概率论与数理统计
第四章 随机变量的数字特征

金玲飞

复旦大学软件学院
Email: lfjin@fudan.edu.cn

2018.11.27

4.3 协方差与相关系数

协方差

定义 (4.3.1 Co-variance)

设(X, Y)为二维随机变量，若

$$E(|(X - E(X))(Y - E(Y))|) < \infty,$$

则称$E((X - E(X))(Y - E(Y)))$为X与Y的协方差，记

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

协方差

定义 (4.3.1 Co-variance)

设(X, Y)为二维随机变量，若

$$E(|(X-E(X))(Y-E(Y))|)<\infty,$$

则称$E((X-E(X))(Y-E(Y)))$为X与Y的协方差，记

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

协方差的计算式

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

协方差

定义 (4.3.1 Co-variance)

设(X, Y)为二维随机变量，若

$$E(|(X - E(X))(Y - E(Y))|) < \infty,$$

则称$E((X - E(X))(Y - E(Y)))$为X与Y的协方差，记

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

协方差的计算式

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

$$D(X \pm Y) = D(X) + D(Y) \pm 2Cov(X, Y)$$

协方差的性质

性质

- X, Y 独立且协方差存在时，$Cov(X, Y) = 0$。
- $Cov(X, Y) = Cov(Y, X)$（对称性）

协方差的性质

性质

- X,Y 独立且协方差存在时，$Cov(X,Y)=0$。
- $Cov(X,Y)=Cov(Y,X)$（对称性）
- $Cov(aX+b,cY+d)=acCov(X,Y)$

协方差的性质

性质

- X,Y 独立且协方差存在时，$Cov(X,Y)=0$。
- $Cov(X,Y)=Cov(Y,X)$（对称性）
- $Cov(aX+b,cY+d)=acCov(X,Y)$
- $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

协方差的性质

性质

- X,Y 独立且协方差存在时，$Cov(X,Y)=0$。
- $Cov(X,Y)=Cov(Y,X)$（对称性）
- $Cov(aX+b,cY+d)=acCov(X,Y)$
- $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

$$Cov(kX,kY)=k^2Cov(X,Y)$$

相关系数

标准化

$$X^* = \frac{X - E(X)}{\sqrt{D(X)}}, Y^* = \frac{Y - E(Y)}{\sqrt{D(Y)}}$$

相关系数

标准化

$$X^* = \frac{X - E(X)}{\sqrt{D(X)}}, Y^* = \frac{Y - E(Y)}{\sqrt{D(Y)}}$$

此时，$E(X^*) = E(Y^*) = 0$，$D(X^*) = D(Y^*) = 1$。

相关系数

标准化

$$X^* = \frac{X - E(X)}{\sqrt{D(X)}}, Y^* = \frac{Y - E(Y)}{\sqrt{D(Y)}}$$

此时，$E(X^*) = E(Y^*) = 0$，$D(X^*) = D(Y^*) = 1$。

$$\mathrm{Cov}(X^*, Y^*) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

定义 (4.3.2 correlation coefficient)

设(X, Y)为二维随机变量。若$D(X) > 0, D(Y) > 0$，则X与Y的相关系数记为

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

因此，相关系数就是随机变量标准化后的协方差。

定义 (4.3.2 correlation coefficient)

设(X, Y)为二维随机变量。若$D(X) > 0, D(Y) > 0$，则X与Y的相关系数记为

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

因此，相关系数就是随机变量标准化后的协方差。

- 相关系数在线性变换下保持不变

$$\rho_{aX+b,cY+d} = \pm\rho_{XY}$$

柯西-施瓦茨不等式

ρ 反映了X,Y之间的什么关系?

