§3 协方差及相关系数

- 协方差的定义
- 协方差的性质
- 相关系数的定义
- 相关系数的性质

问题 对于二维随机变量(X,Y):

已知联合分布 $\underset{\not\leftarrow}{\rightarrow}$ 边缘分布

这说明对于二维随机变量，除了每个随机变量各自的概率特性以外，相互之间可能还有某种联系. 问题是用一个什么样的数去反映这种联系.

数　$E\left[(X-EX)(Y-EY)\right]$

反映了随机变量X,Y之间的某种关系

一、协方差

1）协方差的定义

称

$$Cov(X, Y) = E[(X - EX)(Y - EY)]$$

为随机变量 X, Y 的*协方差*（Covariance）.

特别地, 若 X= Y, 则

$$Cov(X, X) = E(X - EX)^2 = DX.$$

因此, 方差是协方差的特例.

协方差刻画两个随机变量之间的某种关系.

$$Cov(X,Y) = E\left[(X - EX)(Y - EY)\right]$$

若 (X,Y) 为离散型，

$$Cov(X,Y) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(x_i - EX)(y_j - EY)p_{ij}$$

若 (X,Y) 为连续型，

$$Cov(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - EX)(y - EY)f(x,y)dxdy$$

2）相关系数的定义

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

称为随机变量 X, Y 的相关系数（Correlation Coefficient）。

ρ_{XY} 是一个无量纲的量；

若 $\rho_{XY}=0$，称 X,Y *不相关*，此时 $Cov(X,Y)=0$.

3) 计算协方差的常用公式

由定义可得

$$\boldsymbol{D(X \pm Y) = DX + DY \pm 2Cov(X,Y)}$$

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

4) 定理：若X，Y独立，则 X, Y不相关。

(反之，不然)

证明：由数学期望的性质:

若X，Y独立， $E(XY) = E(X)\,E(Y)$

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

注意：若 $E[(X - EX)(Y - EY)] \neq 0$

即 $E(XY) - E(X)\,E(Y) \neq 0$

则X，Y一定相关，且 X，Y一定不独立。

二、协方差的性质

$Cov(X,Y) = E[(X - EX)(Y-EY)]$

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

2) $Cov(aX, bY) = abCov(X, Y)$;

3) $Cov(X+Y, Z) = Cov(X, Z) + Cov(Y, Z)$;

4) $D(aX + bY) = a^2 DX + b^2 DY + 2abCov(X,Y)$

$$D(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 DX_i + 2\sum_{1\le i<j\le n} a_i a_j Cov(X_i, X_j)$$

