概率论与数理统计

第四章　随机变量的数字特征

习题： 在每次试验中，事件A发生的概率为 0.75, 利用切比雪夫不等式求：n需要多么大时，才能使得在n次独立重复试验中, 事件A**出现的频率**在0.74~0.76之间的概率至少为0.90?

解：设X为n 次试验中，事件A出现的次数，

则 $X \sim B(n, 0.75)$

$E(X)=0.75n$, $D(X)=0.75\times 0.25n=0.1875n$

所求为满足 $P(0.74 < \dfrac{X}{n} < 0.76) \geq 0.90$ 的最小的n .

$P(0.74<\frac{X}{n}<0.76)$ 可改写为

$$P(0.74n< X<0.76n\ )$$

$$=P(-0.01n< X-0.75n< 0.01n)$$

$$= P\{\ |X-E(X)|\ <0.01n\}$$

在切比雪夫不等式中取 $\varepsilon=0.01n$，则

$$P(0.74<\frac{X}{n}<0.76) = P\{\ |X-E(X)|\ <0.01n\}$$

$$\geq 1-\frac{D(X)}{(0.01n)^2}=1-\frac{0.1875n}{0.0001n^2}=1-\frac{1875}{n}$$

依题意，取 $1-\frac{1875}{n} \geq 0.9$

解得 $n \geq \frac{1875}{1-0.9} = 18750$

即n 取18750时，可以使得在n次独立重复试验中，事件A出现的频率在0.74~0.76之间的概率至少为0.90 .

当求离散型随机变量 X 的数学期望时，有时直接用定义来求会非常困难和非常复杂.这个时候可以利用数学期望的性质

$$E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)$$

把 X 分解为几个随机变量的和，而这几个随机变量的数学期望很容易求.**一般当X表示的是与计数有关的随机变量时，大部分情形我们可以把它分解成0-1分布的随机变量的和**.

课后第14题　将 n 只球($1\sim n$ 号)随机地放进 n 只盒子($1\sim n$ 号)中去，一只盒子装一只球.若一只球装入与球同号的盒子中，称为一个配对. 记 X 为总的配对个数，求 X 的期望与方差.

解 $X_i = \begin{cases} 1, & i \text{ 号球放入 } i \text{ 号盒} \\ 0, & \text{其它} \end{cases} \quad i = 1, 2, \cdots, n$

则 $X = \sum_{i=1}^{n} X_i$

由于一共有n个盒子，所以第i个球放在第i个盒子的概率为

$$p(x_i) = \frac{1}{n}$$

但 $X_1, X_2, \cdots, X_n$ 不相互独立.

X_i	1	0
P	$\frac{1}{n}$	$1-\frac{1}{n}$

$i = 1, 2, \cdots, n$

$$E(X)=\sum_{i=1}^{n}E(X_i)=n\cdot\frac{1}{n}=1$$

$$E(X^2)=E\left(\sum_{i=1}^{n}X_i\right)^2=E\left(\sum_{i=1}^{n}X_i^2+2\sum_{1\le i<j\le n}^{n}X_iX_j\right)$$ 完全平方公式

$$=\sum_{i=1}^{n}E(X_i^2)+2\sum_{1\le i<j\le n}^{n}E(X_iX_j)$$

X_i^2	1	0
P	$\frac{1}{n}$	$1-\frac{1}{n}$

$i=1,2,\cdots,n$ $\qquad E(X_i^2)=\frac{1}{n}$

X_iX_j	1	0
P	$\frac{1}{n(n-1)}$	$1-\frac{1}{n(n-1)}$

$i,j=1,2,\cdots,n$ $\qquad E(X_iX_j)=\frac{1}{n(n-1)}$

$$E(X^2)=\sum_{i=1}^{n}E(X_i^2)+2\sum_{1\le i<j\le n}^{n}E(X_iX_j)$$

$$=\sum_{i=1}^{n}\frac{1}{n}+2\sum_{1\le i<j\le n}^{n}\frac{1}{n(n-1)}$$

$$=n\cdot\frac{1}{n}+2\cdot C_n^2\cdot\frac{1}{n(n-1)}$$

$$=2$$

$$D(X)=E(X^2)-E^2(X)=1$$

§3 协方差及相关系数

- 协方差
- 相关系数及其意义

前面我们介绍了随机变量的数学期望和方差，对于二维随机变量（X，Y），我们除了讨论X与Y的数学期望和方差以外，还要讨论描述X和Y之间关系的数字特征，这就是本讲要讨论的

