西北工業大學

NORTHWESTERN POLYTECHNICAL UNIVERSITY

概率论与数理统计

徐 爽

西北工业大学理学院概率统计教研室

第三章 随机变量的数字特征

§3.3 协方差与相关系数

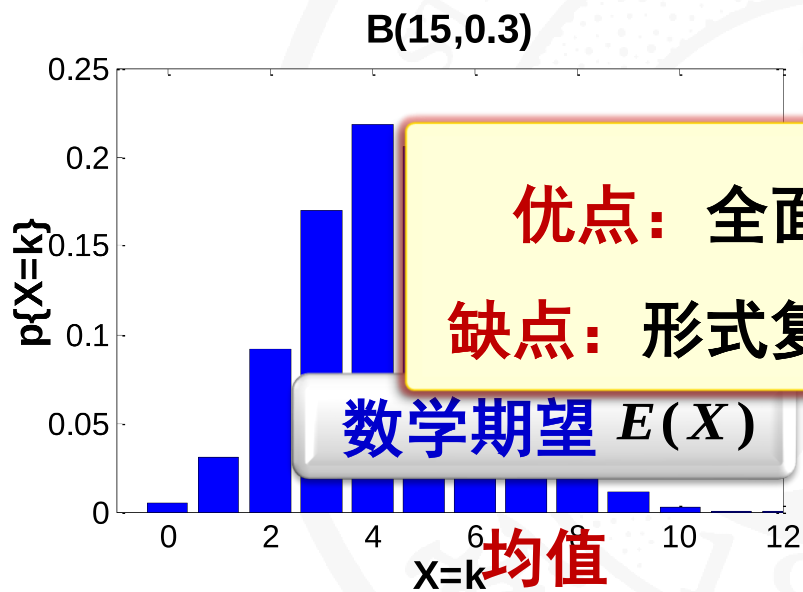
- 一、协方差
- 二、相关系数
- *三、协方差矩阵

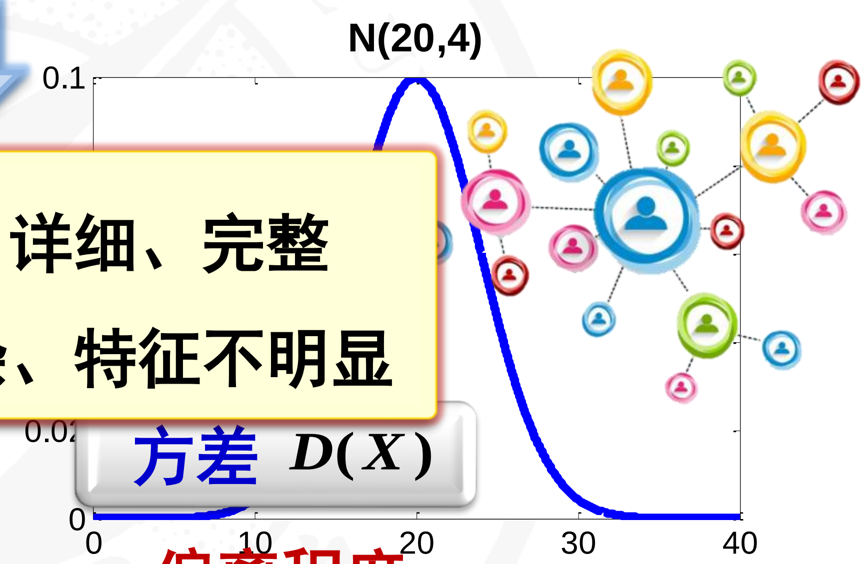
随机变量的统计特性

分布函数 $F(x)=P(X\le x)$

分布律 $P\{X=x_k\}$

概率密度 $p(x)$





优点：全面、详细、完整

缺点：形式复杂、特征不明显

数学期望 $E(X)$ 均值

方差 $D(X)$ 偏离程度

$$X\sim N(\mu,\sigma^2)$$

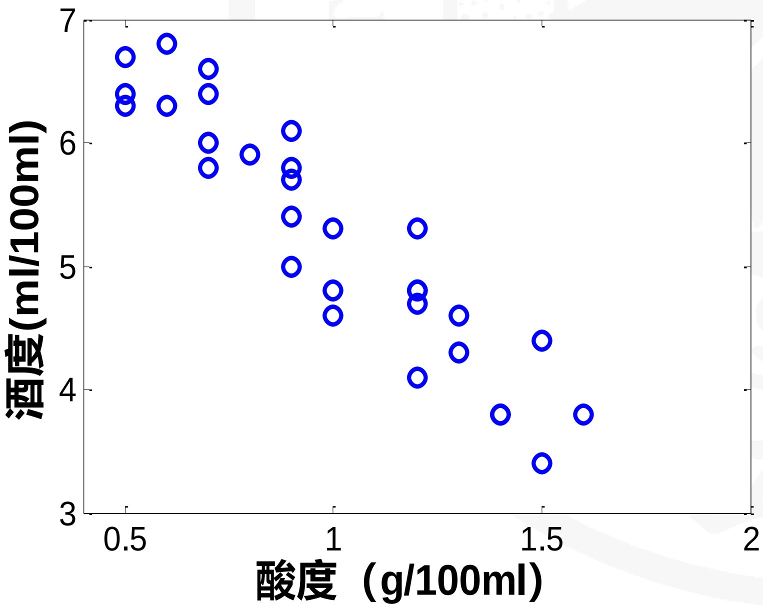
$$\Rightarrow X+X \quad \sim N(2\mu,2\sigma^2) \times \sim N(2\mu,4\sigma^2) \checkmark$$

