前面我们介绍了随机变量的数学期望和方差,对于多维随机变量,反映分量之间关系的数字特征中,最重要的,就是本讲要讨论的

协方差和相关系数



若Var(X), Var Y 存在,则

$$Var(X\pm Y)=Var(X)+Var(Y)\pm 2E[(X-E(X))(Y-E(Y))]$$

若随机变量X, Y相互独立,它们的方差都存在,则 $X\pm Y$ 的方差也存在,且

 $Var(X\pm Y) = Var(X) + Var(Y)$

§3 协方差、相关系数

协方差

设二维随机变量(X,Y),它的分量的数学期望为E(X),E(Y),若E[(X-E(X))(Y-E(Y))]存在,则称它为X,Y的协方差,记为Cov(X,Y),即

Cov(X, Y) = E[(X-E(X))(Y-E(Y))]

计算

(1) 若二维<mark>离散型</mark>随机变量(X, Y)的联合分布 律为

$$p_{ij} = P(X = x_i, Y = y_j), i, j = 1,2,...$$

且C o v(X,Y)存在,则

$$Cov(X,Y) = \sum_{i,j} (x_i - E(X))(y_j - E(Y))p_{ij}$$

若二维连续型随机变量(X,Y)的联合概率密度为 f(x,y),且Cov(X,Y)存在,则

$$Cov(X,Y)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X))(y - E(Y))f(x,y)dxdy$$

(2)
$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

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

$$=E\{XY-XE(Y)-YE(X)+E(X)E(Y)\}$$

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

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

即

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

可见,若X与Y独立,Cov(X,Y)=0.

例1 已知X,Y的联合分布为

$P_{ij}X$	1	0	0 < p <1
1	p	0	p + q = 1
0	0	$oldsymbol{q}$	

求 cov(X,Y)

解

\boldsymbol{X}	1	0	Y	1	0	XY	1	0
P	p	\boldsymbol{q}	P	p	\boldsymbol{q}	P	p	$oldsymbol{q}$

例2 设(X,Y)的联合分布律为

$$P(X = n, Y = m) = \frac{\lambda^{n} p^{m} (1 - p)^{n - m}}{m! (n - m)!} e^{-\lambda}$$

其中

$$\lambda > 0.0$$

求 Cov(X,Y).

$$P(X = n, Y = m) = \frac{\lambda^n}{n!} e^{-\lambda} \cdot \binom{n}{m} p^m (1-p)^{n-m}$$

$$E(g(X,Y)) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} g(x_i, y_j) p_{ij}$$

解:
$$P(X = n, Y = m) = \frac{\lambda^n}{n!} e^{-\lambda} \cdot \binom{n}{m} p^m (1-p)^{n-m}$$
$$E(XY) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} mn \cdot \frac{\lambda^n}{n!} e^{-\lambda} \cdot \binom{n}{m} p^m (1-p)^{n-m}$$

$$E(XY) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} mn \cdot \frac{\lambda^{n}}{n!} e^{-\lambda} \cdot \binom{n}{m} p^{m} (1-p)^{n-m}$$

$$=\sum_{n=0}^{\infty}n\cdot\frac{\lambda^n}{n!}e^{-\lambda}\sum_{m=0}^nm\binom{n}{m}p^m(1-p)^{n-m}$$

$$=p\sum_{n=0}^{\infty}n^2\cdot\frac{\lambda^n}{n!}e^{-\lambda}$$

$$= p\lambda(\lambda+1)$$

二项分布 B(n,p)期望

$$E(X) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i p_{ij}$$

$$E(X) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} n \cdot \frac{\lambda^{n}}{n!} e^{-\lambda} \cdot \binom{n}{m} p^{m} (1-p)^{n-m}$$

$$= \sum_{n=0}^{\infty} n \frac{\lambda^{n}}{n!} e^{-\lambda} \sum_{m=0}^{n} \binom{n}{m} p^{m} (1-p)^{n-m} = \sum_{n=0}^{\infty} n \cdot \frac{\lambda^{n}}{n!} e^{-\lambda} = \lambda$$

$$E(Y) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} m \frac{\lambda^{n}}{n!} e^{-\lambda} \cdot \binom{n}{m} p^{m} (1-p)^{n-m}$$

