第四章 随机变量的数字特征

第一节 随机变量的数学期望

第二节 随机变量的方差与矩

第三节 协方差与相关系数

教学计划：2次课-6学时

复习

随机变量的数学期望与方差

	离散型随机变量	连续型随机变量
X 数学期望	$E(X)=\sum_{k=1}^{\infty}x_k p_k$	$E(X)=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x$
$Y=g(X)$ 函数数学期望	$E(Y)=E[g(X)]$ $=\sum_{k=1}^{\infty}g(x_k)p_k$	$E(Y)=E[g(X)]$ $=\int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x$
$Z=g(X,Y)$ 函数数学期望	$E(Z)=E[g(X,Y)]$ $=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}g(x_i,y_j)p_{ij}$	$E(Z)=E[g(X,Y)]$ $=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$
X 方差	$D(X)=E[X-E(X)]^2=E(X^2)-[E(X)]^2$	

复习

数学期望与方差的性质

$E(X)$性质	$E(c)=c$ $\quad E(cX)=cE(X)$ $\quad E(X+Y)=E(X)+E(Y)$ X,Y 独立， $E(XY)=E(X)E(Y)$
$D(X)$性质	$D(c)=0$ $\quad D(cX)=c^2D(X)$ $\quad D(X)=0 \Leftrightarrow P(X=c)=1$ X,Y 独立， $D(X+Y)=D(X)+D(Y)$

复习

常见分布的数学期望和方差

	概率分布		$E(X)$	$D(X)$
离散型	(0-1)分布	$X \sim B(1,p)$	p	
	二项分布	$X \sim B(n,p)$	np	
	泊松分布	$X \sim P(\lambda)$	λ	
连续型	均匀分布	$X \sim U(a,b)$	$(a+b)/2$	
	指数分布	$X \sim E(\theta)$	θ	
	正态分布	$X \sim N(\mu,\sigma^2)$	μ	

第四章 随机变量的数字特征

第一节 随机变量的数学期望

第二节 随机变量的方差与矩

第三节 协方差与相关系数

教学计划：2次课-6学时

第四章 随机变量的数字特征

第二节 随机变量的方差与矩

- 方差的定义
- 方差的性质
- 离散型随机变量的方差
- 连续型随机变量的方差
- 矩

三. 三种常见离散分布的方差

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

(1) (0−1) 分布 $E(X)=p$ $E(X^2)=p$

X^2	0	1
X	0	1
p	$1-p$	p

$$D(X)=p-p^2=p(1-p)=pq$$

(2) 二项分布 $X\sim B(n,\ p)$ $E(X)=np$

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

$$P(X=k)=C_n^k p^k(1-p)^{n-k}$$

$$E(X^2)=\sum_{k=0}^{n}k^2C_n^k p^k(1-p)^{n-k}=npq+(np)^2$$

(3) 泊松分布 $X\sim P(\lambda)$ $E(X)=\lambda$

$$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$$

$$k=0,1,2,\cdots$$

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

$$E(X^2)=\sum_{k=0}^{\infty}k^2\frac{\lambda^k e^{-\lambda}}{k!}=\lambda+\lambda^2$$

$$E[g(X)]=\sum_{k=1}^{\infty}g(x_k)p_k$$

归纳

(1) (0–1) 分布　$X \sim B(1, p)$　$E(X) = p$　$D(X) = pq$

(2) 二项分布　$X \sim B(n, p)$　$E(X) = np$　$D(X) = npq$

(3) 泊松分布　$X \sim P(\lambda)$　$E(X) = \lambda$　$D(X) = \lambda$

第四章 随机变量的数字特征

第二节 随机变量的方差与矩

- 方差的定义
- 方差的性质
- 离散型随机变量的方差
- 连续型随机变量的方差
- 矩

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

四. 三种常见连续分布的方差

(1) 均匀分布 $X \sim U[a,b]$

$$E(X)=\frac{a+b}{2}$$

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b \\ 0 & \text{其它}\end{cases}$$

$$D(X)=E(X^2)-[E(X)]^2=\frac{a^2+ab+b^2}{3}-\frac{(a+b)^2}{4}=\frac{(b-a)^2}{12}$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2 f(x)\mathrm{d}x=\int_a^b x^2\frac{1}{b-a}\mathrm{d}x=\frac{1}{b-a}\int_a^b x^2\mathrm{d}x$$

$$=\frac{1}{3(b-a)}x^3\Big|_a^b=\frac{b^3-a^3}{3(b-a)}=\frac{a^2+ab+b^2}{3}$$

$$\boxed{E[g(X)]=\int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x}$$

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

(2) 指数分布 $X \sim E(\theta)$

$$f(x)=\begin{cases} \dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0 \\ 0 & \text{其它} \end{cases}$$

$$E(X)=\theta$$

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

分部积分法

$$E(X^2)=\int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x=\int_0^{+\infty} x^2\cdot\frac{1}{\theta}e^{-\frac{x}{\theta}}\mathrm{d}x=-\int_0^{+\infty} x^2\cdot\boxed{e^{-\frac{x}{\theta}}}\mathrm{d}\left(-\frac{x}{\theta}\right)=-\int_0^{+\infty} x^2\,\mathrm{d}e^{-\frac{x}{\theta}}$$

$\mathrm{d}x^2=2x\mathrm{d}x$ 分部积分法

$$=\int_{+\infty}^{0} x^2\,\mathrm{d}e^{-\frac{x}{\theta}}=\underbrace{x^2e^{-\frac{x}{\theta}}\Big|_{+\infty}^{0}}_{0}-\int_{+\infty}^{0} e^{-\frac{x}{\theta}}\mathrm{d}x^2=-2\int_{+\infty}^{0} x\,e^{-\frac{x}{\theta}}\mathrm{d}x=2\theta\int_{+\infty}^{0} x\,\boxed{e^{-\frac{x}{\theta}}}\mathrm{d}\left(-\frac{x}{\theta}\right)$$

$$=2\theta\int_{+\infty}^{0} x\,\mathrm{d}e^{-\frac{x}{\theta}}=2\theta\Big[\underbrace{xe^{-\frac{x}{\theta}}\Big|_{+\infty}^{0}}_{0}-\int_{+\infty}^{0} e^{-\frac{x}{\theta}}\mathrm{d}x\Big]=2\theta^2\int_{+\infty}^{0} e^{-\frac{x}{\theta}}\mathrm{d}\left(-\frac{x}{\theta}\right)=2\theta^2e^{-\frac{x}{\theta}}\Big|_{+\infty}^{0}$$

$$=2\theta^2(1-0)=2\theta^2$$

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

(3) 正态分布 $X \sim N(\mu,\sigma^2)$

$$E(X)=\mu$$

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$D(X)=E(X^2)-[E(X)]^2=\sigma^2+\mu^2-\mu^2=\sigma^2$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2 f(x)\mathrm{d}x=\int_{-\infty}^{+\infty}x^2\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x$$