一、协方差

1. 问题的提出

引例 某酒厂要判定酒醅中酸度和酒度的关系，通过检验测得以下数据:

散点图



序号	酸度 (g/100ml)	酒度 (ml/100ml)
1	0.5	6.3
2	0.9	5.8
3	1.2	4.6
…	…	…
29	0.5	6.7
30	1.2	[illegible]

关系

X：酸度 Y：酒度

关系 → 独立性

X, Y相互独立 $D(X+Y)=D(X)+D(Y)$

X, Y不独立 $D(X+Y)=E\left[(X+Y)-E(X+Y)\right]^2$

$$=E\left\{\left[X-E(X)\right]+\left[Y-E(Y)\right]\right\}$$

相互关系

$$=D(X)+D(Y)+2\boxed{E\left\{\left[X-E(X)\right]\left[Y-E(Y)\right]\right\}}$$

协方差

2. 协方差的定义

定义3.6 对于随机变量X, Y，若

$$E\{[X-E(X)][Y-E(Y)]\}\text{存在，}$$

称之为X与Y的**协方差**(*Covariance*)，记作$Cov(X,Y)$，

即

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

注

1^o $Cov(X,Y)\neq 0 \Rightarrow X,Y$不独立

2^o $Cov(X,Y)=Cov(Y,X)$ 协方差是方差的推广

3° $Cov(X, X)=E\{[X-E(X)][X-E(X)]\}=D(X).$

协方差是描述随机变量相互关系的数字特征

3. 协方差的性质

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

性质(2) $\operatorname{cov}(aX,bY)=ab\operatorname{cov}(X,Y)$, a,b为常数.

性质(3) $\operatorname{cov}(X_1+X_2,Y)=\operatorname{cov}(X_1,Y)+\operatorname{cov}(X_2,Y).$

性质(4) 若X与Y独立, 则$\operatorname{cov}(X,Y)=0.$

性质(5) $D(X\pm Y)=D(X)+D(Y)\pm 2\operatorname{cov}(X,Y).$

推广 $$D\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}D(X_i)+2\sum\sum_{i<j}\operatorname{cov}(X_i,X_j).$$

2. 协方差的计算

(1) 利用定义计算

$f(X,Y)$

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

离散型

$$Cov(X,Y)=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}[x_i-E(X)][y_j-E(Y)]p_{i,j},$$

其中 $p_{i,j}, i,j=1,2,\cdots$ 是 X,Y 的联合分布律.

连续型

$$Cov(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x-E(X)][y-E(Y)]p(x,y)\,\mathrm{d}x\mathrm{d}y,$$

其中 $p(x,y)$ 为 X,Y 的联合密度函数.

(2) 利用公式计算

$f(X,Y)$

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

证明： $Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$

（期望的性质） $=E\{XY-XE(Y)-YE(X)+E(X)E(Y)\}$

$$=E(XY)-2E(X)E(Y)+E(X)E(Y)$$

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

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

例1 设(X,Y)的联合密度函数为 $f(x,y)=\begin{cases}3x, & 0<y<x<1\\ 0, & \text{其他}\end{cases}$，

求：$Cov(X,Y)$.

解： $E(X)=\int_0^1\int_0^x x\cdot\boxed{3x}dydx=\frac{3}{4}$ $\quad E(Y)=\int_0^1\int_0^x y\cdot\boxed{3x}dydx=\frac{3}{8}$

$$E(XY)=\int_0^1\int_0^x xy\cdot\boxed{3x}dydx=\int_0^1 3x^2\cdot\frac{x^2}{2}dx=\frac{3}{10}$$

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

$$=\frac{3}{160}>0$$

正相关

如何度量

三、相关系数

1. 相关系数的定义

相互关系 $\Rightarrow$ 线性相互关系

均方误差 $e = E[Y-(a+bX)]^2$ $(a,\ b \in R$ 为常数)

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

e 的值越小，$a+bX$ 与 Y 的近似程度越好.

确定 a,b 的值,使 e 达到最小.

均方误差 $e=E[Y-(a+bX)]^2$

分别对a,b求偏导数,并令其等于零,得

$$\begin{cases}\dfrac{\partial e}{\partial a}=2a+2bE(X)-2E(Y)=0,\\ \dfrac{\partial e}{\partial b}=2bE(X^2)-2E(XY)+2aE(X)=0.\end{cases} \Rightarrow \begin{cases}a_0=E(Y)-E(X)\dfrac{Cov(X,Y)}{D(X)}\\ b_0=\dfrac{Cov(X,Y)}{D(X)}\end{cases}$$

将a_0,b_0代入$e=E[Y-(a+bX)]^2$中,得

$$\min_{a,b} e=E[Y-(a_0+b_0X)]^2=\left[1-\left(\boxed{\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}}\right)^2\right]\cdot D(Y)$$

定义3.7 若随机变量X与Y的$Cov(X,Y)$ 存在，且有

$D(X)>0$，$D(Y)>0$，称

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

为X与Y的**相关系数**，记作ρ_{XY}.

