第4章 多维随机变量及其分布

4.1 二维随机变量及其联合分布函数

4.2 二维离散型随机变量

4.3 二维连续型随机变量

4.4 随机变量的独立性

4.5 条件分布

4.6 二维随机变量函数的分布

由前所述，随机变量的边际分布完全由它们的联合分布确定. 但反之不成立.

$$(X,Y) \sim F(x,y),\ X \sim F_X(x),\ Y \sim F_Y(y)$$

问 什么情况下，随机变量（X，Y）的联合分布可以用它的边际分布完全确定？

§4.4　随机变量的独立性

- 独立随机变量的概念
- 二维离散随机变量的独立性
- 二维连续随机变量的独立性

回顾：事件的独立性

A, B 相互独立 $\Leftrightarrow$ $P(AB) = P(A)P(B)$

问 **怎样定义 r.v X, Y 之间的独立性？**

分析　　若 X, Y 相互“独立”，从直观上看，X, Y 取任何值之间应该是没有任何关系的，即 $\forall\ x, y \in R^1$

$$\{X \le x\},\ \{Y \le y\}$$

应相互独立，即

$$P\{X \le x, Y \le y\} = P\{X \le x\} \cdot P\{Y \le y\}$$

$$\Downarrow$$

$$F(x, y) = F_X(x) \cdot F_Y(y)$$

独立随机变量的概念

定义 设 $(X,Y)\sim F(x,y), X\sim F_X(x), Y\sim F_Y(y)$

若 $\forall\ x,y\in(-\infty,\infty)$ 有

$$P\{X\le x, Y\le y\}=P\{X\le x\}\cdot P\{Y\le y\}$$

即

$$F(x,y)=F_X(x)\cdot F_Y(y)$$

则称 r.v X,Y **相互独立.**

r. v. X，Y独立性的直观意义：

X的取值与Y的取值互不相干、是相互独立的.

设 X,Y 相互独立，则, $\forall\ x_1 < x_2, y_1 < y_2$ 有

$$P\{x_1 < X \le x_2, y_1 < Y \le y_2\}$$

$$= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$$

$$= F_X(x_2)F_Y(y_2) - F_X(x_1)F_Y(y_2) - F_X(x_2)F_Y(y_1) + F_X(x_1)F_Y(y_1)$$

$$= [F_X(x_2) - F_X(x_1)] \cdot [F_Y(y_2) - F_Y(y_1)]$$

$$= P\{x_1 < X \le x_2\} \cdot P\{y_1 < Y \le y_2\}$$

即事件 $\{x_1 < X \le x_2\}$, $\{y_1 < Y \le y_2\}$ 相互独立.

二维离散随机变量的独立性

设 (X,Y) 的联合分布列为

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

则 X,Y 相互独立等价于 $\forall\, i,j=1,2,\cdots$ 有

$$P\{X=x_i,Y=y_j\}=P\{X=x_i\}\cdot P\{Y=y_j\}$$

即
$$p_{ij}=p_{i\cdot}\cdot p_{\cdot j}\ ,(i,j=1,2,\cdots)$$

例 甲袋中有3个红球，2个白球;乙袋中有4个红球，5个白球.从甲、乙两袋中各任取两球，记 X,Y 分别表示取到白球的个数，问 X,Y 是否独立?

分析 由于从两袋中取球是相互独立的过程,所以 X,Y 的取值是相互独立、互不相干的，故 $X,$相互独立.

判断r.v的独立性的方法
① 按定义判断　② 从直观背景判断

例1 设 r.v X 从 1,2,3,4 四个数中等可能取值, 又设 r.v Y 从 $1\sim X$ 中等可能取值. 问 X,Y 是否独立?

解 由前例知,(X,Y) 联合分布列与边际分布列

$Y \backslash X$	1	2	3	4	$p_{\cdot j}$
1	1/4	1/8	1/12	1/16	25/48
2	0	1/8	1/12	1/16	13/48
3	0	0	1/12	1/16	7/48
4	0	0	0	1/16	3/48
$p_{i\cdot}$	1/4	1/4	1/4	1/4	1

$\because \quad P\{X=i,Y=j\}\neq P\{X=i\}\cdot P\{Y=j\}, \quad i,j=1,2,3,4$

$P(X=2,Y=1)\neq P(X=2)P(y=1)$

从直观上看, X, Y 也不独立。

$\therefore \quad X,Y$ 不独立.

