

## 第三节 协方差及相关系数

- 协方差
- 相关系数
- 课堂练习
- 小结 布置作业

前面我们介绍了随机变量的数学期望和方差，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，还要讨论描述  $X$  和  $Y$  之间关系的数字特征，这就是本讲要讨论的

## 协方差和相关系数



# 一、协方差

1.定义 量 $E\{[X-E(X)][Y-E(Y)]\}$ 称为随机变量 $X$ 和 $Y$ 的协方差,记为 $Cov(X,Y)$  , 即

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

## 2.简单性质

$$(1) \quad Cov(X,Y)=Cov(Y,X)$$

$$(2) \quad Cov(aX,bY) = ab Cov(X,Y) \quad a,b \text{ 是常数}$$

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



### 3. 计算协方差的一个简单公式

由协方差的定义及期望的性质，可得

$$\begin{aligned}Cov(X,Y) &= E\{[X-E(X)][Y-E(Y)]\} \\&= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\&= E(XY) - E(X)E(Y)\end{aligned}$$

即

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

可见，若  $X$  与  $Y$  独立，  $Cov(X,Y) = 0$  .



特别地

$$\text{Cov}(X, X) = E(X^2) - E(X)^2 = D(X)$$

#### 4. 随机变量和的方差与协方差的关系

$$D(X+Y) = D(X) + D(Y) + 2\text{Cov}(X, Y)$$



例1 已知离散型随机向量 $(X, Y)$ 的概率分布为

$X \backslash Y$	-1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

求 $\text{cov}(X, Y)$

解 计算 $E(X)$ ,

$X \backslash Y$	-1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

$x_i p_{ij}$

0	0	0
0.3	0.05	0.1
0.3	0	0.2

求和

$$E(X)=0.95$$



计算 $E(Y)$ ,

$X \backslash Y$	-1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

$y_j p_{ij}$

-0.1	0	0
-0.3	0	0.2
-0.15	0	0.2

求和



$$E(Y) = -0.15$$





计算 $E(XY)$ ,

$X \backslash Y$	-1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

$x_i y_j p_{ij}$

0	0	0
-0.3	0	0.2
-0.3	0	0.4

求和



$$E(XY)=0$$



$$E(X)=0.95, E(Y)=-0.15, E(XY)=0$$

于是

$$\begin{aligned}\text{cov}(X,Y) &= E(XY) - E(X)E(Y) \\ &= 0.95 \times 0.15 = 0.1425\end{aligned}$$



例2 设连续型随机变量 $(X, Y)$ 的密度函数为

$$f(x, y) = \begin{cases} 8xy, & 0 \leq x \leq y \leq 1 \\ 0, & \text{其它} \end{cases}$$

求 $\text{cov}(X, Y)$ 和 $D(X+Y)$ .

解: 由 $(X, Y)$ 的密度函数求边缘密度函数

$$f_X(x) = \begin{cases} 4x(1-x^2), & 0 \leq x \leq 1, \\ 0, & \text{其它.} \end{cases}$$



$$f_Y(y) = \begin{cases} 4y^3, & 0 \leq y \leq 1, \\ 0, & \text{其它.} \end{cases}$$

于是

$$E(X) = \int_{-\infty}^{+\infty} xf_X(x) dx = \int_0^1 x \cdot 4x(1-x^2) dx = 8/15$$

$$E(Y) = \int_{-\infty}^{+\infty} yf_Y(y) dy = \int_0^1 y \cdot 4y^3 dy = 4/5$$

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y) dx dy \\ &= \int_0^1 dx \int_x^1 xy \cdot 8xy dy = 4/5 \end{aligned}$$



从而得  $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$

$$= \frac{4}{9} - \frac{4}{5} \cdot \frac{8}{15} = \frac{4}{225}.$$

又

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_0^1 x^2 4x(1-x^2) dx = \frac{1}{3}$$

$$E(Y^2) = \int_{-\infty}^{+\infty} y^2 f_Y(y) dy = \int_0^1 y^2 4y^3 dy = \frac{2}{3}$$



所以

$$D(X) = E(X^2) - [E(X)]^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}$$

$$D(Y) = E(Y^2) - [E(Y)]^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75}$$

故

$$\begin{aligned} D(X + Y) &= D(X) + D(Y) + 2\text{cov}(X, Y) \\ &= \frac{11}{225} + \frac{2}{75} + \frac{8}{225} = \frac{1}{9}. \end{aligned}$$



协方差的大小在一定程度上反映了 $X$ 和 $Y$ 相互间的关系，但它还受 $X$ 与 $Y$ 本身度量单位的影响。例如：

$$\text{Cov}(kX, kY) = k^2 \text{Cov}(X, Y)$$

为了克服这一缺点，对协方差进行标准化，这就引入了相关系数。



## 二、相关系数

定义： 设 $D(X)>0, D(Y)>0$ ， 称

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

为随机变量  $X$  和  $Y$  的相关系数.

在不致引起混淆时， 记  $\rho_{XY}$  为  $\rho$ .





