概率论与数理统计

第三节　协方差及相关系数

一、协方差概念及性质

二、相关系数

三、相关性和独立性

一、协方差的概念及性质

1. 问题的提出

若随机变量 X 和 Y 相互独立,那么

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

若随机变量 X 和 Y 不相互独立

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

$$D(X \pm Y) = E(X \pm Y)^2 - [E(X \pm Y)]^2$$

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

协方差

2. 定义

设(X,Y)是二维随机变量, 若$E\{[X-E(X)][Y-E(Y)]\}$存在，则称其为随机变量X与Y的协方差. 记为$\mathrm{Cov}(X,Y)$,

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

3. 协方差的计算公式

$$(1)\ \mathrm{Cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\}.$$

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

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

4. 性质

(1) $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$; $\mathrm{Cov}(X,a)=0$;

(2) $\mathrm{Cov}(aX+b,cY+d)=ac\,\mathrm{Cov}(X,Y)$;

a,b,c,d 为常数;

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

二、相关系数

1. 定义

设(X,Y)是二维随机变量,若X与Y的协方差存在,且$DX>0, DY>0$,

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

称为随机变量X与Y的相关系数.

说明 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.

第十三次课结束

2. 相关系数的性质

(1) $|\rho_{XY}| \le 1$; 或 $\mathrm{Cov}^2(X,Y) \le DX \cdot DY$

证：由方差的性质和协方差的定义知，

对任意实数 b, 有

$$0 \le D(Y-tX) = t^2D(X)+D(Y)-2t\,Cov(X,Y)$$

$$= t^2D(X)-2\,Cov(X,Y)\,t + D(Y)$$

当 $t=\dfrac{Cov(X,Y)}{D(X)} \triangleq b$ 时有最小值，

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

2. 相关系数的性质

(2) $|\rho_{XY}| = 1 \Leftrightarrow P\{Y = a + bX\} = 1, a, b(b \neq 0)$ 为常数，并且

$$\rho = \begin{cases} 1 & \text{当} b > 0 \\ -1 & \text{当} b < 0 \end{cases}.$$

例 $X \sim N(0,1), Y \sim N(1,4)$,且相关系数$\rho=1$,则

A $P(Y = -2X - 1) = 1$; B $P(Y = 2X - 1) = 1$

C $P(Y = -2X + 1) = 1$; D $P(Y = 2X + 1) = 1$

3. 相关系数的意义

当 $|\rho_{XY}|$ 较大时，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.

若 $\rho=\pm 1$，Y与X有严格线性关系;

若$\rho=0$，Y 与 X 无线性关系;

若$0<|\rho|<1$

$|\rho|$ 的值越接近于1，Y 与 X 的线性相关程度越高;

$|\rho|$ 的值越接近于0，Y与X的线性相关程度越弱.

三、相关性

1. 定义 当 $\rho_{XY}=0$ 时,称 X 和 Y 不相关;

当 $|\rho_{XY}|=1$ 时,称 X 和 Y 完全相关.

2. 不相关性的判定

以下四个条件等价

(1) $\rho = 0;$

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

(3) $D(X\pm Y)=DX+DY;$

(4) $EXY = EX\cdot EY$

3 .不相关与相互独立的关系

相互独立 $\substack{\Longrightarrow \\ \not\Longleftarrow}$ 不相关

例1 设(X,Y)在圆域$x^2+y^2\le 1$上服从均匀分布，

（1）问X与Y是否独立？ （2）求相关系数ρ

例2 $X\sim N(0,1), Y=X^2$，证明X与Y不相关且不独立

解：$E(XY)=E(XX^2)=\int_{-\infty}^{+\infty}x^3\varphi(x)\mathrm{d}x=0$

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

$$F(1,1)=P\{X\le 1, Y\le 1\}=P\{X\le 1, X^2\le 1\}=P\{X^2\le 1\}$$

$$F_X(1)F_Y(1)=P\{X\le 1\}P\{Y\le 1\}=P\{X\le 1\}P\{X^2\le 1\}$$

$$F(1,1)\ne F_X(1)F_Y(1)$$ 故X与Y不独立

练习设X服从(-1/2, 1/2)内的均匀分布，而Y=cos X,

证明X与Y不相关且不独立

例3 随机变量X的概率密度函数为$f(x)=\frac{1}{2}e^{-|x|}$,

(1) 求EX和DX;

(2) X与$|X|$的协方差，问X与$|X|$是否相关;

(3) X与$|X|$是否独立，为什么?

