第九周 协方差与相关系数

9.3 相关系数

相关系数 二元随机变量(X,Y)，$Var(X)\cdot Var(Y)>0$，则X,Y的相关系数定义为

$$Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)}\cdot\sqrt{Var(Y)}}=\frac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y},$$

将随机变量做方差为 1 的标准化：$Var\left(\frac{X}{\sigma_X}\right)=1$，$Var\left(\frac{Y}{\sigma_Y}\right)=1$，

相关系数是将随机变量做方差为 1 的标准化后的协方差，$Corr(X,Y)=Cov\left(\frac{X}{\sigma_X},\frac{Y}{\sigma_Y}\right)$

性质： $-1\le Corr(X,Y)\le 1$。

****

定理： $[Cov(X,Y)]^2\le Var(X)\cdot Var(Y)=\sigma_X{}^2\cdot\sigma_Y{}^2$

证明：对任意参数$t\in R$， $Var(tX+Y)=Var(X)\cdot t^2+2Cov(X,Y)\cdot t+Var(Y)\ge 0$

二次函数$ax^2+bx+c\ge 0$，则其根的判别式$b^2-4ac\le 0$，

所以，上述关于t的二次函数的判别式小于等于 0，

$\left[2Cov(X,Y)\right]^2-4Var(X)\cdot Var(Y)\le 0$，即

$[Cov(X,Y)]^2\le Var(X)\cdot Var(Y)=\sigma_X{}^2\cdot\sigma_Y{}^2$

****

相关系数的绝对值不会大于 1

$$[Cov(X,Y)]^2\le Var(X)\cdot Var(Y)=\sigma_X{}^2\cdot\sigma_Y{}^2\ \Rightarrow\ \left|\frac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y}\right|\le 1$$

$$\Rightarrow \quad |Corr(X,Y)| = \left|\frac{Cov(X,Y)}{\sqrt{Var(X)}\cdot\sqrt{Var(Y)}}\right| = \left|\frac{Cov(X,Y)}{\sigma_X\cdot\sigma_Y}\right| \le 1$$

$$\Rightarrow \quad -1 \le Corr(X,Y) \le 1$$

$Corr(X,Y)>0$ 正相关，$Corr(X,Y)<0$ 负相关，$Corr(X,Y)=0$ 不相关。

****

相关系数的性质

(1) $Corr(X,a)=0$，　　(2) $Corr(X,Y)=Corr(Y,X)$，

(3) $Corr(c_1X+a,c_2Y+b)=\begin{cases} Corr(X,Y), & c_1c_2>0 \\ -Corr(X,Y), & c_1c_2<0 \\ 0, & c_1c_2=0 \end{cases}$，

$$Corr(c_1X+a,c_2Y+b)=\frac{Cov(c_1X+a,c_2Y+b)}{\sqrt{Var(c_1X+a)}\cdot\sqrt{Var(c_2Y+b)}}$$

$$=\frac{c_1\cdot c_2\cdot Cov(X,Y)}{|c_1\cdot c_2|\cdot\sqrt{Var(X)}\cdot\sqrt{Var(Y)}}=\frac{c_1\cdot c_2}{|c_1\cdot c_2|}Corr(X,Y)$$

****

例 9.3.1 随机变量 $X\sim U[0,1]$，若 $Y=X^2$，试求 $Corr(X,Y)$。

解：由均匀分布的数学期望与方差的结论知 $E(X)=\frac{1}{2}$，$Var(X)=\frac{1}{12}$

且 $E(X^n)=\int_0^1 x^n dx=\frac{1}{n+1}$.

所以 $E(Y)=E(X^2)=\frac{1}{3}$，$E(Y^2)=E(X^4)=\frac{1}{5}$，

于是 $Var(Y)=E(Y^2)-(E(Y))^2=\frac{1}{5}-\frac{1}{9}=\frac{4}{45}$，

$$Corr(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}=\frac{E(XY)-E(X)E(Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}=\frac{E(X^3)-E(X)E(Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

$$=\frac{\frac{1}{4}-\frac{1}{2}\times\frac{1}{3}}{\sqrt{\frac{1}{12}}\times\sqrt{\frac{4}{45}}}=\frac{\sqrt{15}}{4}\approx 0.968\text{。}$$

****

例 9.3.2 计算二维随机变量 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的相关系数 $Corr(X,Y)$。

解： 考虑 (X,Y) 的联合密度函数，

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

若令 $X_1=X-\mu_1$， $Y_1=Y-\mu_2$，则 (X_1,Y_1) 的联合密度函数为

$$f(x_1,y_1)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{x_1^{\;2}}{\sigma_1^2}-2\rho\frac{x_1y_2}{\sigma_1\sigma_2}+\frac{y_2^{\;2}}{\sigma_2^2}\right]\right\}$$

所以 $(X_1,Y_1)\sim N(0,0,\sigma_1^2,\sigma_2^2,\rho)$，根据例 8.4.2，有 $E(X_1Y_1)=\rho\sigma_1\sigma_2$。

$$Cov(X_1,Y_1)=E(X_1Y_1)-E(X_1)E(Y_1)=\rho\sigma_1\sigma_2,$$

$$Corr(X,Y)=Corr(X-\mu_1,Y-\mu_2)=Corr(X_1,Y_1)=\frac{Cov(X_1,Y_1)}{\sqrt{Cov(X_1)}\cdot\sqrt{Cov(Y_1)}}=\frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2}=\rho\text{。}$$

所以，二维正态分布的 5 个参数都有明确的概率意义，μ_1,μ_2 为 X,Y 的期望，σ_1^2,σ_2^2 为 X,Y 的方差，而 ρ 则为 X,Y 的相关系数。

****