$$=\sigma^2+\mu^2$$

归纳

(1) 均匀分布 $X \sim U(a,b)$ $E(X)=\dfrac{a+b}{2}$ $D(X)=\dfrac{(b-a)^2}{12}$

(2) 指数分布 $X \sim E(\theta)$ $E(X)=\theta$ $D(X)=\theta^2$

(3) 正态分布 $X \sim N(\mu,\sigma^2)$ $E(X)=\mu$ $D(X)=\sigma^2$

小结

六种常见分布的数学期望和方差

	概率分布		$E(X)$	$D(X)$
离散型	(0-1)分布	$X \sim B(1, p)$	p	pq
	二项分布	$X \sim B(n, p)$	np	npq
	泊松分布	$X \sim P(\lambda)$	λ	λ
连续型	均匀分布	$X \sim U(a,b)$	$(a+b)/2$	$(b-a)^2/12$
	指数分布	$X \sim E(\theta)$	θ	θ^2
	正态分布	$X \sim N(\mu,\sigma^2)$	μ	σ^2

第四章 随机变量的数字特征

第二节 随机变量的方差与矩

- 方差的定义
- 方差的性质
- 离散型随机变量的方差
- 连续型随机变量的方差
- 矩

五. 矩

矩是随机变量的更为广泛的一种数字特征，前面介绍的数学期望及方差都是某种矩.

定义4 设 X 是随机变量,

(1) 若 $E(X^k)$ 存在, 则称它为 X 的 k 阶原点矩，简称 k 阶矩。

$k = 1, 2, \cdots$

(2) 若 $E\{[X - E(X)]^k\}$ 存在，则称它为 X 的 k 阶中心矩。

注： $E(X)$ 是随机变量 X 的一阶原点矩；

$D(X)$ 是随机变量 X 的二阶中心矩；

第四章 随机变量的数字特征

第二节 随机变量的方差与矩

- 方差的定义
- 方差的性质
- 离散型随机变量的方差
- 连续型随机变量的方差
- 矩

第四章 随机变量的数字特征

第一节 随机变量的数学期望

第二节 随机变量的方差与矩

第三节 协方差与相关系数

第四章 随机变量的数字特征

第三节 协方差与相关系数

- 协方差
- 相关系数

协方差的引出

在实际问题中很多情况下，两个随机变量之间都有着相互关系。如何来描述它们之间的关系呢？

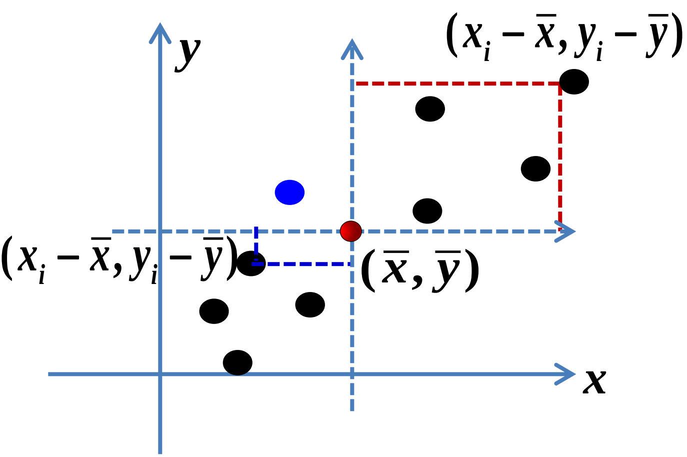
引例1 某医院某天出生了9个男婴，他们体重和母亲怀孕期的数据记录如下表所示：

男婴	1	2	3	4	5	6	7	8	9	平均值
孕期X（周）	36	37	38	40	40	40	41	41	41	39.2
体重Y（kg）	3.10	3.24	3.15	3.12	3.14	3.52	3.40	3.53	3.74	3.35

现在来研究随机变量X与Y之间的关系。

引例1 画出散点图

男婴	1	2	3	4	5	6	7	8	9	平均值
孕期X（周）	36	37	38	40	40	40	41	41	41	39.2
体重Y（kg）	3.10	3.24	3.15	3.12	3.14	3.52	3.40	3.53	3.74	3.35



观察随机变量 X, Y 之间的关系：

x_i 越大，y_i 越大，则称它们正相关。

正相关时，大部分点都落在一、三象限.

正相关时，$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})>0$

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=39.2$$

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=3.35$$

	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$
第1象限	+	+	+
第2象限			
第3象限	-	-	+
第4象限			

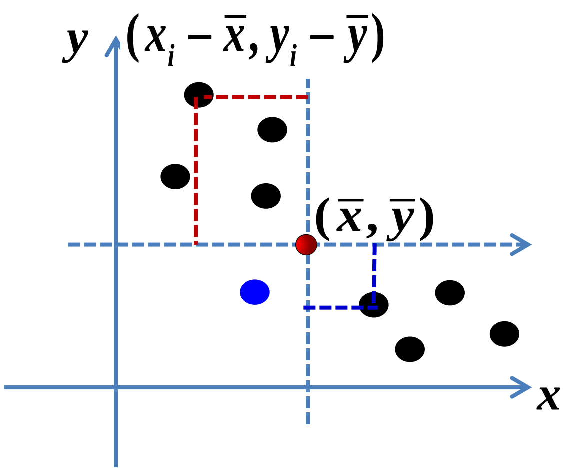
引例2 美国某城市随机抽取9个居民区，经调查，其人口密度与小区到市中心的距离相关数据如下表所示：

居民区	1	2	3	4	5	6	7	8	9	平均值
距离X（公里）	8	14	16	20	22	25	33	47	50	27.3
密度Y（人/公里^2）	3.4	2.4	4.0	3.0	3.8	1.8	2.0	1.5	0.6	2.34

现在来研究随机变量X与Y之间的关系。

引例2 画出散点图

居民区	1	2	3	4	5	6	7	8	9	平均值
距离X（公里）	8	14	16	20	22	25	33	47	50	27.3
密度Y（人/公里^2）	3.4	2.4	4.0	3.0	3.8	1.8	2.0	1.5	0.6	2.34



观察随机变量 X, Y 之间的关系：

x_i 越大，y_i 越小，则称它们负相关。

负相关时，大部分点都落在二、四象限.