解:(1)$E(X)=\int_{-\infty}^{+\infty} x\frac{1}{2}e^{-|x|}\mathrm{d}x=0$

$$DX=EX^2=\int_{-\infty}^{+\infty} x^2\frac{1}{2}e^{-|x|}\mathrm{d}x=\int_{0}^{+\infty} x^2e^{-x}\mathrm{d}x=2$$

$$(2)E(XY)=E(X|X|)=\int_{-\infty}^{+\infty} x|x|\frac{1}{2}e^{-|x|}\mathrm{d}x=0$$

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

$$\forall \alpha > 0$$

$$F(\alpha,\alpha) = P\{X \le \alpha, Y \le \alpha\} = P\{X \le \alpha, |X| \le \alpha\} = P\{|X| \le \alpha\}$$

$$F_X(\alpha)F_Y(\alpha) = P\{X \le \alpha\}P\{Y \le \alpha\} = P\{X \le \alpha\}P\{|X| \le \alpha\}$$

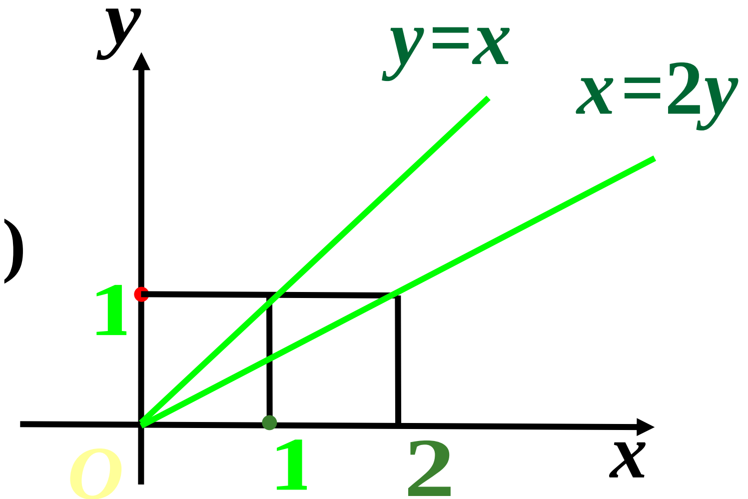
$$F(\alpha,\alpha) \ne F_X(\alpha)F_Y(\alpha)$$ 故X与Y不独立

例4 设(X,Y)在矩形区域$G=\{(x,y)\mid 0\le x\le 2, 0\le y\le 1\}$上服从均匀分布

$$\text{记}U=\begin{cases}0, & X\le Y\\ 1, & X>Y\end{cases}, V=\begin{cases}0, & X\le 2Y\\ 1, & X>2Y\end{cases}$$

(1) 求U和V的联合分布;

(2) 求U和V的相关系数.



