相关系数及
不相关与相互独立
的关系

1. 相关系数的定义

问题： 新生儿身高 X 与体重 Y 之间的协方差，采用单位(cm ，g)与(m ，kg)，前者的协方差是后者的100000倍！

$$\mathrm{cov}(X^*, Y^*) = E(X^* Y^*) \triangleq \rho(X, Y)$$

定义4.4 设 (X, Y) 是随机向量，当 $D(X) > 0, D(Y) > 0$ 时，称

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

为 X 与 Y 的**相关系数**，标准化协方差.

注： (1)

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

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

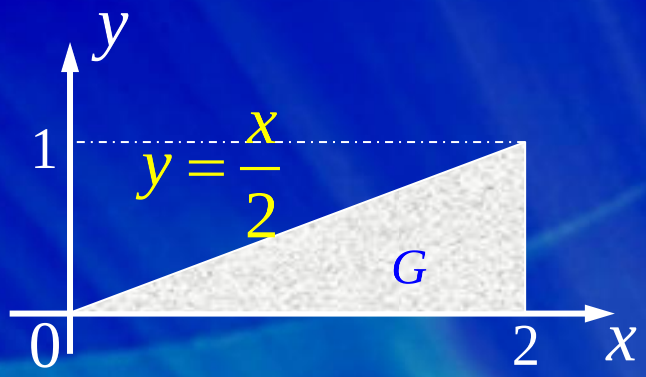
$$\pm 2\rho(X,Y)\sqrt{D(X)D(Y)}.$$

$$\operatorname{cov}(X,Y)$$

例1 设(X,Y)的联合密度函数为

$$f(x,y)=\begin{cases}2xy & 0<2y<x<2\\ 0 & \text{其余}\end{cases}$$

试求相关系数$\rho(X,Y)$.



解 已经求出 $E(X)=\dfrac{8}{5},E(Y)=\dfrac{8}{15},\operatorname{cov}(X,Y)=\dfrac{8}{225}$, 又

$$E(X^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^2\cdot f(x,y)\mathrm{d}x\mathrm{d}y=\int_0^2\mathrm{d}x\int_0^{\frac{x}{2}}x^2\cdot 2xy\mathrm{d}y=\frac{8}{3}$$

$$E(Y^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{\infty}y^2\cdot f(x,y)\mathrm{d}x\mathrm{d}y=\int_0^2 dx\int_0^{\frac{x}{2}}y^2\cdot 2xydy=\frac{1}{3}$$

故 $D(X)=E\left(X^2\right)-E^2(X)=\dfrac{8}{75}$, 同理 $D(Y)=\dfrac{11}{225}$,故

$$\rho(X,Y)=\frac{\frac{8}{225}}{\sqrt{\frac{8}{75}}\cdot\sqrt{\frac{11}{225}}}=\frac{4}{\sqrt{66}}\approx 0.492.$$

$E(X),E(Y),$
$E(X^2),E(Y^2),$
$E(XY)$

例2 设(X,Y)的联合概率函数为

$X \backslash Y$	1	0	2
-1	$1/6$	0	$1/6$
0	0	$1/6$	$1/6$
1	$1/6$	$1/6$	0

试求相关系数$\rho(X,Y)$.

解　已经求出 $E(X)=0, E(Y)=1, \operatorname{cov}(X,Y)=-\dfrac{1}{3}$，又

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

所以 $$D(X)=\frac{2}{3}, D(Y)=\frac{2}{3}$$

$$\Rightarrow \rho(X,Y)=\frac{-\frac{1}{3}}{\sqrt{\frac{2}{3}}\cdot\sqrt{\frac{2}{3}}}=-\frac{1}{2}.$$

经过计算可以得到如下结论

结论　二维正态随机变量 $(X,Y)\sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$ 中 X 与 Y 的相关系数

$$\rho\left(X,Y\right)=\rho.$$

2.相关系数的性质

定理4.5　相关系数的性质： 当 $D(X)>0, D(Y)>0$ 时，

(1) $\rho(X,Y)=\rho(Y,X)$；

(2) $|\rho(X,Y)|\le 1$；

(3) $|\rho(X,Y)|=1$ 的**充分必要条件**是：存在不为零的常数 k 与常数 c，使得 $P(Y=kX+c)=1$.