相关系数的性质:

$$1. |\rho| \leq 1$$

证: 由方差的性质和协方差的定义知,

对任意实数  $b$

$$0 \leq D(Y - bX) =$$

$$\text{令 } b = \frac{\text{Cov}(X, Y)}{D(X)}$$

由于方差  $D(Y)$  是正的, 故必有  
 $1 - \rho^2 \geq 0$ , 所以  $|\rho| \leq 1$ 。

$$D(Y - bX) = D(Y) - \frac{[\text{Cov}(X, Y)]^2}{D(X)}$$

$$= D(Y) \left[ 1 - \frac{[\text{Cov}(X, Y)]^2}{D(X)D(Y)} \right] = D(Y)[1 - \rho^2]$$



2.  $X$ 和 $Y$ 独立时,  $\rho = 0$ , 但其逆不真.

由于当 $X$ 和 $Y$ 独立时,  $Cov(X, Y) = 0$ .

故 
$$\rho = \frac{Cov(X, Y)}{\sqrt{D(X)D(Y)}} = 0$$

但由  $\rho = 0$  并不一定能推出 $X$ 和 $Y$ 独立.

请看下例.



例3 设 $X$ 服从 $(-1/2, 1/2)$ 内的均匀分布, 而 $Y=\cos X$ .

$X$ 的密度函数

$$f(x) = \begin{cases} 1 & -\frac{1}{2} < x < \frac{1}{2} \\ 0 & \text{其它} \end{cases} \quad \text{可得 } E(X) = 0$$

$$E(XY) = E(X \cos X) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x \cos x f(x) dx = 0$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$$

因而  $\rho=0$ , 即 $X$ 和 $Y$ 不相关.

但 $Y$ 与 $X$ 有严格的函数关系, 即 $X$ 和 $Y$ 不独立.



$$3. |\rho| = 1$$

↔ 存在常数  $a, b (b \neq 0)$  , 使  $P\{Y = a + bX\} = 1$ ,

即  $X$  和  $Y$  以概率 1 线性相关.



相关系数刻划了 $X$ 和 $Y$ 间“线性相关”的程度.

考虑以 $X$ 的线性函数 $a+bX$ 来近似表示 $Y$ ,

以均方误差

$$e = E\{[Y-(a+bX)]^2\}$$

来衡量以  $a + b X$  近似表示  $Y$  的好坏程度：

$e$  值越小表示  $a + bX$  与  $Y$  的近似程度越好.

用微积分中求极值的方法，求出使 $e$ 达到最小时的 $a, b$



$$\begin{aligned}
 e &= E\{[Y-(a+bX)]^2\} \\
 &= E(Y^2) + b^2 E(X^2) + a^2 - 2bE(XY) + 2abE(X) \\
 &\quad - 2aE(Y)
 \end{aligned}$$

$$\begin{cases} \frac{\partial e}{\partial a} = 2a + 2bE(X) - 2E(Y) = 0 \\ \frac{\partial e}{\partial b} = 2bE(X^2) - 2E(XY) + 2aE(X) = 0 \end{cases}$$

解得

$$\begin{cases} b_0 = \frac{Cov(X, Y)}{D(X)} \\ a_0 = E(Y) - b_0 E(X) \end{cases}$$

这样求出的  
最佳逼近为

$$L(X) = a_0 + b_0 X$$



这样求出的最佳逼近为 $L(X)=a_0+b_0X$

这一逼近的剩余是

$$E[(Y-L(X))^2]=D(Y)(1-\rho^2)$$

可见, 若  $\rho = \pm 1$ ,  $Y$  与  $X$  有严格线性关系;

若  $\rho = 0$ ,  $Y$  与  $X$  无线性关系;

若  $0 < |\rho| < 1$ ,

$|\rho|$  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.



前面，我们已经看到：

若  $X$  与  $Y$  独立，则  $X$  与  $Y$  不相关，

但由  $X$  与  $Y$  不相关，不一定能推出  $X$  与  $Y$  独立。

但对下述情形，独立与不相关等价

若  $(X, Y)$  服从二维正态分布，则

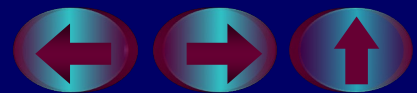
$X$  与  $Y$  独立  $\Leftrightarrow X$  与  $Y$  不相关



例4 设 $(X, Y)$ 服从二维正态分布, 它的概率密度为

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

求 $X$ 和 $Y$ 的相关系数 $\rho_{XY}$ .