解$P(U=0,V=0)=P(X\le Y, X\le 2Y)$

$$=P(X\le Y)$$

$$=\frac{1}{4}$$

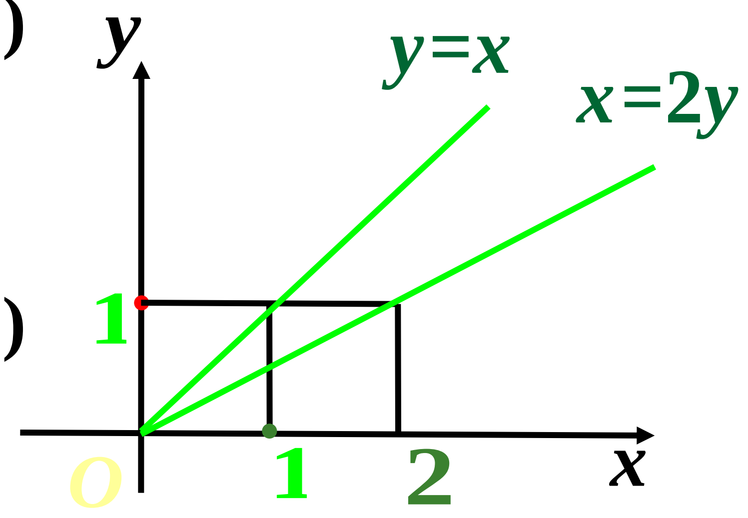
$$P(U=0,V=1)=P(X\le Y, X>2Y)$$

$$=0$$

$$P(U=1,V=0)=P(X>Y, X\le 2Y)$$

$$=P(Y<X\le 2Y)$$

$$=\frac{1}{4}$$



$$P(U=1,V=1)=P(X>Y, X>2Y)=\frac{1}{2}$$

$U \backslash V$	0	1	$P_{i\cdot}$
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
$P_{\cdot j}$	$\frac{1}{2}$	$\frac{1}{2}$	

$$E(UV)=\frac{1}{2}$$

$$E(U)=\frac{3}{4}$$

$$E(V)=\frac{1}{2}$$

$$D(U)=\frac{3}{16}$$

$$D(V)=\frac{1}{4}$$

$$\rho=\frac{1}{\sqrt{3}}$$

例5 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 $$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

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

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

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

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

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

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

$$u=\frac{x-\mu_1}{\sigma_1},$$

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

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

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

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

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

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

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;

若$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2,\rho)$

试求二维正态随机变量的边缘概率密度.

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$$

$$p_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},\quad -\infty<x<+\infty.\quad X\sim N(\mu_1,\sigma_1^2)$$

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

$$p_X(x)p_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$$

$$=\frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$$

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$$

$$p_X(x)p_Y(y)=p(x,y)\Leftrightarrow\rho=0$$

X与Y独立的 $\Leftrightarrow \rho=0$

(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6 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.

解

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

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

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

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

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

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

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

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

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

$$=3-3=0.$$

故 $\rho_{XY}=\operatorname{Cov}(X,Z)\big/\left(\sqrt{D(X)}\sqrt{D(Z)}\right)=0.$

例6 若X与Y满足$D(X+Y)=D(X-Y)$则

A X与Y独立； B X与Y不相关；

C $DY=0$； D $DX\cdot DY=0$

作业题

1． 设 X 与 Y 独立，且 $X\sim P(2), Y\sim N(-3,1)$, 设 $Z=X-2Y-9$, 则 $EZ=$______；

$DZ=$______.

2． 设随机变量X,Y的相关系数$\rho_{XY}=0.7$, 若$Z=X+0.8$, 则$\rho_{YZ}=$______.

2． 设二维随机变量$(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$，则下列结论中错误的是（ ）.

（A）$X\sim N(\mu_1,\sigma_1^2); Y\sim N(\mu_2,\sigma_2^2)$； （B）$D(X+Y)=\sigma_1^2+\sigma_2^2$；

（C）$E(X+Y)=\mu_1+\mu_2$； （D）X与Y独立的充要条件为$\rho=0$.

3. 若随机变量 X 与 Y 满足 $D(X+Y)=D(X-Y)$，则必有（ ）.

（A）X 与 Y 不相关；（B）X 与 Y 不独立；

（C）$DX=DY$；（D）X 与 Y 独立.

设 $X_1 \sim P(2), X_2 \sim U(-2,4), X_3 \sim B(15,0.2)$, 设 $Y=X_1-3X_2+X_3$, 则 $EY=$______.

第四节　矩、协方差矩阵

1.定义

设 X 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.

若 $E\{[X-E(X)]^k\}$, $k=2,3,\cdots$ 存在,称它为 X 的 k 阶中心矩.