$$= \sum_{n=0}^{\infty} \frac{\lambda^{n}}{n!} e^{-\lambda} \sum_{m=0}^{n} m \binom{n}{m} p^{m} (1-p)^{n-m} = p \sum_{n=0}^{\infty} n \frac{\lambda^{n}}{n!} e^{-\lambda} = p \lambda$$

$$E(XY) = p\lambda(\lambda + 1), \quad Cov(X, Y) = E(XY) - E(X)E(Y)$$

= $p\lambda(\lambda + 1) - p\lambda^2 = p\lambda$

简单性质

$$(1) \quad Cov(X,a)=0$$

(2)
$$Cov(X, Y) = Cov(Y,X)$$

(3)
$$Cov(X, X) = Var(X)$$

(4)
$$Cov(aX, bY) = ab \ Cov(X,Y)$$
 a,b是常数

(5)
$$Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$$

(6) 随机变量和的方差与协方差的关系

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i < j} Cov(X_i, X_j)$$

若 $X_1, X_2, ..., X_n$ 两两独立,,上式化为

$$\operatorname{Var}(\sum_{i=1}^{n} X_{i}) = \sum_{i=1}^{n} \operatorname{Var}(X_{i})$$

性质7. 设随机变量X和Y的期望和方差都存在,且DX>0,DY>0,则 $[Cov(X,Y)]^2 \leq D(X)D(Y)$

其中等号成立当且仅当X与Y有严格的线性关系(即存在常数a和b,使得 $P{Y=aX+b}=1$ 成立)

证明:对任意的实数t,有

$$D(tX+Y) = t^2D(X) + D(Y) + 2tCov(X,Y)$$

令
$$g(t) = t^2D(X) + 2Cov(X,Y)t + D(Y)$$
 为关于t的二次三项式。

由方差的性质知 $g(t) \ge 0$,故判别式小于或等于零,即有

$$\Delta = [2Cov(X,Y)]^2 - 4D(X)D(Y) \le 0$$

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

若
$$[Cov(X,Y)]^2 = D(X)D(Y)$$

此时
$$t^2D(X) + 2Cov(X,Y)t + D(Y) = 0$$

有唯一的根, 记为 t_0 , 即 $g(t_0)=0$

所以有 $D(t_0X+Y)=0$

所以有
$$P\{t_0X+Y=E(t_0X+Y)\}=1$$
 即 $P\{Y=-t_0X+E(t_0X+Y)\}=1$

取
$$a = -t_0, b = E(t_0X + Y)$$
 有 $P\{Y = aX + b\} = 1$ 成立。



$$g(t) = D(tX + Y) = t^{2}D(X) + D(Y) + 2tCov(X,Y)$$

反之,若存在常数a和b,使得 $P{Y=aX+b}=1$ 成立

则有 $P\{-aX+Y=b\}=1$

由方差的性质知 D(-aX+Y)=0 即 g(-a)=0

所以有 $\Delta = [2Cov(X,Y)]^2 - 4D(X)D(Y) = 0$

 $[Cov(X,Y)]^2 = D(X)D(Y)$



例3 设(X,Y)的联合密度函数为

$$f(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

解: 由已知条件有

$$X \sim N(0, \sigma^2)$$
, $Y \sim N(0, \sigma^2)$ 且(X,Y)相互独立

$$Cov(X+Y, X-Y) = Cov(X,X) - Cov(X,Y) + Cov(Y,X) - Cov(Y,Y)$$

$$= Var(X) - Var(Y) = 0$$

例4(配对问题)。将n只球(编号为 $1\sim n$ 号)随机放进n个盒子(编号为 $1\sim n$ 号)中去,一个盒子只装一只球。若一只球装入与球同号的盒子中,称为一个配对。记X为总配对数,求EX,DX。

解: 设
$$X_i = \begin{cases} 1, & \text{第} i \text{ 号球放入} \hat{\mathbf{g}} i \text{ 号盒子} \\ 0, & \text{第} i \text{ 号球未放入} \hat{\mathbf{g}} i \text{ 号盒子} \end{cases}$$
, $i = 1, 2, \dots, n$