Y的取值依赖X的取值

例2 设(X,Y)的联合分布列为

$Y \backslash X$	1	2	3
1	1/8	a	1/24
2	b	1/4	1/8

$Y \backslash X$	1	2	3
1	1/8	1/12	1/24
2	3/8	1/4	1/8

① a、b 应满足什么条件？ ② 若 X,Y 独立，求 a、b.

解 ① $\because \sum_{i,j} p_{ij} = 1 \quad ,0 \le a \le 1, 0 \le b \le 1$

$$\therefore \quad a + b = 1 - \left(\frac{1}{8} + \frac{1}{24} + \frac{1}{4} + \frac{1}{8}\right) = \frac{11}{24}$$

② 若 X,Y 相互独立，则

$$a = P\{X=2, Y=1\} = P\{X=2\} \cdot P\{Y=1\} = \left(a + \frac{1}{4}\right)\left(\frac{1}{8} + a + \frac{1}{24}\right)$$

$$b = P\{X=1, Y=2\} = P\{X=1\} \cdot P\{Y=2\} = \left(b + \frac{1}{8}\right)\left(b + \frac{1}{4} + \frac{1}{8}\right)$$

解得 $a = \frac{1}{12}$ 或 $a = \frac{1}{2}$；

$b = \frac{3}{8}$ 或 $b = \frac{1}{8}$.

$\because \quad a + b = \frac{11}{24}$

$\therefore \quad a = \frac{1}{12}, \; b = \frac{3}{8}$

二维连续随机变量的独立性

设 (X,Y) 为连续型 $r.v.$，且

$$(X,Y) \sim f(x,y), \quad X \sim f_X(x), \quad Y \sim f_Y(y)$$

若 X,Y 相互独立，则 $F(x,y) = F_X(x) \cdot F_Y(y)$

$$\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv = \int_{-\infty}^{x} f_X(u)du \cdot \int_{-\infty}^{y} f_Y(v)dv$$

$$= \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u) f_Y(v)dudv$$

从而在 $f(x,y), f_X(x), f_Y(y)$ 的连续点处有

$$f(x,y) = f_X(x) f_Y(y)$$

即有 X,Y 相互独立 $\Leftrightarrow f(x,y) \overset{a.e.}{=} f_X(x) f_Y(y)$

例3 设(X,Y)的密度函数为

$$f(x,y)=\begin{cases}2e^{-(2x+y)}, & x>0, y>0\\ 0, & 其它\end{cases}$$

问 X,Y 是否独立?

解 (参见§4.2例2), X,Y 的边际密度分别为

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)dy=\begin{cases}2e^{-2x}, & x>0,\\ 0, & x\le 0,\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)dx=\begin{cases}e^{-y}, & y>0,\\ 0, & y\le 0,\end{cases}$$

$\because \quad f(x,y)=f_X(x)f_Y(y), \quad -\infty<x,y<\infty$

$\therefore \quad X,Y$ 相互独立.

例4 在某1分钟内，信号进入收信机是**等可能的**. 若收到两个互相独立的信号的时间间隔小于0.5秒, 则信号将相互产生干扰，求两信号相互干扰的概率.

解 设两信号进入收信机时间分别为 X,Y(**分钟**),

则 $X \sim U(0,1), Y \sim U(0,1)$

$\because$ X,Y 独立, 故联合密度为

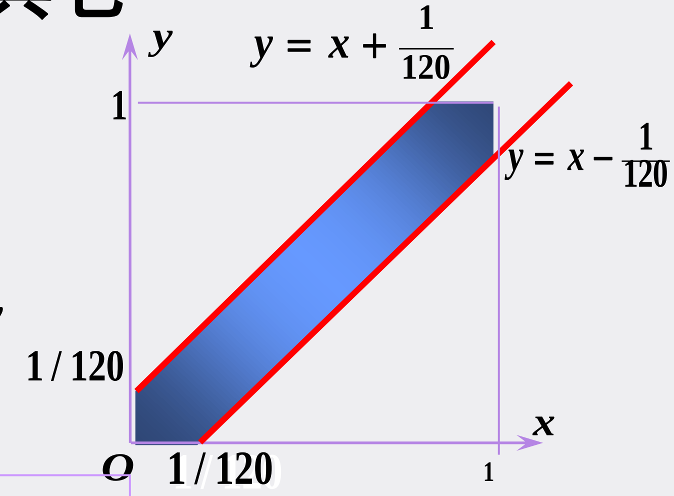
$$f(x,y)=f_X(x)f_Y(y)=\begin{cases}1, & 0<x<1, 0<y<1\\ 0, & \text{其它}\end{cases}$$

