概率论与数理统计

第15讲

协方差及相关系数

一. 协方差

数学期望和方差是两个重要的数字特征，分别表示单个随机变量的平均值和离散程度。

对于多维随机变量，人们自然可以用各个分量的数学期望和方差来表示他们的中心位置和波动性的大小，但是人们除了考虑各个分量自身的特性外，还常常对各个量之间的关系感兴趣。

本节讨论反映两个变量之间关系的数字特征：协方差和相关系数。

一. 协方差

若DX、DY存在，则有

$$D(X\pm Y)=DX+DY\pm 2E[(X-EX)(Y-EY)]$$

且当X和Y独立时，有$D(X\pm Y)=DX+DY$

即：若X和Y独立，$\boxed{\text{则}E[(X-EX)(Y-EY)]=0}$

从而有结论：若$E[(X-EX)(Y-EY)]\neq 0$，则X和Y不独立。

这说明：$E[(X-EX)(Y-EY)]$表达了X与Y之间的某种关系。

协方差

一. 协方差

定义1 设(X,Y)为二维随机变量，它的分量的数学期望$E(X)$和$E(Y)$存在，若$E[(X-EX)(Y-EY)]$存在，则称它为X与Y的协方差，记为$Cov(X,Y)$. 即

$$Cov(X,Y)=E[(X-EX)(Y-EY)]$$

“协”即“协同”的意思，单个随机变量的方差的定义为$DX=E(X-EX)^2$，可以看作$(X-EX)$与$(X-EX)$的乘积的期望，而在协方差的定义中把其中一个$(X-EX)$换成了$(Y-EY)$，其形式与方差的定义形式很接近，而又有X和Y的参与，所以得到了协方差的名称。

协方差的计算

（1）若二维离散型随机变量(X, Y)的联合分布律为

$$P\{X=x_i,Y=y_j\}=p_{ij} \quad i, j =1,2,\dots$$

且$Cov(X,Y)$存在，则

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

（2）若二维连续型随机变量(X, Y)的密度函数为$f(x, y)$

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

协方差的计算

（3）$Cov(X,Y)=E(XY)-E(X)E(Y)$

证明：$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$

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

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

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

可见，若X与Y独立，$Cov(X,Y)=0$。

一. 协方差

例1.设：随机变量X和Y的联合概率分布为

X \ Y	-1	0	1
0	0.06	0.18	0.16
1	0.08	0.32	0.20

求X和Y的协方差。

解：$Cov(X,Y)=E(XY)-E(X)E(Y)$

$$E(XY)=0\times(-1)\times0.06+0\times0\times0.18+0\times1\times0.16$$

$$+1\times(-1)\times0.08+1\times0\times0.32+1\times1\times0.20=0.12$$

另外，X和Y的边缘分布律分别为

X	0	1
P	0.4	0.6

Y	-1	0	1
P	0.14	0.5	0.36

所以 $EX = 0\times0.4+1\times0.6=0.6$

$EY = -1\times0.14+0\times0.5+1\times0.36=0.22$

$Cov(X,Y)=E(XY)-E(X)E(Y)=0.12-0.6\times0.22 = -0.012$

例2.设:(X,Y) 在圆域$D=\{(x,y): x^2+y^2\le r^2\ (r>0)\}$上服从均匀分布，求$Cov(X,Y)$。

解：易知(X,Y)的联合概率密度为 $f(x,y)=\begin{cases}\dfrac{1}{\pi r^2}, & x^2+y^2\le r^2\\ 0, & x^2+y^2>r^2\end{cases}$

所以 $$EX=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy=\iint_{x^2+y^2\le r^2}x\cdot\frac{1}{\pi r^2}dxdy=0$$

$$EY=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy=\iint_{x^2+y^2\le r^2}y\cdot\frac{1}{\pi r^2}dxdy=0$$

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

所以 $Cov(X,Y)=E(XY)-E(X)E(Y)=0$　　此题表明，$Cov(X,Y)$等于0, 但X与Y不独立。

一. 协方差

定理1若X和Y为随机变量，且他们的方差$D(X)$，$D(Y)$ 存在，则$X+Y$的方差存在，且 $D(X+Y)=DX+DY+2Cov(X,Y)$

推论1 若$X_1,X_2,\ldots,X_n$为n（$n\geq 2$）个方差存在随机变量，则$\sum_{i=1}^{n}X_i$的方差存在，且

$$D(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}D(X_i)+2\sum_{i<j}\sum Cov(X_i,X_j)$$