Karl Pearson
(1857～1936)
英国数学家，生物统计学家、统计学之父

注

$$\rho_{XY}=E\left\{\left[\underbrace{\frac{X-E(X)}{\sqrt{D(X)}}}_{X^*}\right]\left[\underbrace{\frac{Y-E(Y)}{\sqrt{D(Y)}}}_{Y^*}\right]\right\}=Cov(X^*,Y^*)$$

相关系数：标准协方差

2. 相关系数的意义

$$\min_{a,b} e = \min_{a,b} E[Y-(a+bX)]^2 = \left[1-\rho_{XY}^2\right]\cdot D(Y) \geq 0$$

$|\rho_{XY}|$越大 ⟹ e越小 ⟹ X,Y的线性关系越强；

$|\rho_{XY}|$越小 ⟹ e越大 ⟹ X,Y的线性关系越弱；

$$|\rho_{XY}| \leq 1 \quad \text{定义} \begin{cases} |\rho_{XY}| = 1 \Leftrightarrow X,Y\text{线性相关} \\ \rho_{XY} = 0 \Leftrightarrow X,Y\text{不相关} \\ 0 < |\rho_{XY}| < 1 \Leftrightarrow X,Y\text{弱相关} \end{cases}$$

相关系数刻画了随机变量的线性相关程度

3、相关系数的性质

性质(1) $|\rho_{XY}| \le 1.$

证 设随机变量 $Z = \dfrac{X-E(X)}{\sqrt{D(X)}} \pm \dfrac{Y-E(Y)}{\sqrt{D(Y)}}$

则
$$D(Z) = D\left(\frac{X-E(X)}{\sqrt{D(X)}}\right) + D\left(\frac{Y-E(Y)}{\sqrt{D(Y)}}\right)$$
$$\pm 2\operatorname{cov}\left(\frac{X-E(X)}{\sqrt{D(X)}}, \frac{Y-E(Y)}{\sqrt{D(Y)}}\right)$$

即 $1+1\pm 2\rho_{XY} \ge 0 \Rightarrow |\rho_{XY}| \le 1.$

性质(2) $|\rho_{XY}|=1$的充要条件是,存在常数a,b

使$P\{Y=aX+b\}=1.$

证 **(⇒)** 若$|\rho_{XY}|=1$,由于

$$D\left(\frac{X-E(X)}{\sqrt{D(X)}}\pm\frac{Y-E(Y)}{\sqrt{D(Y)}}\right)=2(1\pm\rho_{XY}),$$

因而 $\rho_{XY}=1$时, $D\left(\frac{X-E(X)}{\sqrt{D(X)}}-\frac{Y-E(Y)}{\sqrt{D(Y)}}\right)=0,$

有 $P\left(\frac{X-E(X)}{\sqrt{D(X)}}=\frac{Y-E(Y)}{\sqrt{D(Y)}}\right)=1.$

$\because D(X)=0$的充要条件是 $P\{X=C\}=1$

$$\rho_{XY} = -1\text{时}, \; D\left(\frac{X - E(X)}{\sqrt{D(X)}} + \frac{Y - E(Y)}{\sqrt{D(Y)}}\right) = 0.$$

因而 $P\left(\dfrac{X - E(X)}{\sqrt{D(X)}} = -\dfrac{Y - E(Y)}{\sqrt{D(Y)}}\right) = 1.$

故当$|\rho_{XY}| = 1$, 有

$$P\left(\frac{X - E(X)}{\sqrt{D(X)}} = \pm\frac{Y - E(Y)}{\sqrt{D(Y)}}\right) = 1,$$

即以概率1成立$Y = aX + b$, 其中

$$a = \pm\sqrt{\frac{D(Y)}{D(X)}}, \; b = E(Y) \mp \sqrt{\frac{D(Y)}{D(X)}}E(X).$$

$(\Leftarrow)$ $P\{Y=aX+b\}=1$

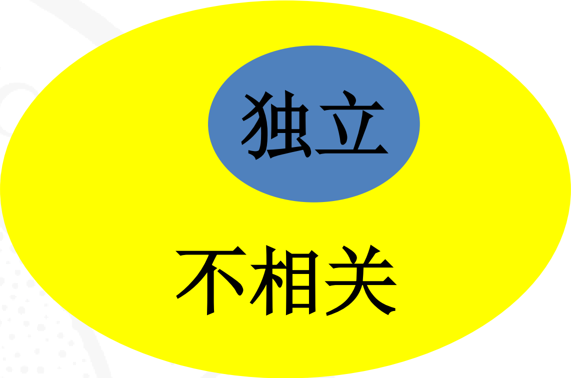
$\therefore\quad E(Y)=aE(X)+b.\quad \therefore\quad D(Y)=a^2E(X).$

$$\therefore\quad \rho=\frac{E[(X-E(X))(Y-E(Y))]}{\sqrt{D(X)}\sqrt{D(Y)}}$$

$$=\frac{aE[X-E(X)]^2}{\sqrt{D(X)}\sqrt{a^2D(X)}}=\frac{a}{|a|}.$$

$\Rightarrow |\rho_{XY}|=1.$

性质(3) 若X与Y相互独立, 则X与Y不相关, 反之不真.



相互独立 $\underset{\not\longleftarrow}{\longrightarrow}$ **不相关**

$$F(x,y)=F_X(x)F_Y(y)$$

$$\rho_{XY}=0;$$

$$Cov(X,Y)=0$$

推论 不相关的充要条件

$$X,Y\text{ 不相关}\Leftrightarrow \rho_{XY}=0;\Leftrightarrow \mathrm{cov}(X,Y)=0;$$

$$\Leftrightarrow E(XY)=E(X)E(Y);$$

$$\Leftrightarrow D(X\pm Y)=D(X)+D(Y).$$

例2 设随机变量X与Y的相关系数为0.5,

$$E(X)=E(Y)=0,\ E(X^2)=E(Y^2)=2,$$

求 $E(X+Y)^2$.

