



大数据的统计学基础——第2周

【声明】 本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

<http://edu.dataguru.cn>

关注炼数成金企业微信



■提供全面的数据价值资讯，涵盖商业智能与数据分析、大数据、企业信息化、数字化技术等，各种高性价比课程信息，赶紧掏出您的手机关注吧！



◆ 总体方差： $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$

◆ 样本方差： $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

```
> a=c(1,2,3,4,5)
> ((1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2)/5
[1] 2
> var(a)
[1] 2.5
```

◆ Excel中求方差也有总体方差与样本方差之分

- ◆ 四分位数的选择具有争议性
- ◆ 分位数的数学定义：
- ◆ 选择四分位的百分比值 y ，及样本总量 n ，分位数的位置可以由下面的公式计算：

$$L_y = n * \left(\frac{y}{100}\right)$$

- ◆ 情况1: 如果 L 是一个整数，则取 第 L 和 第 $L+1$ 的平均值
- ◆ 情况2: 如果 L 不是一个整数，则取下一个最近的整数。（比如1.25, 则取 2 ）
- ◆ 对于1 , 2 , 3 , 4 , 5
- ◆ 下四分位数： $y=25$,所以 $L_{25} = n * \left(\frac{25}{100}\right) = n * 0.25 = 5 * 0.25 = 1.25$ ，故下四分位数应该取位于第二位置的2
- ◆ 1 , 2 , 3 , 4 , 5 , 6 , 7
- ◆ $7 * 0.25 = 1.75$ ——2

- ◆ 如何设计一场赌博的规则，使得对所有参与者都公平？？
- ◆ 赌博的公平性：每个人赢的**概率**都一样
- ◆ 赌大小的公平性：对于一般人来说，赌大小只会押大或小，赔率是1:1，4到10为小,7到11为大，当出现三个骰子点数一样时，庄家大小通吃。赌大小的规则公平吗？——要解决这个问题，先学习概率的基础知识



◆ **试验**：对某种自然现象作一次观察或进行一次科学试验。例如：

E_1 ：抛一枚硬币，观察正面 H 、反面 T 出现的情况。

E_2 ：将一枚硬币抛掷三次，观察正面 H 、反面 T 出现的情况。

E_3 ：将一枚硬币抛掷三次，观察出现正面的次数。

E_4 ：抛一颗骰子，观察出现的点数。

E_5 ：记录某城市 120 急救电话台一昼夜接到的呼唤次数。

E_6 ：在一批灯泡中任意抽取一只，测试它的寿命。

E_7 ：记录某地一昼夜的最高温度和最低温度。

◆ 上面的试验，具有以下特点：

◆ 1. 可以在相同的条件下重复进行

◆ 2. 试验的可能结果不止一个，但在试验前可以知道所有可能结果

◆ 3. 试验前不能确定哪个结果会出现

◆ 拥有以上3个特点的试验称为**随机试验**

- ◆ 对于随机试验E，E的所有可能结果组成的集合称为E的**样本空间**，记为S。其中，S中的元素，即E的每个可能结果，称为**样本点**。

实例1 抛掷一枚硬币,观察字面,花面出现的情况.



$H \rightarrow$ 字面朝上

$S_1 = \{H, T\}.$ $T \rightarrow$ 花面朝上

实例2 抛掷一枚骰子,观察出现的点数.



$S_2 = \{1, 2, 3, 4, 5, 6\}.$

实例3 记录某公共汽车站某日上午某时刻的等车人数.

$$S_4 = \{0, 1, 2, \dots\}.$$



实例4 考察某地区 12 月份的平均气温.

$$S_5 = \{t | T_1 < t < T_2\}.$$

其中 t 为平均温度.



实例5 从一批灯泡中任取一只, 测试其寿命.

$$S_6 = \{t \mid t \geq 0\}.$$

其中 t 为灯泡的寿命.

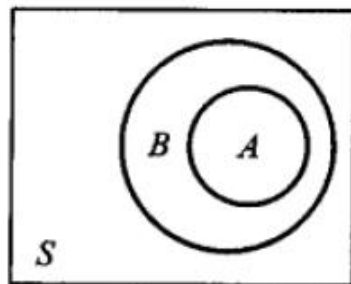


实例6 记录某城市120 急救电话台一昼夜接到的呼唤次数.

$$S_7 = \{0, 1, 2, \dots\}.$$



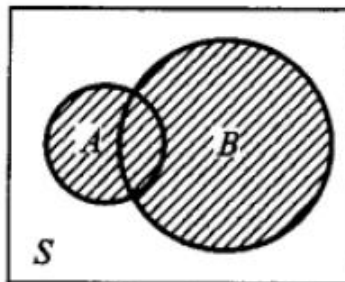
- ◆ 一般地，我们称试验E的样本空间S的某个子集为E的**随机事件**，简称事件。一般用大写字母A，B，C.....表示。
- ◆ 例如，在抛骰子中，“所得点数为偶数”是一个随机事件A，“所得点数为1点”也是一个随机事件B
- ◆ 由一个样本点组成的单点集，称为**基本事件**。抛骰子中，“所得点数是3点”是一个基本事件C。在抛骰子这个试验中，一共有6个基本事件。
- ◆ 在每次试验中，当事件中的某个样本点出现时，称这个**事件发生**。抛骰子中，如果抛得点数为4点，那么我们可以称事件A发生。
- ◆ **必然事件**：在每个试验中一定会发生的事件。抛骰子中，事件D：“点数小于等于6点”是必然事件
- ◆ **不可能事件**：在每个试验中一定不会发生的事件,用 \emptyset 表示。抛骰子中，事件E：“点数大于6点”是不可能事件