则配对总数
$$X = X_1 + X_2 + \cdots + X_n$$

且对任意
$$i$$
 有 $P{X_i = 1} = \frac{1 \times (n-1)!}{n!} = \frac{1}{n}$ $P{X_i = 0} = 1 - \frac{1}{n}$

所以由0—1分布的性质得
$$EX_i = \frac{1}{n}, DX_i = \frac{1}{n}(1 - \frac{1}{n}) = \frac{n-1}{n^2}$$

于是有
$$EX = E(X_1 + X_2 + \dots + X_n) = EX_1 + EX_2 + \dots + EX_n = n \cdot \frac{1}{n} = 1$$

另外,当
$$i\neq j$$
时, $P\{X_iX_j=1\}=P\{X_i=1,\ X_j=1\}=\frac{1\times(n-2)!}{n!}=\frac{1}{n(n-1)}$ $P\{X_iX_j=0\}=1-P\{X_iX_j=1\}=1-\frac{1}{n(n-1)}$

FINA
$$E(X_i X_j) = 1 \times P\{X_i X_j = 1\} = \frac{1}{n(n-1)}$$

从而可得

$$Cov(X_i, X_j) = E(X_i X_j) - EX_i EX_j = \frac{1}{n(n-1)} - \frac{1}{n} \times \frac{1}{n} = \frac{1}{n^2(n-1)}$$

FIFUL
$$DX = D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} DX_i + 2\sum_{1 \le i < j \le n} Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} \frac{n-1}{n^2} + 2\sum_{1 \le i < j \le n} \frac{1}{n^2(n-1)} = n \frac{n-1}{n^2} + 2C_n^2 \frac{1}{n^2(n-1)} = 1$$



例5设随机变量 $X_1, X_2, ..., X_n (n>1)$ 独立同分布,且它们共同的方差为

$$\sigma^2 > 0$$
, 令 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, Y_1 = X_1 - \bar{X}, Y_n = X_n - \bar{X}$. 求 $D(Y_1)$ 以及 $Cov(Y_1, Y_n)$

解:由方差的性质可得:

$$D(\bar{X}) = D(\frac{1}{n} \sum_{i=1}^{n} X_i) = \frac{1}{n^2} \sum_{i=1}^{n} DX_i = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}$$

又易知

$$Cov(X_1, \overline{X}) = Cov(X_1, \frac{1}{n} \sum_{i=1}^n X_i) = Cov(X_1, \frac{1}{n} X_1) = \frac{1}{n} D(X_1) = \frac{\sigma^2}{n}$$

同理有
$$Cov(X_n, \bar{X}) = \frac{\sigma^2}{n}$$

FFILL
$$D(Y_1) = D(X_1 - \overline{X})$$

$$= D(X_1) + D(\overline{X}) - 2Cov(X_1, \overline{X})$$

$$= \sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

由协方差的性质可得

$$Cov(Y_1, Y_n) = Cov(X_1 - \overline{X}, X_n - \overline{X}) = Cov(X_1, X_n) - Cov(X_1, \overline{X}) - Cov(X_n, \overline{X}) + Cov(\overline{X}, \overline{X})$$

$$= 0 - Cov(X_1, \bar{X}) - Cov(X_n, \bar{X}) + D(\bar{X}) = 0 - \frac{\sigma^2}{n} - \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = -\frac{\sigma^2}{n}$$



协方差的大小在一定程度上反映了X和Y相互间的关系,但它还受X与Y本身度量单位的影响.例如:

 $Cov(kX, kY) = k^2 Cov(X, Y)$

为了克服这一缺点,对协方差进行标准化,这就引入了相关系数.





休息片刻继续下一讲

相关系数

若二维随机变量(X, Y)的分量的方差 Var(X), Var(Y)都存在, 且Var(X)>0, Var(Y)>0,

则称
$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

为随机变量X和Y的相关系数.

在不致引起混淆时,记 ρ_{XY} 为 ρ .