负相关时，$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})<0$

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=27.3$$

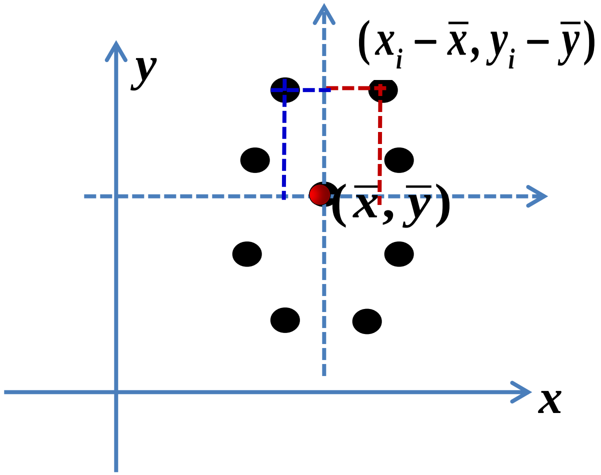
$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=2.34$$

	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$
第1象限	+	+	+
第2象限	-	+	-
第3象限	-	-	+
第4象限	+	-	-

引例3 某医院某天出生了9个男婴，他们体重、头围的数据记录如下表所示：

男婴	1	2	3	4	5	6	7	8	9	平均值
头围X（cm）	31.1	33.3	30.0	35.0	30.2	36.4	37.3	31.4	34.0	33.02
体重Y（kg）	3.10	3.24	3.15	3.12	3.14	3.52	3.40	3.53	3.74	3.35

画出散点图:



观察随机变量 X, Y 之间的关系：

由于随机点平均散落在四个象限里，所以

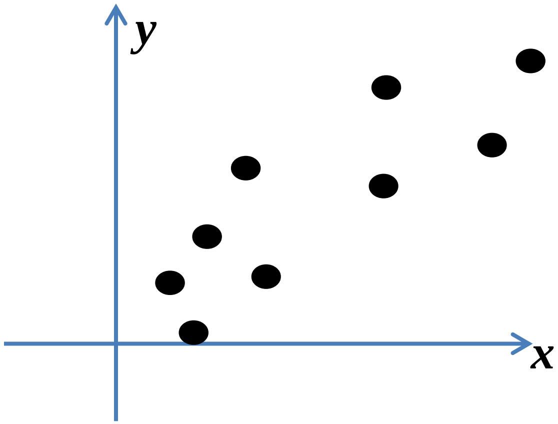
$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=0$$

此时称它们为不相关。

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=33.02 \qquad \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=3.35$$

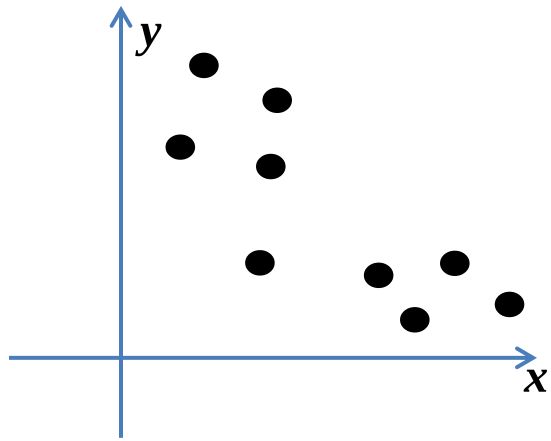
小结：

研究两个随机变量X, Y之间的关系:



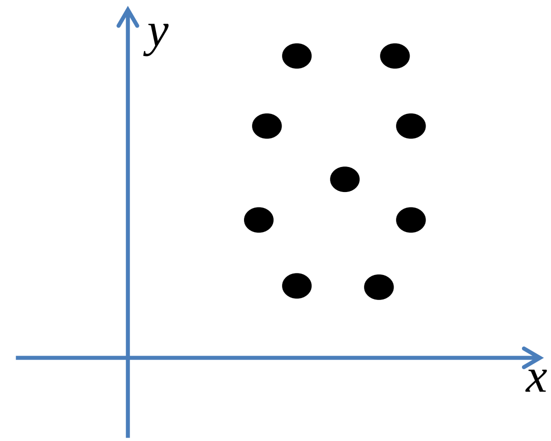
X与Y正相关时,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})>0$$



X与Y负相关时,

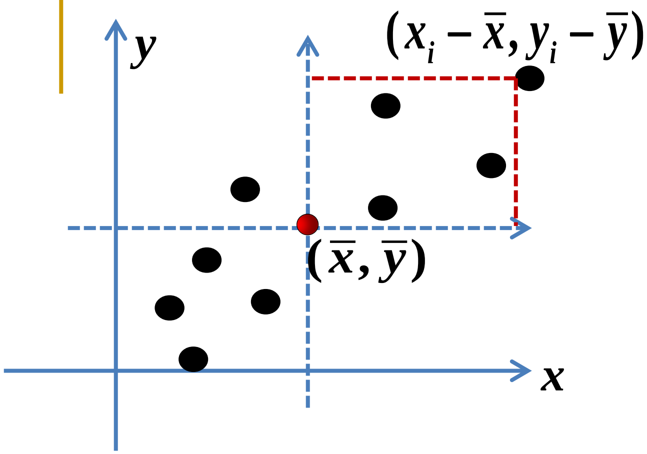
$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})<0$$



X与Y不相关时,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=0$$

结论： 可以用$\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$的取值情况来描述$X$与$Y$的相互关系.



x_i 越大，y_i 越大，则称它们正相关。

正相关时, 大部分点都落在一、三象限.

正相关时, $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})>0$

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

正相关与负相关统称为线性相关

$$\bar{x}=\sum_{i=1}^{n}x_i p_i=E(X)$$

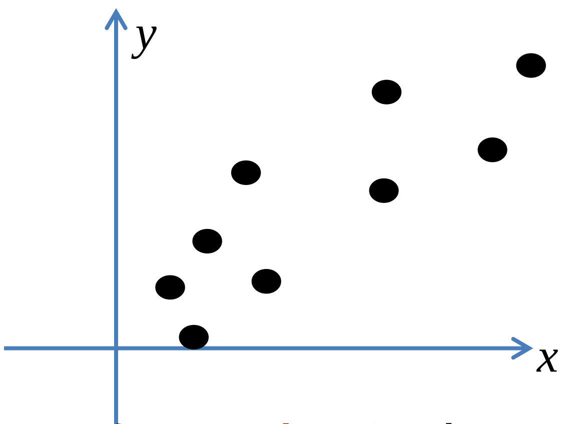
$$\bar{y}=\sum_{j=1}^{n}y_j w_j=E(Y)$$

因此可以用协方差描述两个随机变量之间的线性关系

注意：上述讨论中，为了简化问题，所有平均值都取为简单算术平均值，但实际问题中，随机变量(X,Y)取随机点(x_i, y_j)的概率是不同的，所以平均值应该按概率取为加权平均，即随机变量的数学期望才合理。因此，需要修正上述相关性的指标。