解

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

$$E(X+Y)^2=E(X^2)+2E(XY)+E(Y^2)$$

$$=4+2[\operatorname{cov}(X,Y)+E(X)E(Y)]$$

$$=4+2\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}$$

$$=4+2\times0.5\times2=6.$$

例3 设$(X,Y)\sim N(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2,\rho)$,求$X$与$Y$的相关系数.

解

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

由 $$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp$$

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

$$\Rightarrow E(X)=\mu_1,\ E(Y)=\mu_2,\ D(X)=\sigma_1^2,\ D(Y)=\sigma_2^2.$$

$$\mathrm{cov}(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)p(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

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

$$\mathrm{e}^{-\boxed{\frac{(x-\mu_1)^2}{2\sigma_1^2}}}\,\mathrm{e}^{-\boxed{\frac{1}{2(1-\rho^2)}\left[\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right]^2}}\mathrm{d}y\mathrm{d}x$$

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

$$x-\mu_1=\sigma_1 u,$$

$$y-\mu_2=(t\sqrt{1-\rho^2}+u\rho)\sigma_2,$$

$$dxdy=|J|dudt$$

$$J=\begin{vmatrix} x_u & x_t \\ y_u & y_t \end{vmatrix}=\begin{vmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{vmatrix}$$

$$=\sigma_1\sigma_2\sqrt{1-\rho^2}$$

$$\mathrm{cov}(X,Y)=\qquad (x-\mu_1)(y-\mu_2)\qquad \mathrm{d}x\mathrm{d}y$$

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(\sigma_1\sigma_2\sqrt{1-\rho^2}tu+\rho\sigma_1\sigma_2u^2)\mathrm{e}^{-\frac{t^2+u^2}{2}}\mathrm{d}t\mathrm{d}u$$

$$=\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{2\pi}\left(\int_{-\infty}^{+\infty}u\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u\right)\left(\int_{-\infty}^{+\infty}t\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\right)$$

$$+\frac{\rho\sigma_1\sigma_2}{2\pi}\left(\int_{-\infty}^{+\infty}u^2\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u\right)\left(\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\right)$$

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}\sqrt{2\pi}\cdot\sqrt{2\pi}.$$

故有$\mathrm{cov}(X,Y)=\rho\sigma_1\sigma_2.$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}u^2\mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u=E(U^2),\text{其中}U\sim N(0,1)$$

方法1

$$\because E(U^2)=D(U)+[E(U)]^2=1+0=1$$

$$\therefore \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}u^2\mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u=1$$

$$\therefore \int_{-\infty}^{+\infty}u^2\mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u=\sqrt{2\pi}$$

其中$\int_{-\infty}^{+\infty} u^2 \mathrm{e}^{-\frac{u^2}{2}} \mathrm{d}u = 2\int_0^{+\infty} u^2 \mathrm{e}^{-\frac{u^2}{2}} \mathrm{d}u$

方法2

(令$\frac{u^2}{2} = x, u = \sqrt{2x}, du = \frac{1}{2} \cdot \frac{2}{\sqrt{2x}} dx$)

$$= 2\int_0^{+\infty} 2xe^{-x} \frac{1}{\sqrt{2x}} dx$$

$$= 2\sqrt{2}\boxed{\int_0^{+\infty} x^{\frac{1}{2}} e^{-x} dx}$$

$$= 2\sqrt{2}\Gamma(\frac{3}{2}) = 2\sqrt{2} \cdot \frac{1}{2}\Gamma(\frac{1}{2}) = \sqrt{2\pi}$$

于是 $\rho_{XY}=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\rho.$

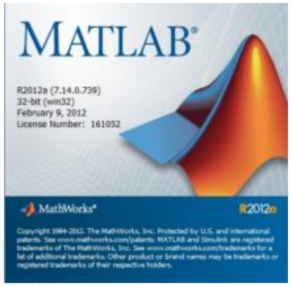
注1° 二维正态分布密度函数中, **参数 ρ 代表了 X 与 Y 的相关系数;**