例3.（配对问题）。将n只球(编号为1~n号)随机放进n个盒子(编号为1~n号)中去，一个盒子只装一只球。若一只球装入与球同号的盒子中，称为一个配对。记X为总配对数，求EX，DX。

解：设 $X_i=\begin{cases}1, & \text{第 } i \text{ 号球放入第} i \text{ 号盒子} \\ 0, & \text{第 } i \text{ 号球未放入第} i \text{ 号盒子}\end{cases}, i=1,2,\cdots,n$

则配对总数 $X=X_1+X_2+\cdots+X_n$

且对任意i有 $P\{X_i=1\}=\dfrac{1\times(n-1)!}{n!}=\dfrac{1}{n}$ $P\{X_i=0\}=1-\dfrac{1}{n}$

所以由0—1分布的性质得 $EX_i=\dfrac{1}{n}, DX_i=\dfrac{1}{n}(1-\dfrac{1}{n})=\dfrac{n-1}{n^2}$

一. 协方差

于是有 $EX=E(X_1+X_2+\cdots+X_n)=EX_1+EX_2+\cdots+EX_n=n\cdot\dfrac{1}{n}=1$

另外，当$i\neq j$时，$P\{X_iX_j=1\}=P\{X_i=1,\ X_j=1\}=\dfrac{1\times(n-2)!}{n!}=\dfrac{1}{n(n-1)}$

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

所以 $E(X_iX_j)=1\times P\{X_iX_j=1\}=\dfrac{1}{n(n-1)}$

从而可得

$$Cov(X_i,X_j)=E(X_iX_j)-EX_iEX_j=\frac{1}{n(n-1)}-\frac{1}{n}\times\frac{1}{n}=\frac{1}{n^2(n-1)}$$

所以 $DX=D(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}DX_i+2\sum_{1\le i<j\le n}Cov(X_i,X_j)$

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

一. 协方差

协方差具有以下简单的性质

性质1 设随机变量X的数学期望EX存在，a为常数，则$Cov(X, a)=0$。

性质2 设随机变量X和Y的期望EX和EY都存在，则$Cov(X, Y)=Cov(Y, X)$。

性质3 设随机变量X的期望和方差都存在，则$Cov(X, X)=D(X)$。

性质4 设随机变量X和Y的期望EX和EY都存在，a和b为常数，则

$$Cov(aX, bY) = abCov(X, Y)$$

性质5 设随机变量X_1, X_2和Y的期望均存在，则

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

一. 协方差

性质6. 设随机变量X和Y的期望和方差都存在，且$DX>0$，$DY>0$，则

$$[Cov(X,Y)]^2 \le D(X)D(Y)$$

其中等号成立当且仅当X与Y有严格的线性关系（即存在常数a和b，使得$P\{Y=aX+b\}=1$成立）

证明：对任意的实数t，有 $D(tX+Y)=t^2D(X)+D(Y)+2tCov(X,Y)$

令 $g(t)=t^2D(X)+2Cov(X,Y)t+D(Y)$ 为关于t的二次三项式。

由方差的性质知$g(t)\ge 0$，故判别式小于或等于零，即有

$\Delta=[2Cov(X,Y)]^2-4D(X)D(Y)\le 0$ $\quad\therefore [Cov(X,Y)]^2 \le D(X)D(Y)$

一. 协方差

若 $[Cov(X,Y)]^2 = D(X)D(Y)$

此时 $t^2D(X)+2Cov(X,Y)t+D(Y)=0$ 有唯一的根，记为t_0，即$g(t_0)=0$ 。

所以有 $D(t_0X+Y)=0$

所以有 $P\{t_0X+Y=E(t_0X+Y)\}=1$ 即 $P\{Y=-t_0X+E(t_0X+Y)\}=1$

取 $a=-t_0, b=E(t_0X+Y)$ 有 $P\{Y=aX+b\}=1$成立。

一. 协方差

$$g(t) = D(tX + Y) = t^2 D(X) + D(Y) + 2tCov(X,Y)$$

反之，若存在常数a和b，使得$P\{Y=aX+b\}=1$ 成立

则有 $P\{-aX+Y=b\}=1$

由方差的性质知 $D(-aX+Y)=0$ 即 $g(-a)=0$

所以有 $\Delta = [2Cov(X,Y)]^2 - 4D(X)D(Y) = 0$

即 $[Cov(X,Y)]^2 = D(X)D(Y)$

例4 设随机变量$X_1, X_2,\dots, X_n(n>1)$独立同分布，且它们共同的方差为$\sigma^2>0$，令 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i, Y_1=X_1-\bar{X}, Y_n=X_n-\bar{X}$. 求$D(Y_1)$以及$Cov(Y_1, Y_n)$。

