第四节　协方差、相关系数与矩

一、协方差

1. 协方差的概念

定义 1　设(X,Y)为二维随机变量，如果

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

存在，就称之为 X 与 Y 的协方差．记为$Cov(X,Y)$，即

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

利用数学期望的性质，得协方差的简化计算公式

$$Cov(X,Y)=E(XY)-EXEY\cdot$$

例 1 设二维随机变量(X,Y)的分布律为

(X,Y)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
P	0.2	0.1	0.4	0.3

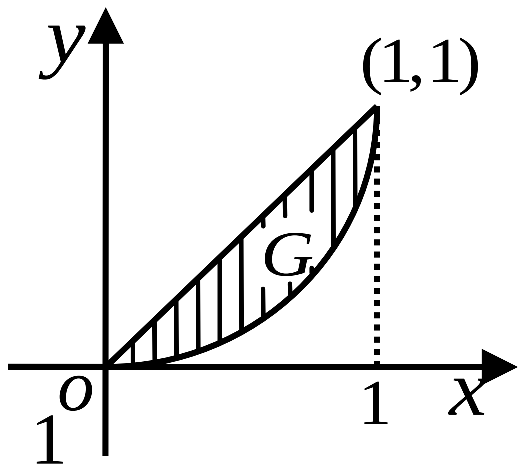
试求$Cov(X,Y)$.

解 经计算有 $EX=0.7$， $EY=0.4$， $E(XY)=0.3$，

所以$Cov(X,Y)=0.3-0.7\times0.4=0.02$.

例 2 设二维随机变量 (X,Y) 在区域 $G=\{(x,y)\mid 0\le x\le 1,\ x^2\le y\le x\}$ 上服从均匀分布．求 $Cov(X,Y)$．

解 由题意知，(X,Y) 的密度函数为 $f(x,y)=\begin{cases}6, & (x,y)\in G,\\ 0, & \text{其它}.\end{cases}$



$$EX=\iint_G x\cdot 6dxdy=6\int_0^1 dx\int_{x^2}^{x} xdy=\frac{1}{2},$$

$$EY=\iint_G y\cdot 6dxdy=6\int_0^1 dx\int_{x^2}^{x} ydy=\frac{2}{5},$$

$$E(XY)=\iint_G xy\cdot 6dxdy=6\int_0^1 dx\int_{x^2}^{x} xydy=\frac{1}{4},$$

所以 $Cov(X,Y)=\dfrac{1}{4}-\dfrac{1}{2}\times\dfrac{2}{5}=\dfrac{1}{20}$．

2. 协方差的性质

性质 1 $Cov(X,X)=DX$.

性质 2 $Cov(X,Y)=Cov(Y,X)$.

性质 3 $Cov(X,c)=0$.

性质 4 $Cov(aX,bY)=abCov(X,Y)$.

性质 5 $Cov(X_1 \pm X_2,Y)=Cov(X_1,Y)\pm Cov(X_2,Y)$.

性质 6 $D(X \pm Y)=DX+DY \pm 2Cov(X,Y)$.

例 3 设随机变量 X 与 Y 的方差均为正. 求

$$Cov\left(\frac{X}{\sqrt{DX}}+\frac{Y}{\sqrt{DY}},\frac{X}{\sqrt{DX}}-\frac{Y}{\sqrt{DY}}\right).$$

解 原式 $=Cov\left(\frac{X}{\sqrt{DX}},\frac{X}{\sqrt{DX}}\right)-Cov\left(\frac{X}{\sqrt{DX}},\frac{Y}{\sqrt{DY}}\right)$

$$+Cov\left(\frac{Y}{\sqrt{DY}},\frac{X}{\sqrt{DX}}\right)-Cov\left(\frac{Y}{\sqrt{DY}},\frac{Y}{\sqrt{DY}}\right)$$

$$=\frac{1}{DX}Cov(X,X)-\frac{1}{\sqrt{DX}\sqrt{DY}}Cov(X,Y)$$

$$+\frac{1}{\sqrt{DY}\sqrt{DX}}Cov(Y,X)-\frac{1}{DY}Cov(Y,Y)$$

$$=\frac{1}{DX}DX-\frac{1}{DY}DY=1-1=0.$$

二、相关系数

定义 2 设 (X,Y) 为二维随机变量，如果 $DX>0, DY>0$，就称 $\frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}}$ 为随机变量 X 与 Y 的**相关系数**．记为 ρ_{XY} 或 ρ，即

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

注 1： 计算相关系数 ρ_{XY}，需要事先计算五个数学期望

$$EX，\quad EY，\quad E(X^2)，\quad E(Y^2) \text{ 和 } E(XY)$$

对于常见分布，EX, EY, DX, DY 可以直接得到.