$$A \subset B$$

图 1-1

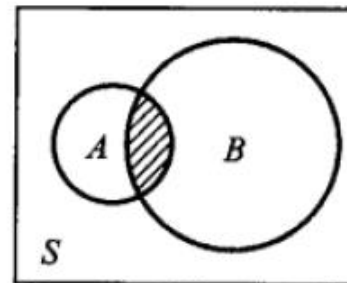
包含



$$A \cup B$$

图 1-2

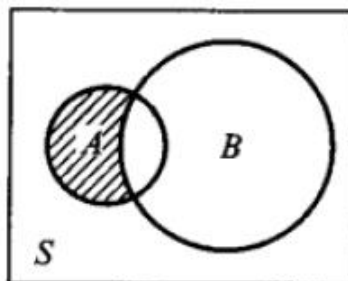
和事件



$$A \cap B$$

图 1-3

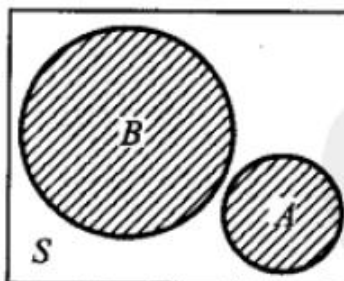
积事件



$$A - B$$

图 1-4

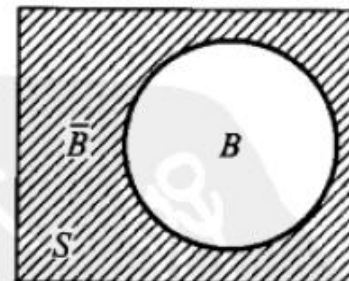
差事件



$$A \cap B = \emptyset$$

图 1-5

互斥事件



$$B \cup \bar{B} = S, B \cap \bar{B} = \emptyset$$

图 1-6

逆事件

- ◆ 抛骰子中，事件A：“出现点数为奇数”，事件B：“出现点数小于等于3”，事件C：“出现点数为2点”，事件D：“出现点数为偶数”，则有
- ◆ 1. 事件B包含事件C
- ◆ 2. $A \cup B$ 表示：“出现点数为奇数或小于等于3点”，即“出现1,2,3,5点”
- ◆ 3. $A \cap B$ 表示：“出现点数为奇数且小于等于3点”，即“出现1,3点”
- ◆ 4. $B - C$ 表示：“出现点数小于等于3但不等于2”，即“出现1,3点”。故 $A \cap B = B - C$
- ◆ 5. 事件A与事件C互斥，但不互逆
- ◆ 6. 事件A与事件D互为逆事件

2. 设 A, B, C 为三个事件, 用 A, B, C 的运算关系表示下列各事件:

(1) A 发生, B 与 C 不发生. $A\bar{B}\bar{C}$

(2) A 与 B 都发生, 而 C 不发生. $AB\bar{C}$

(3) A, B, C 中至少有一个发生. $A \cup B \cup C$

(4) A, B, C 都发生. ABC

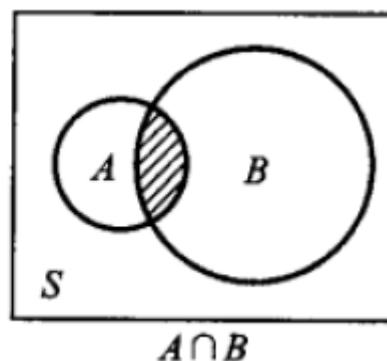
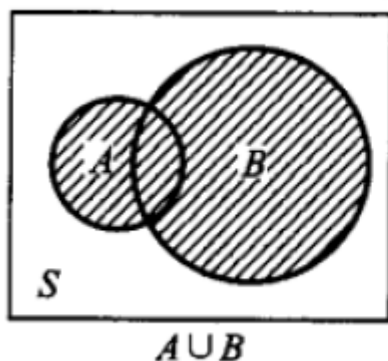
(5) A, B, C 都不发生. $\bar{A}\bar{B}\bar{C}$

(6) A, B, C 中不多于一个发生. $\bar{A}\bar{B} \cup \bar{B}\bar{C} \cup \bar{A}\bar{C}$

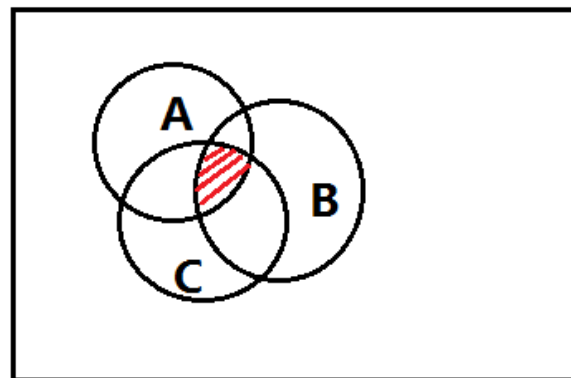
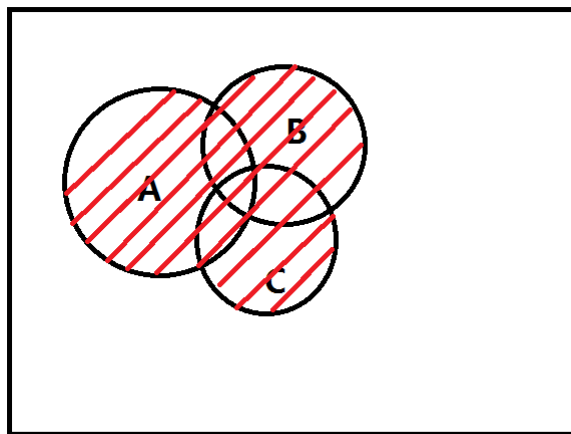
(7) A, B, C 中不多于两个发生. $\bar{A} \cup \bar{B} \cup \bar{C}$