协方差及相关系数

一、协方差

1.定义 量$E\{[X-E(X)][Y-\mathrm{E}(Y)]\}$称为随机变量$X$和$Y$的协方差,记为$Cov(X,Y)$，即

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

2.简单性质

（1） $Cov(X,Y)=\mathrm{Cov}(Y,X)$

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

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

3. 计算协方差的一个简单公式

由协方差的定义及期望的性质，可得

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

$$=E(XY)-E(X)E(Y)-E(Y)E(\mathrm{X})+E(X)E(Y)$$

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

即

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

可见，若 X 与 Y 独立，$Cov(X,Y)=0$.

特别地

$$Cov(X,X)=E(X^2)-[E(X)]^2=D(X)$$

4. 随机变量和的方差与协方差的关系

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

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、相关系数及其意义

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.

1. 问题的提出

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

设 $\quad e=E[(Y-(a+bX))^2]$

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

$$\begin{aligned}\min_{a,b} e &= E[(Y-(a+bX))^2] \\ &= E[(Y-(a_0+b_0X))^2] \\ &= (1-\rho_{XY}^2)D(Y).\end{aligned}$$

$$\begin{aligned}E[(Y-(a_0+b_0X))^2] &= D[Y-a_0-b_0X]+[E(Y-a_0-b_0X)]^2 \\ &= D(Y-b_0X)+\left[-\frac{1}{2}\frac{\partial e}{\partial a}\Big|_{\substack{a=a_0 \\ b=b_0}}\right]^2 \\ &= D(Y-b_0X)+0 \\ &= D(Y)+b_0^2D(X)-2b_0Cov(X,Y)\end{aligned}$$

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

$$= D(Y)\left[1 - \frac{Cov^2(X,Y)}{D(X)D(Y)}\right]$$

$$= \left[1 - \rho_{XY}^2\right] D(Y)$$

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**.

思考　设 θ 服从 $[0,2\pi]$ 的均匀分布, $\xi=\cos\theta$, $\eta=\cos(\theta+a)$, 这里 a 是常数, 求 ξ 和 η 的相关系数?

解　$$E(\xi)=\frac{1}{2\pi}\int_0^{2\pi}\cos x\,\mathrm{d}x=0,$$

$$E(\eta)=\frac{1}{2\pi}\int_0^{2\pi}\cos(x+a)\,\mathrm{d}x=0,$$

$$E(\xi^2)=\frac{1}{2\pi}\int_0^{2\pi}\cos^2 x\,\mathrm{d}x=\frac{1}{2},$$

$$E(\eta^2)=\frac{1}{2\pi}\int_0^{2\pi}\cos^2(x+a)\,\mathrm{d}x=\frac{1}{2},$$

$$E(\xi\eta)=\frac{1}{2\pi}\int_0^{2\pi}\cos x\cdot\cos(x+a)\mathrm{d}x=\frac{1}{2}\cos a,$$

