概率论与数理统计

第十七讲

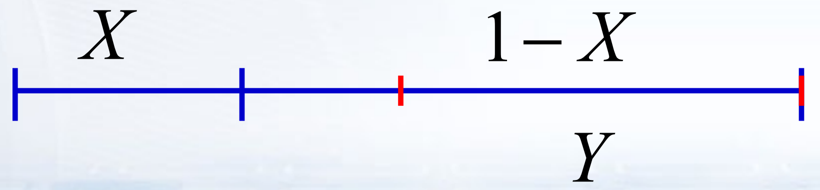
条件分布与独立性（Ⅱ）

1. 利用条件密度函数来表示联合密度

有些实际问题中，往往先已知条件密度. 此时类似乘法公式，可利用它来确定联合密度.

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

例（构成三角形） 设有长为1的木棍，随机截取一段长为X，再在剩下的一段中随机截取一段长为Y，问三段能构成三角形的概率为多少?



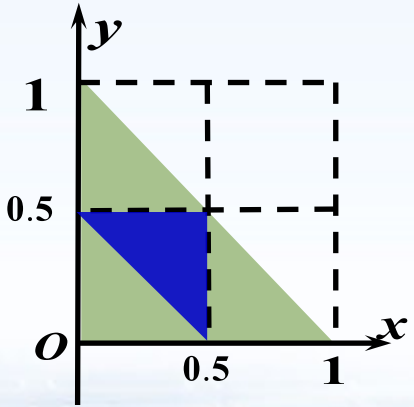
第17讲 条件分布与独立性（II）

分析 随机截取意味着均匀分布，由题意知 $X \sim U(0, 1]$，在 $X=x$ 的条件下，Y 的条件分布为 $U[0,1-x]$，因此 (X, Y) 的联合密度函数为

$$f(x,y) = f_{Y|X}(y|x) f_X(x) = \begin{cases} \dfrac{1}{1-x}, & 0 \le y \le 1-x \le 1 \\ 0, & \text{其他} \end{cases}$$

事件A “可以构成三角形”描述为(x, y)的点集Δ满足

$$\begin{cases} x+y \ge 1-(x+y) \\ x+1-(x+y) \ge y \\ y+1-(x+y) \ge x \end{cases} \Rightarrow \begin{cases} x+y \ge \dfrac{1}{2} \\ x \le \dfrac{1}{2} \\ y \le \dfrac{1}{2} \end{cases}$$



即 $\iint\limits_{\Delta} f_{Y|X}(y|x) f_X(x) \mathrm{d}y \mathrm{d}x$

第17讲 条件分布与独立性(II)

所以 $$P(A)=\iint\limits_{\Delta}\frac{1}{1-x}dydx=\int_0^{0.5}\int_{0.5-x}^{0.5}\frac{1}{1-x}dydx=\ln 2-\frac{1}{2}\approx 0.1931$$

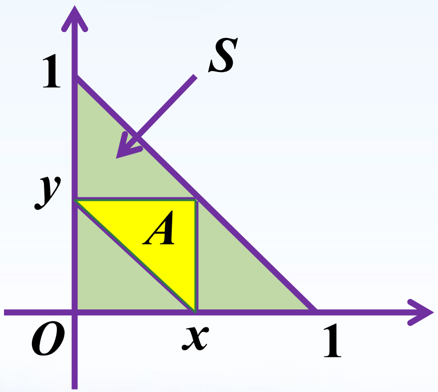
例将一根单位长的木棒随机折成三段，求刚好能构成三角形的概率。

问：样本点是什么？S是什么？

提示：应满足

$$A:\begin{cases}x+y>1-(x+y)\\ x+1-(x+y)>y\\ y+1-(x+y)>x\end{cases}\Rightarrow\begin{cases}x+y>\frac{1}{2}\\ y<\frac{1}{2}\\ x<\frac{1}{2}\end{cases}$$