解：由方差的性质可得：

$$D(\bar{X})=D(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}DX_i=\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=\frac{\sigma^2}{n}$$

又易知 $Cov(X_1,\bar{X})=Cov(X_1,\frac{1}{n}\sum_{i=1}^{n}X_i)=Cov(X_1,\frac{1}{n}X_1+\frac{1}{n}\sum_{i=2}^{n}X_i)$

一. 协方差

$$Cov(X_1,\bar{X}) = Cov(X_1, \frac{1}{n}X_1 + \frac{1}{n}\sum_{i=2}^{n} X_i) = Cov(X_1, \frac{1}{n}X_1) + Cov(X_1, \frac{1}{n}\sum_{i=2}^{n} X_i)$$

$$= \frac{1}{n}Cov(X_1, X_1) + \frac{1}{n}\sum_{i=2}^{n} Cov(X_1, X_i) = \frac{1}{n}D(X_1) + \frac{1}{n}\sum_{i=2}^{n} 0 = \frac{\sigma^2}{n}$$

同理有 $Cov(X_n,\bar{X}) = \dfrac{\sigma^2}{n}$

所以

$$D(Y_1) = D(X_1 - \bar{X}) = D(X_1) + D(\bar{X}) - 2Cov(X_1,\bar{X}) = \sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

一. 协方差

由协方差的性质可得

$$Cov(Y_1,Y_n)=Cov(X_1-\bar{X},X_n-\bar{X})=Cov(X_1,X_n)-Cov(X_1,\bar{X})-Cov(X_n,\bar{X})+Cov(\bar{X},\bar{X})$$

$$=0-Cov(X_1,\bar{X})-Cov(X_n,\bar{X})+D(\bar{X})=0-\frac{\sigma^2}{n}-\frac{\sigma^2}{n}+\frac{\sigma^2}{n}=-\frac{\sigma^2}{n}$$

一. 协方差

协方差衡量了X和Y之间协同变化的关系。但它的大小还受X与Y本身量纲的影响。

例如：变量X和Y之间的关系为：$Y/X=a$，度量单位为米

当度量单位变为厘米时，得到变量$100X$和$100Y$，

其关系依然为：$(100Y)/(100X)=a$

但此时协方差却发生了变化，如下：

$$Cov(100X,\ 100Y)=100^2Cov(X,Y)$$

为了消除度量单位对协方差值的影响，把X、Y标准化后再求协方差，这就相关系数。

随机变量X、Y的标准化变量分别为

$$X^* = \frac{X - E(X)}{\sqrt{D(X)}} \quad 和 \quad Y^* = \frac{Y - E(Y)}{\sqrt{D(Y)}}$$

通过简单的计算可得

$$Cov(X^*, Y^*) = \frac{Cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

以此作为X和Y的相关系数的定义。

二. 相关系数

定义2 设随机变量X和Y的方差存在，且$D(X)>0, D(Y)>0$，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，记为ρ_{XY}，在不致引起混淆时，简记为ρ。

注意：ρ_{XY}是一个无量纲的量。

定义3 若随机变量X与Y的相关系数 $\rho_{XY}=0$，则称X和Y不相关；若$\rho_{XY}>0$，则称X和Y正相关；若$\rho_{XY}<0$，则称X和Y负相关。

二. 相关系数

例5. 设(X,Y)在圆域$D=\{(x, y): x^2+y^2\le r^2\ (r>0)\}$上服从均匀分布，判断$X$和$Y$是否相关。

解：由例2知，$Cov(X,Y)=0$ 所以 $\rho_{XY}=0$ 故X和Y不相关。

例6. 设 (X, Y)服从二维正态分布 $N(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2,\rho)$. 求$X$和$Y$的相关系数。