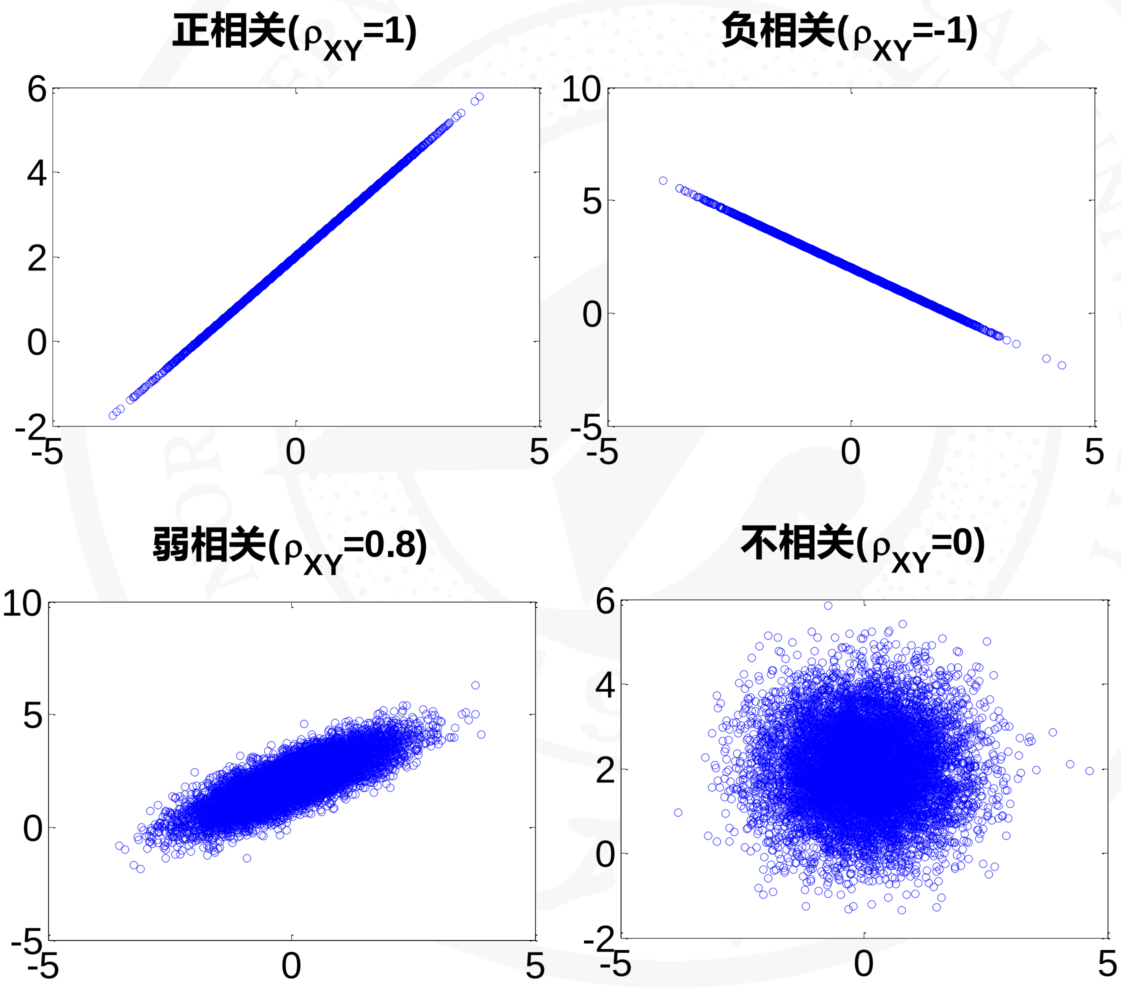
2° 对于二维随机变量(X, Y),

$$\rho=0 \Leftrightarrow X\text{与}Y\text{相互独立}. \Leftrightarrow \rho_{XY}=0.$$

(证明见p41例2.12)

$$(X,Y) \sim N(0,1,0,1,\rho_{XY})$$





例4 设θ服从$[0,2\pi]$上的均匀分布，$X=\cos\theta$，$Y=\cos(\theta+a)$，这里a是定数，求X和Y的相关系数？

解

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

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

$$D(X)=E(X^2)-[E(X)]^2$$

$$\because E(X)=\frac{1}{2\pi}\int_0^{2\pi}\cos\theta\,\mathrm{d}\theta=0,\quad E(X^2)=\frac{1}{2\pi}\int_0^{2\pi}\cos^2\theta\,\mathrm{d}\theta=\frac{1}{2},$$

$$E(Y)=\frac{1}{2\pi}\int_0^{2\pi}\cos(\theta+a)\,\mathrm{d}\theta=0,\ E(Y^2)=\frac{1}{2\pi}\int_0^{2\pi}\cos^2(\theta+a)\,\mathrm{d}\theta=\frac{1}{2}$$

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

$$\therefore \operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)=\frac{1}{2}\cos a,$$

$$D(X)=E(X^2)-[E(X)]^2=\frac{1}{2},$$

$$D(Y)=E(Y^2)-[E(Y)]^2=\frac{1}{2},$$

$$\therefore \rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\cos a.$$

由 $X=\cos\theta, Y=\cos(\theta+a),$

$$\rho_{XY}=\cos a,$$

可知:

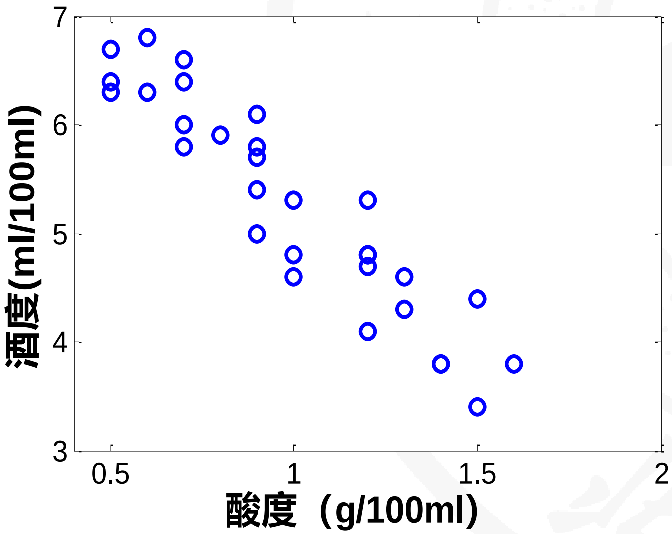
$$\left.\begin{array}{l}\text{当} a=0 \text{时}, \rho=1, X=Y, \\ \text{当} a=\pi \text{时}, \rho=-1, X=-Y,\end{array}\right\} \text{存在**线性关系**.}$$

当 $a=\dfrac{\pi}{2}$ 或 $a=\dfrac{3\pi}{2}$ 时, $\rho=0$, X 与 Y **不相关**;

但 $X^2+Y^2=1$, 因此, X 与 Y 不独立.

