哈爾濱工業大學

第5章随机变量的数字特征与极限定理

第26讲 协方差和相关系数、矩

协方差

方差的性质3.

$$D(X+Y)=D(X)+D(Y)+2E\{[X-E(X)][Y-E(Y)]\};$$

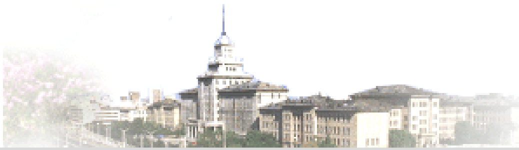
若X与Y独立，则 $D(X+Y)=D(X)+D(Y)$;

即 $E\{[X-E(X)][Y-E(Y)]\}=0$

等价于，若$E\{[X-E(X)][Y-E(Y)]\}\neq 0$，

则X与Y不独立.

X与Y的协方差



协方差

- 定义1 若$E\{[X-E(X)][Y-E(Y)]\}$存在，称它为随机变量X和Y的协方差, 记为Cov(X,Y), 即

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

此时

$$D(X\pm Y)=D(X)+D(Y)\pm 2\mathrm{Cov}(X,Y).$$

- 当Cov(X,Y)>0时，称X与Y正相关；
- 当Cov(X,Y)<0时，称X与Y负相关；
- 当Cov(X,Y)=0时，称X与Y不相关.

协方差的性质

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

1. $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$, $\mathrm{Cov}(X,a)=0$,

2. $D(X)=\mathrm{Cov}(X,X)$;

3. $\mathrm{Cov}(aX,bY)=ab\,\mathrm{Cov}(X,Y)$ a,b是常数;

4. $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$.

协方差的计算公式

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

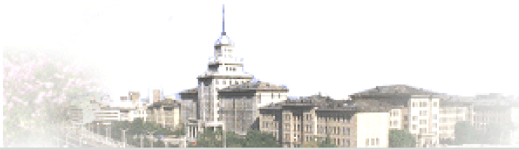
5. 若X与Y独立，则$\mathrm{Cov}(X,Y)=0$. 反之不成立.

协方差

练习

$\mathrm{Cov}(X+2Y, 3X-4)=?$

$$
\begin{aligned}
&=\mathrm{Cov}(X, 3X)+\mathrm{Cov}(X, -4)+\mathrm{Cov}(2Y, 3X)+\mathrm{Cov}(2Y, -4)\\
&=3D(X)+0+2\times 3\mathrm{Cov}(Y, X)+0\\
&=3D(X)+6\mathrm{Cov}(Y, X).
\end{aligned}
$$



例1 设随机变量X的概率密度为

$$f(x)=\frac{1}{2}e^{-|x|},\quad -\infty<x<+\infty,$$

试证：X与$|X|$不相关，但也不独立.

证 $\mathrm{Cov}(X,|X|)=E(X|X|)-E(X)E|X|$

$$E(X|X|)=\int_{-\infty}^{+\infty}x|x|\cdot\frac{1}{2}e^{-|x|}dx=0,$$

$$E(X)=\int_{-\infty}^{+\infty}x\cdot\frac{1}{2}e^{-|x|}dx=0,\quad \mathrm{Cov}(X,|X|)=0,$$

可得X与$|X|$不相关.

例1 设随机变量X的概率密度为

$$f(x)=\frac{1}{2}e^{-|x|},\quad -\infty<x<+\infty,$$

试证：X与$|X|$不相关，但也不独立.

对任意常数a有：

$$\{|X|\le a\}\subset\{X\le a\}\text{且}P(X\le a)<1,$$

$$P(X\le a,|X|\le a)=P(|X|\le a)>P(|X|\le a)P(X\le a).$$

说明X与$|X|$不独立.

相关系数

■定义2 若$\mathrm{Cov}(X,Y)$存在，且$D(X)>0$, $D(Y)>0$,

称

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

为随机变量X和Y的相关系数．

在不致引起混淆时，记ρ_{XY}为ρ．

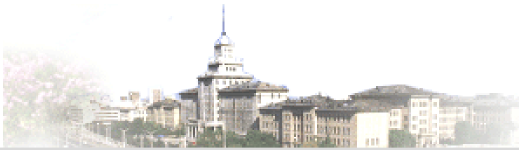
相关系数的性质

1. $|\rho| \le 1$;

2. $|\rho| = 1 \Leftrightarrow$ 存在常数a, b，使$P(Y = a + bX) = 1$.

$\rho = 1$时，$b > 0$; $\rho = -1$时，$b < 0$.

- 相关系数是表示两个随机变量之间线性相关程度的一个数字特征. (无量纲)
- 协方差也是表示两个随机变量之间线性相关程度的一个数字特征. (有量纲)



相关系数的含义

$|\rho|$的值越接近于1, Y与X的线性相关程度越高;

$|\rho|$的值越接近于0, Y与X的线性相关程度越弱;

当$\rho=0$,称X与Y不相关，说明X与Y之间无线性关系.