$$P(A)=\frac{1}{4}$$

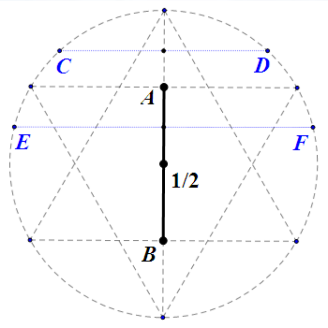


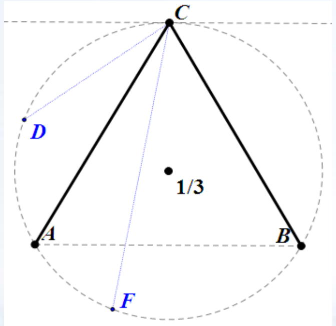
- **问题** 为什么, 同样的问题答案不同？

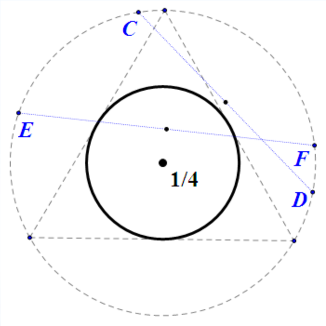
与对“随机”的理解不同, 假设的概率空间有关.

概率论**悖论**(奇论).

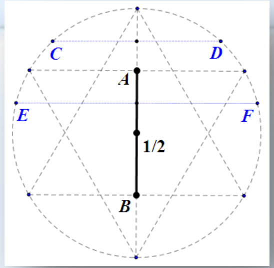
Bertrand悖论 在单位圆内任作一弦，问其长度大于内接正三角形的边长的概率为多少 (1899)



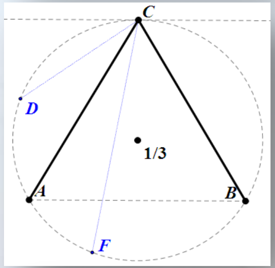




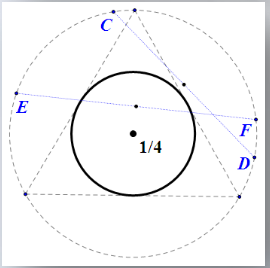
第17讲 条件分布与独立性(II)



AB为直径的1/2，画弦时，CD小于边长，EF大于边长。需要假设中点在直径上等概率分布



∠ACB为 π 的1/3，画弦时，CD小于边长，CF大于边长。需要弦的方向等概率分布，即任意相邻两共端点的弦夹角相等，由于这个夹角是圆周角，圆周角相等意味着弦长相等，弦长相等意味着内接正多边形。所以这个要求可以等效为需要弦的端点在圆周上等概率分布。



小圆面积为大圆面积的1/4，画弦时，CD小于边长，EF大于边长。需要假设中点在圆内部等概率分布

第17讲 条件分布与独立性（II）

同一问题有三种不同答案，究其原因在于圆内“取弦”时规定尚不够具体，不同的“等可能性假定”导致了不同的样本空间，具体如下：

其中“均匀分布”应理解为“等可能取点”。

一中假定弦的中点在直径上均匀分布，直径上的点组成样本空间$\Omega 1$.

二中假定弦的另一端在圆周上均匀分布，圆周上的点组成样本空间$\Omega 2$.

三中假定弦的中点在大圆内均匀分布，大圆内的点组成样本空间$\Omega 3$.

2. 随机变量独立性

回顾随机事件A, B的独立性, 指

$$P(AB)=P(A)P(B)$$

将随机事件具体为由随机变量刻画的事件：

$A=\{X\le x\}$，$B=\{Y\le y\}$，就得到随机变量 X, Y 独立的定义.

第17讲 条件分布与独立性(II)

- **定义** 随机变量 X, Y 独立, 指其联合分布函数与边缘分布函数有下式成立：

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

对于离散型随机变量X, Y，设它们的取值分别是 $x_i, y_i,\ i,j=1,2,\cdots,$ 相应的联合分布律, 边缘分布律记为 $p_{ij}, p_{i\bullet}, p_{\bullet j}$, 则由 $F(x_i,y_j)=F_X(x_i)F_Y(y_j)$ 对一切i,j 成立.