(8) A, B, C 中至少有两个发生. $AB \cup BC \cup AC$

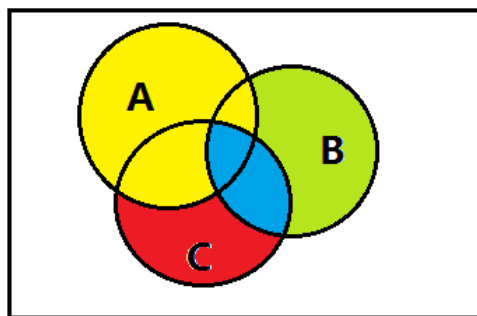
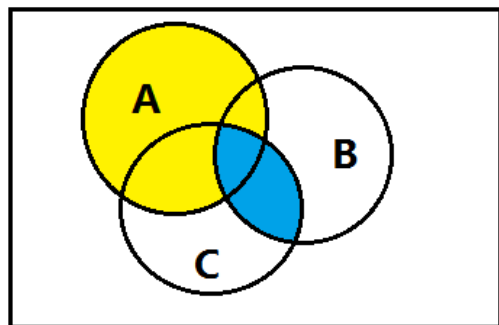
◆ 交换律： $A \cup B = B \cup A$ ； $A \cap B = B \cap A$



◆ 结合律： $A \cup (B \cup C) = (A \cup B) \cup C$ ； $A \cap (B \cap C) = (A \cap B) \cap C$

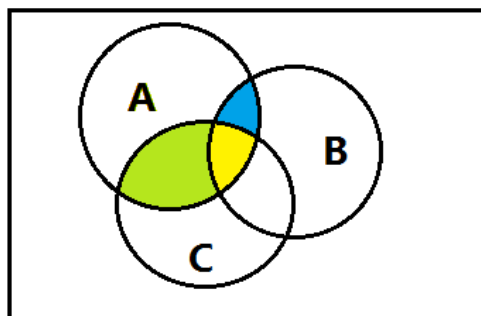
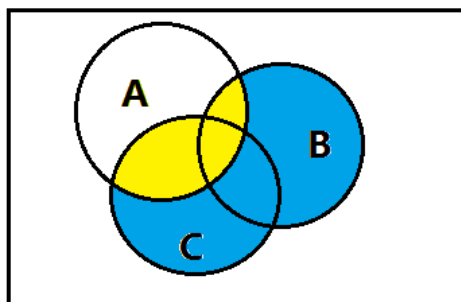


◆ 分配律： $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$



◆ $A \cup (B \cap C)$ 是黄色+蓝色 $(A \cup B)$ 是黄色+蓝色+绿色； $(A \cup C)$ 是黄色+蓝色+红色

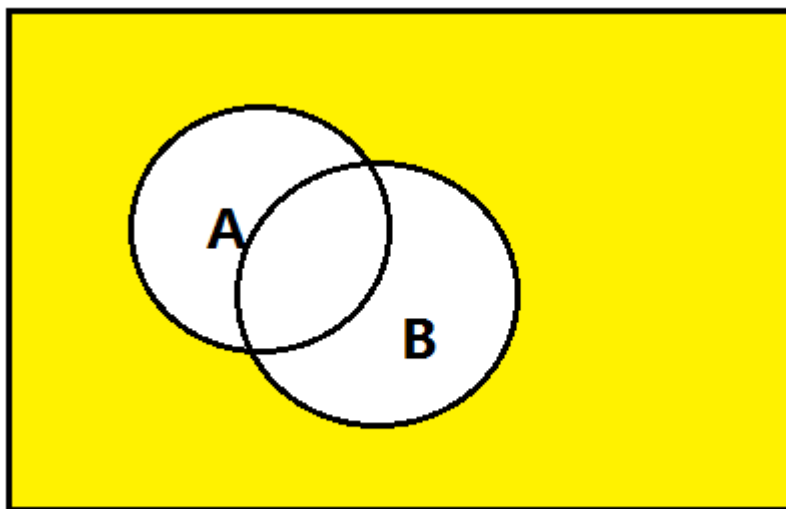
◆ $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$