例 4 设随机变量 $X \sim B(2,\frac{1}{3}), Y = |X-1|$，求 ρ_{XY} .

解 由于 $X \sim B(2,\frac{1}{3})$，故 $EX = \frac{2}{3}, DX = \frac{4}{9}$.

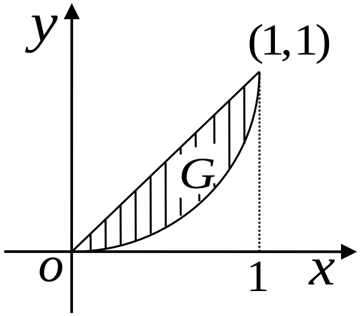
又 $X \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{4}{9} & \frac{4}{9} & \frac{1}{9} \end{pmatrix}$,故 $Y \sim \begin{pmatrix} 0 & 1 \\ \frac{4}{9} & \frac{5}{9} \end{pmatrix}$, $XY = X|X-1| \sim \begin{pmatrix} 0 & 2 \\ \frac{8}{9} & \frac{1}{9} \end{pmatrix}$,

得 $EY = \frac{5}{9}, DY = \frac{20}{81}$, $E(XY) = \frac{2}{9}$，所以

$$\rho_{XY} = \frac{\frac{2}{9} - \frac{2}{3} \times \frac{5}{9}}{\sqrt{\frac{4}{9}} \times \sqrt{\frac{20}{81}}} = \frac{-\frac{4}{27}}{\sqrt{\frac{4}{9}} \times \sqrt{\frac{20}{81}}} = -\frac{\sqrt{5}}{5} .$$

例 5 设二维随机变量(X,Y)在区域$G=\{(x,y)\mid 0\le x\le 1,$ $x^2\le y\le x\}$上服从均匀分布．求ρ_{XY}．

解 由本节例 2 知，(X,Y)的密度函数为 $f(x,y)=\begin{cases}6,(x,y)\in G,\\ 0,\text{其它}.\end{cases}$



且已计算得 $EX=\dfrac{1}{2}$， $EY=\dfrac{2}{5}$， $Cov(X,Y)=\dfrac{1}{20}$，

$$E(X^2)=\iint\limits_G x^2\cdot 6dxdy=6\int_0^1 dx\int_{x^2}^{x}x^2dy=\frac{3}{10},$$

$E(Y^2)=\iint\limits_G y^2\cdot 6dxdy=\dfrac{3}{14}$， 所以 $DX=\dfrac{3}{10}-(\dfrac{1}{2})^2=\dfrac{1}{20}$，

$DY=\dfrac{3}{14}-(\dfrac{2}{5})^2=\dfrac{19}{350}$． 故 $\rho_{XY}=\dfrac{\dfrac{1}{20}}{\sqrt{\dfrac{1}{20}}\times\sqrt{\dfrac{19}{350}}}=\sqrt{\dfrac{35}{38}}$．

例 6 设随机变量 X 与 Y 的相关系数 $\rho_{XY}=0.5$，且 $DX=1$，$DY=4$，求 $U=2X+Y$ 与 $V=X-Y$ 的相关系数 ρ_{UV}．

解 由于 $Cov(X,Y)=\rho_{XY}\sqrt{DX}\sqrt{DY}=0.5\times\sqrt{1}\times\sqrt{4}=1$，所以

$$DU=D(2X+Y)=4DX+DY+4Cov(X,Y)$$

$$=4\times1+4+4\times1=12,$$

$$DV=D(X-Y)=DX+DY-2Cov(X,Y)=1+4-2\times1=3,$$

$$Cov(U,V)=Cov(2X+Y,X-Y)=2DX-DY-Cov(X,Y)$$

$$=2\times1-4-1=-3,$$

所以

$$\rho_{UV}=\frac{-3}{\sqrt{3}\sqrt{12}}=-0.5.$$

2. 相关系数的性质

性质 7 $|\rho_{XY}| \le 1$，即 $\rho_{XY} \in [-1,1]$.

