协方差及相关系数

- 协方差
  - 定义
  - 性质
- 相关系数
  - 定义
  - 性质
  - 意义

协方差

回忆方差的一个性质：

随机变量X 和 Y的方差存在，则 $X+Y$ 的方差也存在, 且

$$D(X+Y)=D(X)+D(Y)+2E([X-E(X)][Y-E(Y)])$$

当随机变量X, Y 相互独立时，$E([X-E(X)][Y-E(Y)])=0$

协方差的定义

设(X,Y)是二维随机变量，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，并记作$\mathrm{Cov}(X,Y)$，即

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

- $\mathrm{Cov}(X,X)=D(X)$

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

- 当随机变量X, Y 相互独立时，$\mathrm{Cov}(X,Y) = 0$
- 对任意随机变量X 和 Y, 有 $D(X \pm Y) = D(X) + D(Y) \pm 2\,\mathrm{Cov}\,(X,Y)$
- $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$

协方差的性质

性质1. 对称性 $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$

性质2. 若a, b, c, d为常数，则

$$\mathrm{Cov}(aX+c,bY+d)=ab\mathrm{Cov}(X,Y)$$

性质3. $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$

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

令 $X^* = \dfrac{X-E(X)}{\sqrt{D(X)}}$ (X 的标准化) $Y^* = \dfrac{Y-E(Y)}{\sqrt{D(Y)}}$ (Y 的标准化)

则 $E(X^*) = E(Y^*) = 0$ $D(X^*) = D(Y^*) = 1$

故 $\mathrm{Cov}(X^*,Y^*) = E(X^*Y^*) - E(X^*)E(Y^*) = E(X^*Y^*)$

$$= E\left(\frac{X-E(X)}{\sqrt{D(X)}}\frac{Y-E(Y)}{\sqrt{D(Y)}}\right) = \frac{E([X-E(X)][Y-E(Y)])}{\sqrt{D(X)}\sqrt{D(Y)}}$$

相关系数的定义

设(X,Y)是二维随机变量，若$D(X)>0, D(Y)>0$, 则称

$\dfrac{\operatorname{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$ 为X与Y的相关系数，记为ρ_{XY} 或 ρ,

即 $$\rho_{XY}=\frac{\operatorname{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{E([X-E(X)][Y-E(Y)])}{\sqrt{D(X)}\sqrt{D(Y)}}$$

