§4.3 协方差和相关系数

- 协方差的定义
- 协方差的计算
- 协方差的性质
- 相关系数的定义
- 相关系数的性质

协方差的定义

定义3 设(X,Y)是一个二维随机变量，若 $E\{[X-E(X)][Y-E(Y)]\}$存在，

则称其为X与Y的协方差，记为$\mathrm{cov}(X,Y)$，

即$\mathrm{cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$。 (4.3.1)

$$=E[XY-E(X)Y-XE(Y)+E(X)E(Y)]$$

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

$$\mathrm{cov}(X,Y)=E(XY)-E(X)E(Y) \qquad (4.3.2)$$

协方差的性质:

(1).$\operatorname{cov}(X,Y)=\operatorname{cov}(Y,X)$;$\operatorname{cov}(X,X)=D(X)$;

(2).设 X ,Y是 $r.v.$, a ,b 为常数，则

$$\operatorname{cov}(aX,bY)=ab\operatorname{cov}(X,Y);$$

(3).设X和 Y是两个 $r.v.$, 且相互独立,

则 $\operatorname{cov}(X,Y)=0$;

$(4).\operatorname{cov}(X_1+X_2,Y)=\operatorname{cov}(X_1,Y)+\operatorname{cov}(X_2,Y);$

$(5).D(aX+bY)=a^2D(X)+b^2D(Y)+2ab\operatorname{cov}(X,Y)$。

例11 设$r.v.X\sim E(2)$, $Y\sim N(1,4)$, $\operatorname{cov}(X,Y)=\frac{1}{2}$,

求$\operatorname{cov}(2X-3Y,X+Y)$。

解: $\operatorname{cov}(2X-3Y,X+Y)$

$=2\operatorname{cov}(X,X)-3\operatorname{cov}(Y,X)+2\operatorname{cov}(X,Y)-3\operatorname{cov}(Y,Y)$

$=2D(X)-\operatorname{cov}(X,Y)-3D(Y)$

$=2\times\frac{1}{2^2}-\frac{1}{2}-3\times4=-12.$

例12 设$X_1, X_2, \cdots X_{10}$相互独立， 且都服从$P(\lambda)$,

令$X = \sum_{i=1}^{7} X_i, Y = \sum_{i=4}^{10} X_i$, 求 $\mathrm{cov}(X, Y)$。

解：

$$\mathrm{cov}(X, Y) = \mathrm{cov}\left(\sum_{i=1}^{3} X_i + \sum_{i=4}^{7} X_i, \sum_{i=4}^{7} X_i + \sum_{i=8}^{10} X_i\right)$$

$$= \mathrm{cov}\left(\sum_{i=1}^{3} X_i, \sum_{i=4}^{7} X_i\right) + \mathrm{cov}\left(\sum_{i=1}^{3} X_i, \sum_{i=8}^{10} X_i\right)$$

$$+ \mathrm{cov}\left(\sum_{i=4}^{7} X_i, \sum_{i=4}^{7} X_i\right) + \mathrm{cov}\left(\sum_{i=4}^{7} X_i, \sum_{i=8}^{10} X_i\right)$$

$$= 0 + 0 + D(\sum_{i=4}^{7} X_i) + 0 = \sum_{i=4}^{7} D(X_i) = 4\lambda.$$

相关系数的定义

定义 4 设(X,Y)是一个二维随机变量， 称

$$\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$$为X与Y的相关系数，记为ρ_{XY}，

即$$\boxed{\rho_{XY} = \frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}}。$$

相关系数的性质

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

(2). $|\rho_{XY}| = 1 \Leftrightarrow \exists$ 常数 a, b, 使 $P\{Y = aX + b\} = 1$.

证明: (1).考虑实变量 t 的二次函数

$$g(t) = D(tX + Y)$$

$$= t^2 D(X) + D(Y) + 2t \cdot \operatorname{cov}(X, Y)$$

二次函数 $g(t) \ge 0 \Longleftrightarrow \Delta \le 0$

即 $\Delta = 4\operatorname{cov}(X, Y)^2 - 4D(X)D(Y) \le 0$

即 $\dfrac{\operatorname{cov}(X,Y)^2}{D(X)D(Y)} \le 1$，　即 $|\rho_{XY}| \le 1.$

(2). $|\rho_{XY}| = 1 \iff \Delta = 0 \iff$ 方程 $g(t) = 0$ 有重根 t_0

$$g(t_0) = D(t_0 X + Y) = 0 \iff P\{t_0 X + Y = C\} = 1.$$

性质(2)说明当 $|\rho_{XY}| = 1$ 时，X 与 Y 概率为1的存在线性关系。

当X与Y有线性关系，即$Y=aX+b$，　则$|\rho_{XY}|=1$,

具体地，$\rho_{XY}=\begin{cases}1, & a>0\\ -1, & a<0\end{cases}$

定义5　当$\rho_{XY}=0$时，称X与Y不相关。

1°若X与Y相互独立，则$\rho_{XY}=0$，即X与Y不相关。

反之，若X与Y不相关，则X与Y不一定相互独立。

例13 设连续型随机变量 X的密度函数为

$$f(x)=\begin{cases}\frac{3}{2}x^2, -1<x<1\\ 0, \quad \text{其他}\end{cases}, \quad Y=X^2,$$

求 (1).$E(X), D(X)$;　(2).$\mathrm{cov}(X,Y)$, X与Y是否相关?