解:

$$\text{cov}(X, Y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_1)(y - \mu_2) \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[ \frac{(x - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x - \mu_1)(y - \mu_2)}{\sigma_1\sigma_2} + \frac{(y - \mu_2)^2}{\sigma_2^2} \right] \right\} dx dy$$

在上式中令

$$t = \frac{x - \mu_1}{\sigma_1}, u = \frac{y - \mu_2}{\sigma_2}, dx = \sigma_1 dt, dy = \sigma_2 du$$



则

$$\text{cov}(X, Y) = \frac{\sigma_1 \sigma_2}{2\pi \sqrt{1 - \rho^2}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} t u e^{\frac{-1}{2(1-\rho^2)}(t^2 - 2\rho t u + u^2)} dt du$$

而指数上的

$$\begin{aligned} t^2 - 2\rho t u + u^2 &= t^2 - 2\rho t u + (\rho u)^2 + u^2 - (\rho u)^2 \\ &= (t - \rho u)^2 + u^2(1 - \rho^2), \text{ 因此} \end{aligned}$$

$$\begin{aligned} &\frac{-1}{2(1-\rho^2)} \left[ (t - \rho u)^2 + u^2(1 - \rho^2) \right] \\ &= \frac{-(t - \rho u)^2}{2(1 - \rho^2)} - \frac{u^2}{2} \end{aligned}$$



因此,

$$\text{cov}(X, Y) = \frac{\sigma_1 \sigma_2}{2\pi \sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} t u e^{\frac{-1}{2(1-\rho^2)}(t^2 - 2\rho t u + u^2)} dt du$$

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

$$= \frac{\sigma_1 \sigma_2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u(u\rho) e^{-\frac{u^2}{2}} du = \rho \sigma_1 \sigma_2$$

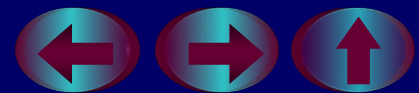
于是

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



因此二维正态随机变量 $(X, Y)$ 的概率密度中的参数 $\rho$ 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 $X, Y$ 各自的数学期望, 方差以及它们的相关系数所确定.

注: 若 $(X, Y)$ 服从二维正态分布, 则 $X$ 与 $Y$ 相互独立, 当且仅当 $X$ 与 $Y$ 不相关.



### 三、 原点矩      中心矩

定义    设 $X$ 和 $Y$ 是随机变量，若

$$E(X^k), k = 1, 2, \dots$$

存在，称它为 $X$ 的 $k$ 阶原点矩，简称  $k$ 阶矩

若  $E\{[X - E(X)]^k\}, k = 2, 3, \dots$

存在，称它为 $X$ 的 $k$ 阶中心矩

可见，均值  $E(X)$  是 $X$ 一阶原点矩，方差 $D(X)$ 是 $X$ 的二阶中心矩。



设  $X$  和  $Y$  是随机变量，若

$$E(X^k Y^L) \quad k, L=1, 2, \dots \quad \text{存在,}$$

称它为  $X$  和  $Y$  的  $k+L$  阶混合（原点）矩.

若  $E\{[X - E(X)]^k [Y - E(Y)]^L\}$  存在,

称它为  $X$  和  $Y$  的  $k+L$  阶混合中心矩.

可见,

协方差  $Cov(X, Y)$  是  $X$  和  $Y$  的二阶混合中心矩.



## 四、课堂练习

1、设随机变量  $(X, Y)$  具有概率密度

$$f(x, y) = \begin{cases} \frac{1}{8}(x + y) & 0 \leq x \leq 2, 0 \leq y \leq 2 \\ 0 & \text{其它} \end{cases}$$

求  $E(X), E(Y), Cov(X, Y), D(X + Y)$ 。

2、设  $X \sim N(\mu, \sigma^2), Y \sim N(\mu, \sigma^2)$ , 且设  $X, Y$  相互独立  
试求  $Z_1 = \alpha X + \beta Y$  和  $Z_2 = \alpha X - \beta Y$  的相关系数(其中  $\alpha, \beta$  是不全为零的常数)。





$$1、\text{解 } E(X) = E(Y) = \frac{7}{6}, Cov(X, Y) = -\frac{1}{36}, D(X + Y) = \frac{5}{9}$$

$$2、\text{解 } D(X) = D(Y) = \sigma^2$$

$$D(Z_1) = D(\alpha X + \beta Y) = \alpha^2 D(X) + \beta^2 D(Y) = (\alpha^2 + \beta^2) \sigma^2$$

$$D(Z_2) = D(\alpha X - \beta Y) = \alpha^2 D(X) + \beta^2 D(Y) = (\alpha^2 + \beta^2) \sigma^2$$

$$Cov(Z_1, Z_2) = Cov(\alpha X + \beta Y, \alpha X - \beta Y)$$

$$= \alpha^2 Cov(X, X) - \beta^2 Cov(Y, Y) = \alpha^2 D(X) - \beta^2 D(Y)$$

$$= (\alpha^2 - \beta^2) \sigma^2$$

$$\rho_{Z_1 Z_2} = \frac{Cov(Z_1, Z_2)}{\sqrt{D(Z_1)} \sqrt{D(Z_2)}} = \frac{\alpha^2 - \beta^2}{\alpha^2 + \beta^2}$$



## 五、小结

这一节我们介绍了协方差、相关系数、  
相关系数是刻画两个变量间线性相关程度的一个重要的数字特征.

注意独立与不相关并不是等价的.

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



## 六、 布置作业

习题4-3 (p110) 3、 6、 10