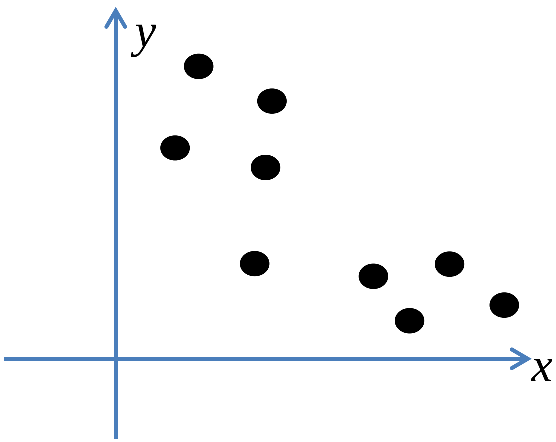
$$Cov(X,Y)=E[(X-E(X)(Y-E(Y)]$$

可以用**协方差**描述两个随机变量X, Y之间的线性关系:



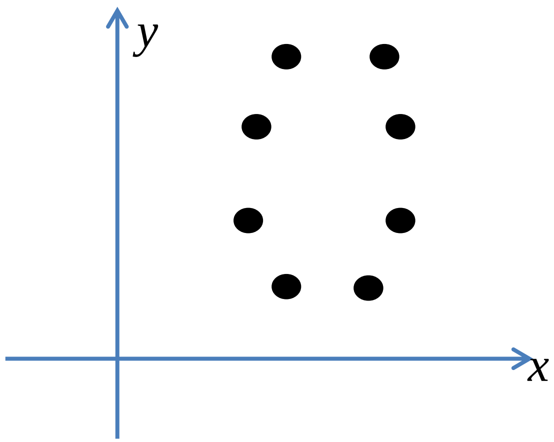
X与Y**正相关**时,

$Cov(X,Y)>0$



X与Y**负相关**时,

$Cov(X,Y)<0$



X与Y**不相关**时,

$Cov(X,Y)=0$

一. 协方差

$$\boxed{E(X\cdot Y)=E(X)\cdot E(Y)}$$

1. 定义　若 $E\{[X-E(X)][(Y-E(Y)]\}$ 存在，则称它为随机变量 X 与 Y 的协方差，记为：$Cov(X,Y)$.

即：$Cov(X,Y)=E\{[X-E(X)][(Y-E(Y)]\}$

注: ➢协方差中当 $X=Y$ 时为方差的定义，即：

$Cov(X,X)=D(X)$, 故方差是协方差的特例。

➢协方差是反映 X,Y 的线性关系的数字特征，当 X,Y 相互独立时，协方差为 0.

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

$$=E[X-E(X)]\cdot E[Y-E(Y)]$$

$$=[\underbrace{E(X)-E(X)}_{0}]\cdot[\underbrace{E(Y)-E(Y)}_{0}]=0$$

一. 协方差

1. 定义 若 $E\{[X-E(X)][(Y-E(Y)]\}$ 存在，则称它为随机变量 X 与 Y 的协方差，记为：$Cov(X,Y)$.

即：$Cov(X,Y)=E\{[X-E(X)][(Y-E(Y)]\}$

注：

- 协方差中当 $X=Y$ 时为方差的定义，即：

  $Cov(X,X)=D(X)$ 故方差是协方差的特例。

- 协方差是反映 X,Y 的线性关系的数字特征，当 X,Y 相互独立时，协方差为 0.

2. 协方差的简单性质

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

(2) $Cov(X,c)=0$ （c是常数）

一. 协方差

1.定义. 若 $E\{[X-E(X)][(Y-E(Y)]\}$ 存在，则称它为随机变量 X 与 Y 的协方差，记为：$Cov(X,Y)$.

即：$Cov(X,Y)=E\{[X-E(X)][(Y-E(Y)]\}$

2.协方差的简单性质

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

(2) $Cov(X,c)=0$

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

证明：
$$Cov(aX,bY)=E\{[aX-E(aX)][(bY-E(bY)]\}$$
$$=abE\{[X-E(X)][(Y-E(Y)]\}$$
$$=abCov(X,Y)$$

一. 协方差

1.定义. 若 $E\{[X-E(X)][(Y-E(Y)]\}$ 存在，则称它为随机变量 X 与 Y 的协方差，记为：$Cov(X,Y)$.

即：$Cov(X,Y)=E\{[X-E(X)][(Y-E(Y)]\}$

2.协方差的简单性质

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

(2) $Cov(X,c)=0$

(3) $Cov(aX,bY)=ab\,Cov(X,Y)$

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

3. 计算协方差的公式

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

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

证明：

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

$$=E[XY-XE(Y)-YE(X)+E(X)E(Y)]$$

$$=E(XY)+E[-XE(Y)]+E[-YE(X)]+E[E(X)E(Y)]$$

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

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

$$\boxed{E(X_1+X_2+X_3+X_4)=E(X_1)+E(X_2)+E(X_3)+E(X_4)}$$

$$\boxed{E(c\,X)=c\,E(X)} \qquad \boxed{E(c)=c}$$

3. 计算协方差的公式

$$\boxed{E(X\cdot Y)=E(X)\cdot E(Y)}$$

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

注： 显然，若X与Y相互独立，则: $Cov(X, Y)= 0$

4. 方差与协方差的关系

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

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

证明：

$$\boxed{D(X)=E[X-E(X)]^2}$$

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

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

$$=E\{[X-E(X)]+[Y-E(Y)]\}^2$$

$$=E[X-E(X)]^2+E[Y-E(Y)]^2+2E\{[X-E(X)][(Y-E(Y)]\}$$

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

小结　随机变量的数字特征

$E(X)$性质	$E(c)=c \quad E(cX)=cE(X) \quad E(X+Y)=E(X)+E(Y)$ X,Y独立 $E(XY)=E(X)E(Y)$
$D(X)$性质	$D(X)=E(X^2)-[E(X)]^2 \quad D(c)=0 \quad D(cX)=c^2D(X)$ X,Y独立, $D(X+Y)=D(X)+D(Y) \quad D(X)=0 \Leftrightarrow P(X=c)=1$
协方差性质	$Cov(X,Y)=E\{[X-E(X)][(Y-E(Y)]\}$ $=E(XY)-E(X)E(Y)$ $D(X+Y)=D(X)+D(Y)+2\operatorname{cov}(X,Y)\overset{独立}{=}D(X)+D(Y)$ (1) $Cov(X,Y)=Cov(Y,X)$ (2) $Cov(X,c)=0$ (3) $Cov(aX,bY)=ab\,Cov(X,Y)$ (4) $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

例1 设二维随机变量(X, Y)的联合分布律为

$X \backslash Y$	-1	0	1
0	0.07	0.18	0.15
1	0.08	0.32	0.20

求$Cov(X^2+3, Y^2-5)$

解： $Cov(X^2+3, Y^2-5)$

$$=Cov(X^2,Y^2)+\underbrace{Cov(X^2,-5)}_{0}+\underbrace{Cov(3,Y^2)}_{0}+\underbrace{Cov(3,-5)}_{0}$$