X与Y独立 $\begin{array}{c}\longrightarrow\\ \not\longleftarrow\end{array}$ X与Y不相关

当相关系数存在时，有

$$\rho=0 \Leftrightarrow \mathrm{Cov}(X,Y)=0 \Leftrightarrow E(XY)=E(X)E(Y)$$

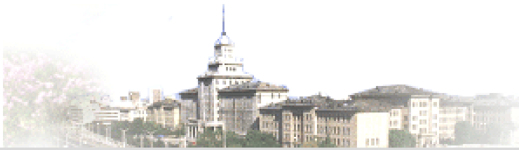
$$\Leftrightarrow D(X\pm Y)=D(X)+D(Y).$$

例2 某箱装有100件产品，其中一、二和三等品分别为80、10和10件，现在从中随机抽取一件，记

$$X_i=\begin{cases}1, & \text{抽到}i\text{等品}, \\ 0, & \text{其他}.\end{cases} \quad (i=1,2,3).$$

求（1）(X_1,X_2)的分布列及边缘分布列；

（2）X_1与X_2的相关系数 ρ .



(2) 在第四章中我们已经求得了(X_1,X_2)的分布列及边缘分布列为

$X_2 \backslash X_1$	0	1	$p_{\cdot j}$
0	0.1	0.8	0.9
1	0.1	0	0.1
$p_{i\cdot}$	0.2	0.8	1

$$E(X_1)=0.8, E(X_2)=0.1$$

$$D(X_1)=0.8\times 0.2=0.16,$$

$$D(X_2)=0.1\times 0.9=0.09.$$

$$E(X_1X_2)=0\times 0\times 0.1+0\times 1\times 0.1$$

$$+1\times 0.8\times 0+1\times 1\times 0=0,$$

$$\mathrm{Cov}(X_1,X_2)=E(X_1X_2)-E(X_1)E(X_2)=-0.08.$$

$$\rho=\frac{\mathrm{Cov}(X_1,X_2)}{\sqrt{D(X_1)}\sqrt{D(X_2)}}=\frac{-0.08}{\sqrt{0.16}\sqrt{0.09}}=-\frac{2}{3}.$$

例3. 设r,v X，Y的相关系数为 $1/\sqrt{3}$ ，X在[0，1]上服从均匀分布，Y 服从参数λ=1的指数分布，$Z=2X+Y$ ， 求 $E(Z),D(Z),\rho_{YZ}$

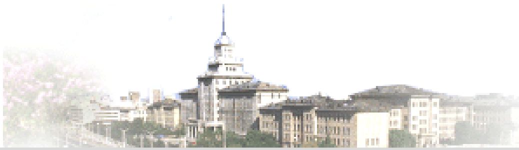
解

$$X \sim U[0,1], E(X)=1/2, D(X)=1/12$$

$$Y \sim E(1), E(Y)=D(Y)=1$$

$$E(Z)=2E(X)+E(Y)=2,$$

$$D(Z)=4D(X)+D(Y)+4\rho\sqrt{D(X)D(Y)}=2,$$



例3. 设r,v X，Y的相关系数为 $1/\sqrt{3}$，X在[0，1]上服从均匀分布，Y 服从参数$\lambda=1$的指数分布，$Z=2X+Y$，求 $E(Z),D(Z),\rho_{YZ}$

解

$$\rho_{YZ}=\frac{Cov(Y,Z)}{\sqrt{D(Y)D(Z)}}=\frac{Cov(Y,2X+Y)}{\sqrt{D(Y)D(Z)}}$$

$$=\frac{2Cov(Y,X)+D(Y)}{\sqrt{D(Y)D(Z)}}$$

$$=\frac{2\sqrt{2}}{3}$$

$$E(X)=1/2,$$

$$D(X)=1/12$$

$$E(Y)=D(Y)=1$$

$$E(Z)=D(Z)=2,$$

设$(X,Y) \sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;\rho)$,可以求得$X$与$Y$的协方差为

$$\mathrm{Cov}(X,Y) = E\{[X - E(X)][Y - E(Y)]\} = \sigma_1\sigma_2\rho.$$

则 $\rho_{XY} = \rho.$

- 当(X,Y)服从二维正态分布时，有

  X与Y独立 $\Leftrightarrow$ X与Y不相关

一般情况下，独立与不相关并不是等价的.

原点矩、中心矩

- **定义3** 若$E(X^k)(k=1,2,\ldots)$存在，则称$E(X^k)$为X的**k阶原点矩**，记为$\alpha_k=E(X^k)$；

  若$E[X-E(X)]^k\ (k=1,2,\ldots)$存在，则称$E[X-E(X)]^k$为$X$的**$k$阶中心矩**，记为

$$\beta_k=E[X-E(X)]^k;$$

$E(X)$为1阶原点矩；$D(X)$为2阶中心矩.

原点矩、中心矩

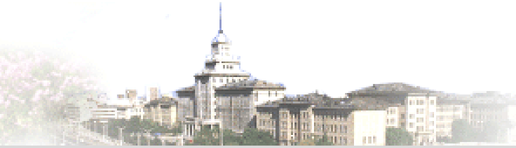
- **定义4** 若$E(X^kY^l)(k, l=1,2, \ldots)$存在，则称$E(X^kY^l)$为$X$和$Y$的**$k+l$阶混合原点矩**，记为 $\alpha_{k,l}=E(X^kY^l)$;

  若$E\{[X-E(X)]^k[Y-E(Y)]^l\}(k,l=1,2, \ldots)$存在，则称

$E\{[X-E(X)]^k[Y-E(Y)]^l\}$为$X$与$Y$的**$k+l$阶混合中心矩**，记为

$$\beta_{k,l}=E\{[X-E(X)]^k[Y-E(Y)]^l\}.$$

- 协方差为1+1阶混合中心矩.



谢　　谢！