(3).X与Y是否相互独立,为什么?

解: (1).$E(X)=\int_{-1}^{1}x\cdot\frac{3}{2}x^2dx=0$

$$E(Y)=E(X^2)=\int_{-1}^{1}x^2\cdot\frac{3}{2}x^2dx=\frac{3}{5}$$

$$D(X) = E(X^2) - E^2(X) = \tfrac{3}{5} - 0 = \tfrac{3}{5}$$

$$(2).\ E(XY) = E(X^3) = \int_{-1}^{1} x^3 \cdot \tfrac{3}{2} x^2 dx = 0$$

$$\operatorname{cov}(X,Y) = E(XY) - E(X)E(Y) = 0$$

所以 X 与 Y 不相关；

(3). X 与 Y 不相互独立。

假设 X 与 Y 相互独立， 在 X^2 与 Y 也相互独立，

必有 $E(X^2Y) = E(X^2)E(Y)$

$$E(X^2Y) = E(X^4) = \int_{-1}^{1} x^4 \cdot \frac{3}{2} x^2 dx = \frac{3}{7}$$

$E(X^2)E(Y) = \frac{3}{5} \times \frac{3}{5} = \frac{9}{25} \neq \frac{3}{7}$，矛盾，

X与Y不相互独立。

2°.对二维正态分布的随机变量(X,Y),

X与Y不相关 $\Leftrightarrow$ X与Y相互独立。

$$\rho = 0$$

二维$r.v.(X,Y) \sim N(\mu_1,\mu_2,\sigma_1{}^2,\sigma_2{}^2,\rho)$,

$X \sim N(\mu_1,\sigma_1{}^2)$, $\quad Y \sim N(\mu_2,\sigma_2{}^2)$, $\qquad \rho_{XY} = \rho$

$$\mathrm{cov}(X,Y) = E[(X-\mu_1)(Y-\mu_2)] = \rho\sigma_1\sigma_2$$

$$\operatorname{cov}(X,Y)=E[(X-\mu_1)(Y-\mu_2)]$$

$$=\int_{-\infty}^{+\infty}dx\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)f(x,y)dy$$

$$=\int_{-\infty}^{+\infty}dx\int_{-\infty}^{+\infty}\frac{(x-\mu_1)(y-\mu_2)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

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

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

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

$$=\int_{-\infty}^{+\infty}\sigma_1 dv\int_{-\infty}^{+\infty}\frac{\cancel{\sigma_1}v(\sqrt{1-\rho^2}u+\rho v)\sigma_2}{2\pi\cancel{\sigma_1}\cancel{\sigma_2}\sqrt{1-\cancel{\rho^2}}}\cdot e^{-\frac{v^2}{2}}e^{-\frac{u^2}{2}}\cancel{\sigma_2}\sqrt{1-\cancel{\rho^2}}\,du$$

$$=\frac{\sigma_1\sigma_2}{2\pi}\left[\sqrt{1-\rho^2}\int_{-\infty}^{+\infty}ve^{-\frac{v^2}{2}}dv\int_{-\infty}^{+\infty}ue^{-\frac{u^2}{2}}du+\rho\int_{-\infty}^{+\infty}e^{-\frac{u^2}{2}}du\int_{-\infty}^{+\infty}v^2e^{-\frac{v^2}{2}}dv\right]=\rho\sigma_1\sigma_2$$

§4.4 其他数字特征

矩

定义 6 设X为随机变量，如果 $E|X|^k<+\infty$，$k=1,2,\cdots$，则称$E(X^k)$为随机变量X的k阶原点矩，称$E[(X-E(X))^k]$为X的k阶中心矩。