5) X，Y不相关 $\Leftrightarrow$ $Cov(X,Y) = 0$ $\Leftrightarrow$

$$E(XY) = E(X)E(Y) \Leftrightarrow D(aX + bY) = a^2 DX + b^2 DY.$$

三、相关系数的性质

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

2) $|\rho_{XY}| = 1 \Leftrightarrow$ 存在常数a,b使 $P\{Y = a + bX\} = 1.$

证明：1) 考虑以线性函数 $a+bX$ 来近似表示Y. 令

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

$$= E(Y^2) + b^2E(X^2) + a^2 - 2bE(XY) + 2abE(X) - 2aE(Y)$$

求a，b 使 e 达到最小。

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

将 $a=E(Y)-bE(X)$，代入第二个方程得

$$2bE(X^2)-2E(XY)+2aE(X)=0$$

$$bE(X^2)-E(XY)+\big(E(Y)-bE(X)\big)E(X)=0,$$

故 $b=\dfrac{E(XY)-E(X)E(Y)}{E(X^2)-(E(X))^2}=\dfrac{Cov(X,Y)}{DX},$

得： $$b_0=\frac{Cov(X,Y)}{DX};$$

$$a_0=EY-b_0EX=EY-EX\cdot\frac{Cov(X,Y)}{DX}.$$

$$\min_{a,b} E[Y-(a+bX)]^2 = E[Y-(a_0+b_0X)]^2$$

$$= E(Y-EY+EX\frac{Cov(X,Y)}{DX}-X\cdot\frac{Cov(X,Y)}{DX})^2$$

$$= E((Y-EY)-(X-EX)\cdot\frac{Cov(X,Y)}{DX})^2$$

$$= DY + DX\cdot\frac{Cov^2(X,Y)}{(DX)^2} - 2Cov(X,Y)\cdot\frac{Cov(X,Y)}{DX}$$

$$= DY+\frac{Cov^2(X,Y)}{DX}-2\frac{Cov^2(X,Y)}{DX}$$

$$= DY - \frac{Cov^2(X,Y)}{DX} = DY - \frac{\rho_{XY}^2 \cdot DX \cdot DY}{DX}$$

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

$$= (1-\rho_{XY}^2)DY \quad = \min_{a,b} E[Y-(a+bX)]^2$$

即 $\min\limits_{a,b} E[Y-(a+bX)]^2 = (1-\rho_{XY}^2)DY$

由上式得:

1) $1-\rho_{XY}^2 \geq 0,$ 　　即 $|\rho_{XY}| \leq 1.$

2）“充分性”：

若 $|\rho_{XY}| = 1$ $\Rightarrow$ 存在常数a,b使 $P\{Y = a + bX\} = 1$

由上面知此时 $E[Y-(a_0+b_0X)]^2 = 0$

从而

$$D[Y-(a_0+b_0X)] + (E[Y-(a_0+b_0X)])^2$$

$$=E[Y-(a_0+b_0X)]^2=0$$

所以

$$D[Y-(a_0+b_0X)]=0,\quad E[Y-(a_0+b_0X)]=0$$

由于 $DX=0 \Leftrightarrow P\{X=c\}=1, c=EX,$

故 $P\{Y-a_0-b_0X=0\}=1$

即 $P\{Y=a_0+b_0X\}=1.$

“必要性”：反之，若存在a^*,b^*使，

$$P\{Y=a^*+b^*X\}=1\Rightarrow|\rho_{XY}|=1.$$

这时 $$P\{Y-(a^*+b^*X)=0\}=1,$$

故 $$E[Y-(a^*+b^*X)]^2=0$$

则

$$0=E[Y-(a^*+b^*X)]^2\geq\min_{a,b}E[Y-(a+bX)]^2$$

$$=(1-\rho_{XY}^2)DY\geq 0.$$

故 $$1-\rho_{XY}^2=0,$$ 即$|\rho_{XY}|=1.$

说　明

相关系数是表征随机变量 X 与 Y 之间线性关系紧密程度的量.

当 $|\rho_{XY}|=1$ 时， X 与 Y 之间以概率 1存在着线性关系;

当 $|\rho_{XY}|$越接近于 0 时，X 与 Y 之间的线性关系越弱;

当 $|\rho_{XY}|=0$ 时，X 与 Y 之间不存在线性关系(不相关).

注：X与Y之间没有线性关系并不表示它们之间没有关系。

X 与 Y 不相关，但不一定相互独立。

例1　设 X，Y是二个随机变量，已知 $DX=1$，$DY=4$，$Cov(X, Y)=1$，记

$$\xi = X-2Y, \quad \eta = 2X-Y$$

$$\rho_{\xi\eta} = \frac{COV(\xi,\eta)}{\sqrt{D\xi}\sqrt{D\eta}}$$

试求：$\rho_{\xi\eta}$.