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

$$=E(X^2Y^2)-E(X^2)E(Y^2)$$

$$\boxed{Cov(X,c)=0}$$

$$\boxed{Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)}$$

$$\boxed{Cov(X,Y)=E(XY)-E(X)E(Y)}$$

例1 设二维随机变量(X,Y)的联合分布律为 $E[g(X)]=\sum_{k=1}^{\infty}g(x_k)p_k$

$$E(X^2)=\sum_{k=1}^{\infty}x_k^2p_k$$

$$E(Y^2)=\sum_{k=1}^{\infty}y_k^2p_k$$

X^2	$X \backslash Y$	-1	0	1	$P(X=x_i)$
0	0	0.07	0.18	0.15	0.4
1	1	0.08	0.32	0.20	0.6
	$P(Y=y_j)$	0.15	0.50	0.35	

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

$$E(X^2Y^2)=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}x_i^2y_j^2p_{ij}$$

求 $Cov(X^2+3,Y^2-5)$

解：$Cov(X^2+3,Y^2-5)=\overset{0.28}{E(X^2Y^2)}-\overset{0.6}{E(X^2)}\overset{0.5}{E(Y^2)}=-0.02$

$$E(X^2)=0^2\times0.4+1^2\times0.6=0.6$$

$$E(Y^2)=(-1)^2\times0.15+0^2\times0.5+1^2\times0.35=0.5$$

$$E(X^2Y^2)=0^2\cdot(-1)^2\times0.07+0^2\cdot0^2\times0.18+0^2\cdot1^2\times0.15$$

$$+1^2\cdot(-1)^2\times0.08+1^2\cdot0^2\times0.32+1^2\cdot1^2\times0.20=0.28$$

第四章 随机变量的数字特征

第三节 协方差与相关系数

- 协方差

相关系数

问题的引出

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

协方差的大小在一定程度上反映了X 和 Y 之间的相互关系，但它受 X与Y 本身度量单位的影响.

例如： $Cov(kX,kY)=k^2Cov(X,Y)$

若 X与Y 表示长度，单位是厘米，当$k=100$时，kX与kY 的单位是米.

例如：

居民区	1	2	3	4	5	6	7	8	9	平均值
距离X（公里）	8	14	16	20	22	25	33	47	50	27.3
密度Y（人/公里^2）	3.4	2.4	4.0	3.0	3.8	1.8	2.0	1.5	0.6	2.34

问题的引出

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

协方差的大小在一定程度上反映了X 和 Y 之间的相互关系，但它受 X与Y 本身度量单位的影响.

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的概念：

X数据标准化　　Y数据标准化

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}=E\{[\boxed{\frac{X-E(X)}{\sqrt{D(X)}}}][\boxed{\frac{(Y-E(Y)}{\sqrt{D(Y)}}}]\}$$ ---无量纲

它与 X,Y 本身度量单位无关(无量纲).

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

$$\frac{X-\mu}{\sigma}\sim N(0,1)$$

二. 相关系数

1. 定义 量 $\dfrac{Cov(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$ 称为随机变量X, Y 的相关系数,

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

2. 相关系数的性质

(1) $|\rho_{XY}|\le 1$

X 和 Y 以概率 1 存在着线性关系

(2) $|\rho_{XY}|=1 \Leftrightarrow$ 存在常数a, b, 使得: $P(Y=aX+b)=1$

分析: 考虑以 X 的线性函数 $aX+b$ 来近似表示 Y, 以均方误差

$e(a,b)=e=E\{[Y-(aX+b)]^2\}$ 来衡量以 $aX+b$ 近似表示 Y 的好坏程度。显然, e 值越小, $aX+b$ 与 Y 的近似程度越好.

最小均方误差是: $e=E\{[Y-(a_0X+b_0)]^2\}=D(Y)(1-\rho_{XY}^2)$

2.相关系数的性质

$$E\{[Y-(a_0X+b_0)]^2\}=D(Y)(1-\rho_{XY}^2)$$

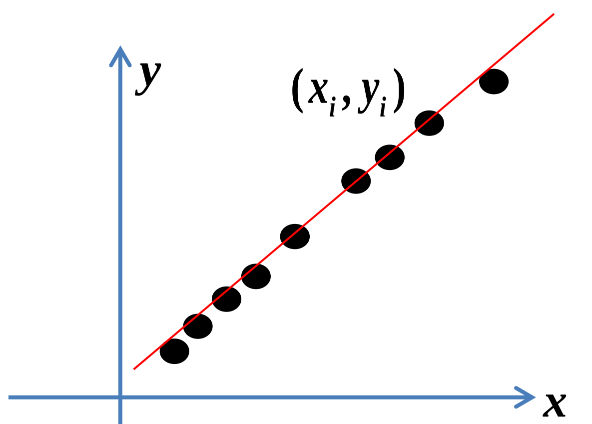
(1) $|\rho_{XY}|\le 1$

(2) $|\rho_{XY}|=1 \Leftrightarrow$ 存在常数 a, b使得：$P(Y=aX+b)=1$

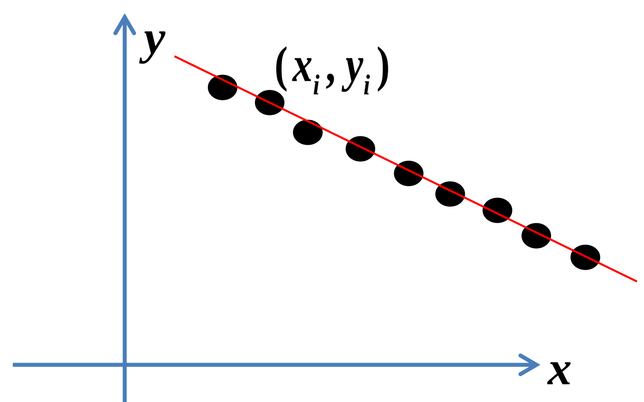
3. 相关系数的意义：

若 $\rho_{XY}=\pm 1$, 则 Y 与 X 概率为1有线性关系;

$$E\{[Y-(a_0X+b_0)]^2\}=0 \Rightarrow P(Y=a_0X+b_0)=1$$



$a_0>0$ $\rho_{XY}=1$ 严格正相关



$a_0<0$ $\rho_{XY}=-1$ 严格负相关

2.相关系数的性质

$$E\{[Y-(a_0X+b_0)]^2\}=D(Y)(1-\rho_{XY}^2)$$

(1) $|\rho_{XY}|\le 1$ $\quad D(Y)(1-\rho_{XY}^2)\ge 0 \to 1-\rho_{XY}^2\ge 0$

(2) $|\rho_{XY}|=1 \Leftrightarrow$ 存在常数 a, b使得：$P(Y=aX+b)=1$

3. 相关系数的意义：

若 $\rho_{XY}=\pm 1$, 则 Y 与 X 概率为1有线性关系;

