§4.协方差及相关系数

1、定义

称COV(X,Y)= E[(X-EX)(Y-EY)]=EXY-EXEY

为随机变量X,Y的协方差. 而 COV(X,X)=DX.

$$\rho_{XY}=\frac{COV(X,Y)}{\sqrt{DX}.\sqrt{DY}}$$ 为随机变量X,Y的相关系数。

ρ_{XY}是一个无量纲的量；若ρ_{XY}=0，

称 X,Y 不相关,此时 COV(X,Y)=0。

定理： 若X，Y独立，则X,Y不相关。

证明： 由数学期望的性质有

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

又 E(X-EX)=0,　　E(Y-EY)=0

所以　E(X-EX)(Y-EY)=0。

即　　COV(X,Y)=0

返回主目录

注意：若E(X-EX)(Y-EY)≠0，即EXY-EXEY ≠0，　则X，Y一定相关，且X，Y一定不独立。

2、协方差的性质

1) COV(X,Y)=COV(Y,X);

2) COV(aX，bY)=abCOV(X,Y);

3) COV(X+Y,Z)=COV(X,Z)+COV(Y,Z);

4) 若 X,Y 不相关，则：EXY=EXEY, D(aX+bY)=a^2DX+b^2DY

由方差的性质 3 ）知：

$$D(aX+bY)= a^2DX+b^2DY+2abCOV(X,Y)$$

3、相关系数的性质

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

2) $|\rho_{XY}| = 1 \Leftrightarrow$ 存在常数 a,b 使 P{Y=a+bX}=1.

证明：

令：$e = E[Y-(a+bX)]^2$

$$= EY^2 + b^2EX^2 + a^2 - 2aEY - 2bEXY + 2abEX$$

求 a,b 使 e 达到最小

令 $$\begin{cases} \dfrac{\partial e}{\partial a} = 2a + 2bEX - 2EY = 0 \\ \dfrac{\partial e}{\partial b} = 2bEX^2 - 2EXY + 2aEX = 0 \end{cases}$$

将 $a = EY - bEX$，代入第二个方程得

$bEX^2 - EXY + (EY - bEX)EX = 0$，故 $b = \dfrac{EXY - EXEY}{EX^2 - (EX)^2}$

解得

$$b_0 = \frac{COV(X,Y)}{DX};$$

$$a_0 = EY - b_0 EX = EY - EX \cdot \frac{COV(X,Y)}{DX}$$

$$\min_{a,b} E[Y-(a+bX)]^2 = E[Y-(a_0+b_0X)]^2$$

$$= E(Y - EY + EX\frac{COV(X,Y)}{DX} - X\cdot\frac{COV(X,Y)}{DX})^2$$

$$= E((Y-EY)-(X-EX)\cdot\frac{COV(X,Y)}{DX})^2$$

$$= DY + DX\cdot\frac{COV^2(X,Y)}{(DX)^2} - 2COV(X,Y)\cdot\frac{COV(X,Y)}{DX}$$

$$= DY + \frac{COV^2(X,Y)}{DX} - 2\frac{COV^2(X,Y)}{DX}$$

返回主目录

$$= DY - \frac{COV^2(X,Y)}{DX} = DY - \frac{\rho_{XY}^2 \cdot DX \cdot DY}{DX} = (1-\rho_{XY}^2)DY$$

即：$\min\limits_{a,b} E[Y-(a+bX)]^2 = (1-\rho_{XY}^2)DY$

由上式得:

1) $1 - \rho_{XY}^2 \geq 0$, $|\rho_{XY}| \leq 1$。

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

从而 $D[Y-(a_0+b_0X)] + (E[Y-(a_0+b_0X)])^2 = E[Y-(a_0+b_0X)]^2 = 0$

所以 $D[Y-(a_0+b_0X)] = 0$, $E[Y-(a_0+b_0X)] = 0$

故 $P\{Y-(a_0+b_0X)=0\}=1$.