定理 (4.3.1 Cauchy-Schwarz inequality)

设X, Y为任意两个随机变量，若$E(X^2) < \infty, E(Y^2) < \infty$，则有

$$|E(XY)|^2 \leq E(X^2)E(Y^2)$$

且等式成立的充要条件是存在常数t_0，使得

$$P(Y = t_0 X) = 1$$

柯西-施瓦茨不等式

ρ 反映了X,Y之间的什么关系？

定理 (4.3.1 Cauchy-Schwarz inequality)

设X, Y为任意两个随机变量，若$E(X^2) < \infty, E(Y^2) < \infty$，则有

$$|E(XY)|^2 \leq E(X^2)E(Y^2)$$

且等式成立的充要条件是存在常数t_0，使得

$$P(Y = t_0 X) = 1$$

类似的，我们有

$$(\mathrm{Cov}(X, Y))^2 \leq D(X)D(Y).$$

由柯西-施瓦茨不等式，我们有以下定理

定理 (4.3.2)

设随机变量X, Y的相关系数为ρ，则有

1. $|\rho| \leq 1$
2. $|\rho| = 1$的充要条件是X和Y之间线性相关，即存在常数a, b，使得$P(Y = aX + b) = 1$。

ρ_{XY}衡量了X, Y之间的线性关系的强弱。

- $\rho > 0$，则称X, Y正相关
- $\rho < 0$，则称X, Y负相关
- $\rho = 0$，则称X, Y不相关（不线性相关）

ρ_{XY}衡量了X, Y之间的线性关系的强弱。

- $\rho > 0$，则称X, Y正相关
- $\rho < 0$，则称X, Y负相关
- $\rho = 0$，则称X, Y不相关（不线性相关）

- X和Y独立，则X和Y不相关。反之不一定成立。

X和Y不相关，但可能不独立。

例子 (4.3.2)

设θ服从均匀分布$U[0, 2\pi]$， $X = \cos\theta$， $Y = \cos(\theta + \alpha)$，其中$\alpha$是常数。考虑$\alpha$取不同值时$X$和$Y$的关系。

X和Y不相关，但可能不独立。

例子 (4.3.2)

设θ服从均匀分布$U[0, 2\pi]$，$X = \cos\theta$，$Y = \cos(\theta + \alpha)$，其中$\alpha$是常数。考虑$\alpha$取不同值时$X$和$Y$的关系。

若X和Y服从二维正态分布，则不相关性等价于独立性。
此时，$\mathrm{Cov}(X, Y) = \sigma_1\sigma_2\rho$。

定理 (4.3.3)

对随机变量X, Y，一下命题是等价的：

1. $Cov(X, Y) = 0$；
2. X和Y不相关；
3. $E(XY) = E(X)E(Y)$；
4. $D(X + Y) = D(X) + D(Y)$。

协方差矩阵

定义 (4.3.2 covariance matrix)

设$(X_1, \dots, X_n)$为n维随机变量$(n \geq 2)$。记

$$b_{ij} = Cov(X_i, X_j) = E((X_i - E(X_i))(X_j - E(X_j)), i, j = 1, 2, \dots, n$$

称矩阵

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix}$$

为$(X_1, \dots, X_n)$的协方差矩阵。

设(X_1, X_2)为二维正态随机变量，已知联合概率密度 函数

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

协方差矩阵

$$B = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$$

我们可以写成

$$f(x_1, x_2) = \frac{1}{(2\pi)|B|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T B^{-1}(x-\mu)}$$

其中

$$x = (x_1, x_2)^T, \quad \mu = (\mu_1, \mu_2)^T$$

n维正态随机变量$(X_1, \cdots, X_n)$的联合密度函数是

$$f(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}|B|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T B^{-1}(x-\mu)}$$

n维正态随机变量$(X_1, \cdots, X_n)$的联合密度函数是

$$f(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}|B|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T B^{-1}(x-\mu)}$$

- $X_1, \cdots, X_n$ 相互独立的充要条件是它们两两不相关。

例子

将一颗骰子投n次，求1点出现的次数与6点出现的次数的协方差和相关系数分别是多少？