解：见参考书，X和Y的相关系数为ρ。即：X和Y不相关$\Leftrightarrow\rho=0$

二维正态分布的5个参数的含义为:

$$EX=\mu_1, DX=\sigma_1^2, EY=\mu_2, DY=\sigma_2^2, \rho_{XY}=\rho$$

二维正态分布的随机变量独立和不相关是等价的，充要条件都是$\rho=0$

定理2 设随机变量X和Y的相关系数ρ_{XY}存在，则有

（1）若X与Y相互独立，则$\rho_{XY}=0$，但反之不一定成立。

（2）$|\rho_{XY}|\le 1$，其中等号成立当且仅当X与Y有严格的线性关系（即存在常数a, b，使得$P\{Y=aX+b\}=1$成立）

证明：（1）因为X与Y独立，所以$E(XY)=E(X)E(Y)$，于是$\mathrm{Cov}(X,Y)=0$，所以$\rho_{XY}=0$. 例5表明$\rho_{XY}=0$，但X与Y不独立。

（2）由协方差的性质6，$[Cov(X, Y)]^2\le DXDY$，即有

$$\rho_{XY}^2=\frac{[Cov(X,Y)]^2}{D(X)\cdot D(Y)}\le 1 \quad \text{所以} \quad |\rho_{XY}|\le 1$$

二. 相关系数

易知当且仅当$|\rho_{XY}|=1$时，$[Cov(X, Y)]^2=DXDY$。

故由协方差的性质6知 $|\rho_{XY}|=1$当且仅当X与Y有严格的线性关系（即存在常数a, b，使得$P\{Y=aX+b\}=1$成立）。

由此可以看出，相关系数和协方差反映的不是X与Y之间“一般”的关系，而是反映二者之间的线性关系的密切程度。

定义4 当随机变量X和Y的相关系数$|\rho_{XY}|=1$时，称X和Y完全相关，而当$\rho_{XY}=1$时，称X与Y完全正相关，当$\rho_{XY}=-1$时，称X与Y完全负相关。

还可以从最小二乘法的角度来进一步加深理解相关系数的含义。

考虑以X的线性函数$a+bX$来近似Y，近似的误差为 $e=E[Y-(a+bX)]^2$

求a，b使e最小

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

将e分别关于a, b求偏导数，令其等于0，得

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

二. 相关系数

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

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

将a_0，b_0代入e，用a_0+b_0X来近似Y，最小误差为

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

二. 相关系数

用a_0+b_0X来近似Y，误差$e=D(Y)(1-\rho_{XY}^2)$为$|\rho_{XY}|$的严格减函数。因此可以用$|\rho_{XY}|$来刻画X和Y的线性相关的程度。

若$|\rho_{XY}|=1$，则$e=0$，说明Y与X之间以概率1有严格线性关系；

若$|\rho_{XY}|$越接近于1，则e越小，说明X与Y之间越近似有线性关系；即：X与Y的线性相关的程度越高；

若$|\rho_{XY}|$越接近于0，则e越大，说明X与Y的线性相关的程度越弱；

若$\rho_{XY}=0$，说明X与Y之间没有线性关系，此时X与Y之间的关系较复杂，可能相互独立，可能在平面上的某个区域内服从均匀分布，也可能有其他某种非线性的函数关系。