期望$E(X)$是一阶原点矩；方差$D(X)$是二阶中心矩。

例14 设随机变量$X \sim E(\lambda)$，求X的三阶原点矩。

解：

$$E(X^3) = \int_0^{+\infty} x^3 \cdot \lambda e^{-\lambda x} dx$$

$$= \int_0^{+\infty} x^3 d(-e^{-\lambda x})$$

$$= \frac{6}{\lambda^3}.$$

例15 设随机变量$X\sim N(0,\sigma^2)$,求X的k阶原点矩。

解: $E(X^{2k-1})=\int_{-\infty}^{+\infty}x^{2k-1}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}dx=0,\quad k=1,2,\cdots$

$$E(X^{2k})=\int_{-\infty}^{+\infty}x^{2k}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}dx$$

$$=2\int_{0}^{+\infty}x^{2k}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}dx\ =2\int_{0}^{+\infty}x^{2k-1}\cdot\frac{\sigma^2}{\sqrt{2\pi}\sigma}d(-e^{-\frac{x^2}{2\sigma^2}})$$

$$=(2k-1)\sigma^2\cdot 2\int_{0}^{+\infty}x^{2k-2}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}dx\ =\sigma^{2k}(2k-1)!!.$$

协方差矩阵

定义7 设$X_1, X_2, \cdots, X_n$为n个随机变量，令$\sigma_{ij} = Cov(X_i, X_j)$,

其中$i, j = 1,2,\cdots,n$, 称矩阵$\Sigma = (\sigma_{ij})_{n\times n}$

为$X_1, X_2, \cdots, X_n$的协方差矩阵。

协方差矩阵的性质：

(1) 协方差矩阵为对称阵

(2) 协方差矩阵为半正定阵

对任意的非零向量$\mathbf{a} = (a_1, a_2, \cdots, a_n)$,都有 $\mathbf{a}^T \sum \mathbf{a} \geq 0.$

设$Y=\sum_{i=1}^{n}a_iX_i$, 则$D(Y)=\operatorname{cov}(\sum_{i=1}^{n}a_iX_i,\sum_{j=1}^{n}a_jX_j)$

$$
\begin{aligned}
&=\sum_{i=1}^{n}a_i\operatorname{cov}(X_i,\sum_{j=1}^{n}a_jX_j)\\
&=\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_j\operatorname{cov}(X_i,X_j)\\
&=\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_j\sigma_{ij}\quad=\mathbf{a}^T\Sigma\mathbf{a}\ge 0.
\end{aligned}
$$

§ 5.5 综合例题

例15 设二维随机变量 $(X,Y)\sim N(1,1,2^2,3^2,0.5)$，

求：(1). $\mathrm{cov}(X,Y), E(XY), \mathrm{cov}(X,2X-3Y), D(X-2Y)$;

(2)求k使得$X+kY$与$X-Y$相互独立；

(3).求 $Z=3X-2Y$的分布；

(4).求概率 $P(3X-2Y>7)$。

解： (1). $\mathrm{cov}(X,Y)=\rho\sigma_1\sigma_2=0.5\times2\times3=3$

$$E(XY)=\mathrm{cov}(X,Y)+E(X)E(Y)$$

$$=3+1\times1=4$$

$$\mathrm{cov}(X,2X-3Y)=2D(X)-3\mathrm{cov}(X,Y)$$

$$=2\times 2^2-3\times 3=-1$$

$$D(X-2Y)=D(X)+D(-2Y)-2\mathrm{cov}(X,2Y)$$

$$=2^2+4\times 3^2-2\times 2\times 3=28$$

(2).若二维 $r.v.(X,Y)\sim N(\mu_1,\mu_2,\sigma_1{}^2,\sigma_2{}^2,\rho)$,则

① $aX+bY\sim N(a\mu_1+b\mu_2,D(aX+bY))$.

② $\begin{cases} Z=aX+bY \\ W=cX+dY \end{cases}$,则 $(Z,W)\sim$

(2)求k使得$X+kY$与$X-Y$相互独立;