即 $P\{Y=a_0+b_0X\}=1$。

返回主目录

反之，若存在 a^*, b^* 使 P{Y＝a^*+b^*X }=1，则

$$P\{Y-(a^*+b^*X)=0\}=1,$$

故 $E[Y-(a^*+b^*X)]^2=0$ 而

$$0=E[Y-(a^*+b^*X)]^2 \geq \min_{a,b} E[Y-(a+bX)]^2=(1-\rho_{XY}^2)DY$$

则 $1-\rho_{XY}^2=0, |\rho_{XY}|=1$。

说　明

相关系数是表征随机变量 X 与 Y 之间线性关系紧密程度的量.

当 $|\rho_{X,Y}|=1$ 时，X 与 Y 之间以概率1存在着线性关系；

当 $|\rho_{X,Y}|$ 越接近于0时，X 与 Y 之间的线性关系越弱；

当 $|\rho_{X,Y}|=0$ 时，X 与 Y 之间不存在线性关系(不相关).

X与Y之间没有线性关系并不表示它们之间没有关系。

5、例子

设 X，Y 是二个随机变量，已知 $DX=1$，$DY=4$，$\operatorname{cov}(X,\ Y)=1$，记

$$\xi = X-2Y,\quad \eta = 2X-Y$$

试求：$\rho_{\xi,\ \eta}$.

解：

$$D\xi = D(X-2Y) = DX+4DY-4\operatorname{cov}(X,\ Y)$$

$$= 1+4\times 4-4\times 1=13$$

$$D\eta = D(2X-Y) = 4DX+DY-4\operatorname{cov}(X,\ Y)$$

$$= 4\times 1+4-4\times 1\quad =4$$

$$\operatorname{cov}(\xi,\ \eta) = \operatorname{cov}(X-2Y,\ 2X-Y)$$

$$= 2\operatorname{cov}(X,\ X) - 4\operatorname{cov}(Y,\ X) - \operatorname{cov}(X,\ Y) + 2\operatorname{cov}(Y,\ Y)$$

$$= 2DX - 5\operatorname{cov}(X,\ Y) + 2DY$$

$$= 2 \times 1 - 5 \times 1 + 2 \times 4$$

$$= 5$$

所以，

$$\rho_{\xi,\ \eta} = \frac{\operatorname{cov}(\xi,\ \eta)}{\sqrt{D\xi}\sqrt{D\eta}} = \frac{5}{\sqrt{13}\sqrt{4}} = \frac{5\sqrt{13}}{26}$$

设(X,Y)服从二维正态分布，求：ρ_{XY}

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

由上述知：$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$，$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$

$$EX=\mu_1,DX=\sigma_1^2,EY=\mu_2,DY=\sigma_2^2,$$

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

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

令 $t=\frac{1}{\sqrt{1-\rho^2}}[\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}]$， $u=\frac{x-\mu_1}{\sigma_1}$，

则 $x-\mu_1=\sigma_1 u$， $y-\mu_2=(t\sqrt{1-\rho^2}+\rho u)\sigma_2$

$$J=\begin{vmatrix}\frac{\partial t}{\partial x} & \frac{\partial t}{\partial y}\\ \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y}\end{vmatrix}^{-1}=\begin{vmatrix}-\frac{1}{\sqrt{1-\rho^2}}\frac{\rho}{\sigma_1} & \frac{1}{\sqrt{1-\rho^2}}\frac{1}{\sigma_2}\\ \frac{1}{\sigma_1} & 0\end{vmatrix}^{-1}$$

$$=(-\frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}})^{-1}=-\sigma_1\sigma_2\sqrt{1-\rho^2}$$

$$x-\mu_1=\sigma_1 u,\qquad J=-\sigma_1\sigma_2\sqrt{1-\rho^2}$$
$$y-\mu_2=(t\sqrt{1-\rho^2}+\rho u)\sigma_2$$

$$COV(X,Y)=$$

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

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

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty}u^2e^{-\frac{u^2}{2}}du\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt+\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{2\pi}\int_{-\infty}^{\infty}ue^{-\frac{u^2}{2}}du\int_{-\infty}^{\infty}te^{-\frac{t^2}{2}}dt$$