引例 某酒厂要判定酒醅中酸度和酒度的关系，通过检验测得以下数据：

散点图



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

$$=\frac{E(XY)-E(X)E(Y)}{\sqrt{EX^2-[E(X)]^2}\cdot\sqrt{EY^2-[E(Y)]^2}}$$

$$p(x,y)$$

$$E(X)=\int_{-\infty}^{+\infty}xp(x,y)dy;\quad E(Y)=\int_{-\infty}^{+\infty}yp(x,y)dx$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2p(x,y)dy;\quad E(Y^2)=\int_{-\infty}^{+\infty}y^2p(x,y)dx$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyp(x,y)dxdy$$

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

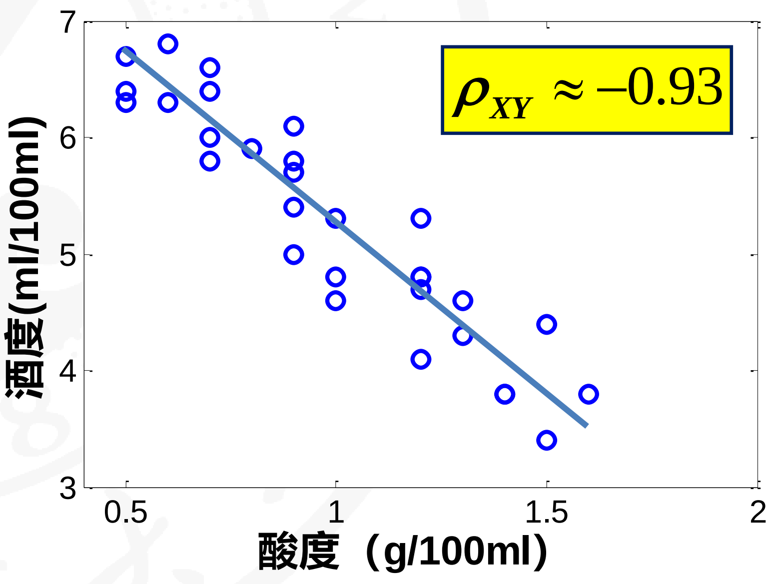
$$= \frac{E(XY) - E(X)E(Y)}{\sqrt{EX^2 - [E(X)]^2}\cdot\sqrt{EY^2 - [E(Y)]^2}}$$

$$\rho = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i y_i - \overline{x}\,\overline{y}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \overline{x}^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \overline{y}^2}} \approx -0.93$$

------- *Pearson*相关系数

酸度和酒度近似呈现负线性关系

散点图



$$Y = -0.93X + a_0$$

三、协方差矩阵

二维随机变量的协方差矩阵

设(X_1, X_2)为二维随机变量,其协方差矩阵为

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} DX_1 & \text{cov}(X_1, X_2) \\ \text{cov}(X_2, X_1) & DX_2 \end{pmatrix},$$

其中

$$\sigma_{11} = E[X_1 - E(X_1)]^2 = D(X_1),$$

$$\sigma_{21} = E\{[X_2 - E(X_2)][X_1 - E(X_1)]\} = \sigma_{12},$$

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

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

本节测试

1.

2.

3.

课后思考

搜集100位患者的化验单，分析红细胞数和血小板的相关性。

撰写一篇科技小论文——告诉大家你的结论！

内容小结

1. 协方差与相关系数的定义及计算

X与Y的协方差：

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

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

离散 $$\operatorname{cov}(X,Y)=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}x_i y_j p_{i,j}-E(X)E(Y),$$

连续 $$\operatorname{cov}(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyp(x,y)\mathrm{d}x\mathrm{d}y-E(X)E(Y)$$

X与Y的相关系数：

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

$$X'=\frac{X-EX}{\sqrt{DX}},\quad Y'=\frac{Y-EY}{\sqrt{DY}}$$

$$\Rightarrow\quad E(X')=0,\quad D(X')=1$$

2、协方差的性质

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

性质(2) $\operatorname{cov}(aX,bY)=ab\operatorname{cov}(X,Y)$，$a,b$为常数.

性质(3) $\operatorname{cov}(X_1+X_2,Y)=\operatorname{cov}(X_1,Y)+\operatorname{cov}(X_2,Y).$

性质(4) 若X与Y独立，则$\operatorname{cov}(X,Y)=0.$

性质(5) $D(X\pm Y)=D(X)+D(Y)\pm 2\operatorname{cov}(X,Y).$

推广 $$D\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}D(X_i)+2\sum_{i<j}\sum\operatorname{cov}(X_i,X_j).$$

结论：$(X,Y)\sim N(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2,\rho)\Rightarrow Cov(X,Y)=\rho$

3. 相关系数的意义与性质

衡量X与Y线性相关密切程度的指标

当$|\rho_{XY}|$接近1时,表明X,Y的线性关系联系较紧密.

当$|\rho_{XY}|$接近0时,X,Y线性相关的程度较差.

$\rho_{XY}=0$,则称X和Y不相关.

性质(1) $|\rho_{XY}|\le 1.$

性质(2) $|\rho_{XY}|=1$的充要条件是,存在常数a,b使$P\{Y=aX+b\}=1.$

性质(3) 相互独立 $\underset{\not\longleftarrow}{\longrightarrow}$ 不相关

特例：$(X,Y) \sim N(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2,\rho)$

$$\rho = 0 \Leftrightarrow X\ Y\text{相互独立} \Leftrightarrow \text{不相关}$$

4.不相关的充要条件

$$X, Y\text{ 不相关} \Leftrightarrow \rho_{XY} = 0;$$

$$\Leftrightarrow \operatorname{cov}(X,Y) = 0;$$

$$\Leftrightarrow E(XY) = E(X)E(Y);$$

$$\Leftrightarrow D(X+Y) = D(X) + D(Y).$$

3-3 协方差与相关函数

Thank You !