$$N(a\mu_1 + b\mu_2, c\mu_1 + d\mu_2, D(Z), D(W), \rho_{ZW}).$$

$X+kY$与$X-Y$相互独立 $\Longleftrightarrow$ $X+kY$与$X-Y$不相关

$$\operatorname{cov}(X+kY, X-Y)$$

$$= D(X) + k\operatorname{cov}(X,Y) - \operatorname{cov}(X,Y) - kD(Y)$$

$$= 2^2 + (k-1)\times 3 - k\times 3^2$$

$$= 1 - 6k = 0$$

$k=\frac{1}{6}$时，$X+kY$与$X-Y$相互独立。

（3） $E(Z)=3E(X)-2E(Y)=3\times 1-2\times 1=1$

$$D(Z)=D(3X-2Y)$$

$$=9D(X)+D(-2Y)-2\operatorname{cov}(3X,2Y)$$

$$=9\times 2^2+4\times 3^2-2\times 3\times 2\times 3=36$$

$$Z\sim N(1,36)$$

(4). $P(3X-2Y>7)=P\left(\dfrac{3X-2Y-1}{6}>\dfrac{7-1}{6}\right)$

$$=1-\Phi(1)=1-0.8413=0.1587$$

(3).求 $Z=3X-2Y$ 的分布；

例16 设二维随机变量$(X,Y)\sim N(1,1,2^2,3^2,0.5)$，求$E\max\{X,Y\}$。

解： $\max\{X,Y\}=\frac{1}{2}(X+Y+|X-Y|)$

$$E\max\{X,Y\}=E[\tfrac{1}{2}(X+Y+|X-Y|)]$$

$$=\tfrac{1}{2}[E(X)+E(Y)+E|X-Y|]$$

$X-Y\sim N(0,7)$

$$E|X-Y|=E|Z|=\int_{-\infty}^{+\infty}|z|\cdot\frac{1}{\sqrt{2\pi}\cdot\sqrt{7}}e^{-\frac{z^2}{14}}dz=\frac{14}{\sqrt{14\pi}}$$

$$E\max\{X,Y\} = 1+\frac{7}{\sqrt{14\pi}}$$

$$E\min\{X,Y\}?$$

$$E\min\{X,Y\} = E[\tfrac{1}{2}(X+Y-|X-Y|)]$$

$$= 1-\frac{7}{\sqrt{14\pi}}$$

例17 将n个编号为$1,2,\cdots$，n的球随机地放入编号为$1,2,\cdots$，n的n个盒子中，一个盒子只 能装一个球。如果第i号球正好放入第i号盒子，称为一个配对。求配对数的期望与方差。

解： 记$X_i=\begin{cases}1, & 第i个盒子配对\\ 0, & 第i个盒子没有配对\end{cases}$，$i=1,2,\cdots,n$,

则配对数为$X=\sum_{i=1}^{n}X_i$. $E(X_i)=\frac{1}{n},i=1,2,\cdots,n$

$$E(X)=\sum_{i=1}^{n}E(X_i)=1$$

$$D(X)=\sum_{i=1}^{n}D(X_i)+\sum_{i\neq j}\operatorname{cov}(X_i,X_j)$$

$$D(X_i)=E(X_i{}^2)-E^2(X_i)=\frac{1}{n}-\frac{1}{n^2}=\frac{n-1}{n^2}$$

$$\operatorname{cov}(X_i,X_j)=E(X_iX_j)-E(X_i)E(X_j)$$

$$=\frac{1}{n(n-1)}-\frac{1}{n^2}=\frac{1}{n^2(n-1)}$$

$$D(X)=\sum_{i=1}^{n} D(X_i)+\sum_{i\neq j}\operatorname{cov}(X_i,X_j)$$

$$=n\times\frac{n-1}{n^2}+A_n^2\cdot\frac{1}{n^2(n-1)}$$

$$=1$$

例18 设r.v. $X \sim E(4)$, $Y=\min\{2, X\}$, 求$E(Y)$。

解： $E(Y) = E\min\{2, X\}$

$$= \int_0^{+\infty} \min\{2, x\} 4e^{-4x} dx$$

$$= \int_0^{2} x \cdot 4e^{-4x} dx + \int_2^{+\infty} 2 \times 4e^{-4x} dx$$

$$= \frac{1}{4} - \frac{e^{-8}}{4}.$$

例19 从区间[0,1]上任取n个点，求最大点与最小点 之间距离的数学期望。

解： 设$X_1, X_2, \cdots X_n$为这n个点，则$X_1, X_2, \cdots X_n$相互独立且都服从$U[0,1]$，令$Y = \max\{X_1, X_2, \cdots X_n\}$，$Z = \min\{X_1, X_2, \cdots X_n\}$，则最大点与最小点的距 离为$Y - Z$，由第三章例38知

$$f_Y(y)=\begin{cases} ny^{n-1}, 0<y<1 \\ 0, \qquad 其他 \end{cases}$$

$$f_Z(z)=\begin{cases} n(1-z)^{n-1}, 0<z<1 \\ 0, \qquad 其他 \end{cases}$$

$$\mathrm{E(Y)}=\int_0^1 y\cdot ny^{n-1}dy$$

$$= n\cdot\int_0^1 y^n dy = \frac{n}{n+1}$$

$$E(Z)=\int_0^1 z\cdot n(1-z)^{n-1}dz$$

$$=n\int_0^1(1-t)t^{n-1}dt$$

$$=n[\int_0^1 t^{n-1}dt-\int_0^1 t^n dt]$$

$$=n[\frac{1}{n}-\frac{1}{n+1}]=\frac{1}{n+1}$$

$$\mathrm{E}(\mathrm{Y}-\mathrm{Z})=\mathrm{E}(\mathrm{Y})-\mathrm{E}(\mathrm{Z})=\frac{\mathrm{n}}{\mathrm{n}+1}-\frac{1}{\mathrm{n}+1}=\frac{\mathrm{n}\text{-}1}{\mathrm{n}+1}$$