若 $0\le|\rho_{XY}|\le 1$,

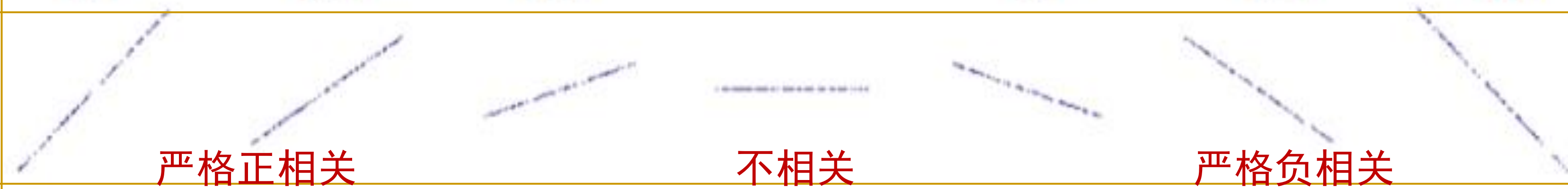
$|\rho_{XY}|$ 的值越接近于1，Y 与 X 线性关系的程度越高;

$|\rho_{XY}|$ 的值越接近于0，Y 与 X 线性关系的程度越弱.

若 $\rho_{XY}=0$, 则称 Y 与 X 不相关(没有线性关系)

结论：相关系数刻划了X 和 Y 之间线性关系的紧密程度

4. 相关系数的计算机模拟：

ρ_{XY}	1.0	0.8	0.4	0.0	−0.4	−0.8	−1.0
散点图 	严格正相关	正相关	正相关	不相关	负相关	负相关	严格负相关
ρ_{XY}	1.0	1.0	1.0		−1.0	−1.0	−1.0
散点图 		严格正相关		不相关		严格负相关	
ρ_{XY}	0.0	0.0	0.0	0.0	0.0	0.0	0.0
散点图 		不相关		不相关		不相关	

$$P(Y = aX + b) = 1 \begin{cases} a > 0, & \rho_{XY} = 1 \\ a < 0, & \rho_{XY} = -1 \end{cases}$$

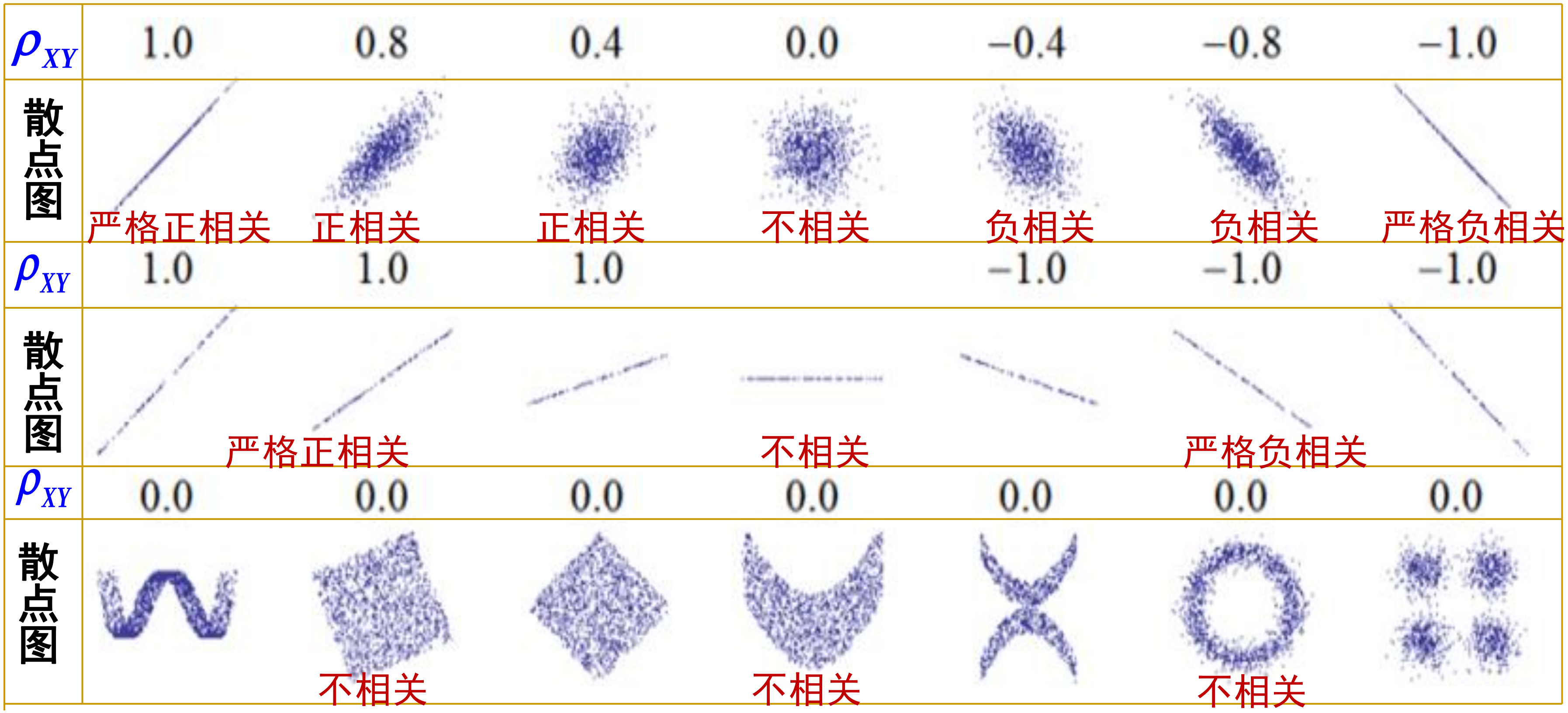
5. 独立与不相关的关系：

X 和 Y 独立 $\underset{\text{不一定}}{\rightleftarrows}$ $\rho_{XY}=0$ X 与 Y 不相关

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

结论： 设 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则 X 和 Y 是相互独立 $\Leftrightarrow \rho=0$

4. 相关系数的计算机模拟：



第四章 随机变量的数字特征

第三节 协方差与相关系数

- 协方差
- 相关系数

第四章 随机变量的数字特征

小结

随机变量的数学期望与方差

	离散型随机变量	连续型随机变量
X 数学期望	$E(X)=\sum_{k=1}^{\infty} x_k p_k$	$E(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$
$Y=g(X)$ 函数数学期望	$E(Y)=E[g(X)]$ $=\sum_{k=1}^{\infty} g(x_k)p_k$	$E(Y)=E[g(X)]$ $=\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x$
$Z=g(X,Y)$ 函数数学期望	$E(Z)=E[g(X,Y)]$ $=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} g(x_i,y_j)p_{ij}$	$E(Z)=E[g(X,Y)]$ $=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$
X 方差	$D(X)=E[X-E(X)]^2=E(X^2)-[E(X)]^2$	