性质 8 $|\rho_{XY}| = 1$ 的充要条件为存在常数 $a,b\,(a \ne 0)$，使得 $P\{Y = aX + b\} = 1$.

注 2: $|\rho_{XY}|$ 越大（越小），X 与 Y 线性关系越强（越弱）.

性质 9 若 $Y = aX + b$，有 $\rho_{XY} = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}$.

定义 3 如果 $\rho_{XY}=0$，就称随机变量 X 与 Y **不相关**.

定理 1 设随机变量 X 与 Y 的相关系数 ρ_{XY} 存在，则下列结论是等价的

(1) X 与 Y 不相关； (2) $\rho_{XY}=0$； (3) $Cov(X,Y)=0$；

(4) $E(XY)=EXEY$； (5) $D(X\pm Y)=DX+DY$.

定理 2 如果随机变量 X 与 Y 相互独立，且 X 与 Y 的相关系数 ρ_{XY} 存在，则 X 与 Y 不相关.

注 3： 如果 X 与 Y 不相关，则 X 与 Y 未必相互独立，即定理 2 的逆命题不成立.

例 8 设二维随机变量$(X,Y)\sim U(G)$，其中平面区域$G=\{(x,y)\,|\,x^2+y^2\le 1\}$，问 X 与 Y 是否相互独立？是否不相关？

解 由题意知，(X,Y)的密度函数 $f(x,y)=\begin{cases}\dfrac{1}{\pi},(x,y)\in G,\\ 0,\ 其它.\end{cases}$

可求得 $f_X(x)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-x^2},|x|\le 1,\\ 0, \qquad 其它,\end{cases}$ $f_Y(y)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-y^2},|y|\le 1,\\ 0, \qquad 其它,\end{cases}$

由于 $f(x,y)\ne f_X(x)f_Y(y)$，所以 X 与 Y 不相互独立.

利用二重积分的奇偶对称性，有 $EX=\iint\limits_G x\cdot\dfrac{1}{\pi}dxdy=0$，

$$EY=\iint\limits_G y\cdot\frac{1}{\pi}dxdy=0,\quad E(XY)=\iint\limits_G xy\cdot\frac{1}{\pi}dxdy=0,$$

故有 $E(XY)=EXEY$，所以 X 与 Y 不相关.

例 9 设二维随机变量 (X,Y) 的分布律为

$Y \backslash X$	-1	0	1	$p_{\bullet j}$
0	0	1/3	0	1/3
1	1/3	0	1/3	2/3
$p_{i\bullet}$	1/3	1/3	1/3	1

问 X 与 Y 是否相互独立？又是否不相关？

解 由于 $P\{X=-1,Y=0\}=0 \neq P\{X=-1\}P\{Y=0\}=\dfrac{1}{9}$，

所以 X 与 Y 不相互独立.

又可计算得 $EX=0,\quad EY=\dfrac{2}{3},\quad E(XY)=0$，

故有 $E(XY)=EXEY$，所以 X 与 Y 不相关.

定理 3 如果二维随机变量 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则

(1) X 与 Y 的相关系数 $\rho_{XY}=\rho$；

(2) X 与 Y 相互独立 $\Leftrightarrow\ \rho=0\ \Leftrightarrow\ Cov\left(X,Y\right)=0$.