相关系数的性质

性质1. Cauchy-Schwarz 不等式

设X,Y是任意的两个随机变量，若$E\left(X^2\right) < \infty, E(Y^2) < \infty$，则有

$$b^2 \quad \boxed{|E(XY)|^2} \le E\left(X^2\right)E(Y^2) \qquad a \quad c$$

且等式成立的充要条件是存在常数t_0, 使得 $P(Y = t_0X) = 1$

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

证：考虑实变量t的二次函数

$$g(t) = t^2 E(X^2) - 2tE(XY) + E(Y^2) = E[(tX - Y)^2]$$

因为对一切t, 有$g(t) = E[(tX - Y)^2] \geq 0$, 故

$$4|E(XY)|^2 - 4E(X^2)E(Y^2) \leq 0$$

即 $|E(XY)|^2 \leq E(X^2)E(Y^2)$

$$|E(XY)|^2 = E(X^2)E(Y^2) \Longleftrightarrow P(Y = t_0 X) = 1$$

此外，等式 $|E(XY)|^2 = E(X^2)E(Y^2)$ 成立的充要条件是

方程 $t^2E(X^2) - 2tE(XY) + E(Y^2) = 0$有一个重根

即存在常数t_0，使得 $E[(t_0X - Y)^2] = 0$

由数学期望的性质知 $P(Y = t_0X) = 1$

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

- 由Cauchy-Schwarz 不等式可得 $(\mathrm{Cov}(X,Y))^2 \leq D(X)D(Y)$

  且等式成立的充要条件是存在常数 a,b 使得 $P(Y=aX+b)=1$

证： $[\mathrm{Cov}(X,Y)]^2 = [E([X-E(X)][Y-E(Y)])]^2$

$$\leq E([X-E(X)]^2)\,E([Y-E(Y)]^2) = D(X)D(Y)$$

故等式成立的充要条件是 即存在常数t_0, 使得

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

令 $a=t_0,\ b=E(Y)-t_0E(X)$ 得 $P(Y=aX+b)=1$

$$(Cov(X,Y))^2 \leq D(X)D(Y)$$

- 进一步可得 $|\mathrm{Cov}(X,Y)| \leq \sqrt{D(X)}\sqrt{D(Y)}$

即 $|\rho_{XY}| = \dfrac{|\mathrm{Cov}\,(X,\,Y)|}{\sqrt{D(X)}\sqrt{D(Y)}} \leq 1$

且$|\rho_{XY}|=1$的充要条件是存在常数a, b, 使得

$$P(Y = aX + b) = 1$$

相关系数的性质

性质2. 设随机变量X,Y的相关系数为ρ，，则有

(1) $|\rho| \le 1$；

(2) $|\rho| = 1$的充分必要条件是X与Y线性相关，

即存在常数a, b, 使得 $P(Y = aX + b) = 1$

问题： 如何度量随机变量X与Y线性关系的强弱?

如何度量随机变量X与Y线性关系的强弱？

考虑关于实变量a与b的二元函数

$$g(a,b)=E\left([Y-(aX+b)]^2\right)$$

可求得 $$L(X)=aX+b=\frac{\sqrt{D(Y)}}{\sqrt{D(X)}}\rho_{XY}X+E(Y)-\frac{\sqrt{D(Y)}}{\sqrt{D(X)}}\rho_{XY}E(X)$$

$$\inf_{-\infty<a,\,b<\infty} E\left([Y-(aX+b)]^2\right)=D(Y)(1-\rho_{XY}^2)$$

$$\inf_{-\infty<a,b<\infty} E\left([Y-(aX+b)]^2\right) = D(Y)(1-\rho_{XY}^2)$$

$$L(X) = aX + b = \frac{\sqrt{D(Y)}}{\sqrt{D(X)}}\rho_{XY}X + E(Y) - \frac{\sqrt{D(Y)}}{\sqrt{D(X)}}\rho_{XY}E(X)$$

- 当$|\rho_{XY}| = 1$时， X与Y线性相关；
- 当$|\rho_{XY}| < 1$时，线性关联程度随着ρ_{XY}减小而减弱；
- 当 $|\rho_{XY}| = 0$ 时， X与Y不存在线性关系，此时称X与Y不相关；
- 当$\rho_{XY} > 0$时，X与Y正相关； 当$\rho_{XY} < 0$时，X与Y负相关

$\rho_{XY} = 0 \Longleftrightarrow$ X 与 Y 不相关

对随机变量 X 与 Y,下列命题等价:

- X 与 Y 不相关
- $\rho_{XY} = 0$
- $\mathrm{Cov}(X, Y) = 0$
- $D(X+Y) = D(X) + D(Y)$
- $E(XY) = E(X)E(Y)$

相互独立 $\underset{\not\Longleftarrow}{\Longrightarrow}$ 不相关

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

例1. 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.

求X与Y的相关系数，并判断其相关性和独立性。

解： $E(X)=\int_0^{2\pi}\frac{1}{2\pi}\cos\theta\,d\theta=0$ $E(Y)=\int_0^{2\pi}\frac{1}{2\pi}\cos(\theta+\alpha)\,d\theta=0$

$$E(XY)=\int_0^{2\pi}\frac{1}{2\pi}\cos\theta\cos(\theta+\alpha)\,d\theta=\frac{1}{2}\cos\alpha$$

