§4.3　协方差和相关系数

对于多维随机变量 $X=(X_1,X_2,\cdots,X_n)$，也可以定义它的期望与方差为

$$E(X)=(E(X_1),E(X_2),\cdots,E(X_n)).$$

$$D(X)=(D(X_1),D(X_2),\cdots,D(X_n)).$$

对于二维随机变量（X,Y）而言，方差 $D(X),D(Y)$分别反映了X，Y相对于各自数学期望的偏离程度，但并没有反映随机变量X，Y之间的关系。由方差的性质知，当X与Y**相互独立**时有 $D(X+Y)=D(X)+D(Y)$，在X 与Y不相互独立时却多出了一项 $2E\left[(X-E(X))(Y-E(Y))\right]$，它在某种程度上反映了X与Y之间的某种关系，值得我们进一步考察，由此引出了概念：**协方差**。

协方差

定义4.6 设（X, Y）为 r. v.，若 $E\{[X-E(X)][Y-E(Y)]\}$ 存在，称其为X与Y 的**协方差**，记为$Cov(X,Y)$，即

$$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}。$$

注 （1）协方差是一种二维随机变量特殊函数的数学期望。它并不是一定存在的。

（2）当X=Y 时，显然有

$$Cov(X,X)=E[X-E(X)]^2=D(X)。$$

所以，方差只是协方差的一个特例。

（3）在实际计算协方差时，常用公式

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

这是因为

$$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$$
$$=E[XY-XE(Y)-YE(X)+E(X)E(Y)]$$
$$=E(XY)-E(Y)E(X)-E(X)E(Y)+E(X)E(Y)$$
$$=E(XY)-E(X)E(Y).$$

协方差具有如下一些性质：

定理 设 X, Y, Z 为随机变量，a, b, c为常数，则协方差具有以下性质（设此处所涉协方差均存在）：

协方差的性质

（1） $Cov(X,Y)=Cov(Y,X)$ (由注3显然)

（2） $Cov(X,a)=0$

（3） $Cov(aX,bY)=abCov(X,Y)$

（4） $Cov(X\pm Y,Z)=Cov(X,Z)\pm Cov(Y,Z)$

（5） $D(X\pm Y)=D(X)+D(Y)\pm 2Cov(X,Y)$

（6）若X与Y相互独立，则 $Cov(X,Y)=0$

由性质（3）和（4），我们有如下推论：

推论：设对任意 $i,j=1,2,\cdots$ 为常数 a_i,b_j 为随机变量 X_i,Y_j，则有

$$Cov\left(\sum_{i=1}^{n}a_iX_i,\sum_{j=1}^{m}b_jY_j\right)=\sum_{j=1}^{m}\sum_{i=1}^{n}a_ib_jCov(X_i,Y_j)。$$

协方差阵

协方差阵 设$(X_1, X_2, \cdots, X_n)$是 n 维随机变量，称矩阵

$$V=\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

协方差阵显然是对称阵。

为$(X_1, X_2, \cdots, X_n)$的协方差阵。其中

$$\sigma_{ij}=Cov(X_i, X_j), i, j=1,2,\cdots,n.$$

特别地，当n=2时，即 (X,Y) 的协方差阵为

$$V=\begin{bmatrix} D(X) & Cov(X,Y) \\ Cov(Y,X) & D(Y) \end{bmatrix}$$

例：设 (X, Y) 的联合分布律为见右，求（X, Y）的协方差阵及均值向量。

$Y \backslash X$	-1	0	1
-1	$1/8$	$1/8$	$1/8$
0	$1/8$	0	$1/8$
1	$1/8$	$1/8$	$1/8$

解：显然，X与Y有相同的分布律，且有

$$X, \ Y \sim \begin{bmatrix} -1 & 0 & 1 \\ 3/8 & 2/8 & 3/8 \end{bmatrix}; \quad XY \sim \begin{bmatrix} -1 & 0 & 1 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$$

于是 $E(X) = E(Y) = E(XY) = 0; \quad D(X) = D(Y) = \frac{3}{4};$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 0。$$

从而X与Y的协方差阵及(X,Y)的均值向量分别为

$$V = \begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{4} \end{bmatrix}; \quad E(X,Y) = (E(X), E(Y)) = (0,0)。$$

例4.17 设$X \sim U(3,6), Y \sim e(3), Z \sim \Gamma(2,3)$, 且$X$与Y独立，

$E(YZ) = 1/3$, Cov（X， Z）=2， 求D(3X- 2Y+Z) .

