# 第九周 协方差与相关系数

## 9.4 相关与独立

### 相关与独立

随机变量 $X,Y$ 独立时，$Cov(X,Y)=E(XY)-E(X)E(Y)=E(X)E(Y)-E(X)E(Y)=0$

随机变量 $X,Y$ 独立 $\Rightarrow$ $X,Y$ 不相关

随机变量 $X,Y$ 不相关 $\not\Rightarrow$ $X,Y$ 相互独立

*************************************************************

例 9.4.1 把一枚均匀硬币抛掷三次，设 $X$ 为三次抛掷中正面出现的次数，而 $Y$ 为正面出现的次数与反面出现的次数之差的绝对值。试求 $X$ 与 $Y$ 的联合分布律以及 $X$ 与 $Y$ 的相关系数，并判断 $X$ 与 $Y$ 是否独立？

解： $X$ 可能取值为 0，1，2，3，而 $Y$ 的可能取值为 1，3，且

$P(X=0,Y=3)=\frac{1}{8}$，$P(X=1,Y=1)=\frac{3}{8}$，$P(X=2,Y=1)=\frac{3}{8}$，$P(X=3,Y=3)=\frac{1}{8}$，

其余的均为 0，因此，$X$ 与 $Y$ 联合分布律和边缘分布律为：

| $Y\backslash X$ | 0 | 1 | 2 | 3 | $P(Y=y)$ |
|---|---|---|---|---|---|
| 1 | 0 | 3/8 | 3/8 | 0 | 3/4 |
| 3 | 1/8 | 0 | 0 | 1/8 | 1/4 |
| $P(X=x)$ | 1/8 | 3/8 | 3/8 | 1/8 | 1 |

从而

$$E(X)=0\times\frac{1}{8}+1\times\frac{3}{8}+2\times\frac{3}{8}+3\times\frac{1}{8}=\frac{3}{2}，\quad E(X^2)=0^2\times\frac{1}{8}+1^2\times\frac{3}{8}+2^2\times\frac{3}{8}+3^2\times\frac{1}{8}=3，$$

$$Var(X)=E(X^2)-E(X)^2=3-\left(\frac{3}{2}\right)^2=\frac{3}{4}。$$

$$E(Y)=1\times\frac{6}{8}+3\times\frac{2}{8}=\frac{3}{2}，\quad E(Y^2)=1^2\times\frac{6}{8}+3^2\times\frac{2}{8}=3，$$

$Var\left(Y\right)=E\left(Y^2\right)-E\left(Y\right)^2=3-(\frac{3}{2})^2=\frac{3}{4}$

而 $E(XY)=1\times1\times\frac{3}{8}+2\times1\times\frac{3}{8}+0\times3\times\frac{1}{8}+3\times3\times\frac{1}{8}=\frac{9}{4}$。

故 $Corr\left(X,Y\right)=\frac{E(XY)-E\left(X\right)E\left(Y\right)}{\sqrt{Var\left(X\right)\cdot Var\left(Y\right)}}=\frac{\frac{9}{4}-\frac{3}{2}\times\frac{3}{2}}{\sqrt{\frac{3}{4}\times\frac{3}{4}}}=0$，即 $X$ 与 $Y$ 不相关。

又由于 $0=P(X=0,Y=1)\neq P(X=0)P(Y=1)=\frac{1}{8}\times\frac{6}{8}$，所以 $X$ 与 $Y$ 不独立。

*************************************************************

例 9.4.2 考虑 $X\sim N\left(0,1\right)$ 与 $Y=X^2$ 的相关性和独立性。

解：显然 $X,Y$ 不独立 （思考：对不独立的关系是否有直观上的理解），

利用概率定义验证：$\left(X,Y\right)$ 的联合密度函数 $f\left(x,y\right)=\begin{cases}\varphi\left(x\right), & x\in R, y=x^2\\ 0, & \text{其他}\end{cases}$，

抛物线 $y=x^2$ 以外的点 $\left(x,y\right)$，联合密度 $f\left(x,y\right)=0$；

而 $X,Y$ 的边缘密度 $f_X\left(x\right)$ 和 $f_Y\left(y\right)$ 均不为 0，对抛物线 $y=x^2$ 以外的点

$f\left(x,y\right)=f_X\left(x\right)\cdot f_Y\left(y\right)$ 均不成立，所以 $X,Y$ 不独立。

$Cov\left(X,Y\right)=Cov\left(X,X^2\right)=E\left(X^3\right)-E\left(X\right)E\left(X^2\right)=0$， $X,Y$ 不相关。