解：$D\xi = D(X-2Y) = DX + 4DY - 4Cov(X, Y)$

$$=1+4\times4-4\times1=13$$

$$D\eta = D(2X-Y) = 4DX + DY - 4Cov(X, Y)$$

$$=4\times1+4-4\times1=4$$

$$\mathrm{Cov}(\xi,\ \eta) = \mathrm{Cov}(X-2Y,\ 2X-Y)$$

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

$$- \mathrm{Cov}(X,\ Y) + 2\mathrm{Cov}(Y,\ Y)$$

$$= 2DX - 5\mathrm{Cov}(X,\ Y) + 2DY$$

$$= 2\times1 - 5\times1 + 2\times4 = 5$$

所以，$\rho_{\xi\eta} = \dfrac{\mathrm{Cov}(\xi,\ \eta)}{\sqrt{D\xi}\sqrt{D\eta}} = \dfrac{5}{\sqrt{13}\sqrt{4}} = \dfrac{5\sqrt{13}}{26}$

例3 设 A, B 是二随机事件；随机变量

$$X=\begin{cases}1, & \text{若}A\text{出现},\\ -1, & \text{若}A\text{不出现},\end{cases}\qquad Y=\begin{cases}1, & \text{若}B\text{出现},\\ -1, & \text{若}B\text{不出现}.\end{cases}$$

试证明随机变量X 和 Y 不相关的充分必要条件是A与B相互独立.

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

$$E(X)=P(A)-P(\overline{A})=2P(A)-1,\quad E(Y)=2P(B)-1.$$

$$E(X)E(Y)=[2P(A)-1][2P(B)-1]$$

$$=4P(A)P(B)-2P(A)-2P(B)+1$$

$$E(XY)=\mathbf{1\times 1\times P\{X=1,Y=1\}}$$

$$+1\times(-1)\times P\{X=1,Y=-1\}$$

$$+(-1)\times 1\times P\{X=-1,Y=1\}$$

$$+(-1)\times(-1)\times P\{X=-1,Y=-1\}$$

$$=P(AB)-P(A\bar{B})-P(\bar{A}B)+P(\bar{A}\bar{B})$$

$$=\boldsymbol{P(AB)}-[P(A)-P(AB)]$$

$$-[P(B)-P(AB)]\quad+[P(\bar{A})-P(\bar{A}B)]$$

$$=P(AB)-[P(A)-P(AB)]-[P(B)-P(AB)]$$

$$+[1-P(A)-P(B)+P(AB)]$$

$$=\boldsymbol{4P(AB)-2P(A)-2P(B)+1}$$

$$\therefore \mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=0\Leftrightarrow \boldsymbol{P(AB)=P(A)P(B)}$$

第四节　矩、协方差矩阵

一、基本概念

二、*n* 维正态变量的性质

三、小结

一、基本概念

1.定义

设 X 和 Y 是随机变量,若 $E(X^k)$, $k=1,2,\cdots$ 存在,称它为 X 的 k 阶原点矩,简称 k 阶矩.

若 $E\{[X-E(X)]^k\}$, $k=2,3,\cdots$ 存在,称它为 X 的 k 阶中心矩.

若 $E(X^kY^l)$, $k,l=1,2,\cdots$ 存在,称它为 X 和 Y 的 $k+l$ 阶混合矩.

若 $E\{[X-E(X)]^k[Y-E(Y)]^l\}$, $k,l=1,2,\cdots$ 存在,称它为 X 和 Y 的 $k+l$ 阶混合中心矩.

2. 说明

(1) 以上数字特征都是随机变量函数的数学期望;

(2) 随机变量 X 的数学期望 $E(X)$ 是 X 的一阶原点矩,方差为二阶中心矩, 协方差 $\mathrm{Cov}(X,Y)$是 X 与 Y 的二阶混合中心矩;

(3) 在实际应用中,高于 4 阶的矩很少使用.

三阶中心矩$E\{[X-E(X)]^3\}$主要用来衡量随机变量的分布是否有偏.

四阶中心矩 $E\{[X-E(X)]^4\}$主要用来衡量随机变量的分布在均值附近的陡峭程度如何.

3. 协方差矩阵

设 n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 的二阶混合中心矩

$$c_{ij} = \mathrm{Cov}(X_i, X_j) = E\{[X_i - E(X_i)][X_j - E(X_j)]\}$$

$$i, j = 1,2,\cdots,n$$

都存在,则称矩阵

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

为 n 维随机变量的**协方差矩阵**.