例 10 已知随机变量 (X,Y) 服从二维正态分布，并且 X 和 Y 分别服从正态分布 $N(1,3^2)$ 和 $N(0,4^2)$， X 与 Y 的相关系数 $\rho_{XY}=-\dfrac{1}{2}$，设 $Z=\dfrac{X}{3}+\dfrac{Y}{2}$. (1) 求 X 与 Z 的相关系数 ρ_{XZ}；(2) 问 X 和 Z 是否相互独立？为什么？

解 (1) $Cov(X,Z)=Cov(X,\dfrac{X}{3}+\dfrac{Y}{2})=\dfrac{1}{3}Cov(X,X)+\dfrac{1}{2}Cov(X,Y)$

$$=\frac{1}{3}DX+\frac{1}{2}\rho_{XY}\sqrt{DX}\sqrt{DY}=\frac{1}{3}\times 9+\frac{1}{2}\times(-\frac{1}{2})\times 3\times 4=0,$$

（续解）因此 $\rho_{XZ}=\dfrac{Cov(X,Z)}{\sqrt{DX}\sqrt{DZ}}=\dfrac{0}{\sqrt{DX}\sqrt{DZ}}=0$.

(2) 由于 (X,Y) 服从二维正态分布，而 $\begin{cases} X=X, \\ Z=\dfrac{X}{3}+\dfrac{Y}{2}, \end{cases}$ 且系数行列式 $\begin{vmatrix} 1 & 0 \\ \dfrac{1}{3} & \dfrac{1}{2} \end{vmatrix}=\dfrac{1}{2}\neq 0$，由第三章结论知，$(X,Z)$ 也服从二维正态分布. 又由(1)知，X 与 Z 不相关，所以利用定理 3(2)可得 X 和 Z 相互独立.

三、矩

定义 4 设有随机变量 X，如果对于正整数 k，$E(X^k)$ 存在，就称 $E(X^k)$ 为 X 的 k **阶原点矩**.

如果 $E[(X-EX)^k]$ 存在，就称 $E[(X-EX)^k]$ 为 X 的 k **阶中心矩**. $k=1,2,\cdots$.

定义 5 设有二维随机变量 (X,Y)，如果对于正整数 k,l，$E(X^kY^l)$ 存在，就称 $E(X^kY^l)$ 为 X 和 Y 的 $k+l$ **阶混合原点矩**.

如果 $E[(X-EX)^k(Y-EY)^l]$ 存在，就称

$$E[(X-EX)^k(Y-EY)^l]$$

为 X 和 Y 的 $k+l$ **阶混合中心矩**. $k=1,2,\cdots,l=1,2,\cdots$.

注 4：①X 的一阶原点矩即为 X 的数学期望 EX．

②X 的一阶中心矩 $E(X-EX)=0$．

③X 的二阶中心矩即为 X 的方差 DX．

④(X,Y) 的$1+1$阶混合中心矩为(X,Y)的协方差$Cov(X,Y)$．

练习：

1. 将长度为1m的木棒随机地截成两段，求两段长度 X 与 Y 的相关系数 ρ_{XY}.

答案： -1.

2. 设 A,B 为两个随机事件，$P(A)=p$，$P(B)=q$，$P(AB)=r$，且 $p,q\in(0,1)$，记 $X=\begin{cases}0, & A\text{不发生},\\ 1, & A\text{发生},\end{cases}$ $Y=\begin{cases}0, & B\text{不发生},\\ 1, & B\text{发生},\end{cases}$，

(1) 求相关系数 ρ_{XY}；

(2) 证明 X 与 Y 不相关的充要条件为 X 与 Y 相互独立.

解（1）可求得 (X,Y) 的分布律为

$Y \backslash X$	0	1	$p_{\bullet j}$
0	$1-p-q+r$	$p-r$	$1-q$
1	$q-r$	r	q
$p_{i\bullet}$	$1-p$	p	1

（续解） $EX=p, EY=q, DX=p(1-p)$， $DY=q(1-q)$

$E(XY)=r$，所以 $\rho_{XY}=\dfrac{r-pq}{\sqrt{p(1-p)q(1-q)}}$.

(2) **证** 由(1)知，

X 与 Y 不相关 $\Leftrightarrow r=pq \Leftrightarrow (X,Y)$ 分布律为

$Y \backslash X$	0	1	$p_{\bullet j}$
0	$(1-p)(1-q)$	$p(1-q)$	$1-q$
1	$(1-p)q$	pq	q
$p_{i\bullet}$	$1-p$	p	1

$\Leftrightarrow$ X 与 Y 相互独立.