*************************************************************

$X\sim N\left(0,1\right)$ 和 $Y=X^2$， $X,Y$ 不相关，但它们显然具有很强的关联。

实际上，相关系数反映的是随见变量之间在线性关系意义下的相关程度。

定理： $Corr\left(X,Y\right)=\pm1$ 的充要条件是 $X,Y$ 之间几乎处处有线性关系，即存在常数 $a,b$，使得 $P\left(Y=aX+b\right)=1$。

所以也称（线性）相关系数，不相关指的是不存在线性相关的关系，相关系数并不能

有效地表达非线性的相关关系。

************************************************************

**定理：** $Corr(X,Y)=\pm 1$ 的充要条件是 $X,Y$ 之间几乎处处有线性关系，即存在常数 $a,b$，使得 $P(Y=aX+b)=1$。

**充分性** $Y=aX+b \Rightarrow Var(Y)=a^2 Var(X) \Rightarrow \sigma_Y=|a|\cdot\sigma_X$

$$Y=aX+b \Rightarrow Cov(X,Y)=a\cdot Cov(X,X)=a\cdot Var(X)=a\cdot{\sigma_X}^2$$

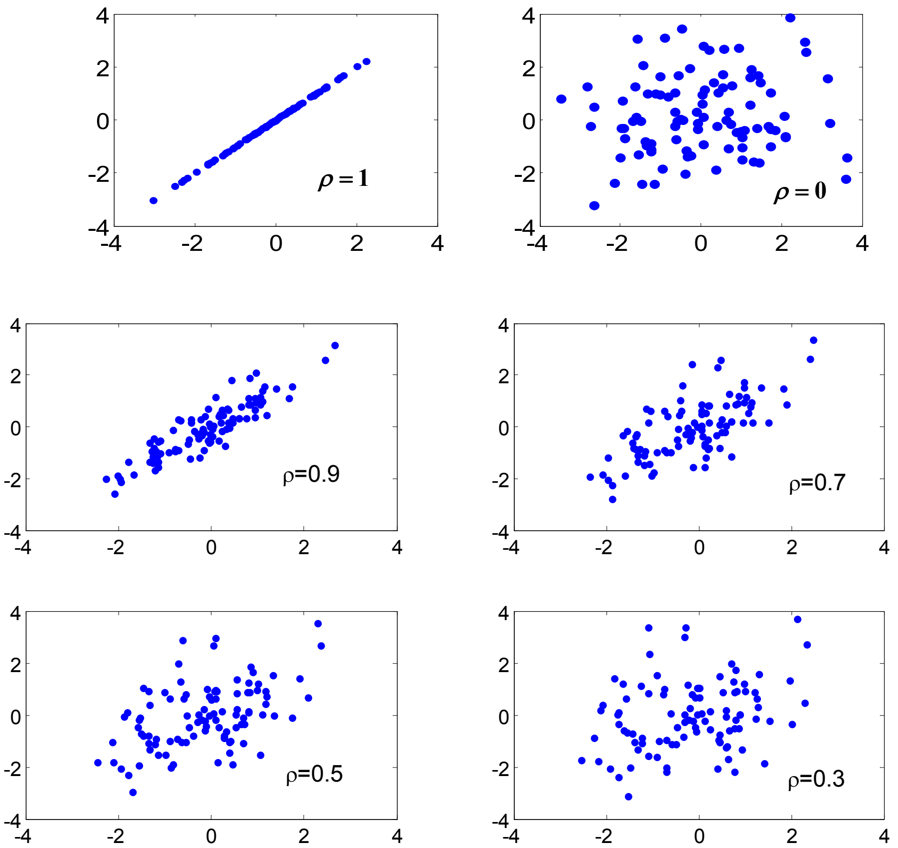
$$Corr(X,Y)=\frac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y}=\frac{a\cdot{\sigma_X}^2}{|a|\cdot{\sigma_X}^2}=\frac{a}{|a|}=\pm 1。$$

**必要性** $Var\left(\frac{X}{\sigma_X}\pm\frac{Y}{\sigma_Y}\right)=Var\left(\frac{X}{\sigma_X}\right)+Var\left(\frac{Y}{\sigma_Y}\right)\pm 2Cov\left(\frac{X}{\sigma_X},\frac{Y}{\sigma_Y}\right)=2(1\pm Corr(X,Y))$

$$Corr(X,Y)=\pm 1\Rightarrow Var\left(\frac{X}{\sigma_X}\mp\frac{Y}{\sigma_Y}\right)=0\Rightarrow P\left(\frac{X}{\sigma_X}\mp\frac{Y}{\sigma_Y}=c\right)=1$$

************************************************************

不同相关程度的示意



**********************************************************