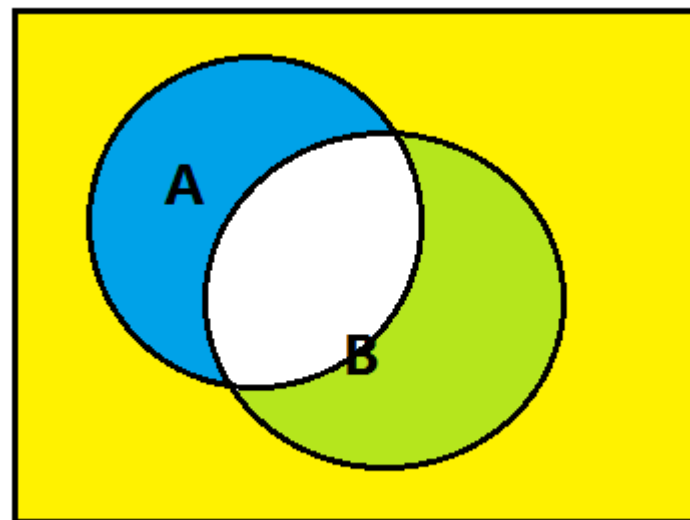
◆ $A \cap (B \cup C)$ 是黄色

$(A \cap B)$ 是蓝色+黄色； $(A \cap C)$ 是绿色+黄色

◆ 德摩根律： $\overline{A \cup B} = \bar{A} \cap \bar{B}$

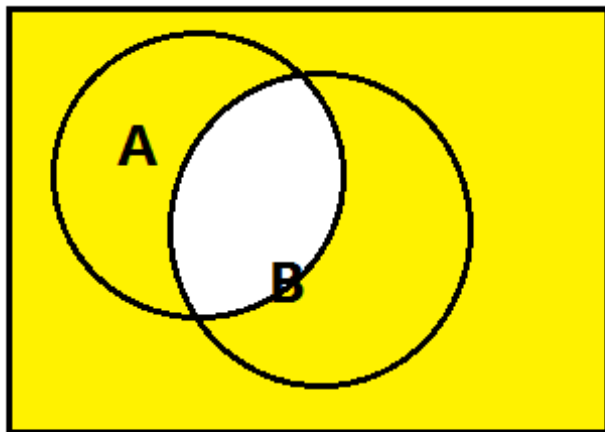


黄色部分表示 $\overline{A \cup B}$

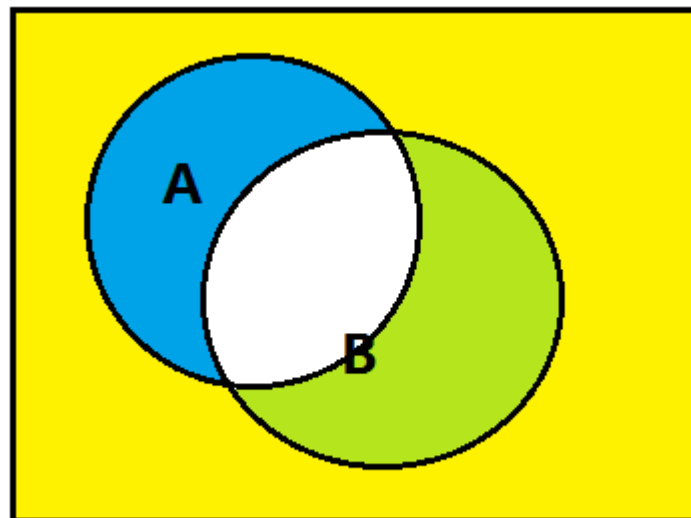


黄色+绿色部分表示 \bar{A}
黄色+蓝色部分表示 \bar{B}

◆ 德摩根律： $\overline{A \cap B} = \bar{A} \cup \bar{B}$



黄色部分表示 $\overline{A \cap B}$



黄色+绿色部分表示 \bar{A}
黄色+蓝色部分表示 \bar{B}

例 1 在 E_2 中事件 A_1 : “第一次出现的是 H ”, 即

$$A_1 = \{HHH, HHT, HTH, HTT\}.$$

事件 A_2 : “三次出现同一面”, 即

$$A_2 = \{HHH, TTT\}.$$

在 E_6 中, 事件 A_3 : “寿命小于 1000 小时”, 即

$$A_3 = \{t \mid 0 \leq t < 1000\}.$$

在 E_7 中, 事件 A_4 : “最高温度与最低温度相差 10 摄氏度”, 即

$$A_4 = \{(x, y) \mid y - x = 10, T_0 \leq x \leq y \leq T_1\}.$$

例 2 在例 1 中有

$$A_1 \cup A_2 = \{HHH, HHT, HTH, HTT, TTT\},$$

$$A_1 \cap A_2 = \{HHH\},$$

$$A_2 - A_1 = \{TTT\},$$

$$\overline{A_1 \cup A_2} = \{THT, TTH, THH\}.$$

- ◆ 在相同的条件下，重复 n 次试验，事件 A 发生的次数 n_A 称为 A 发生的频数， $\frac{n_A}{n}$ 称为事件 A 发生的频率。

例 1 考虑“抛硬币”这个试验，我们将一枚硬币抛掷 5 次、50 次、500 次，各做 10 遍，得到数据如表 1-1 所示（其中 n_H 表示 H 发生的频数， $f_n(H)$ 表示 H 发生的频率）。

表 1-1

实验序号	$n=5$		$n=50$		$n=500$	
	n_H	$f_n(H)$	n_H	$f_n(H)$	n_H	$f_n(H)$
1	2	0.4	22	0.44	251	0.502
2	3	0.6	25	0.50	249	0.498
3	1	0.2	21	0.42	256	0.512
4	5	1.0	25	0.50	253	0.506
5	1	0.2	24	0.48	251	0.502
6	2	0.4	21	0.42	246	0.492
7	4	0.8	18	0.36	244	0.488
8	2	0.4	24	0.48	258	0.516
9	3	0.6	27	0.54	262	0.524
10	3	0.6	31	0.62	247	0.494

- ◆ 从抛硬币的多次试验结果可以看出，当试验重复次数较少时，事件H（正面向上）发生的频率在0到1之间随机浮动。但是，当试验重复次数较多时，事件H发生的频率却围绕着0.5上下波动，并逐步地稳定于0.5。

实验者	n	n_H	$f_n(H)$
德摩根	2 048	1 061	0.518 1
蒲 丰	4 040	2 048	0.506 9
K·皮尔逊	12 000	6 019	0.501 6
K·皮尔逊	24 000	12 012	0.500 5