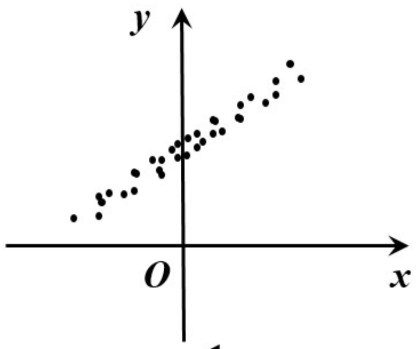
相关性与独立性的关系

X与Y独立 $\Rightarrow$ X与Y不相关；

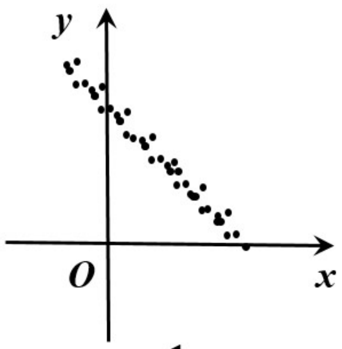
X与Y不相关 $\Rightarrow$ X与Y不一定独立；

X与Y相关 $\Rightarrow$ X与Y不独立。

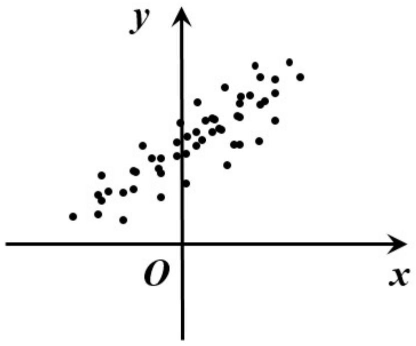
二. 相关系数



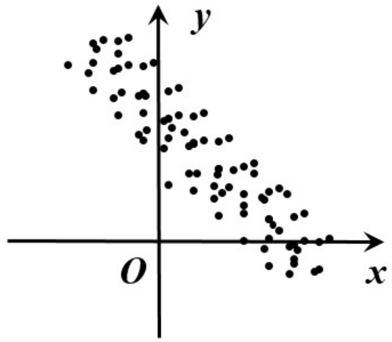
$\rho_{XY}=1$



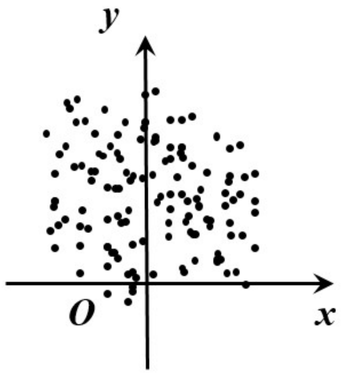
$\rho_{XY}=-1$



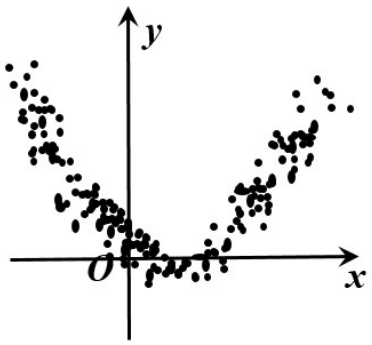
$0<\rho_{XY}<1$



$-1<\rho_{XY}<0$



$\rho_{XY}=0$



$\rho_{XY}=0$

例7.设随机变量Θ 服从上的均匀分布$U(0,2\pi)$，$X=\cos\Theta$，$Y=\cos(\Theta+\alpha)$，这里α是常数，求X和Y的相关系数。

解：易知Θ的密度函数为 $f_\Theta(\theta)=\begin{cases}\dfrac{1}{2\pi}, & 0<\theta<2\pi\\ 0 & \text{其他}\end{cases}$

所以 $E(X)=\int_{-\infty}^{+\infty}\cos\theta f_\Theta(\theta)d\theta=\int_0^{2\pi}\cos\theta\frac{1}{2\pi}d\theta=\frac{1}{2\pi}\int_0^{2\pi}\cos\theta d\theta=0$

$$E(Y)=\int_{-\infty}^{+\infty}\cos(\theta+\alpha)f_\Theta(\theta)d\theta=\int_0^{2\pi}\cos(\theta+\alpha)\frac{1}{2\pi}d\theta=\frac{1}{2\pi}\int_0^{2\pi}\cos(\theta+\alpha)d\theta=0$$

$$E(XY)=E[\cos(\Theta)\cos(\Theta+\alpha)]=\int_{-\infty}^{+\infty}\cos(\theta)\cos(\theta+\alpha)f_{\Theta}(\theta)d\theta$$

$$=\int_{0}^{2\pi}\cos(\theta)\cos(\theta+\alpha)\frac{1}{2\pi}d\theta=\frac{1}{2\pi}\int_{0}^{2\pi}\frac{1}{2}[\cos(\theta+\alpha-\theta)+\cos(\theta+\alpha+\theta)]d\theta$$

$$=\frac{1}{4\pi}\int_{0}^{2\pi}[\cos(\alpha)+\cos(2\theta+\alpha)]d\theta=\frac{1}{4\pi}[\int_{0}^{2\pi}\cos(\alpha)d\theta+\int_{0}^{2\pi}\cos(2\theta+\alpha)d\theta]$$

$$=\frac{1}{4\pi}[2\pi\cos(\alpha)+\frac{1}{2}(-\sin(2\theta+\alpha))\Big|_{0}^{2\pi}]=\frac{1}{2}\cos\alpha$$