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}\sqrt{2\pi}\cdot\sqrt{2\pi}+0=\rho\sigma_1\sigma_2$$

返回主目录

故 $\rho_{XY} = \rho$。

X，Y独立⇔ $\rho_{XY} = \rho$ =0⇔X，Y不相关。

注意，上面的结论只对于X,Y是服从二维正态分布时的情况成立，一般情况不成立。接下去的例3可以说明这一点。

【小结】五个充要条件：

$$\rho_{XY}=0 \Leftrightarrow Cov(X,Y)=0$$

$$\Leftrightarrow EXY=EX\cdot EY$$

$$\Leftrightarrow D(X+Y)=DX+DY$$

$$\Leftrightarrow D(X-Y)=DX+DY$$

若 X,Y 相互独立，则可推出上述五个结论；但反过来，不成立.

若 (X,Y) 服从二维正态分布，X,Y 相互独立 $\Leftrightarrow \rho_{XY}=0$.

一般来说，X,Y相互独立可以得出$\rho_{XY}=0$;

但反过来，$\rho_{XY}=0$并不能得出X,Y相互独立.

例3　设二维随机变量(X,Y)的密度函数为

$$f(x,y)=\frac{1}{2}[\varphi_1(x,y)+\varphi_2(x,y)],$$

其中$\varphi_1(x,y)$和$\varphi_2(x,y)$都是二维正态密度函数，且它们对应的二维随机变量的相关系数分别为$\frac{1}{3}$和$-\frac{1}{3}$，它们的边缘密度函数所对应的随机变量的数学期望都是零，方差都是1.

(1) 求随机变量 X 和 Y 的密度函数$f_X(x)$和$f_Y(y)$，及 X 和 Y 的相关系数

(2)问 X 和 Y 是否独立？为什么？

解　（1）由于二维正态密度函数的两个边缘密度都是正态密度函数，因此有

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy=\frac{1}{2}\left[\int_{-\infty}^{\infty}\varphi_1(x,y)dy+\int_{-\infty}^{\infty}\varphi_2(x,y)dy\right]$$

$$=\frac{1}{2}\left[\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}+\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right]=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}};$$

同理，

$$f_Y(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}};$$

则　$X\sim N(0,1), Y\sim N(0,1),$

所以　$EX=EY=0,\ DX=DY=1.$

随机变量 X 和 Y 的相关系数

$$\rho = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy$$

$$= \frac{1}{2}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\varphi_1(x,y)dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\varphi_2(x,y)dxdy\right]$$

$$= \frac{1}{2}\left[\frac{1}{3} - \frac{1}{3}\right] = 0.$$

（2）由题设

$$f(x,y) = \frac{3\sqrt{2}}{8\pi}\left[e^{-\frac{9}{16}(x^2 - \frac{2}{3}xy + y^2)} + e^{-\frac{9}{16}(x^2 + \frac{2}{3}xy + y^2)}\right],$$

$$f(x,y)=\frac{3\sqrt{2}}{8\pi}\left[e^{-\frac{9}{16}(x^2-\frac{2}{3}xy+y^2)}+e^{-\frac{9}{16}(x^2+\frac{2}{3}xy+y^2)}\right],$$

$$f_X(x)\cdot f_Y(y)=\frac{1}{2\pi}e^{-\frac{x^2}{2}}\cdot e^{-\frac{y^2}{2}}=\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}},$$

$$f(x,y)\neq f_X(x)\cdot f_Y(y).$$

所以 X 与 Y 不独立.

该例子说明，虽然$\rho_{XY}=0$，但X,Y并不相互独立.