随着试验次数的增加，事件H的频率与0.5之间的差距越来越小

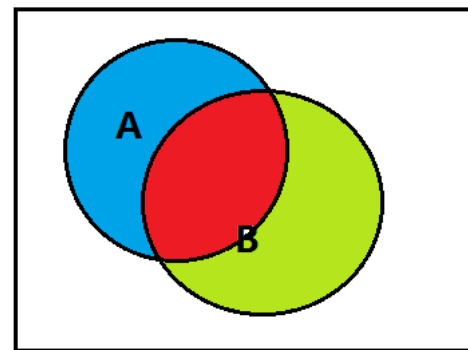
- ◆ 大量的试验证明，当试验的重复次数 n 逐渐增大时，事件 A 发生的频会逐渐稳定于某个常数 p 。这个 p 就是事件 A 发生的概率，用于表示在一次试验中，事件 A 发生的可能性大小。记事件 A 的概率为 $P(A)$ 。
- ◆ 概率需要满足的条件：
- ◆ 1. **非负性**： $P(A) \geq 0$;
- ◆ 2. **规范性**：对于必然事件 S ，有 $P(S)=1$
- ◆ 3. **可列可加性**：对于两两互不相容的事件 A_1, A_2, A_3, \dots ，即 $A_i \cap A_j = \emptyset, i \neq j, i, j=1, 2, \dots$ ，有 $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$
- ◆ 例：抛3次硬币中，事件 A ：“正面向上次数是1”，事件 B ：“正面向上次数是2”，事件 C ：“正面向上次数是3”，事件 D ：“正面向上次数至少是1”。
- ◆ $D=A \cup B \cup C$ ，且 A, B, C 互不相容，则 $P(D) = P(A \cup B \cup C) = P(A) + P(B) + P(C)$

◆ 性质：

- ◆ 1. $P(\emptyset)=0$ 不可能事件发生的概率为0
- ◆ 2. 有限可加性：对于两两互不相容的事件 $A_1, A_2, A_3, \dots, A_n$ ，即 $A_i \cap A_j = \emptyset, i \neq j, i, j=1, 2, \dots, n$ ，有 $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$
- ◆ 3. 对于A，B两个事件，若 $A \supset B$ ，则 $P(A-B) = P(A) - P(B); P(A) \geq P(B)$
- ◆ 4. 对于任一事件A，有 $P(A) \leq 1$
- ◆ 5. 对于任一事件A，有 $P(\bar{A}) = 1 - P(A)$ $P(A \cup \bar{A}) = P(A) + P(\bar{A}) = P(S) = 1$
- ◆ 6. 对于A，B两个事件，有 $P(A \cup B) = P(A) + P(B) - P(AB)$ （称为**加法公式**）

推广：

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 A_2) - P(A_1 A_3) - P(A_2 A_3) + P(A_1 A_2 A_3)$$



◆ 例：从一所高中中随机抽取一名学生，已知抽到一名女生的概率是0.5，抽到一名高二学生的概率是0.3，抽到一名高二女生的概率是0.2。那么抽到一名高二学生或一名女生的概率是？

◆ 使用公式 $P(A \cup B) = P(A) + P(B) - P(AB)$ ，记事件 $A = \{\text{抽到一名女生}\}$ ，事件 $B = \{\text{抽到一名高二学生}\}$ ，则 $P(A) = 0.5$ ， $P(B) = 0.3$ ， $P(AB) = 0.2$ ，所以

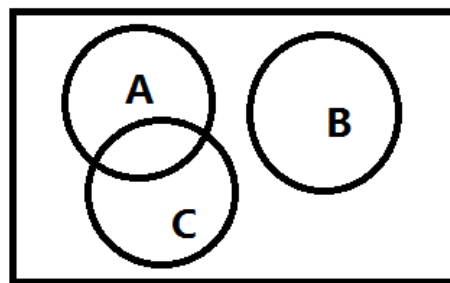
$$P(A \cup B) = P(A) + P(B) - P(AB) = 0.5 + 0.3 - 0.2 = 0.6$$

◆ 例：设 A ， B ， C 是三个事件，且 $P(A) = P(B) = P(C) = 1/4$ ， $P(AB) = P(BC) = 0$ ， $P(AC) = 1/8$ ，求 A ， B ， C 至少有一个发生的概率。

◆ $0 \leq P(ABC) \leq P(AB) = 0$ ，所以 $P(ABC) = 0$

◆ 故 $P(A \cup B \cup C)$

$$= P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC) = 3/8$$



- ◆ 对于试验E，若满足：
- ◆ 1. 试验的样本空间只包含有限个元素
- ◆ 2. 试验中每个基本事件发生的可能性相同，即每个基本事件发生的概率相等
- ◆ 则称这样的试验E为**古典概型**，也叫等可能概型
- ◆ 例子：抛硬币，抛骰子等



- ◆ 设古典概型的样本空间为 $S=\{e_1, e_2, e_3, \dots, e_n\}$ ，根据古典概型的定义，有

$P(e_1)=P(e_2)=\dots=P(e_n)$ ，而基本事件是互斥的，所以有

$1=P(S)=P(e_1)+P(e_2)+\dots+P(e_n)$ 。从而有 $P(e_1)=P(e_2)=\dots=P(e_n)=1/n$

- ◆ 对于事件A，若A包含k个基本事件，即 $A=\{e_1\}+\{e_2\}+\dots+\{e_n\}$ 则有

$$P(A) = \sum_{j=1}^k P(\{e_{i_j}\}) = \frac{k}{n} = \frac{A \text{ 包含的基本事件数}}{S \text{ 中基本事件的总数}}$$

- ◆ 例：抛骰子中，事件A：“出现点数为奇数”，则A包含3个基本事件，分别是“出现的点数是1点”，“出现的点数是3点”，“出现的点数是5点”。所以 $P(A)=3/6=1/2$

- ◆ 例1：一袋中有8个球，编号为1 - 8，其中1 - 3号为红球，4 - 8号为黄球，设摸到每一球的可能性相等，从中随机摸一球，记 $A = \{ \text{摸到红球} \}$ ，求 $P(A)$ 。