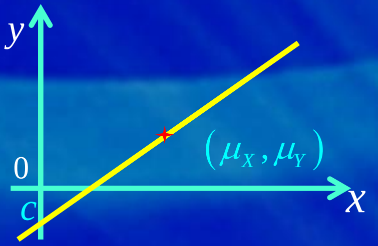
证明: (2) $D\left(X^*\pm Y^*\right)=D\left(X^*\right)+D\left(Y^*\right)\pm 2\operatorname{cov}\left(X^*,Y^*\right)$

$$=2\pm 2\rho\left(X,Y\right)\ge 0\Rightarrow\left|\rho\left(X,Y\right)\right|\le 1$$

(3) $\rho\left(X,Y\right)=1\Leftrightarrow D\left(X^*-Y^*\right)=0\Leftrightarrow P\left(X^*-Y^*=0\right)=1$

$$\Leftrightarrow P\left(\frac{X-\mu_X}{\sigma_X}-\frac{Y-\mu_Y}{\sigma_Y}=0\right)=1\Leftrightarrow P\left(Y=\frac{\sigma_Y}{\sigma_X}\left(X-\mu_X\right)+\mu_Y\right)=1$$

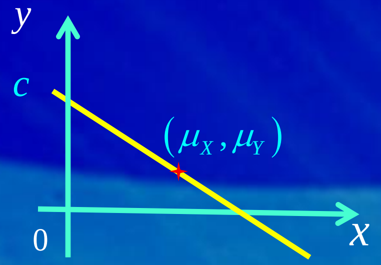
$k=\dfrac{\sigma_Y}{\sigma_X}>0,\ c=-\dfrac{\sigma_Y}{\sigma_X}\mu_X+\mu_Y$ 该直线经过点 $\left(\mu_X,\mu_Y\right)$



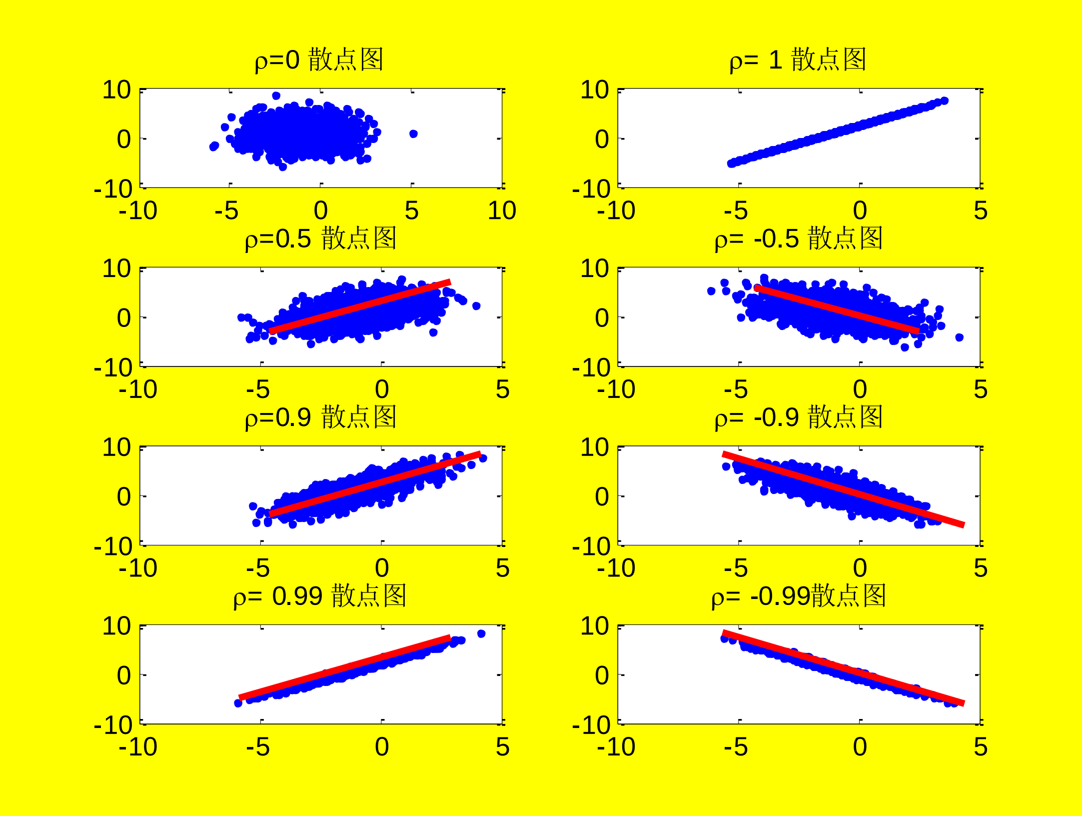
同理可证

$$\rho(X,Y)=-1 \Leftrightarrow P\left(Y=-\frac{\sigma_Y}{\sigma_X}(X-\mu_X)+\mu_Y\right)=1$$