小结

六种常见分布的数学期望和方差

	概率分布		$E(X)$	$D(X)$
离散型	(0-1)分布	$X \sim B(1, p)$	p	pq
	二项分布	$X \sim B(n, p)$	np	npq
	泊松分布	$X \sim P(\lambda)$	λ	λ
连续型	均匀分布	$X \sim U(a,b)$	$(a+b)/2$	$(b-a)^2/12$
	指数分布	$X \sim Exp(\theta)$	θ	θ^2
	正态分布	$X \sim N(\mu,\sigma^2)$	μ	σ^2

小结　随机变量的数字特征

$E(X)$性质	$E(c)=c$　$E(cX)=cE(X)$　$E(X+Y)=E(X)+E(Y)$ X,Y独立　$E(XY)=E(X)E(Y)$
$D(X)$性质	$D(X)=E(X^2)-[E(X)]^2$　$D(c)=0$　$D(cX)=c^2D(X)$ X,Y独立，$D(X+Y)=D(X)+D(Y)$　$D(X)=0 \Longleftrightarrow P(X=c)=1$
协方差	$Cov(X,Y)=E\{[X-E(X)][(Y-E(Y)]\}$ $=E(XY)-E(X)E(Y)$ $D(X+Y)=D(X)+D(Y)+2\operatorname{cov}(X,Y)\overset{独立}{=}D(X)+D(Y)$
相关系数	$\rho_{XY}=\dfrac{Cov(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$ (1) $\lvert\rho_{XY}\rvert\le 1$ (2) $\lvert\rho_{XY}\rvert=1 \Leftrightarrow$ 存在常数 a,b，使得：$P(Y=aX+b)=1$

作 业

授课内容	习题四
4.1 数学期望	2, 6(1), 8离散, 5, 7(1), 9(1), 11, 12连续
4.2 方差	13, 21, 22连续
4.3 协方差及相关系数	26,29,31,32

04, 9分

设A,B是随机事件，且 $P(A)=\frac{1}{4}, P(B|A)=\frac{1}{3}, P(A|B)=\frac{1}{2}$,

令$X=\begin{cases}1, & A\text{发生}\\ 0, & A\text{不发生}\end{cases}$ $Y=\begin{cases}1, & B\text{发生}\\ 0, & B\text{不发生}\end{cases}$

求：(1) (X,Y)的概率分布；(2) (X,Y)的相关系数.

解：$\frac{1}{3}=P(B|A)=\frac{P(AB)}{P(A)} \Rightarrow P(AB)=\frac{1}{12}$,

$\frac{1}{2}=P(A|B)=\frac{P(AB)}{P(B)} \Rightarrow P(B)=\frac{1}{6}$

04, 9分

设A,B是随机事件，且$P(A)=\dfrac{1}{4}, P(B|A)=\dfrac{1}{3}, P(A|B)=\dfrac{1}{2}$

令$X=\begin{cases}1, & A\text{发生}\\ 0, & A\text{不发生}\end{cases}$ $Y=\begin{cases}1, & B\text{发生}\\ 0, & B\text{不发生}\end{cases}$ $P(B)=\dfrac{1}{6},\ P(AB)=\dfrac{1}{12}$

求：(1) (X,Y)的概率分布；

Y \ X	0	1	
0	2/3	1/6	5/6
1	1/12	1/12	1/6
	3/4	1/4	1

$$P(X=0,Y=0)=P(\bar{A}\bar{B})=P(\overline{A\cup B})$$
$$=1-P(A\cup B)$$
$$=1-[P(A)+P(B)-P(AB)]=\frac{2}{3}$$

$$P(X=1,Y=0)=P(A\bar{B})$$
$$=P(A)-P(AB)=1/6$$

$$P(X=0,Y=1)=P(\bar{A}B)$$
$$=P(B)-P(AB)=1/12$$

$$P(X=1,Y=1)=P(AB)=1/12$$

04, 9分

设A,B是随机事件，且 $P(A)=\frac{1}{4}, P(B|A)=\frac{1}{3}, P(A|B)=\frac{1}{2}$

令$X=\begin{cases}1, & A\text{发生}\\ 0, & A\text{不发生}\end{cases}$ $\quad Y=\begin{cases}1, & B\text{发生}\\ 0, & B\text{不发生}\end{cases}$

求：(2) (X,Y)的相关系数. $\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)$

$=1/12-1/4\cdot 1/6=1/24$

$Y \backslash X$	0	1	
0	2/3	1/6	5/6
1	1/12	1/12	1/6
	3/4	1/4	1

$E(X)=1/4,\qquad E(Y)=1/6,$

$E(XY)=1/12,\qquad \boxed{E(XY)=\sum x_i y_j p_{ij}}$

$E(X^2)=1/4,\qquad E(Y^2)=1/6,$

$D(X)=E(X^2)-E^2(X)=3/16,$

$D(Y)=5/36,$

$$\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{1/24}{\sqrt{3/4}\cdot\sqrt{5/6}}=\frac{1}{\sqrt{15}}$$

12, 11分

设二维离散型随机变量(X,Y)的概率分布为

$X \backslash Y$	0	1	2
0	1/4	0	1/4
1	0	1/3	0
2	1/12	0	1/12

(1) 求 $P\{X=2Y\}$;

(2) 求 $\mathrm{cov}(X-Y,Y)$

12, 11分

设二维离散型随机变量(X,Y)的概率分别为

$X \backslash Y$	0	1	2
0	1/4	0	1/4
1	0	1/3	0
2	1/12	0	1/12

(1) 求 $P\{X=2Y\}$;

解：$P\{X=2Y\}=P\{X=Y=0\}+P\{X=2,Y=1\}$

$$=\frac{1}{4}+0=\frac{1}{4}$$

12, 11分

设二维离散型随机变量(X,Y)的概率分别为

(2) 求 $\mathrm{cov}(X-Y,Y)$

解：$\mathrm{cov}(X-Y,Y)=\mathrm{cov}(X,Y)-\mathrm{cov}(Y,Y)=\mathrm{cov}(X,Y)-DY=0-\frac{2}{3}=-\frac{2}{3}$

$$\mathrm{cov}(X,Y)=EXY-EX\cdot EY=\frac{2}{3}-\frac{2}{3}=0$$

$$DY=EY^2-(EY)^2=\frac{5}{3}-1=\frac{2}{3}$$

$$EX=2/3$$

$$EY=1$$

$$EY^2=0^2\times\frac{1}{3}+1^2\times\frac{1}{3}+2^2\times\frac{1}{3}=\frac{5}{3}$$

$$EXY=\sum x_i y_j p_{ij}$$

$$=1\times1\times\frac{1}{3}+2\times2\times\frac{1}{12}=\frac{2}{3}$$

X \ Y	0	1	2	
0	1/4	0	1/4	1/2
1	0	1/3	0	1/3
2	1/12	0	1/12	1/6
	1/3	1/3	1/3	

04, 4分

设 $X_1, X_2, \cdots, X_n$ 独立同分布，且其方差为 $\sigma^2 > 0$.