故两信号互相干扰的概率为

$$P\{|X-Y|<\frac{1}{120}\}$$

$$=\iint\limits_{|x-y|<1/120} f(x,y)dxdy=\iint\limits_{\substack{|x-y|<1/120\\ 0<x<1,0<y<1}} dxdy$$

$$=1-(1-\frac{1}{120})^2 \approx 0.016$$



会面问题

回忆：二维正态分布

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

密度函数为

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}$$

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

其中各参数满足

$$-\infty<\mu_1<\infty,-\infty<\mu_2<\infty,\sigma_1>0,\sigma_2>0,|\rho|<1$$

重要结论

$$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho) \Rightarrow \begin{cases} X\sim N(\mu_1,\sigma_1^2) \\ Y\sim N(\mu_2,\sigma_2^2) \end{cases}$$

问 参数ρ与独立性有什么关系？

定理 设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 则

$$X,Y\text{ 相互独立} \Leftrightarrow \rho=0$$

证 若(X,Y)相互独立, 则 $f(x,y)=f_X(x)f_Y(y)$

, 即

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

$$=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\}\cdot\frac{1}{\sqrt{2\pi}\sigma_2}\exp\{-\frac{(y-\mu_2)^2}{\sigma_2^2}\}$$

令$x=\mu_1, y=\mu_2$, 则有

$$\sqrt{1-\rho^2}=1 \qquad \therefore\ \rho=0$$

若 $\rho=0$, 显然

$$f(x,y)\equiv f_X(x)f_Y(y)$$

即 X,Y相互独立.

§4.5　条件分布

- 二维离散随机变量的条件分布
- 二维连续随机变量的条件分布

条件分布

回顾：条件概率

$$P(A|B)=\frac{P(AB)}{P(B)} \qquad (P(B)>0)$$

设(X,Y)为二维 r.v ,$\forall\, y\in R^1$考虑条件概率

$$P\{X\le x\,|\,Y=y\} \qquad (x\in R^1)$$

这可视为在 $\{Y=y\}$ 发生的条件下 r.v X 的概率分布 —— 条件分布

问 如何由条件概率定义并计算条件分布

$$P\{X\le x\,|\,Y=y\}=\frac{P\{X\le x, Y=y\}}{P\{Y=y\}}$$

?

二维离散随机变量的条件分布列

分析 设 (X,Y) 的频率函数为

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

(X,Y) 的关于X和Y边际分布列为分别为

$$P\{X=x_i)=p_{i\cdot}=\sum_{j=1}^{\infty}p_{ij},\quad(i=1,2,\cdots)$$

$$P\{Y=y_j)=p_{\cdot j}=\sum_{i=1}^{\infty}p_{ij},\quad(j=1,2,\cdots)$$

考虑在$\{Y=y_j\}$已发生的条件下,$\{X=x_i\}$发生的条件概率

$$P\{X=x_i\mid Y=y_j\}\quad(i=1,2,\cdots)$$

当 $p_{\cdot j}>0$ 时，由条件概率公式，有

$$P\{X=x_i\mid Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\cdot j}}\quad(i=1,2,\cdots)$$

二维离散随机变量的条件分布列

定义 设 (X,Y) 的联合分布列为

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

对于固定的 j，若 $P\{Y=y_j\}=p_{\cdot j}>0$，则称

$$P\{X=x_i\mid Y=y_j\}=\frac{p_{ij}}{p_{\cdot j}}\quad(i=1,2,\cdots)$$

为在 $Y=y_j$ 的条件下，r.v X 的条件分布列.

注：若 $p_{\cdot j}=0$，定义此概率为零.