得到 $\sum_{k=1}^{i}\sum_{l=1}^{j} p_{kl} = \sum_{k=1}^{i} p_{k\bullet}\sum_{l=1}^{j} p_{\bullet l}$

依次取$i,\ j = 1, 2, \ldots,$ 递推得到

$$p_{ij} = p_{i\bullet}p_{\bullet j}, \quad i, j = 1, 2, \cdots$$

上式对于离散型随机变量使用更方便.

例 设 X, Y 的联合分布律如下表

若 X, Y 独立, 试确定未知数 $a,\ b, c$.

$Y \backslash X$	1	2	3
1	0.08	a	0.2
2	b	0.18	c

解：由 $p_{21} = p_{2\bullet} p_{\bullet 1} \Longrightarrow (a+0.18)(a+0.28) = a$

$\Longrightarrow a^2 - 0.54a + 0.18 \times 0.28 = 0$

得 $a = 0.12$, 或 $a = 0.42$

当 a =0.12 时, 由 $p_{11} = p_{1\bullet} p_{\bullet 1}$

$\Longrightarrow (b+0.08) \times 0.4 = 0.08 \Longrightarrow b = 0.12$

由 $p_{32} = p_{3\bullet} p_{\bullet 2} \Longrightarrow (c+0.2) \times 0.4 = 0.2 \Longrightarrow c = 0.3$

当 a =0.42 时, 可算出 $b = \dfrac{6}{175}$, $c = \dfrac{3}{35}$

两组解答均合题意.

对于连续型随机变量, 设 (X, Y) 的密度函数与相应的边缘密度函数分别为 $f(x,y)$, $f_X(x)$, $f_Y(y)$, 则在它们连续点处, 关于分布函数等式两边求偏导：

$$\frac{\partial^2}{\partial x \partial y} F(x,y) = \frac{\partial^2}{\partial x \partial y}[F_X(x)F_Y(y)]$$

得到 $$f(x,y) = f_X(x) f_Y(y)$$

同样对于连续型随机变量, 后一式更常用.

例 设 (X, Y) 的密度函数如下：

$$f(x,y)=\begin{cases} e^{-(x+y)}, & x\ge 0,\ y\ge 0 \\ 0, & \text{其他} \end{cases}$$

问 X, Y 是否独立?

解： 因 $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y=\begin{cases} \int_0^{+\infty} e^{-(x+y)}\mathrm{d}y, & x\ge 0 \\ 0, & x<0 \end{cases}$

$$=\begin{cases} e^{-x}, & x\ge 0 \\ 0, & x<0 \end{cases}$$

同理 $f_Y(y)=\begin{cases} e^{-y}, & y\ge 0 \\ 0, & y<0 \end{cases}$

故 $f(x,y)=f_X(x)f_Y(y)$, 即 X, Y 独立.

可见, 若 (X, Y) 为独立的连续型随机变量, 则其联合密度函数 $f(x, y)$ 可以表成分离形式

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

反之, 若有 (X, Y) 的密度函数可表成分离形式

$$f(x,y)=g(x)h(y)$$

则 X, Y 是独立的.