由以上数据可得相关系数为 $\rho=\cos a.$

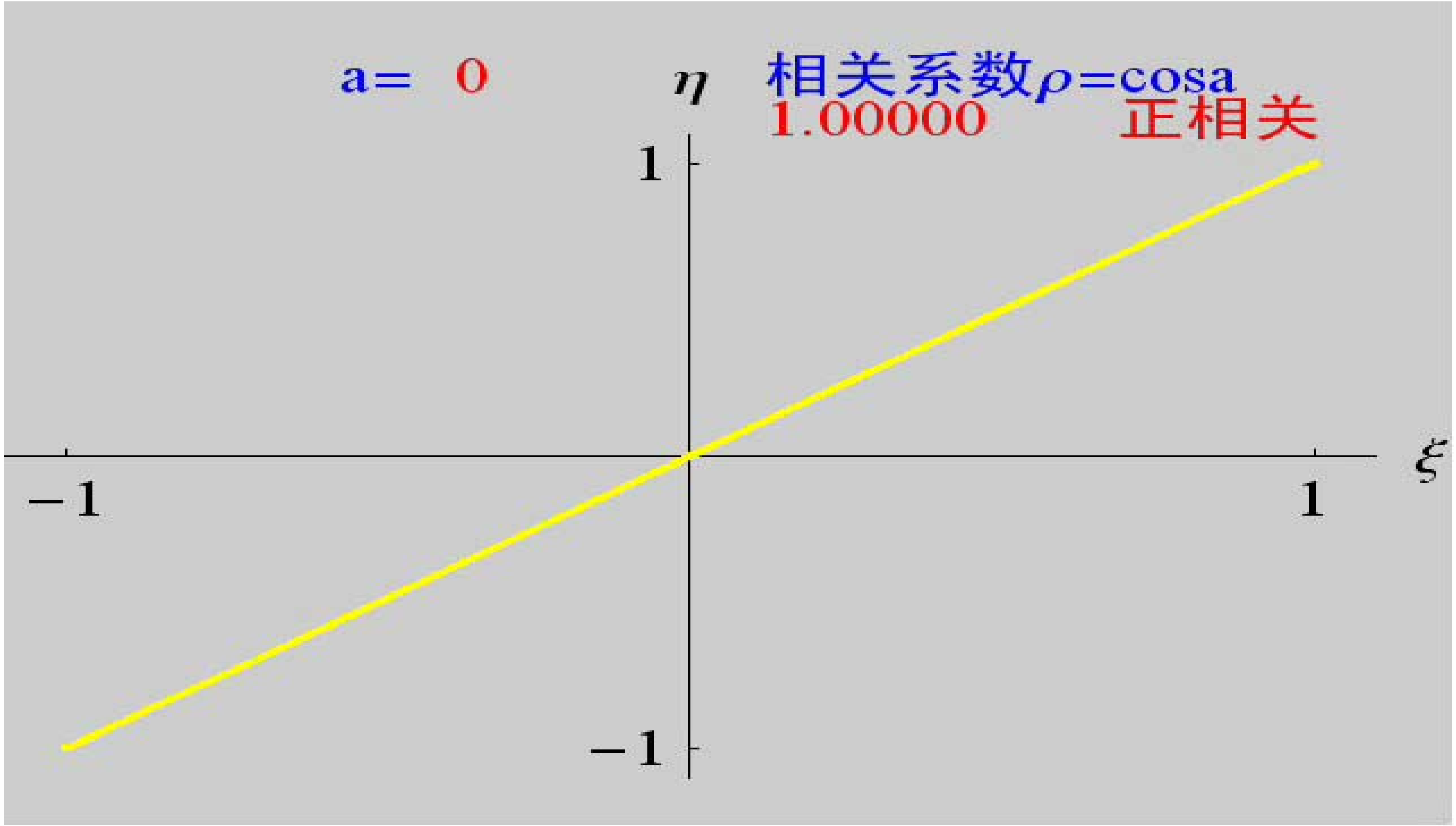
$$\left.\begin{array}{l}\text{当}a=0\text{时},\rho=1,\xi=\eta,\\ \text{当}a=\pi\text{时},\rho=-1,\xi=-\eta,\end{array}\right\}\text{存在线性关系.}$$

当 $a=\frac{\pi}{2}$ 或 $a=\frac{3\pi}{2}$ 时，$\rho=0,$ ξ 与 η 不相关.

但 $\xi^2+\eta^2=1,$ 因此 ξ 与 η 不独立.

动画演示 ξ 与 η 的相关关系.