- ◆ 解： $S = \{1, 2, \dots, 8\}$ $A = \{1, 2, 3\}$

$$\Rightarrow P(A) = \frac{3}{8}$$

- ◆ 例2：将一个硬币抛三次，记 $A = \{ \text{恰有一次出现正面} \}$ ， $B = \{ \text{至少一次出现正面} \}$ ，则
 $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$; $A = \{HTT, THT, TTH\}$;
 $B = \{HHH, HHT, HTH, THH, HTT, THT, TTH\}$ 。故 $P(A) = 3/8$; $P(B) = 7/8$

- ◆ 排列：从n个不同元素中,任取m个元素,按照一定的顺序排成一列,叫做从n个不同元素中取出m个元素的一个排列.与顺序有关
- ◆ 组合：从n个不同元素中,任取m个元素,并成一组,叫做从n个不同元素中取出m个元素的一个组合.与顺序无关
- ◆ 例：
- ◆ 排列问题：从1，.....9这9个数字中任意取出3个不同数字组成一个三位数，问可以组成多少个不同的三位数？与顺序有关，123与321是不同的 $9*8*7$
- ◆ 组合问题：从1，.....9这9个数字中任意取出3个不同数字，问可以有多少种不同的取法？与顺序无关，123与321是一样的 $9*8*7/$

- ◆ 从n个不同的元素中取出m个元素，并按照一定的顺序排成一行，则共有

$$A_n^m = n(n-1)(n-2) \cdots (n-m+1) = \frac{n!}{(n-m)!} \text{ 种排列方法。}$$

- ◆ 从n个不同的元素中取出m个元素，则共有

- ◆ $C_n^m = \frac{n(n-1)(n-2) \cdots (n-m+1)}{m!} = \frac{n!}{(n-m)!m!}$ 种取法。

- ◆ 对于前面例子的两个问题：

- ◆ 排列问题： $A_9^3 = 9 \times 8 \times 7 = 504$ 个不同的三位数

- ◆ 组合问题： $C_9^3 = \frac{9 \times 8 \times 7}{3 \times 2 \times 1} = 84$ 种取法

例 2 一个口袋装有 6 只球, 其中 4 只白球、2 只红球. 从袋中取球两次, 每次随机地取一只. 考虑两种取球方式: (a) 第一次取一只球, 观察其颜色后放回袋中, 搅匀后再取一球. 这种取球方式叫做放回抽样. (b) 第一次取一球不放回袋中, 第二次从剩余的球中再取一球. 这种取球方式叫做不放回抽样. 试分别就上面两种情况求 (1) 取到的两只球都是白球的概率; (2) 取到的两只球颜色相同的概率; (3) 取到的两只球中至少有一只是白球的概率.

以 A 、 B 、 C 分别表示事件“取到的两只球都是白球”, “取到的两只球都是红球”, “取到的两只球中至少有一只是白球”. 易知“取到两只颜色相同的球”这一事件即为 $A \cup B$, 而 $C = \bar{B}$.

解 (a) 放回抽样的情况.

第一次从袋中取球有 6 只球可供抽取, 第二次也有 6 只球可供抽取. 由组合法的乘法原理, 共有 6×6 种取法. 即样本空间中元素总数为 6×6 . 对于事件 A 而言, 由于第一次有 4 只白球可供抽取, 第二次也有 4 只白球可供抽取, 由乘法原理共有 4×4 种取法, 即 A 中包含 4×4 个元素. 同理, B 中包含 2×2 个元素. 于是

$$P(A) = \frac{4 \times 4}{6 \times 6} = \frac{4}{9}.$$

$$P(B) = \frac{2 \times 2}{6 \times 6} = \frac{1}{9}.$$

由于 $AB = \emptyset$, 得

$$P(A \cup B) = P(A) + P(B) = \frac{5}{9}.$$

$$P(C) = P(\bar{B}) = 1 - P(B) = \frac{8}{9}.$$

- ◆ 不放回的情形：
- ◆ 第一次从袋中取球有6个球可以取，第二次取球只有5个球可以取，所以样本空间中一共有 $6 \times 5 = 30$ 个元素。
- ◆ 第一次从袋中取白球有4个球可以取，第二次取白球只有3个球可以取，所以共有 4×3 种抽法，故 $P(A) = 12/30$
- ◆ 第一次从袋中取红球有2个球可以取，第二次取红球只有1个球可以取，所以共有 2×1 种，故 $P(B) = 2/30$
- ◆ $P(A \cup B) = 14/30 = 7/15$
- ◆ $P(C) = 1 - P(B) = 14/15$

例 4 设有 N 件产品, 其中有 D 件次品, 今从中任取 n 件, 问其中恰有 k ($k \leq D$) 件次品的概率是多少?

解 在 N 件产品中抽取 n 件(这里是指不放回抽样), 所有可能的取法共有

$\binom{N}{n}$ ^① 种, 每一种取法为一基本事件, 且由于对称性知每个基本事件发生的可能

性相同. 又因在 D 件次品中取 k 件, 所有可能的取法有 $\binom{D}{k}$ 种. 在 $N-D$ 件正品

中取 $n-k$ 件所有可能的取法有 $\binom{N-D}{n-k}$ 种, 由乘法原理知在 N 件产品中取 n

件, 其中恰有 k 件次品的取法共有 $\binom{D}{k} \binom{N-D}{n-k}$ 种, 于是所求概率为

$$p = \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}}. \quad (4.2)$$

$\binom{N}{n}$ 表示 $C_N^n = \frac{n(n-1)(n-2)\dots(n-m+1)}{m!}$

例 7 将 15 名新生随机地平均分配到三个班级中去, 这 15 名新生中有 3 名是优秀生. 问 (1) 每个班级各分配到一名优秀生的概率是多少? (2) 3 名优秀生分配在同一班级的概率是多少?