对于固定的 i，若 $P\{X=x_i\}=p_{i\cdot}>0$，定义

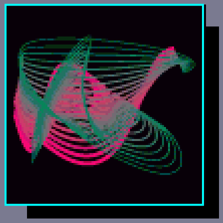
$$P\{Y=y_j\mid X=x_i\}=\frac{p_{ij}}{p_{i\cdot}}\quad(j=1,2,\cdots)$$

为在 $X=x_i$ 的条件下 r.v Y 的条件分布列.

条件分布列的性质

① $P\{X = x_i \mid Y = y_j\} \geq 0 \quad (i = 1, 2, \cdots)$

② $$\sum_{i=1}^{\infty} P\{X = x_i \mid Y = y_j\} = \sum_{i=1}^{\infty} \frac{p_{ij}}{p_{\cdot j}} = \frac{1}{p_{\cdot j}} \sum_{i=1}^{\infty} p_{ij} = \frac{1}{p_{\cdot j}} p_{\cdot j} = 1$$



这两条性质说明:

条件分布列也是一种分布列.

例1 设 r.v X 从1, 2, 3, 4 四个数中等可能取值, 又设r.v Y 从 $1 \sim X$ 中等可能取值. 问当第二次取到数字 3时第一次取四个数字的概率各是多少?

解 由§3例 ,(的联合分布列及边际分布列为

$Y \backslash X$	1	2	3	4	$p_{\cdot j}$
1	1 / 4	1 / 8	1 / 12	1 / 16	25/48
2	0	1 / 8	1 / 12	1 / 16	13/48
3	0	0	1 / 12	1 / 16	7/48
4	0	0	0	1 / 16	3/48
$p_{i\cdot}$	1/4	1/4	1/4	1/4	

故在 $Y =$ 的条件下 取到四个数字的概率是

$$P\{X=1 \mid Y=3\} = \frac{p_{13}}{p_{\cdot 3}} = \frac{0}{7/48} = 0$$

$$P\{X=2 \mid Y=3\} = \frac{p_{23}}{p_{\cdot 3}} = \frac{0}{7/48} = 0$$

$$P\{X=3 \mid Y=3\} = \frac{p_{33}}{p_{\cdot 3}} = \frac{1/12}{7/48} = \frac{4}{7}$$

$$P\{X=4 \mid Y=3\} = \frac{p_{43}}{p_{\cdot 3}} = \frac{1/16}{7/48} = \frac{3}{7}$$

例2 假设粒子计数器有缺陷，以概率 p 独立地探测到每个到达粒子. 如果单位时间内到达粒子数N（真粒子数）服从参数为λ的泊松分布，问被记录的粒子数X的分布是什么?

解 由题意，$N \sim P(\lambda)$，在给定$N=n$的条件下，X 的条件分布是试验次数为 n，成功概率为 p 的二项分布$B(n,p)$.

$$P\{X=k\}=P\{X=k,\bigcup_{n=0}^{\infty}(N=n)\}=P\{\bigcup_{n=0}^{\infty}(X=k,N=n)\}$$

$$=\sum_{n=0}^{\infty}P\{N=n\}P\{X=k\mid N=n\}=\sum_{n=k}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}C_n^k p^k(1-p)^{n-k}$$

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

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

$$\boxed{\therefore X \sim P(\lambda p)}$$

$$\boxed{e^{\lambda(1-p)}}$$

二维连续随机变量的条件概率密度

问 如何定义连续随机变量的条件分布

$P\{X \le x \mid Y = y\}$?

$\forall\ \varepsilon > 0$, 考虑条件概率

$$P\{X \le x \mid y < Y \le y+\varepsilon\} = \frac{P\{X \le x, y < Y \le y+\varepsilon\}}{P\{y < Y \le y+\varepsilon\}}$$

称为条件分布

$$= \frac{\int_{-\infty}^{x} \int_{y}^{y+\varepsilon} f(u,v)\,dv\,du}{\int_{y}^{y+\varepsilon} f_Y(y)\,dy}$$

应用积分中值定理

$$= \frac{\varepsilon \int_{-\infty}^{x} f(u, y_\varepsilon)\,du}{\varepsilon\, f_Y(\tilde{y}_\varepsilon)} \to \int_{-\infty}^{x} \frac{f(u,y)}{f_Y(y)}\,du \quad (\varepsilon \to 0)$$