单击图形播放/暂停 ESC键退出



3. 注意

(1) 不相关与相互独立的关系

$$\text{相互独立} \;\begin{array}{c}\Longrightarrow \\ \not\Longleftarrow\end{array}\; \text{不相关}$$

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

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

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

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

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

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

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

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

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

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

两项分别等于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

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

$$= E\{[Y - (a_0 + b_0 X)]^2\} = (1 - \rho_{XY}^2) D(Y)$$

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

ρ_{XY} 的含义：

ρ_{XY} 是一个用来表征 X,Y 之间线性关系紧密程度的量.

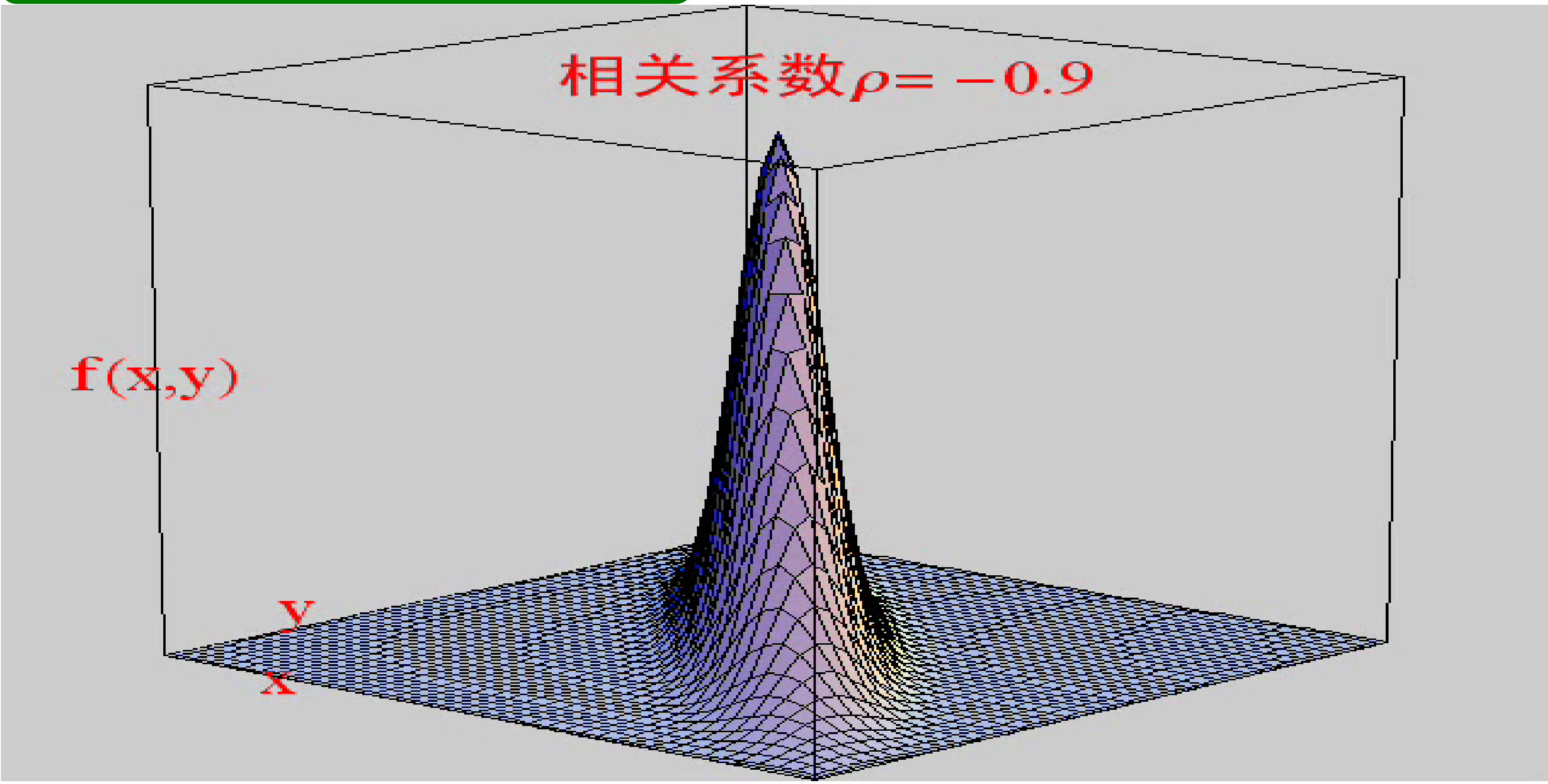
当 $|\rho_{XY}|$ 较大时，X,Y 线性相关的程度较好；

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.

