§4.3 协方差及相关系数

一、协方差及相关系数的定义

定义. 如果$E(X)$, $E(Y)$ 以及 $E\{[X-E(X)][Y-E(Y)]\}$都存在，则称 $E\{[X-E(X)][Y-E(Y)]\}$为随机变量X 与Y的协方差，记为 $\mathrm{Cov}(X,Y)$ 。即

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

当$+\infty>\sqrt{D(X)}\sqrt{D(Y)}>0$ 时，称 $\rho_{XY}=\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$

为随机变量 X 与 Y 的相关系数.

由定义即知

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

一般情况下, 由上述定义及（2.5 ）式知:

$$D(X+Y)=D(X)+D(Y)+2\mathrm{Cov}(X,Y) \qquad (3.1)$$

$\mathrm{Cov}(X,Y)$ 的计算公式:

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y) \qquad (3.2)$$

协方差的性质:

1、$Cov(X,X)=DX$

2、$Cov(aX,bY)=abCov\left(X,Y\right)$. a,b 是常数

3、$Cov\ (X_1+X_2,Y)=\ Cov\ (X_1,Y)+\ Cov\ (X_2,Y)$

4、当 X, Y 相互独立时，$Cov\left(X,Y\right)=0$

定理 1^0 $|\rho_{XY}| \leq 1$

2^0 $|\rho_{XY}| = 1$的充要条件是,存在常数 $b \neq 0$ 及a , 使

$$P\{Y = a + bX\} = 1$$

证:充分性显然成立。考虑必要性，算一算

$E\{[Y - EY - b(X - EX)]^2\}$。由数学期望的性质得

$$0 \leq E\{[(Y - EY) - b(X - EX)]^2\}$$

$$= E(Y - EY)^2 + b^2 E(X - EX)^2 - 2bE[(Y - EY)(X - EX)]$$

$\therefore$ $D(Y) + b^2 D(X) - 2bCov(X, Y) \geq 0$ 对一切 b 成立.

$$\therefore \quad [Cov(X, Y)]^2 \leq D(Y)D(X)$$

所以　$a)\ [Cov(X,Y)]^2 \le D(Y)D(X)$

b）$|\rho_{XY}|=1 \Leftrightarrow [Cov(X,Y)]^2 = D(Y)D(X)$

$\Leftrightarrow$ 存在常数$b \ne 0$使得$D(Y)+b^2D(X)-2bCov(X,Y)=0$

$\Leftrightarrow$ 存在常数$b \ne 0$使得$E\{[Y-EY-b(X-EX)]^2\}=0$

$\Leftrightarrow$ 存在常数$b \ne 0$使得$P(Y=EY+b(X-EX))=1$

$\Leftrightarrow$ 存在常数$b \ne 0$及a使得$P(Y=bX+a)=1$

其中常数b是

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

1. $|\rho_{XY}| \leq 1.$

2. 当$|\rho_{XY}| = 1$时, X与Y有完全的线性关系,

即存在常数$b = \rho_{XY} \dfrac{\sqrt{D(Y)}}{\sqrt{D(X)}}$和$a = EY - bEX$,

使$P\{Y = bX + a\} = 1$。

性质：

3．当$|\rho_{XY}|$逐渐减小时, X与Y的线性关系越来越差; 当$|\rho_{XY}|=0$时, X与Y完全没有线性关系, 此时称X与Y不相关.

4．若X与Y相互独立, 则X与Y不相关。但反之未必成立。因此“不相关”不是“没有关系”！

例1. 设(X,Y)的联合分布律为

$X \backslash Y$	-2	-1	1	2
1	0	$\frac{1}{4}$	$\frac{1}{4}$	0
4	$\frac{1}{4}$	0	0	$\frac{1}{4}$

此时有 $X = Y^2$

而且 $E(Y) = 0, E(X) = 5/2, E(XY) = 0,$

于是$\rho_{XY} = 0,$

例2　设(X,Y)服从二维正态分布,它的概率密度如下，求X和Y的相关系数.

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

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

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

$$D(X)=\sigma_1^2,D(Y)=\sigma_2^2,\rho_{XY}=\rho$$

详细计算过程见111页。

作业：27（4）（5），31，32