$k=-\frac{\sigma_Y}{\sigma_X}<0,\ c=\frac{\sigma_Y}{\sigma_X}\mu_X+\mu_Y$ 该直线经过点 (μ_X,μ_Y)



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



相关系数反映
随机变量
线性关系的
数字特征

当$\rho(X,Y)=\pm1$时，X 与Y之间以概率 1成立线性关系

当$\rho(X,Y)=1$ 时，称X与Y正线性相关；

当$\rho(X,Y)=-1$时，称X与Y负线性相关.

X与Y之间的线性联系的程度随着$|\rho(X,Y)|$的减小而减弱. 特别地有下面定义:

定义4.5　当$\rho(X,Y)=0$时，称X与Y **(线性)**不相关.

当 $D(X)>0, D(Y)>0$ 时, 有

X 与 Y 不相关 $\Leftrightarrow$ $\rho(X,Y)=0$

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

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

3.不相关与相互独立的关系

定理4.6 如果 X 与 Y 相互独立，那么 X 与 Y 必定不相关；反之则不然（见**例3**）．独立时，$E(XY)=E(X)E(Y)$

如果 X 与 Y 相关，则它们一定不独立.

独立 $\Leftrightarrow F(x,y)=F_X(x)F_Y(y), \forall x,y\in R$

独立意味着随机变量之间没有任何关系

例3 设随机变量 X 的概率函数为

X	-1	0	1
Pr.	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

又 $Y=X^2$，试说明 X 与 Y 不相关且不相互独立.

解 计算可知

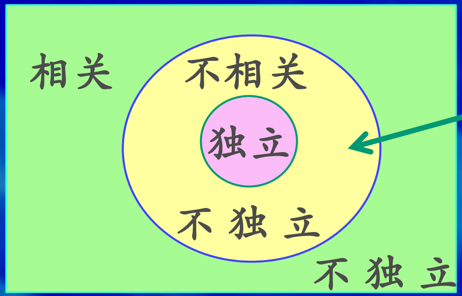
$$E(X)=0, E(Y)=E\left(X^2\right)=\frac{2}{3}, E(XY)=E\left(X^3\right)=0$$

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

即 X 与 Y 不相关. 又因为 $Y = X^2$, 即 X 与 Y 存在非线性关系，所以 X 与 Y 不独立.

独立意味着 X 与 Y 无任何关系，不相关仅意味着无线性关系，不能排除具有非线性关系，因此 X 与 Y 不一定相互独立.

随机变量的关系



无线性关系但有非线性关系

例4 设有 $X \sim N(1,9), Y \sim N(0,16), \rho(X,Y) = -\frac{1}{2}$,

又 $Z = \frac{X}{3} + \frac{Y}{2}$. 试求: $E(Z), D(Z), \rho(X,Z)$.

解 由期望和方差的性质以及常见分布的数字特征即得

$$E(Z) = E\left(\frac{X}{3} + \frac{Y}{2}\right) = \frac{1}{3}E(X) + \frac{1}{2}E(Y) = \frac{1}{3}$$

$$\operatorname{cov}(X,Y) = \rho(X,Y)\sqrt{D(X)D(Y)} = -\frac{1}{2}\times 3\times 4 = -6$$

$$D(Z)=D\left(\frac{X}{3}+\frac{Y}{2}\right)=\frac{1}{9}D(X)+\frac{1}{4}D(Y)+\frac{1}{3}\operatorname{cov}(X,Y)$$

$$=\frac{1}{9}\times 9+\frac{1}{4}\times 16+\frac{1}{3}\times(-6)=3$$

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

$$=\frac{1}{3}D(X)+\frac{1}{2}\operatorname{cov}(X,Y)=\frac{1}{3}\times 9+\frac{1}{2}\times(-6)=0$$

$$\rho(X,Z)=0.$$

定理4.7 如果(X,Y)服从二维正态分布，则X与Y相互独立等价于X与Y不相关.

如果(X,Y)服从二维正态分布，则X与Y相互独立等价于$\rho=0$.

相互独立 $\underset{\not\Leftarrow}{\Rightarrow}$ 不相关

相互独立 $\Leftrightarrow$ 不相关

(二维正态分布)

本节内容

1. 相关系数定义
2. 相关系数的性质
3. 不相关与相互独立的关系

谢　谢

同济大学数学科学学院概率统计教学团队