若 $\rho_{XY}=0$ 则称X,Y不相关;

若 $\rho_{XY}>0$ 称X,Y 正相关;

若 ρ_{XY} <0 则称X, Y负相关



例 6 设 $X\sim N(0,4)$, $Y\sim \pi$ (2) , $\rho_{XY}=1/2$, $E(X+Y)^2$

解:由已知条件有

$$E(X) = 0, D(X) = 4, \quad E(X^{2}) = D(X) + [E(X)]^{2} = 4$$

$$E(Y) = D(Y) = 2, \quad E(Y^{2}) = D(Y) + [E(Y)]^{2} = 6$$

$$Cov(X, Y) = \rho_{XY} \sqrt{Var(X)Var(Y)} = \frac{1}{2} \sqrt{4 \cdot 2} = \sqrt{2}$$

$$E(XY) = Cov(X, Y) + E(X)E(Y) = \sqrt{2}$$

$$\text{Fig.} \quad E[(X + Y)^{2}] = E(X^{2}) + 2E(XY) + E(Y^{2})$$

$$= 4 + 2\sqrt{2} + 6 = 10 + 2\sqrt{2}$$

例7 设 $X\sim U[0,2\pi]$, Y=cos(X), 求 ρ_{XY}



例8 设(X,Y)在圆域

$$D = \{(x, y); x^2 + y^2 \le r^2\} (r > 0)$$

上服从均匀分布,判断X,Y是否不相关,是否独立。

解: 由题意 (X,Y)的联合密度函数

$$f(x,y) = \begin{cases} \frac{1}{\pi r^2} & (x,y) \in D \\ 0 & 其他 \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dx dy = \iint_{x^2 + y^2 \le r^2} x \frac{1}{\pi r^2} dx dy = 0$$
同理E(Y)=0

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy = \iint_{x^2 + y^2 \le r^2} xy \frac{1}{\pi r^2} dxdy = 0$$

$$cov(X,Y) = E(XY) - E(X)E(Y) = 0$$

$$f(x,y) = \begin{cases} \frac{1}{\pi r^2} & (x,y) \in D\\ 0 & \text{ 其他} \end{cases}$$

$$f_{X}(x) = \int_{-\infty}^{\infty} f(x, y) dy \qquad f_{X}(x) = \int_{-\sqrt{r^{2} - x^{2}}}^{\sqrt{r^{2} - x^{2}}} \frac{1}{\pi r^{2}} dy$$

$$= \frac{2\sqrt{r^{2} - x^{2}}}{\pi r^{2}}$$

$$f_{X}(x) = \begin{cases} \frac{2\sqrt{r^{2} - x^{2}}}{\pi r^{2}} & |x| \leq r \\ 0 & |x| \leq r \end{cases}$$

$$f_{Y}(y) = \begin{cases} \frac{2\sqrt{r^{2} - y^{2}}}{\pi r^{2}} & |y| \leq r \\ 0 & |x| \leq r \end{cases}$$

$$f_{Y}(y) = \begin{cases} \frac{2\sqrt{r^{2} - y^{2}}}{\pi r^{2}} & |y| \leq r \\ 0 & |x| \leq r \end{cases}$$

$$f_{Y}(y) = \begin{cases} \frac{2\sqrt{r^{2} - y^{2}}}{\pi r^{2}} & |y| \leq r \\ 0 & |x| \end{cases}$$

$$D = \{(x, y) : x^2 + y^2 \le r^2\}$$

$$f(x, y) \neq f_X(x) f_Y(y), \qquad (x, y) \in D$$

X,Y不独立



例9 设(X, Y)服从二维正态分布

$$N(\mu_{1}, \sigma_{1}^{2}; \mu_{2}, \sigma_{2}^{2}, \rho)$$

求X和Y的相关系数



$$Cov(X,Y) = E[(X - \mu_1)(Y - \mu_2)]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_1)(y - \mu_2) f(x, y) dx dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_{1}\sigma_{2}\sqrt{1-\rho^{2}}} (x-\mu_{1})(y-\mu_{2}) \exp\{-\frac{1}{2(1-\rho^{2})}$$

