1$的充要条件是,存在常数 a,b 使

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

事实上,$|\rho_{XY}|=1 \Rightarrow E[(Y-(a_0+b_0X))^2]=0$

$$\Rightarrow 0=E[(Y-(a_0+b_0X))^2]$$

$$=D[Y-(a_0+b_0X)]+[E(Y-(a_0+b_0X))]^2$$

$$\Rightarrow D[Y-(a_0+b_0X)]=0,$$

$$E[Y-(a_0+b_0X)]=0.$$

由方差性质知

$$P\{Y-(a_0+b_0X)=0\}=1, \text{或 } P\{Y=a_0+b_0X\}=1.$$

反之,若存在常数 a^*,b^* 使

$$P\{Y=a^*+b^*X\}=1 \Leftrightarrow P\{Y-(a^*+b^*X)=0\}=1,$$

$$\Rightarrow P\{[Y-(a^*+b^*X)]^2=0\}=1,$$

$$\Rightarrow E\{[Y-(a^*+b^*X)]^2\}=0.$$

故有

$$0=E\{[Y-(a^*+b^*X)]^2\}\geq \min_{a,b} E[(Y-(a+bX))^2]$$

$$=E\{[Y-(a_0+b_0X)]^2\}=(1-\rho_{XY}^2)D(Y)$$

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

例2:设二维随机变量（X,Y）的概率密度为

$$f(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2\le 1\\ 0, & 其他\end{cases}$$

试验证 X 和 Y 不相关，但 X 和 Y 不是相互独立的。

解：

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy=\frac{1}{\pi}y\Big|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}=\frac{2}{\pi}\sqrt{1-x^2}\quad\left(|x|\le 1\right)$$

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}\frac{1}{\pi}dx=\frac{1}{\pi}x\Big|_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}=\frac{2}{\pi}\sqrt{1-y^2}\quad\left(|y|\le 1\right)$$

$$E(X)=\int_{-\infty}^{+\infty}xf_X(x)dx=\int_{-1}^{1}\frac{2x}{\pi}\sqrt{1-x^2}dx=0$$

$$E(Y)=\int_{-\infty}^{+\infty}yf_Y(y)dy=\int_{-1}^{1}\frac{2y}{\pi}\sqrt{1-y^2}dy=0$$

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

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

$$\rho_{XY}=0$$

例3:设二维随机变量（X,Y）的概率密度为

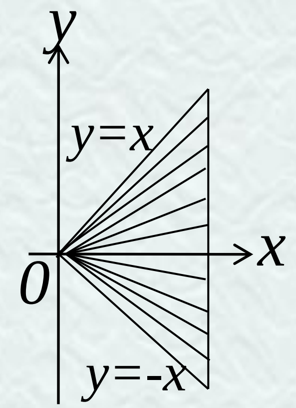
$$f(x,y)=\begin{cases}1\ , & |y|<x\ ,\ 0<x<1\\ 0\ , & \text{其他}\end{cases}$$

求 $Cov\ (X\ ,Y)$

解：由于 $Cov(X,Y)=E(XY)-E(X)E(Y)$

$$E(X)=\int_{-\infty}^{+\infty}xf_X(x)dx$$

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{-x}^{x}1dy=2x\left(0<x<1\right)$$



$$E(X)=\int_0^1 x2xdx=\int_0^1 2x^2dx=\frac{2}{3}x^3\Big|_0^1=\frac{2}{3}$$

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx \xlongequal{|y|<1} \begin{cases} \int_y^1 1dx = 1-y\,, & 0<y<1 \\ \int_{-y}^1 1dx = 1+y\,, & -1<y\le 0 \end{cases}$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy$$

$$=\int_0^1 xdx\int_{-x}^{x} ydy=\int_0^1 x\cdot\frac{1}{2}\left(x^2-(-x)^2\right)dx=0$$

$$Cov\ (XY)=0$$

例4　已知离散型随机向量 (X,Y) 的概率分布如右表，求 $Cov(X,Y)$

X \ Y	-1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

解　容易求得X 的概率分布为:

$$P\{X=0\}=0.3,$$
$$P\{X=1\}=0.45,$$
$$P\{X=2\}=0.25;$$

Y 的概率分布为:

$$P\{Y=-1\}=0.55,$$
$$P\{Y=0\}=0.25,$$
$$P\{Y=2\}=0.2,$$

$X \backslash Y$	-1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

$$E(X)=0\times 0.3+1\times 0.45+2\times 0.25=0.95,$$

$$E(Y)=(-1)\times 0.55+0\times 0.25+2\times 0.2=-0.15.$$

计算得

$$\begin{aligned}E(XY)&=0\times(-1)\times 0.1+0\times 0\times 0.2+0\times 2\times 0\\&\quad+1\times(-1)\times 0.3+1\times 0\times 0.5+1\times 2\times 0.1\\&\quad+2\times(-1)\times 0.15+2\times 0\times 0+2\times 2\times 0.1=0.\end{aligned}$$

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

$$=0.95\times 0.15=0.1425.$$

例5 设 (X,Y) 的分布律为

$Y \backslash X$	-2	-1	1	2	$P\{Y=y_i\}$
1	0	1/4	1/4	0	1/2
4	1/4	0	0	1/4	1/2
$P\{Y=x_i\}$	1/4	1/4	1/4	1/4	1

易知 $E(X)=0,\ E(Y)=5/2,\ E(XY)=0,$

于是 $\rho_{XY}=0$ X,Y 不相关. 这表示 X,Y 不存在线性关系，但

$$P\{X=-2,Y=1\}=0\neq P\{X=-2\}P\{Y=1\},$$

这表示 X,Y 不存在线性关系，但

$$P\{X=-2,Y=1\}=0\neq P\{X=-2\}P\{Y=1\},$$

知 X,Y 不是相互独立的.

事实上, X 和 Y 具有关系: $Y=X^2$, Y 的值完全可由 X 的值所确定.

例6:已知三个随机随机变量X，Y，Z中

$$E(X)=E(Y)=1, E(Z)=-1, D(X)=D(Y)=D(Z)=1$$

$$\rho_{XY}=0\ ,\ \rho_{XZ}=\frac{1}{2}, \rho_{YZ}=-\frac{1}{2}$$

求 $E(X+Y+Z), D(X+Y+Z)$

解： $E(X+Y+Z)=E(X)+E(Y)+E(Z)=1+1-1=1$

$$D(X+Y+Z)=D(X+Y)+D(Z)+2Cov(X+Y,Z)$$
$$=D(X)+D(Y)+2Cov(X,Y)+D(Z)+2Cov(X,Z)+2Cov(Y,Z)$$

$$Cov(X,Y)=\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}=0$$

$$Cov(X,Z)=\rho_{XZ}\sqrt{D(X)}\sqrt{D(Z)}=\frac{1}{2}$$

$$Cov(Y,Z)=\rho_{YZ}\sqrt{D(Y)}\sqrt{D(Z)}=-\frac{1}{2}$$

$$D(X+Y+Z)=3$$

例7 $X \sim N(0,1)$ ， $Y = X^2$ 求 ρ_{XY}

解:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}}$$

$$Cov(X,Y) = EXY - EXEY$$

$$EX = 0$$

$$EXY = E(X \cdot X^2) = E(X^3)$$

$$= \int_{-\infty}^{\infty} X^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}} = 0$$

例8　将一枚硬币投掷 n 次，以 X 和 Y 分别表示正面向上和反面向上的次数，则X 和 Y 的相关系数等于（　　）.

A.　-1　　B.　0　　C.　½　　D.　1

$$n = X + Y$$

$$Y = -X + n$$

$$P\{Y = -X + n\} = 1 \Leftrightarrow |\rho_{XY}| = 1$$

$$\because -X$$

$$\therefore \rho_{XY} = -1$$

例9 若随机变量 $X \sim N(0,1)$ ，$Y \sim N(1,4)$

且 $\rho_{XY} = 1$ **，则（ D ）.**

A. $\rho\{Y = -2X - 1\} = 1$

B. $\rho\{Y = 2X - 1\} = 1$

C. $\rho\{Y = -2X + 1\} = 1$

D. $\rho\{Y = 2X + 1\} = 1$

三、小结

相关系数的意义

当 $|\rho_{XY}|$ 较大时，X, Y 的线性相关程度较高.

当 $|\rho_{XY}|$ 较小时，X, Y 的线性相关程度较差.

当 $\rho_{XY} = 0$ 时，X 和 Y 不相关.