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

$$D(X)=E\left(X^2\right)-E(X)=E\left(X^2\right)=\frac{1}{2\pi}\int_0^{2\pi}(\cos\theta)^2 d\theta=\frac{1}{2}$$

$$D(Y)=E\left(Y^2\right)-E(Y)=E\left(Y^2\right)=\frac{1}{2\pi}\int_0^{2\pi}(\cos(\theta+\alpha))^2 d\theta=\frac{1}{2}$$

$$\Longrightarrow \rho_{XY}=\cos\alpha$$

当 $\alpha=0$ 或 π 时，$|\rho_{XY}|=1 \Longrightarrow X$ 与 Y 不独立但线性相关

当 $\alpha=\frac{\pi}{2}$ 或 $\frac{3\pi}{2}$ 时，$\rho_{XY}=0 \Longrightarrow X$ 与 Y 不相关

$X=\cos\theta \quad Y=\cos(\theta\pm\frac{\pi}{2}) \Longrightarrow X^2+Y^2=1 \Longrightarrow$ 不独立.

$$\rho_{XY}=\frac{\text{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{E([X-E(X)][Y-E(Y)])}{\sqrt{D(X)}\sqrt{D(Y)}}$$

例2. 设 $(X,Y)\sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$, 求$X,Y$的相关系数.

解: $E(X)=\mu_1,\quad E(Y)=\mu_2,\quad D(X)=\sigma_1^2,\quad D(Y)=\sigma_2^2,$

$$\text{Cov}(X,Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_1)(y-\mu_2)f(x,y)dxdy$$

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

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

$\xrightarrow{\hspace{6em}}$ 配方换元

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

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

$$\mathrm{Cov}(X, Y) = \frac{\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty} v e^{-\frac{v^2}{2}} \underline{\left(\int_{-\infty}^{\infty}(\rho v + u\sqrt{1-\rho^2}\,)e^{-\frac{u^2}{2}}du\right)} dv$$

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

$$= \frac{\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty} v e^{-\frac{v^2}{2}}(\rho v\sqrt{2\pi} + 0)dv = \sigma_1\sigma_2\rho\boxed{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}v^2 e^{-\frac{v^2}{2}}dv}$$

$$= \sigma_1\sigma_2\rho$$

$E(Z^2)$ 其中$Z \sim N(0, 1)$

$$(X,Y)\sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right) \qquad \rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

于是可得 $\mathrm{Cov}(X,Y)=\sigma_1\sigma_2\rho$

进一步得 $\rho_{XY}=\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\dfrac{\sigma_1\sigma_2\rho}{\sigma_1\sigma_2}=\rho$

若 $(X,Y)\sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$，则$X$与$Y$相互独立的充分必要条件是 $\rho=0$

若 $(X,Y)\sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$，则$X$与$Y$相互独立等价于$X$与$Y$不相关

$$E(X)=\int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x$$

例3. 设随机变量(X, Y)的联合概率密度为

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

求 $E(X), E(Y), D(X), D(Y), \mathrm{Cov}(X, Y), \rho_{XY}$

并判断X与Y的相关性，独立性.

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

解: $E(X)=\int_0^2\left(\int_0^2 x\frac{1}{8}(x+y)dy\right)dx=\int_0^2 x\left(\frac{x}{4}+\frac{1}{4}\right)dx=\frac{7}{6}$

$$E(X^2)=\int_0^2\left(\int_0^2 x^2\frac{1}{8}(x+y)dy\right)dx=\int_0^2 x^2\left(\frac{x}{4}+\frac{1}{4}\right)dx=\frac{5}{3}$$

故 $D(X)=E(X^2)-[E(X)]^2=\frac{11}{36}$

同理 $E(Y)=\frac{7}{6}$ $D(Y)=\frac{11}{36}$

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

$$E(XY)=\int_0^2\left(\int_0^2 xy\frac{1}{8}(x+y)dy\right)dx=\frac{4}{3}$$

$$\text{Cov}(X,Y)=E(XY)-E(X)E(Y)=-\frac{1}{36}$$

由此可知X与Y不独立

$$\rho_{XY}=-\frac{1}{11}$$

由此可知X与Y负相关，线性关系不强

小 结