于是 $Cov(X,Y)=E(XY)-E(X)E(Y)=\dfrac{1}{2}\cos\alpha$

二. 相关系数

另外 $E(X^2)=\int_0^{2\pi}\cos^2\theta\frac{1}{2\pi}d\theta=\frac{1}{2}$ $E(Y^2)=\int_0^{2\pi}\cos^2(\theta+\alpha)\cdot\frac{1}{2\pi}d\theta=\frac{1}{2}$

所以 $DX=1/2$ $DY=1/2$ 于是 $\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)D(Y)}}=\cos\alpha$

$X=\cos\Theta$, $Y=\cos(\Theta+\alpha)$

当$\alpha=0$ 时，$\rho=1$， $X=Y$

当$\alpha=\pi$时，$\rho=-1$， $X=-Y$

完全线性相关

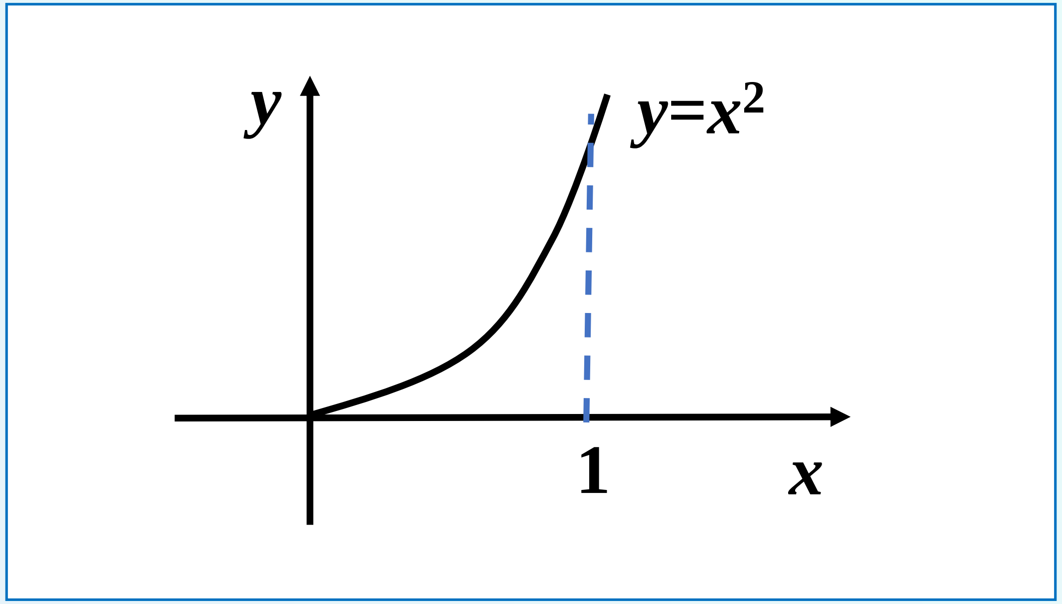
当$\alpha=\pi/2$ 或$\alpha=3\pi/2$时，$\rho=0$, X与Y不相关。 但$X^2+Y^2=1$，因此X与Y不独立。

例4.设(X,Y)在$D=\{(x,y): 0< y<x^2, 0<x<1\}$上服从均匀分布, 求$(X,Y)$的相关系数$\rho_{XY}$。

解：易知D的面积为：$S_D=\int_0^1 x^2dx=\frac{1}{3}$

$\therefore$ (X,Y)的联合密度函数

$$f(x,y)=\begin{cases}3, & (x,y)\in D\\ 0, & 其他\end{cases}$$

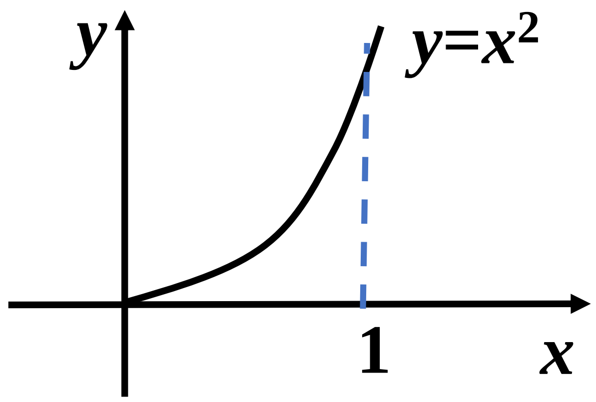


$$f(x,y)=\begin{cases}3, & (x,y)\in D\\ 0, & 其他\end{cases}$$

$$EX=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy=\int_0^1 dx\int_0^{x^2}3xdy=\int_0^1 3x^3dx=\frac{3}{4}$$