解：$X \sim U(3,6) \Rightarrow \mathrm{D}(X) = (6-3)^2/12$,

$Y \sim e(3) \Rightarrow \mathrm{D}(Y) = 1/3^2$,

$Z \sim \Gamma(2,3) \Rightarrow \mathrm{D}(Z) = 2/3^2$,

D(3X- 2Y+Z)=D(3X-2Y)+D(Z)+2Cov(3X-2Y,Z).

2Cov(3X-2Y,Z)=6Cov(X,Z)-4Cov(Y,Z),

Cov(Y,Z)=E(YZ)-E(Y)E(Z)= $1/3 - 1/3 \times 2/3 = 1/9$,

X与Y独立$\Rightarrow$

D(3X-2Y)=9D(X)+4D(Y)=9$\times 9/12 + 4 \times 1/9$.

相关系数

协方差反映了两个随机变量之间的关系，但其值没有固定的范围。为消除这种影响，我们引入了另一种数字特征：相关系数。

定义4.7：设随机变量X与Y的数学期望及方差都存在，且D(X)>0，D(Y)>0，称

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

为X与Y的相关系数。

若令

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

即将X与Y标准化则 $R(X,Y)=Cov(X^*,Y^*)$。因此，X 与Y 的相关系数也称为X 与Y 的标准化的协方差。此时有

$$D(X^*\pm Y^*)=2(1\pm R(X,Y)).$$

相关系数的性质

定理4.4 设X与Y的相关系数存在，则

(1) $R(X,Y)=R(Y,X)$; $R(X,Y)=0 \Leftrightarrow Cov(X,Y)=0$;

(2) $|R(X,Y)|\leq 1$;

(3) $R(X,Y)=1 \Leftrightarrow \exists a>0, b,$ 使得 $P(aX+b=Y)=1.$

$R(X,Y)=\text{-}1 \Leftrightarrow \exists a<0, b,$ 使得 $P(aX+b=Y)=1.$

证明：（1）由相关系数的定义$R(X,Y)=\dfrac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$=Cov（$X^*$，$Y^*$）.

（2）$0\leq D(X^*\pm Y^*)=2(1\pm R(X,Y))\Rightarrow \pm R(X,Y)\geq -1$

$\Rightarrow -1\leq R(X,Y)\leq 1\Rightarrow \left|R(X,Y)\right|\leq 1.$

（3）$R(X,Y)=1 \Leftrightarrow D(X^*-Y^*)=2(1-R(X,Y))=0$

$\Leftrightarrow$ 存在常数c使得1=P（X^*-Y^*=c）=P（$\frac{X-E(X)}{\sqrt{D(X)}}-\frac{Y-E(Y)}{\sqrt{D(Y)}}$=c）

$\Leftrightarrow$ 存在常数a=$\frac{\sqrt{D(Y)}}{D(X)}$>0, b=E(Y)-c$\sqrt{D(Y)}$-$\frac{E(X)\sqrt{D(Y)}}{D(X)}$,

使得P（X^*-Y^*=c）=1

类似的利用$R(X,Y)=-1 \Leftrightarrow D(X^*+Y^*)=2(1+R(X,Y))=0$

$\Leftrightarrow$ 存在常数a<0, b 使得P（Y=aX+b）=1

正负线性相关

定义:设X与Y为两随机变量，

若R(X,Y)=1,则称X与Y完全正线性相关，

此时(X, Y)落在某条正斜率直线上的概率为1

若R(X,Y)=-1,则称X与Y完全负线性相关，

此时(X, Y)落在某条负斜率直线上的概率为1

|R(X,Y)|=1时， 统称为X与Y完全线性相关。

若R(X,Y)>0,则称X与Y正相关; 若R(X,Y)<0,则称X与Y负相关， R(X,Y)=0时， 称X与Y不相关。

不相关与独立性

值得注意的是，这里的X与Y不相关与前面介绍的X与Y相互独立是两个不同的概念，前者仅仅表明X与Y在线性关系上是不相关的，所以这里的“不相关”意指“线性不相关”，并非是指相互之间没有任何关系，即“独立”，因此，若X与Y 独立，则X与Y当然不相关；反之，若X与Y不相关，仅指X与Y 之间没有线性关系，但它们之间有可能有其它关系，如平方关系 等，则此时X与Y是不独立的。即不相关与独立有入如下关系。