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

$$= \frac{\sigma_1 \sigma_2}{2\pi \sqrt{1 - \rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv \cdot \exp\{-\frac{1}{2(1 - \rho^2)}(u^2 - 2\rho uv + v^2)\} dv du$$

$$u^{2} - 2\rho uv + v^{2}$$

$$= (v - \rho u)^{2} + (1 - \rho^{2})u^{2}$$

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

$$[(v - \rho u)^{2} + (1 - \rho^{2})u^{2}] dv du$$

 $= \frac{\sigma_1 \sigma_2}{2\pi \sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv \exp\{-\frac{1}{2(1-\rho^2)}[(v-\rho u)^2 + (1-\rho^2)u^2]\}dvdu$

$$= \frac{\sigma_1 \sigma_2}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(\rho u + \sqrt{1 - \rho^2} t) \exp(-\frac{1}{2}t^2) \exp(-\frac{1}{2}u^2) dt du$$

$$= \rho \sigma_1 \sigma_2 (\int_{-\infty}^{\infty} u^2 \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2) du) (\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2) dt)$$

$$+\sqrt{1-\rho^{2}}\sigma_{1}\sigma_{2}(\int_{-\infty}^{\infty}u\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^{2})du)(\int_{-\infty}^{\infty}t\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}t^{2})dt)$$

 $= \rho \sigma_1 \sigma_2$

独立与相关的关系

(1) X, Y相互独立 \longrightarrow X, Y不相关



$$X$$
, Y不相关 $\rightarrow \rho_{XY} = 0$ $\rightarrow \operatorname{cov}(X,Y) = 0$ $\rightarrow E(XY) = E(X)E(Y)$ $\rightarrow \operatorname{Var}(X \pm Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$

(2) 若X,Y服从二维正态分布,

X, Y相互独立 \longrightarrow X, Y不相关

相关系数的性质

$$(1) \mid \rho \mid \leq 1$$

(2)
$$|\rho| = 1 \Longrightarrow$$

存在常数 $a,b(b\neq 0)$,

使 $P{Y=a+bX}=1$.

 $\rho_{XY} = 1$ 的充要条件是,P(Y=a+bX)=1(b>0)

这时称X与Y完全正相关;

 ρ_{XY} = -1的充要条件是,P(Y=a+bX)=1(b<0) 这时称X=Y完全负相关。

相关系数的意义

它是用来刻 画X, Y线性相关程度的一个量。

 $\Xi | \rho |$ 的值越接近于1,X与Y的 线性相关程度越高。

若 $|\rho|$ 的值越接近于0, X与Y的线性相关程度越弱。

矩

设二维随机变量(X,Y), k, l 为非负整数。

若 $E(X^k)$ 存在,则称它为X的k阶原点矩,记作 m_k ,即

$$m_k = E(X^k)$$

若 $E(X-E(X))^k$ 存在,则称它为X的k阶中心矩,记作 c_k ,

若 $E(X^kY^l)$ 存在,则称它为X和Y的(k,l)阶混合矩,记作 $m_{k,l}$,即 $m_k = E(X^kY^l)$

若 $E[(X-E(X))^k(Y-E(Y))^l]$ 存在,

则称它为X和Y的 (k,l)阶混合中心矩,记作 c_{kl} ,即 $c_{kl} = E[(X-E(X))^k(Y-E(Y))^l].$

关于矩有下述结论: 设k为正整数。

(1) 若 $E(X^k)$ 存在,则对小于k的一切非负整数l, $E(X^l)$ 存在.

(2) 原点矩与中心矩可相互表示。



例10 求标准正态分布的各阶矩.

没X~N(0,1),
$$m_k = c_k = E(X^k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-\frac{x^2}{2}} dx$$

 $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \quad x > 0$

 $\Gamma(\frac{n}{2}) = 1 \cdot 3 \cdot 5 \cdots (n-2) \quad 2^{-(n-1)/2} \sqrt{\pi}$

 $\Gamma(n+1) = n!$

若
$$k$$
为奇数, $c_k = 0$

若k为偶数,

$$c_{k} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k} e^{-\frac{x^{2}}{2}} dx = \frac{\sqrt{2}}{\sqrt{\pi}} \int_{0}^{\infty} x^{k} e^{-\frac{x^{2}}{2}} dx$$

$$\frac{x}{2} = t \sqrt{2} \sqrt{2} \int_{0}^{\infty} t^{\frac{k-1}{2}} e^{-t} dt = \frac{\sqrt{2}}{\sqrt{\pi}} 2^{\frac{k-1}{2}} \Gamma(\frac{k+1}{2})$$

$$= 1 \cdot 3 \cdot 5 \cdots (k-1)$$

第四章

作业3: 36,40,42,44,45