例如　二维随机变量(X_1, X_2)的协方差矩阵为

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

其中　$c_{11} = E\{[X_1 - E(X_1)]^2\},$

$$c_{12} = E\{[X_1 - E(X_1)][X_2 - E(X_2)]\},$$

$$c_{21} = E\{[X_2 - E(X_2)][X_1 - E(X_1)]\},$$

$$c_{22} = E\{[X_2 - E(X_2)]^2\}.$$

由于 $c_{ij}=c_{ji}\ (i,j=1,2,\cdots,n)$,所以协方差矩阵为对称的非负定矩阵.

协方差矩阵的应用

协方差矩阵可用来表示多维随机变量的概率密度，从而可通过协方差矩阵达到对多维随机变量的研究

以二维随机变量 (X_1, X_2) 为例.

由于

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}.$$

引入矩阵 $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}.$

及 (X_1, X_2) 的协方差矩阵 $C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix},$

$$C=\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}=\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

由此可得

$$C^{-1}=\frac{1}{\det C}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

$$=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

由于

$$(X-\mu)^{\mathrm{T}}C^{-1}(X-\mu)$$

$$=\frac{1}{\det C}(x_1-\mu_1, x_2-\mu_2)\begin{pmatrix}\sigma_2^2 & -\rho\sigma_1\sigma_2\\ -\rho\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix}\begin{pmatrix}x_1-\mu_1\\ x_2-\mu_2\end{pmatrix}$$

$$=\frac{1}{1-\rho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right].$$

于是(X_1, X_2)的概率密度可写成

$$f(x_1, x_2)$$

$$= \frac{1}{(2\pi)^{2/2}(\det C)^{1/2}} \exp\left\{-\frac{1}{2}(X-\mu)^T C^{-1}(X-\mu)\right\}.$$

推广

n 维随机变量$(X_1, X_2, \cdots, X_n)$的概率密度可表示为 $f(x_1, x_2, \cdots, x_n)$

$$= \frac{1}{(2\pi)^{n/2}(\det C)^{1/2}} \exp\left\{-\frac{1}{2}(X-\mu)^T C^{-1}(X-\mu)\right\}.$$

其中$X = (x_1, x_2, \cdots, x_n)^T$,

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix}, \quad C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}.$$

二、n 维正态变量的性质

1. n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 的每一个分量 $X_i, i = 1, 2, \cdots, n$ 都是正态变量;

反之,若 $X_1, X_2, \cdots, X_n$ 都是正态变量,且相互独立, 则 $(X_1, X_2, \cdots, X_n)$ 是 n 维正态变量.

2. n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布的充要条件是 $X_1, X_2, \cdots, X_n$ 的任意的线性组合 $l_1X_1 + l_2X_2 + \cdots + l_nX_n$ 服从一维正态分布(其中 $l_1, l_2, \cdots, l_n$ 不全为零).

3.若$(X_1, X_2, \cdots, X_n)$服从n维正态分布,设$Y_1, \cdots, Y_k$是$X_j (j = 1,2,\cdots,n)$的线性函数,则$(Y_1, Y_2, \cdots, Y_k)$也服从多维正态分布. 线性变换不变性

4.设$(X_1, \cdots, X_n)$服从n维正态分布,则“$X_1, X_2, \cdots, X_n$相互独立”与“$X_1, X_2, \cdots, X_n$两两不相关”是等价的.

三、小结

1. 矩是随机变量的数字特征.

随机变量 X 的

$$\begin{cases} \text{数学期望 } E(X) \text{ 是 } X \text{ 的一阶原点矩;} \\ \text{方差 } D(X) \text{ 是 } X \text{ 的二阶中心矩;} \\ \text{协方差 } \mathrm{Cov}(X,Y) \text{ 是 } X \text{ 与 } Y \text{ 的二阶混合中心矩.} \end{cases}$$

2.正态变量是最重要的随机变量，其性质一定要熟练掌握.