$$EX^2=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^2f(x,y)dxdy=\int_0^1 dx\int_0^{x^2}3x^2dy=\int_0^1 3x^4dx=\frac{3}{5}$$

$$DX=EX^2-(EX)^2=\frac{3}{5}-(\frac{3}{4})^2=\frac{3}{80}$$ 同理可得 $$DY=\frac{37}{700}$$



二. 相关系数

$$f(x,y)=\begin{cases}3, & (x,y)\in D\\ 0, & \text{其他}\end{cases}\qquad EX=\frac{3}{4},DX=\frac{3}{80},EY=\frac{3}{10},DY=\frac{37}{700}$$

$$EXY=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy=\int_0^1 dx\int_0^{x^2}3xydy=\int_0^1\frac{3}{2}x^5dx=\frac{3}{12}$$

$$Cov(X,Y)=EXY-EX\cdot EY=\frac{3}{12}-\frac{3}{4}\cdot\frac{3}{10}=\frac{1}{40}$$

$$\therefore \rho_{XY}=\frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}}=\frac{1/40}{\sqrt{3/80}\sqrt{37/700}}=\sqrt{\frac{35}{101}}$$

例5.已知X与Y服从正态分布$N(1,3^2)$和$N(0,4^2)$，且$\rho_{XY}=-1/2$，设$Z=X/3-Y/2$

(1)求Z的数学期望$E(Z)$和方差$D(Z)$；(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)易知 $EX=1,\ EY=0,\ DX=9,\ DY=16$ 。

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

$$\because \rho_{XY}=\frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}} \qquad \therefore Cov(X,Y)=\rho_{XY}\sqrt{DX}\sqrt{DY}=-\frac{1}{2}\times 3\times 4=-6$$

二. 相关系数

$$EX=1,\ EY=0,\ DX=9,\ DY=16 \qquad Cov(X,Y)=-6$$

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

$$=\frac{1}{9}D(X)+\frac{1}{4}D(Y)-\frac{1}{3}Cov(X,Y)=\frac{1}{9}\times 9+\frac{1}{4}\times 16-\frac{1}{3}\times(-6)=7$$

$$Cov(X,Z)=Cov(X,\frac{X}{3}-\frac{Y}{2})=Cov(X,\frac{X}{3})-Cov(X,\frac{Y}{2})=\frac{1}{3}D(X)-\frac{1}{2}Cov(X,Y)=6$$

$$\rho_{XZ}=\frac{Cov(X,Z)}{\sqrt{DX}\sqrt{DZ}}=\frac{2}{\sqrt{7}}$$

$\because \rho_{XZ}\neq 0$ $\therefore X$与Z相关，$\therefore$不独立

1.定义　设X和Y是随机变量

若$E(X^k)$, $k=1,2,\ldots$存在，称它为X的k阶原点矩，简称k阶矩。

如：期望$E(X)$为X的一阶原点矩。

若$E[X-E(X)]^k$, $k=2,3,\ldots$ 存在，称它为X的k阶中心矩。

方差$D(X)$为二阶中心矩。

若$E(X^kY^l)$, $k, l=1,2,\ldots$存在，称它为X和Y的$k+l$ 阶混合矩。

若$E\{[X-E(X)]^k[Y-E(Y)]^l\}$，$k$、$l=1,2,\ldots$ 存在，称它为X和Y的$k+l$阶混合中心矩。

协方差$Cov(X, Y)$是X与Y的二阶混合中心矩。

2.说 明

(1)以上数字特征都是随机变量函数的数学期望；

(2)若$E(X^k)$存在,则对小于k的一切非负整数l，$E(X^l)$存在。

(3)原点矩与中心矩可相互表示。如：$D(X)=E(X^2)-[E(X)]^2$

(4)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。

作业：36,40,42,44,45

第 15 讲

谢谢聆听