称为条件密度

定义 设(X,Y)的概率密度为$f(x,y)$,若对于固定的y,(X,Y)关于Y的边际密度$f_Y(y)>0$,则称

$$f_{X|Y}(x\mid y)\triangleq\frac{f(x,y)}{f_Y(y)}\quad(-\infty<x<\infty)$$

为在$Y=y$的条件下,X的条件密度. 称

$$F_{X|Y}(x\mid y)\triangleq\int_{-\infty}^{x}f_{X|Y}(u\mid y)du\quad(-\infty<x<\infty)$$

为在$Y=y$的条件下,X的条件分布(函数).

类似地,可定义

$$f_{Y|X}(y\mid x)\triangleq\frac{f(x,y)}{f_X(x)}\quad(-\infty<y<\infty)$$

$$F_{Y|X}(y\mid x)\triangleq\int_{-\infty}^{y}f_{Y|X}(v\mid x)dv\quad(-\infty<y<\infty)$$

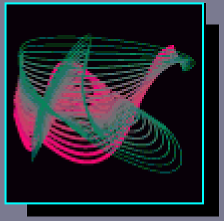
条件密度的性质

① $f_{X|Y}(x\,|\,y) \geq 0$

② $\int_{-\infty}^{\infty} f_{X|Y}(u\,|\,y)du = 1$

$$= \int_{-\infty}^{\infty} \frac{f(u,y)}{f_Y(y)} du \quad = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(u,y)du$$

$$= \frac{1}{f_Y(y)} \cdot f_Y(y) \quad = 1$$



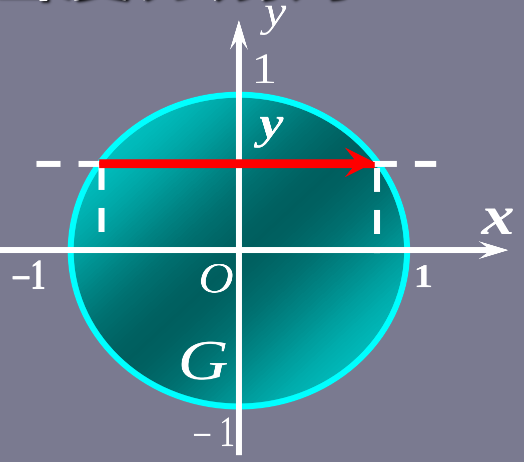
这两条性质说明:

条件密度也是一种密度.

例3 设(X,Y)服从圆域$G: x^2+y^2\le 1$上的均匀分布. 求条件概率密度$f_{X|Y}(x|y)$.

解 由于(X,Y)的联合密度及Y的边际密度分别为

$$f(x,y)=\begin{cases}1/\pi, & x^2+y^2\le 1\\ 0, & 其它\end{cases}$$



$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)dx=\begin{cases}\frac{2}{\pi}\sqrt{1-y^2}, & |y|<1\\ 0, & |y|\ge 1\end{cases}$$

故当$-1<y<1$时有

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\begin{cases}\dfrac{1}{2\sqrt{1-y^2}}, & -\sqrt{1-y^2}<x<\sqrt{1-y^2}\\ 0\ , & 其它\ x\end{cases}$$

结果表明，在 $Y=y(|y|<1)$ 条件下，X在区间 $(-\sqrt{1-y^2},\sqrt{1-y^2})$ 上的是均匀分布的.

例4 设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,

则经过计算可得

$$f_{Y|X}(y\,|\,x)=\frac{f(x,y)}{f_X(x)}$$

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

$$\sim N(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2)).$$

即：

二维正态分布，给定 X 时 Y 的条件密度是一维(单变量)正态分布.

由条件密度的定义

$$f_{Y|X}(y|x) \triangleq \frac{f(x,y)}{f_X(x)} \quad (-\infty < y < \infty)$$

得

$$f(x,y) = f_{Y|X}(y|x) f_X(x)$$

即：联合密度可以用边际密度和条件密度表示.