定理：若D(X)、D(Y)存在且都大于零，则

$$X\text{与}Y\text{独立}\underset{\not\Leftarrow}{\Rightarrow} X\text{与}Y\text{不相关}\Leftrightarrow R(X,Y)=0\text{。}$$

例4.19 设X~U(-1,1),Y=X²则X,Y不相关.

$X \sim U(-1,1)$, $Y=X^2$

解: $\because EX=0, E(X^3)=\int_{-1}^{1} x^3/2dx=0$

$\therefore Cov(X,Y)=E(X^3)-E(X)E(X^2)=0$

$$\therefore R(X,Y)=\frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DX}}=0$$

X与Y不相关，但X与Y不独立。

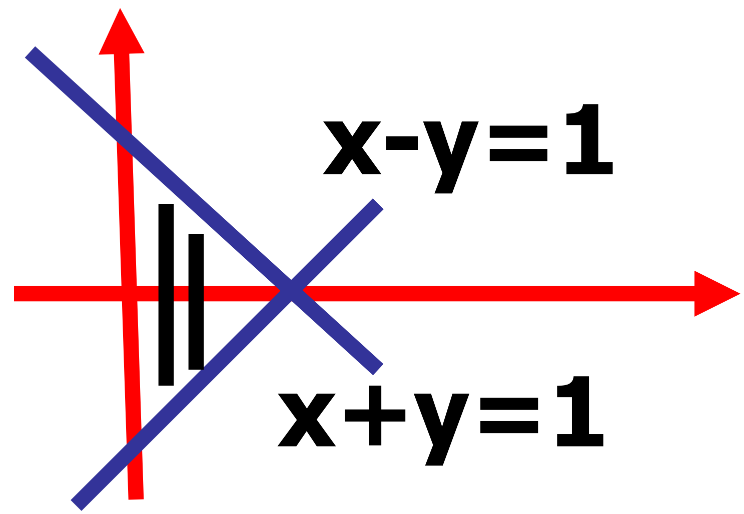
例4.20设二维随机变量(X,Y)在G上服从均匀分布其中

$$G=\{(x,y)\,|\,0\le x\le 1, x+y\le 1, x-y\le 1\}$$

(1)求X,Y的期望与方差;(2)证明:X与Y不相关,不独立;

解:写出(X,Y)的联合密度函数

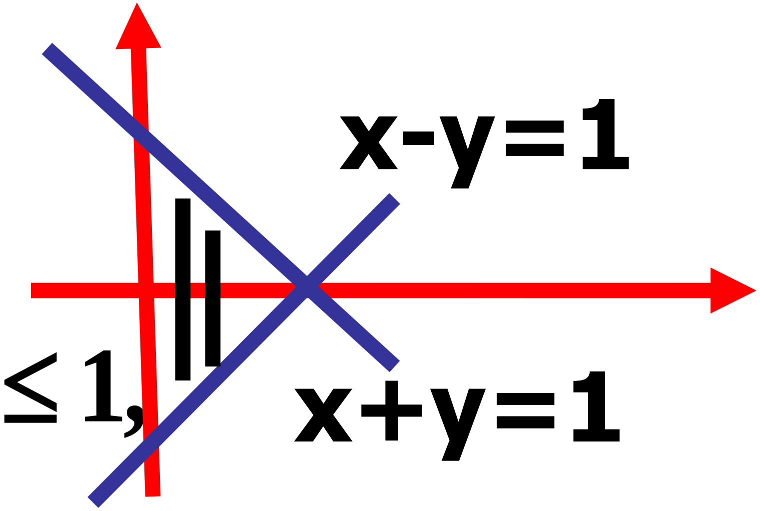
$$f(x,y)=\begin{cases}1,(x,y)\in G\\ 0,\text{其它}.\end{cases}$$