解 15 名新生平均分配到三个班级中的分法总数为

$$\binom{15}{5} \binom{10}{5} \binom{5}{5} = \frac{15!}{5! 5! 5!}.$$

每一种分配法为一基本事件, 且由对称性易知每个基本事件发生的可能性相同.

(1) 将 3 名优秀生分配到三个班级使每个班级都有一名优秀生的分法共 $3!$ 种. 对于这每一种分法, 其余 12 名新生平均分配到三个班级中的分法共有 $\frac{12!}{4! 4! 4!}$ 种. 因此, 每一班级各分配到一名优秀生的分法共有 $\frac{3! 12!}{4! 4! 4!}$ 种. 于是所求概率为

$$p_1 = \frac{3! \times 12!}{4! 4! 4!} \bigg/ \frac{15!}{5! 5! 5!} = \frac{25}{91}.$$

(2) 将 3 名优秀生分配在同一班级的分法共有 3 种. 对于这每一种分法, 其余 12 名新生的分法(一个班级 2 名, 另两个班级各 5 名)有 $\frac{12!}{2! 5! 5!}$ 种. 因此 3 名优秀生分配在同一班级的分法共有 $\frac{3 \times 12!}{2! 5! 5!}$ 种, 于是, 所求概率为

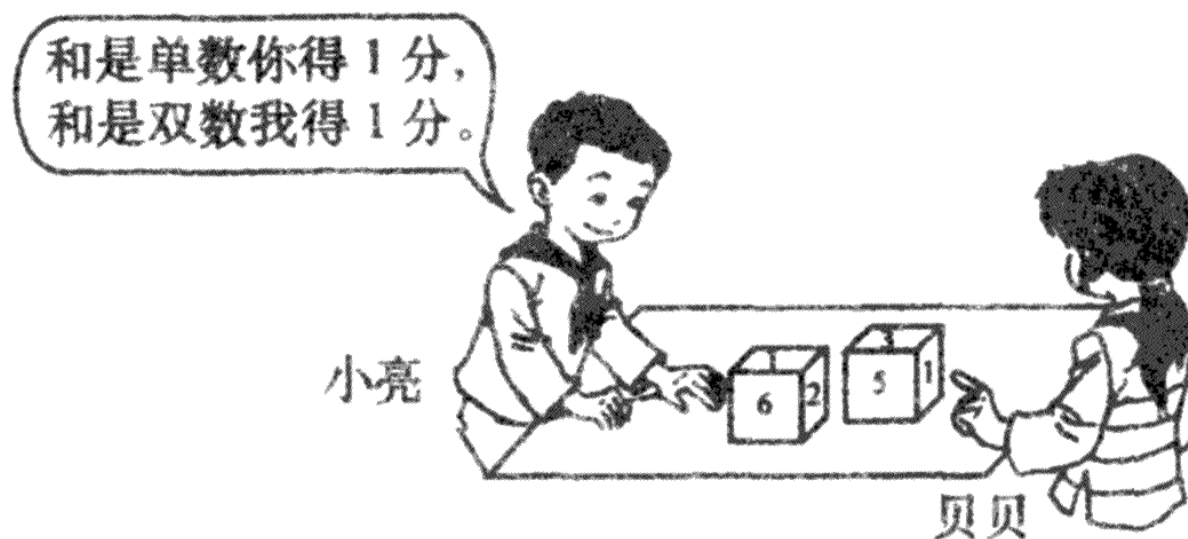
$$p_2 = \frac{3 \times 12!}{2! 5! 5!} / \frac{15!}{5! 5! 5!} = \frac{6}{91}.$$

□

- ◆ 例：某接待站在某一周曾接待12次来访，已知所有这12次接待都是在周二和周四进行的，问是否可以推断接待时间是有规定的？
- ◆ 解：假设接待站的接待时间没有规定，而各来访者在一周的任一天中去接待站是等可能的，那么，12次接待来访者都是在周二、周四的概率为 $2^{12}/7^{12} = 0.000\ 000\ 3$.
- ◆ 人们在长期的实践中总结得到“**概率很小的事件在一次试验中实际上几乎是不发生的**”（称之为**实际推断原理**）。例如，买一次彩票就中奖了几乎是不可能发生的事情。
- ◆ 现在概率很小的事件在一次试验中竟然发生了，因此有理由怀疑假设的正确性，从而推断接待站不是每天都接待来访者，即认为其接待时间是有规定的。

- ◆ 赌大小的公平性：对于一般人来说，赌大小只会押大或小，赔率是1:1，4到10为小,11到17为大，当出现三个骰子点数一样时，庄家大小通吃。赌大小的规则公平吗？
- ◆ 三颗骰子出现的点数共有 $6*6*6=216$ （假设骰子的点数有顺序）种情况，每种情况出现的概率都一样
- ◆ $P(\text{小}) = 0.486111$
- ◆ $P(\text{大}) = 0.486111$
- ◆ $P(\text{三个点数一样}) = 6/216 = 0.02778$
- ◆ 所以 $P(\text{庄家赢}) = 0.486111 + 0.02778 = 0.513891 > 0.486111 = P(\text{玩家赢})$