两边关于 x 积分，得Y的边际密度可表示为

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) \mathrm{d}x$$

连续情形的全概率公式

特别地，如果随机变量*X*,*Y*相互独立，则

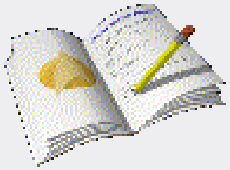
$$f(x,y) \overset{a.e.}{=} f_X(x) f_Y(y)$$

从而有

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} \overset{a.e.}{=} \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x)$$

即

$$f_{X|Y}(x \mid y) \overset{a.e.}{=} f_X(x)$$



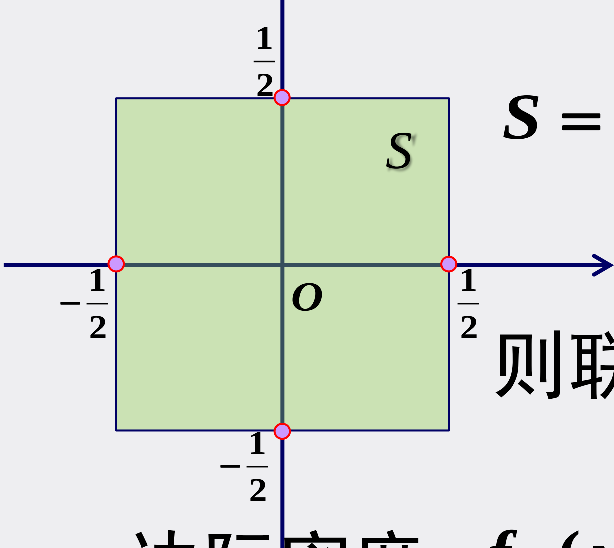
课后作业

P89： 习题4

8，9，10，12，13，14，31，35。

思考与练习

设 (X,Y) 服从正方形 S 上的均匀分布, 其中



$$S=\{(x,y)\,|\,-\frac{1}{2}\le x\le \frac{1}{2},-\frac{1}{2}\le y\le \frac{1}{2}\}$$

则联合密度 $f_{XY}(x,y)=\begin{cases}1, & (x,y)\in S\\ 0, & (x,y)\notin S\end{cases}$

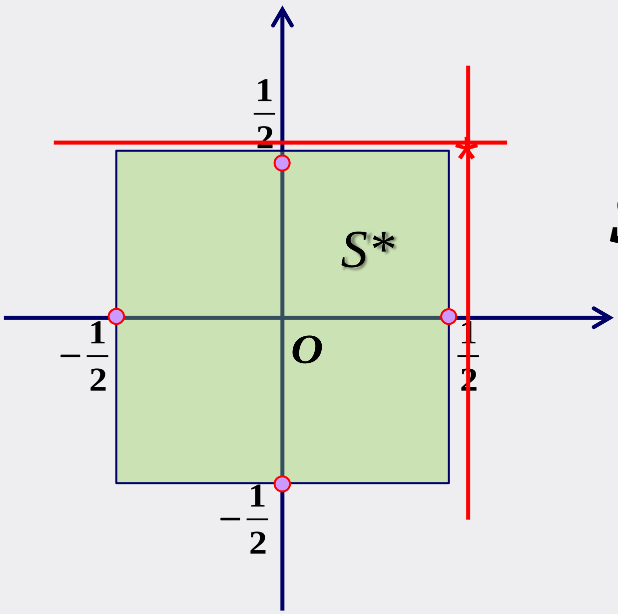
边际密度 $f_X(x)=1,\quad -1/2\le x\le 1/2$

$$f_Y(y)=1,\quad -1/2\le y\le 1/2$$

易知 $f_{XY}(x,y)=f_X(x)f_Y(y)$,

$\therefore (X,Y)$独立.

现考虑将 S 旋转45°, 正方形S → 菱形S*.



$$S=\{(x,y)\mid -\frac{1}{2}\le x\le\frac{1}{2},-\frac{1}{2}\le y\le\frac{1}{2}\}$$

$$S^*=\{(x,y)\mid |x+y|\le\frac{\sqrt{2}}{2}\}$$

$$f(x,y)=\begin{cases}1, & (x,y)\in S^*\\ 0, & (x,y)\notin S^*\end{cases}$$

验证 $f_X(x), f_Y(y)$ 不再是均匀分布密度.

且X,Y 也不再独立.

(如: $f_X(0.7)>0,\ f_Y(0.7)>0,\ f_{XY}(0.7,0.7)=0.$)