4.3 协方差及相关系数

北京化工大学数学系

苏贵福

对于二维随机变量(X,Y), 我们除了讨论X与Y的数学期望和方差以外, 还需要讨论描述X与Y之间相互关系的数字特征.

本节讨论有关这方面的数学特征. 在第二节关于方差性质的证明中发现: 如果两个随机变量X和Y是相互独立的, 则

$$E\{[X-E(X)][Y-E(Y)]\}=0$$

这就意味着当$E\{[X-E(X)][Y-E(Y)]\}\neq 0$时, X与Y不相互独立. 那么此时它们之间存在着什么关系?

定义1 设(X,Y)为二维随机变量, 若$E\{[X-E(X)][Y-E(Y)]\}$存在, 则称它为X与Y的协方差, 记作

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

而

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

称为随机变量X与Y的相关系数.

定理1 设X与Y是随机变量, 则

(1) $D(X+Y)=D(X)+D(Y)+2Cov(X,Y)$.

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

(3) $Cov(X,Y)=Cov(Y,X)$.

(4) $Cov(aX,bY)=abCov(X,Y)$, 其中a,b为常数.

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

定理2 设X与Y是随机变量, 则

(1) $\rho_{XY} \leq 1$.

(2) 若X与Y相互独立, 则$\rho_{XY} = 0$, 并称X与Y不相关.

(3) $\rho_{XY} = 1 \Leftrightarrow X$与$Y$依概率1线性相关, 即存在两个常数$a$和$b \neq 0$, 使得$P\{Y = bX + a\} = 1$.

假设随机变量X与Y的相关系数ρ_{XY}存在. 当X和Y相互独立时有$\rho_{XY}=0$, 即X与Y不相关. 但当X和Y不相关时, X与Y并非相互独立.

"不相关"只是讨论随机变量的线性关系, 而"相互独立"考察随机变量之间的一般关系.

例1 设二维随机变量(X, Y)的联合分布律为

$Y \setminus X$	-2	-1	1	2	$P\{Y=i\}$
1	0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
4	$\frac{1}{4}$	0	0	$\frac{1}{4}$	$\frac{1}{2}$
$P\{X=i\}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

讨论X与Y的关系.

解 计算得$E(X)=0$, $E(Y)=\frac{5}{2}$, 而$E(XY)=0$. 因此$\rho_{XY}=0$, 即X和Y不相关, 也就是说它们不存在线性关系.

又$P\{X=-2, Y=1\}=0 \neq \frac{1}{16}=P\{X=-2\}P\{Y=1\}$, 说明$X$与$Y$不相互独立.

例2 设二维随机变量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2\le 1\\ 0, & \text{其他}\end{cases}$$

试验证X与Y不相关，但X与Y不是相互独立的.

解 已知在单位圆盘上有$f(x,y)=\frac{1}{\pi}$，其他地方有$f(x,y)=0$. 那么

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dx=\frac{2}{\pi}\sqrt{1-x^2}.$$

同理可得

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

根据期望的概念

$$E(X)=\int_{-\infty}^{+\infty} x f_X(x)dx=\int_{-1}^{1}\frac{2x}{\pi}\sqrt{1-x^2}dx=0.$$

$$E(Y)=\int_{-\infty}^{+\infty} y f_Y(y)dy=\int_{-1}^{1}\frac{2y}{\pi}\sqrt{1-y^2}dy=0.$$

另外XY的数学期望为

$$E(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy=\int_{-1}^{1}xdx\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}y\frac{1}{\pi}dy=0.$$

于是有

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 0.$$

故根据相关系数的概念

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = 0.$$

这说明随机变量X与Y不相关. 另外显然$f(x,y) \neq f_X(x)f_Y(y)$, 这说明X与Y不相互独立. ■

例3 设二维随机变量(X, Y)的概率密度为

$$f(x,y)=\begin{cases}1, & |y|<x, 0<x<1\\ 0, & \text{其他}\end{cases}$$

试求$Cov(X, Y)$.

解 已知X的概率密度为

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)dy=\int_{-x}^{x} 1dy=2x \quad (0<x<1).$$

于是有

$$E(X)=\int_{-\infty}^{\infty} xf_X(x)dx=\int_0^1 2x^2dx=\frac{2}{3}.$$

同理可知Y的概率密度为

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx=\begin{cases}\displaystyle\int_y^1 dx=1-y, & 0<y<1\\ \displaystyle\int_{-y}^1 dx=1+y, & -1<y\le 0\end{cases}$$

于是有

$$E(Y)=\int_{-\infty}^{\infty}yf_Y(y)dy=\int_{-1}^{0}y(1+y)dy+\int_0^1 y(1-y)dy=0.$$

又因为

$$E(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)dxdy=\int_0^1 xdx\int_{-x}^{x}ydy=0.$$

故$Cov(X,Y)=0.$ ■

例4 设二维随机变量(X, Y)服从正态分布, 其联合分密度为

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

试求X与Y的相关系数.

解 已知(X, Y)的边缘密度为

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \quad -\infty < x < +\infty$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}, \quad -\infty < y < +\infty.$$

故知$E(X)=\mu_1$, $E(Y)=\mu_2$, $D(X)=\sigma_1^2$, $D(Y)=\sigma_2^2$. 而

$$
\begin{aligned}
Cov(X,Y)=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_1)(y-\mu_2)f(x,y)dxdy\\
=&\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_1)(y-\mu_2)\\
&\times exp\left[\frac{-1}{2(1-\rho^2)}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)^2-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right]dydx.
\end{aligned}
$$

令$t=\frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)$, $u=\frac{x-\mu_1}{\sigma_1}$. 则有

$$
\begin{aligned}
Cov(X,Y) =& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_1\sigma_2\sqrt{1-\rho^2}tu+\rho\sigma_1\sigma_2u^2)e^{-\frac{u^2+t^2}{2}}dtdu \\
=&\frac{\rho\sigma_1\sigma_2}{2\pi}\left(\int_{-\infty}^{\infty}u^2e^{-\frac{u^2}{2}}du\right)\left(\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt\right) \\
+&\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{2\pi}\left(\int_{-\infty}^{\infty}ue^{-\frac{u^2}{2}}du\right)\left(\int_{-\infty}^{\infty}et^{-\frac{t^2}{2}}\right) \\
=&\frac{\rho\sigma_1\sigma_2}{2\pi}\cdot\sqrt{2\pi}\cdot\sqrt{2\pi}=\rho\sigma_1\sigma_2.
\end{aligned}
$$

于是

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

♠ 二维正态随机变量(X, Y)的概率密度中的参数ρ就是 X和Y的相关系数. 因而二维正态随机变量的分布完全可由X和Y各自的数学期望, 方差以及它们的相关系数所确定.

♠ 若(X, Y)服从二维正态分布, 则X与Y相互独立的 充要条件是$\rho = 0$. 现在证得$\rho = \rho_{XY}$. 因此对于二维正态随机变量(X, Y)而言: X与Y不相关$\Leftrightarrow$$X$与$Y$相互独立.