令 $Y = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ 则：<u>(A)</u>

$(A)\ \mathrm{cov}(X_1, Y) = \dfrac{\sigma^2}{n}$

$(B)\ \mathrm{cov}(X_1, Y) = \sigma^2$

$(C)\ D(X_1 + Y) = \dfrac{(n+2)\sigma^2}{n}$

$(D)\ D(X_1 - Y) = \dfrac{(n+1)\sigma^2}{n}$

解：

$$\mathrm{cov}(X_1, Y) = \mathrm{cov}(X_1, \frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\mathrm{cov}(X_1, \sum_{i=1}^{n} X_i)$$

$$= \frac{1}{n}[\underset{\sigma^2}{\mathrm{cov}(X_1, X_1)} + \underset{0}{\mathrm{cov}(X_1, X_2)} + \cdots + \underset{0}{\mathrm{cov}(X_1, X_n)}] = \frac{\sigma^2}{n}$$

01, 3分

将一枚硬币重复掷n次，以X,Y分别表示出现正面和反面的次数，则X与Y的相关系数= $\underline{\quad (A) \quad}$

(A) -1　　　　(B) 0

(C) 1/2　　　　(D) 1

$a=-1$ $\rho_{XY}=-1$ 严格负相关

解：$X+Y=n \rightarrow Y=n-X \rightarrow Y=-X+n$ $\boxed{P(Y=aX+b)=1}$

$$\operatorname{cov}(X,Y)=\operatorname{cov}(X,n-X)=\underbrace{\operatorname{cov}(X,n)}_{0}-\operatorname{cov}(X,X)=-DX$$

$$DY=D(n-X)=\underbrace{D(n)}_{0}+D(-X)=(-1)^2D(X)=DX$$

$$\rho_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{DX}\sqrt{DY}}=\frac{-DX}{\sqrt{DX}\sqrt{DX}}=-1$$

故选(A)

08, 4分

设随机变量 $X\sim N(0,1), Y\sim N(1,4)$，且相关系数 $\rho_{XY}=1$，

$\downarrow$

$a>0$

则 (D)

(A) $P\{Y=-2X-1\}=1$ ×

(B) $P\{Y=2X-1\}=1$ ×

(C) $P\{Y=-2X+1\}=1$ ×

(D) $P\{Y=2X+1\}=1$

解： 因为 $\rho_{XY}=1$，所以Y与X概率为1有线性关系，即

$$P\{Y=aX+b\}=1\rightarrow Y=aX+b$$

$$\rightarrow EY=E(aX+b)$$

$$\rightarrow EY=aEX+b$$

$$\rightarrow b=1$$

$X\sim N(0,1), Y\sim N(1,4)$

$$\rightarrow EX=0, EY=1$$

故选(D)

练习1

将掷一均匀硬币的实验独立重复地进行100次，用X表示出现正面的次数。求$E(X^2)$

解:

由题意， $X \sim B(100, \frac{1}{2})$

$$E(X) = np = 100 \times \frac{1}{2} = 50,$$

$$D(X) = npq = 100 \times \frac{1}{2} \times \frac{1}{2} = 25$$

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

$$E(X^2) = D(X) + [E(X)]^2 = 25 + 50^2 = 2525$$

练习2

设随机变量 X 服从拉普拉斯分布，求常数 $A, E(X), D(X)$

$$f(x) = Ae^{-\lambda|x|}, -\infty < x < +\infty, \lambda > 0$$

解：

$$1 = \int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_{-\infty}^{0} Ae^{\lambda x}\mathrm{d}x + \int_{0}^{+\infty} Ae^{-\lambda x}\mathrm{d}x$$

$$= \frac{A}{\lambda}\int_{-\infty}^{0} e^{\lambda x}\mathrm{d}(\lambda x) - \frac{A}{\lambda}\int_{0}^{+\infty} e^{-\lambda x}\mathrm{d}(-\lambda x)$$

$$= \frac{A}{\lambda} e^{\lambda x}\Big|_{-\infty}^{0} - \frac{A}{\lambda} e^{-\lambda x}\Big|_{0}^{+\infty}] = \frac{A}{\lambda}(e^{0} - e^{-\lambda\infty}) - \frac{A}{\lambda}(e^{-\lambda\infty} - e^{0})$$

$$= A\frac{2}{\lambda} \qquad \therefore A = \frac{\lambda}{2}$$

$$E(X) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x = \frac{\lambda}{2}\int_{-\infty}^{+\infty} xe^{-\lambda|x|}dx = 0$$

奇函数在对称
区间积分为零

练习2

设随机变量 X 服从拉普拉斯分布，求常数 $A, E(X), D(X)$

$$f(x)=Ae^{-\lambda|x|},-\infty<x<+\infty,\lambda>0$$

解： $E(X)=0$

$$D(X)=E(X^2)-[E(X)]^2=E(X^2)=\int_{-\infty}^{+\infty}x^2f(x)\mathrm{d}x=\frac{\lambda}{2}\int_{-\infty}^{+\infty}x^2e^{-\lambda|x|}\mathrm{d}x$$

$$=\lambda\int_0^{+\infty}x^2e^{-\lambda x}\mathrm{d}x=-\int_0^{+\infty}x^2e^{-\lambda x}\mathrm{d}(-\lambda x)=-\int_0^{+\infty}x^2\mathrm{d}e^{-\lambda x}$$ 分部积分法

$$\mathrm{d}x^2=2x\mathrm{d}x$$

$$=-[\underbrace{x^2e^{-\lambda x}\Big|_0^{+\infty}}_{0}-\int_0^{+\infty}e^{-\lambda x}\mathrm{d}x^2]=2\int_0^{+\infty}e^{-\lambda x}x\mathrm{d}x]=-\frac{2}{\lambda}\int_0^{+\infty}e^{-\lambda x}x\mathrm{d}(-\lambda x)$$

$$=-\frac{2}{\lambda}[\int_0^{+\infty}x\mathrm{d}e^{-\lambda x}]=-\frac{2}{\lambda}[\underbrace{xe^{-\lambda x}\Big|_0^{+\infty}}_{0}-\int_0^{+\infty}e^{-\lambda x}\mathrm{d}x]$$ 分部积分法

$$=\frac{2}{\lambda}\int_0^{+\infty}e^{-\lambda x}\mathrm{d}x=-\frac{2}{\lambda^2}\int_0^{+\infty}e^{-\lambda x}\mathrm{d}(-\lambda x)=-\frac{2}{\lambda^2}e^{-\lambda x}\Big|_0^{+\infty}=\frac{2}{\lambda^2}$$