因为, 此时

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y = \int_{-\infty}^{+\infty} g(x)h(y)\mathrm{d}y = c_1 g(x)$$

同理 $f_Y(y)=c_2 h(y)$, 显然 $c_1 c_2 = 1.$

故 $f(x,y)=g(x)h(y)=f_X(x)f_Y(y).$

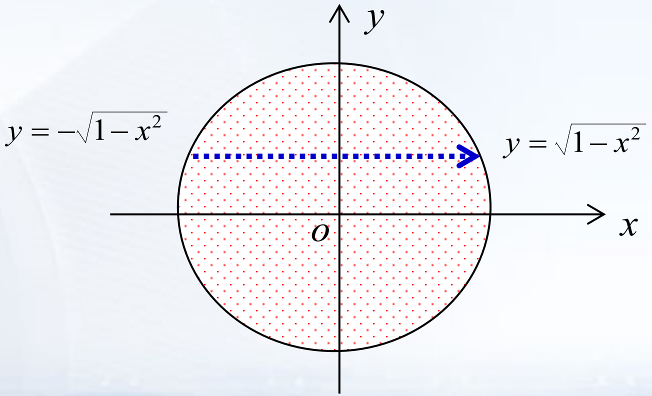
- **注** 所谓“分离形式”, 应连同其支撑区域.

  见下例.

例 设 (X, Y) 的密度函数如下：

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

求 $f_X(x)$, $f_Y(y)$, 并问二者是否独立？



解： $f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}\mathrm{d}y=\frac{2\sqrt{1-x^2}}{\pi},\ -1\le x\le 1$

$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x=\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}\frac{1}{\pi}\mathrm{d}x=\frac{2\sqrt{1-y^2}}{\pi},\ -1\le y\le 1$

故 $f(x,y)\neq f_X(x)f_Y(y)$, 可见, 两者不独立.

- **分析** 密度函数形式可分离, 但支撑区域不可分离.

此外, $\forall\, x_1, x_2, \cdots, x_n \in R$, 若有

$$F(x_1, x_2, \cdots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

则称 $X_1, X_2, \cdots, X_n$ 相互独立.

设 $(X_1, X_2, \cdots, X_m) \sim F_1(x_1, x_2, \cdots, x_m)$

$(Y_1, Y_2, \cdots, Y_n) \sim F_2(y_1, y_2, \cdots, y_n)$

$(X_1, X_2, \cdots, X_m = Y_1, Y_2, \cdots, Y_n) \sim F(x_1, x_2, \cdots, x_m = y_1, y_2, \cdots, y_n)$

$\forall\, x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_n \in R$ 若有

$$F(x_1, x_2, \cdots, x_m = y_1, y_2, \cdots, y_n) = F_1(x_1, x_2, \cdots, x_m) \cdot F_2(y_1, y_2, \cdots, y_n)$$

则称 $(X_1, X_2, \cdots, X_m)$, $(Y_1, Y_2, \cdots, Y_n)$ 相互独立.

第17讲 条件分布与独立性（II）

定理 $(X_1,X_2,\cdots,X_m)$, $(Y_1,Y_2,\cdots,Y_n)$ 相互独立, 则

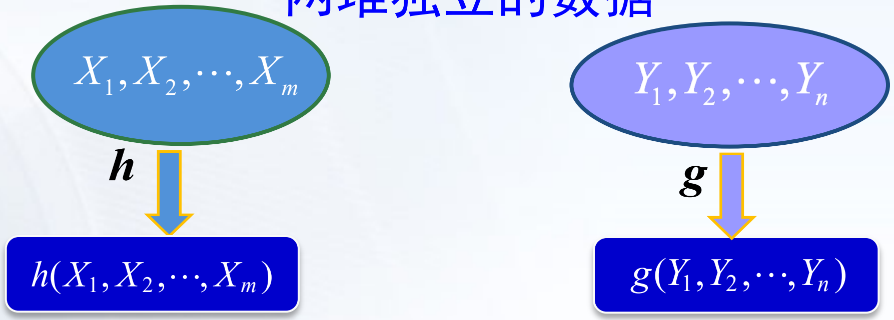
① X_i, Y_j 相互独立 $(i=1,2,\cdots,m\ ;j=1,2,\cdots,n)$

② 设 h,g 是两个通常的函数,则

$$h(X_1,X_2,\cdots,X_m),\ g(Y_1,Y_2,\cdots,Y_n)$$

相互独立.

两堆独立的数据



处理后的数据仍是独立的