二维正态随机变量 (X,Y) 的概率密度曲面与相关系数 $\rho_{XY} = \rho$ 的关系.

单击图形播放/暂停　ESC键退出



例1 设（X，Y）的分布律为

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

易知 $E(X)=0, E(Y)=5/2, E(XY)=0,$

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

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

所以 X,Y 不相互独立.

事实上，$Y=X^2$, Y 的值完全可由 X 的值所确定.

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

例3 已知随机变量X,Y分别服从$N(1,3^2),N(0,4^2)$, $\rho_{XY}=-1/2$,设 $Z=X/3+Y/2$.

(1) 求 Z 的数学期望和方差.

(2) 求 X 与 Z 的相关系数.

(3) 问 X 与 Z 是否相互独立?为什么?

解 (1)由$E(X)=1,\ D(X)=9,\ E(Y)=0,\ D(Y)=16.$

得
$$E(Z)=E\left(\frac{X}{3}+\frac{Y}{2}\right)=\frac{1}{3}E(X)+\frac{1}{2}E(Y)$$
$$=\frac{1}{3}.$$

$$D(Z) = D(\frac{X}{3}) + D(\frac{Y}{2}) + 2\text{Cov}(\frac{X}{3}, \frac{Y}{2})$$

$$= \frac{1}{9}D(X) + \frac{1}{4}D(Y) + \frac{1}{3}\text{Cov}(X, Y)$$

$$= \frac{1}{9}D(X) + \frac{1}{4}D(Y) + \frac{1}{3}\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}$$

$$= 1 + 4 - 2 = 3.$$

(2) $\mathrm{Cov}(X,Z)=\mathrm{Cov}(X,\frac{X}{3}+\frac{Y}{2})$

$$=\frac{1}{3}\mathrm{Cov}(X,X)+\frac{1}{2}\mathrm{Cov}(X,Y)$$

$$=\frac{1}{3}D(X)+\frac{1}{2}\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}=3-3=0.$$

故 $\rho_{XY}=\mathrm{Cov}(X,Z)/(\sqrt{D(X)}\sqrt{D(Z)})=0.$

(3)由二维正态随机变量相关系数为零和相互独立两者是等价的结论,可知:X与Z是相互独立的.

练习

设$X \sim N(\mu,\sigma^2), Y \sim N(\mu,\sigma^2)$，且设$X$，$Y$相互独立

试求（1）$Z_1 = \alpha X + \beta Y$和$Z_2 = \alpha X - \beta Y$的相关系数(其中α，β是不全为零的常数）.

解 $$D(X)=D(Y)=\sigma^2$$

$$D(Z_1)=D(\alpha X+\beta Y)=\alpha^2 D(X)+\beta^2 D(Y)=(\alpha^2+\beta^2)\sigma^2$$

$$D(Z_2)=D(\alpha X-\beta Y)=\alpha^2 D(X)+\beta^2 D(Y)=(\alpha^2+\beta^2)\sigma^2$$

$$Cov(Z_1,Z_2)=Cov(\alpha X+\beta Y,\alpha X-\beta Y)$$

$$=\alpha^2 Cov(X,X)-\beta^2 Cov(Y,Y)=\boldsymbol{\alpha^2 D(X)-\beta^2 D(Y)}$$

$$\boldsymbol{=(\alpha^2-\beta^2)\sigma^2}$$

$$\rho_{Z_1Z_2}=\frac{Cov(Z_1,Z_2)}{\sqrt{D(Z_1)}\sqrt{D(Z_2)}}=\frac{\alpha^2-\beta^2}{\alpha^2+\beta^2}$$

作业：课后习题28 、31、32、37