备用题

例1-1 设X与Y是两个随机变量, 且$D(X)=1$, $D(Y)=3$, $\text{cov}(X,Y)=-0.3$, 求方差$D(X+Y)$与$D(2X-3Y)$.

解

$$D(X+Y)=D(X)+D(Y)+2\text{cov}(X,Y)=3.4,$$

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

$$=4D(X)+9D(Y)-12\text{cov}(X,Y)$$

$$=34.6.$$

例1-2 设随机变量X和Y均服从参数$\lambda=1/2$的指数分布，且相关系数$\rho_{XY}=1/2$，令函数$U=2X$，$V=X-Y$，求U与V的协方差 $\operatorname{cov}(U,V)$.

解 由随机变量X和Y均服从参数$\lambda=1/2$的指数分布，则

$$D(X)=4,\ D(Y)=4.$$

而

$$\operatorname{cov}(X,Y)=\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}=2,$$

$$\operatorname{cov}(U,V)=\operatorname{cov}(2X,X-Y)$$

$$=2D(X)-2\operatorname{cov}(X,Y)=4.$$

例2-1 设二维连续型随机变量(X, Y)的联合密度函数为

$$p(x,y)=\begin{cases}\frac{1}{3}(x+y), & 0\le x\le 1, 0\le y\le 2\\ 0, & \text{其他}\end{cases}$$

试计算$D(2X-3Y+8)$.

解 由**性质3.16**得

$$D(2X-3Y+8)=D(2X)+D(3Y)-2\operatorname{cov}(2X,3Y)$$
$$=4D(X)+9D(Y)-12\operatorname{cov}(X,Y)$$

为了计算上述方差和协方差, 需要先计算$E(X)$, $E(X^2)$, $E(Y)$, $E(Y^2)$和$E(XY)$. 为此, 先计算X和Y的边缘分布.

$$p_Y(y)=\int_0^1 \frac{1}{3}(x+y)\mathrm{d}x=\frac{1}{3}\left(y+\frac{1}{2}\right)\quad (0\le y\le 2)$$

$$p_X(x)=\int_0^2 \frac{1}{3}(x+y)\mathrm{d}y=\frac{2}{3}(x+1)\quad (0\le x\le 1)$$

由此计算得

$$E(X)=\int_0^1 \frac{2}{3}x(x+1)\mathrm{d}x=\frac{2}{3}\times\left(\frac{1}{3}+\frac{1}{2}\right)=\frac{5}{9}$$

$$E\left(X^2\right)=\int_0^1 \frac{2}{3}x^2(x+1)\mathrm{d}x=\frac{2}{3}\times\left(\frac{1}{4}+\frac{1}{3}\right)=\frac{7}{18}$$

$$D(X)=\frac{7}{18}-\frac{25}{81}=\frac{13}{162}$$

$$E(Y)=\int_0^2 \frac{1}{3}y\left(\frac{1}{2}+y\right)\mathrm{d}y=\frac{11}{9}$$

$$E\left(Y^2\right)=\int_0^2 \frac{1}{3}y^2\left(\frac{1}{2}+y\right)\mathrm{d}y=\frac{16}{9}$$

$$D(Y)=\frac{16}{9}-\left(\frac{11}{9}\right)^2=\frac{23}{81}$$

$$E(XY)=\frac{1}{3}\int_0^1\int_0^2 xy(x+y)\mathrm{d}y\mathrm{d}x$$

$$=\frac{1}{3}\int_0^1 xy\left(2x^2+\frac{8}{3}y\right)\mathrm{d}x=\frac{2}{3}$$

于是可得协方差

$$\operatorname{cov}(X,Y)=E(XY)-EX\cdot EY=\frac{2}{3}-\frac{5}{9}\times\frac{11}{9}=-\frac{1}{81}$$

代回原式, 可得

$$D(2X-3Y+8)$$

$$=4\times\frac{13}{162}+9\times\frac{23}{81}-12\times\left(-\frac{1}{81}\right)=\frac{245}{81}\approx 3$$

例2-2 设二维连续型随机变量(X, Y)的联合密度函数为

$$p(x,y)=\begin{cases}\dfrac{2}{x^3}\mathrm{e}^{1-y}, & x\ge 1, y\ge 1\\ 0, & \text{其他}\end{cases}$$

求函数$W=XY$与$Z=\dfrac{Y}{X}$的协方差$\mathrm{cov}(W,Z)$.

解

$$\mathrm{cov}(W,Z)=E(WZ)-E(W)E(Z)$$

$$E(W)=\int_1^{+\infty}\mathrm{d}x\int_1^{+\infty}xy\frac{2}{x^3}\mathrm{e}^{1-y}\mathrm{d}y=4,$$

$$E(Z)=\int_1^{+\infty}\mathrm{d}x\int_1^{+\infty}\frac{y}{x}\frac{2}{x^3}\mathrm{e}^{1-y}\mathrm{d}y=\frac{4}{3},$$

$$E(WZ)=\int_1^{+\infty}\mathrm{d}x\int_1^{+\infty}y^2\frac{2}{x^3}\mathrm{e}^{1-y}\mathrm{d}y=5.$$

由协方差公式得

$$\operatorname{cov}(W,Z)=E(WZ)-E(W)E(Z)=-\frac{1}{3}.$$

例2-3 已知随机变量X,Y分别服从$N(1,3^2)$, $N(0,4^2)$, 且$\rho_{XY}=-1/2$, 设 $Z=X/3+Y/2$.