再分别求出**X,Y**的边缘密度函数:

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy$$

$$=\int_{x-1}^{1-x} 1dy = 2(1-x), 0\le x\le 1,$$



$$f_Y(y)=1-|y|, -1\le y\le 1$$

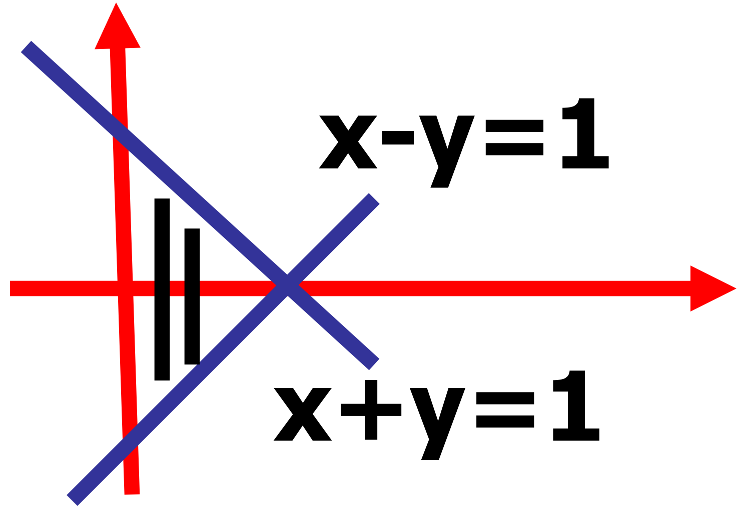
$$EX=\int_0^1 x2(1-x)dx = 1/3,$$

$$EX^2=\int_0^1 x^2 2(1-x)dx = 1/6$$

$$EY=0, EY^2=1/6$$

$$E(XY)=\iint\limits_{R^2} xyf(x,y)dxdy$$

$$=\int_0^1\{\int_{x-1}^{1-x} xy1dy\}dx \quad =0$$



$$\therefore Cov(X,Y)=E(XY)-EX\bullet EY=0$$

可见,**X,Y**不相关,

但是在G上$f(x,y)\neq f_X(x)\bullet f_Y(y)$

所以,**X,Y**不独立.

例: 设（X，Y）有联合密度函数

$$f(x,y)=\begin{cases} x+y, & 0\le x\le 1, 0\le y\le 1 \\ 0, & else \end{cases},$$

求 $E(X), E(Y), D(X), D(Y), Cov(X,Y), \rho_{XY}$, 并问X与Y是否相互独立?

解:由（X，Y）的联合密度函数得X，Y 的（边缘）密度函数为

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\begin{cases} \int_0^1 (x+y)dy = x+\frac{1}{2}, & 0\le x\le 1 \\ 0, & else \end{cases};$$

所以
$$E(X)=\int_{-\infty}^{+\infty} xf_X(x)dx=\int_0^1 x(x+\frac{1}{2})dx=\frac{7}{12};$$
$$E(X^2)=\int_{-\infty}^{+\infty} x^2 f_X(x)dx=\int_0^1 x^2(x+\frac{1}{2})dx=\frac{5}{12};$$
$$D(X)=E(X^2)-(E(X))^2=\frac{5}{12}-(\frac{7}{12})^2=\frac{11}{144}.$$

由X与Y的对称性有
$$E(Y)=\frac{7}{12};\quad D(Y)=\frac{11}{144}.$$

又
$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy=\int_0^1\int_0^1 xy(x+y)dxdy=\frac{1}{3},$$

所以

$$Cov(X,Y)=E(XY)-E(X)E(Y)=\frac{1}{3}-\frac{7}{12}\times\frac{7}{12}=-\frac{1}{144}.$$

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{-\frac{1}{144}}{\sqrt{\frac{11}{144}}\sqrt{\frac{11}{144}}}=-\frac{1}{11}.$$

由于 $\rho_{XY}\neq 0$ ，所以X与Y不相互独立。