(1)求Z的数学期望和方差;

(2) 求X与Z的相关系数;

解 (1) 由 $E(X)=1,\ D(X)=9,\ E(Y)=0,\ D(Y)=16.$

文件中找不到关系 ID 为 rId7 的图像部件。

$$=\frac{1}{3}E(X)+\frac{1}{2}E(Y)=\frac{1}{3}.$$

$$D(Z)=D\left(\frac{X}{3}\right)+D\left(\frac{Y}{2}\right)+2\mathrm{cov}\left(\frac{X}{3},\frac{Y}{2}\right)$$

$$=\frac{1}{9}D(X)+\frac{1}{4}D(Y)+\frac{1}{3}\mathrm{cov}(X,Y)$$

$$=\frac{1}{9}D(X)+\frac{1}{4}D(Y)+\frac{1}{3}\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}$$

$$=1+4-2=3.$$

$$\text{(2)}\ \operatorname{cov}(X,Z)=\operatorname{cov}\left(X,\frac{X}{3}+\frac{Y}{2}\right)$$

$$=\frac{1}{3}\operatorname{cov}(X,X)+\frac{1}{2}\operatorname{cov}(X,Y)$$

$$=\frac{1}{3}D(X)+\frac{1}{2}\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}$$

$$=0$$

故 $\rho_{XZ}=\dfrac{\operatorname{cov}(X,Z)}{\sqrt{D(X)}\sqrt{D(Z)}}=0.$

例2-4 设 $E(X)=-2,\ E(Y)=2,\ D(X)=1,$ $D(Y)=4, \rho_{XY}=-0.5,$ 试根据切比谢夫不等式估计：$P\{|X+Y|\geq 6\}.$

解 $E(X+Y)=E(X)+E(Y)=0,$

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

$$=D(X)+D(Y)+2\sqrt{D(X)}\sqrt{D(Y)}\rho_{XY}$$

$$=1+4+2\times 1\times 2\times(-0.5)=3,$$

$$P\{|X+Y|\geq 6\}=P\{|(X+Y)-E(X+Y)|\geq 6\}$$

$$\leq \frac{D(X+Y)}{6^2}=\frac{1}{12}.$$

$$\boxed{P\{|X-E(X)|\geq \varepsilon\}\leq \frac{D(X)}{\varepsilon^2}}$$

例4-1 设随机变量X, Y有方差, 求证: 随机变量$U=X+Y$与$V=X-Y$不相关的充分必要条件为$D(X)=D(Y)$.

解 因为
$$
\begin{aligned}
\operatorname{cov}(U,V) &= \operatorname{cov}(X+Y, X-Y) \\
&= \operatorname{cov}(X,X)-\operatorname{cov}(Y,Y) \\
&= D(X)-D(Y).
\end{aligned}
$$

因此, $\operatorname{cov}(U,V)=0$ 的充要条件是 $D(X)=D(Y)$.

例4-2 设掷三次均匀硬币，随机变量X表示出现的正面次数，Y表示正面次数与反面次数的差的绝对值，

(1) X与Y是否不相关?

(2) X与Y是否相互独立?

解 计算概率，得联合分布律

Y \ X	0	1	2	3
1	0	3/8	3/8	0
3	1/8	0	0	1/8

计算行和与列和, 得边缘分布律

$$X\sim\begin{pmatrix}0 & 1 & 2 & 3\\ 1/8 & 3/8 & 3/8 & 1/8\end{pmatrix},\ Y\sim\begin{pmatrix}1 & 3\\ 3/4 & 1/4\end{pmatrix}.$$

由于 $P\{X=0,Y=1\}\neq P\{X=0\}P\{Y=1\}$,

因此, X与Y不相互独立. 又因为

$$E(X)=3/2, E(Y)=3/2, E(XY)=9/4.$$

由协方差公式得

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

于是X与Y不相关.

一、协方差的概念与性质

1. 问题的提出

数学期望：随机变量的均值

方差：随机变量的对期望的离散程度

$$X \sim N(\mu, \sigma^2)$$

$$\Rightarrow X + X \quad \sim N(2\mu, 2\sigma^2) \times \quad \sim N(2\mu, 4\sigma^2) \ \surd$$

$$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$$

若X和Y相互独立，$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

二、相关系数的意义与性质

1. 问题的提出

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

分析 设 $\mathrm{e} = E[Y-(a+bX)]^2$

则 e 可用来衡量 $a+bX$ 近似表达 Y 的好坏程度.

当 e 的值越小,表示 $a+bX$ 与 Y 的近似程度越好.

确定 a,b 的值,使 e 达到最小.

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

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

$$- 2aE(Y).$$

将 e 分别关于 a,b 求偏导数,并令它们等于零,得

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

解之得 $b_0=\dfrac{\operatorname{cov}(X,Y)}{D(X)}$,

$$a_0=E(Y)-E(X)\frac{\operatorname{cov}(X,Y)}{D(X)}.$$

将 a_0,b_0 代入 $\mathrm{e}=E[Y-(a+bX)]^2$ 中,得

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

$$=D(Y)-\frac{\operatorname{cov}^2(X,Y)}{D(X)}=[1-\frac{\operatorname{cov}^2(X,Y)}{D(X)D(Y)}]\cdot D(Y)$$

$$=(1-\rho_{XY}^2)D(Y).$$