- ◆ 例3：小亮和贝贝两人各抛一个骰子，规定：若两个骰子的点数和为奇数，则贝贝赢；若点数和为偶数，则小亮赢。这个规定公平吗？

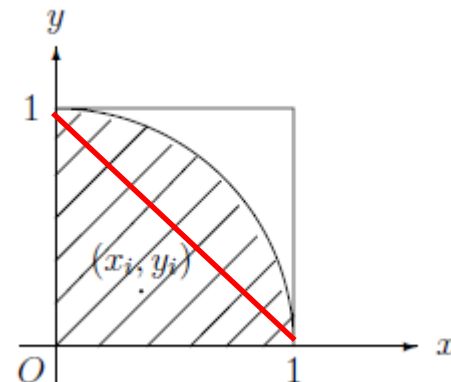


小亮的点数	贝贝的点数	点数和	奇数为1，偶数为0
1	1	2	0
1	2	3	1
1	3	4	0
1	4	5	1
1	5	6	0
1	6	7	1
2	1	3	1
2	2	4	0
2	3	5	1
2	4	6	0
2	5	7	1
2	6	8	0
3	1	4	0
3	2	5	1
3	3	6	0
3	4	7	1
3	5	8	0
3	6	9	1

4	1	5	1
4	2	6	0
4	3	7	1
4	4	8	0
4	5	9	1
4	6	10	0
5	1	6	0
5	2	7	1
5	3	8	0
5	4	9	1
5	5	10	0
5	6	11	1
6	1	7	1
6	2	8	0
6	3	9	1
6	4	10	0
6	5	11	1
6	6	12	0
点数和为奇数的组合个数			18

关于 π 的估算

- ◆ 如图，在一个边长为1的正方形中有一个1/4圆，假设点 (x_i, y_i) 是该正方形内随机的一个点，问点 (x_i, y_i) 落在1/4圆内的概率是？
- ◆ (x_i, y_i) 有无限种可能取值，每种取值都是等可能的
- ◆ 点 (x_i, y_i) 落在1/4圆内的概率与1/4圆的面积有关——面积越大，落入圆中的可能性越大
- ◆ 1/4圆的面积是： $\frac{\pi r^2}{4} = \frac{\pi}{4}$
- ◆ 正方形的面积是：1
- ◆ $P(\text{点}(x_i, y_i)\text{落在1/4圆内}) = \frac{\frac{\pi}{4}}{1} = \frac{\pi}{4}$



- ◆ 对于试验E，若满足：
- ◆ 1. 试验的样本空间包含无限个元素
- ◆ 2. 试验中每个基本事件发生的可能性相同，即每个基本事件发生的概率相等
- ◆ 这样的试验E称为几何概型
- ◆
$$P(A) = \frac{\text{构成事件A的区域长度(面积或体积)}}{\text{实验的全部结果所构成的区域长度(面积或体积)}}$$

- ◆ 取一根长为3米的绳子，拉直后在任意位置剪断，那么剪得两段的长都不少于1米的概率有多大？

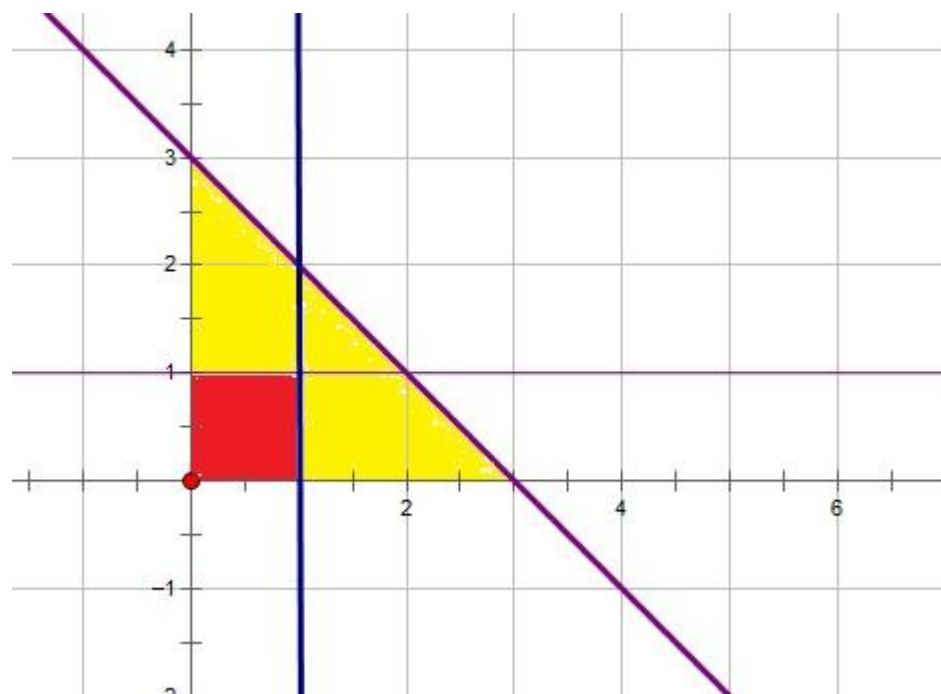


- ◆ 记 x （米）为拉断后第一段绳子的长度， y （米）为拉断后第二段绳子的长度，则 $x+y=3$
- ◆ 若要使两段的长都不少于1米，则 $x \geq 1; y \geq 1$
- ◆ $P(\text{剪得两段的长都不少于1米}) = 1/3$

例子

- ◆ 取一根长为3米的绳子，拉直后在任意位置剪断剪成3段，那么剪得其中两段的长都不少于1米的概率有多大？
- ◆ 记 x （米）为拉断后第一段绳子的长度， y （米）为拉断后第二段绳子的长度，则 $x+y<3, x>0, y>0$
- ◆ 若要使两段的长都不少于1米，则 $x \geq 1; y \geq 1$
- ◆ 故 P （剪得两段的长都不少于1米）

$$= \frac{\frac{1}{2} \times 1 \times 1}{\frac{1}{2} (3 \times 3)} = \frac{1}{9}$$



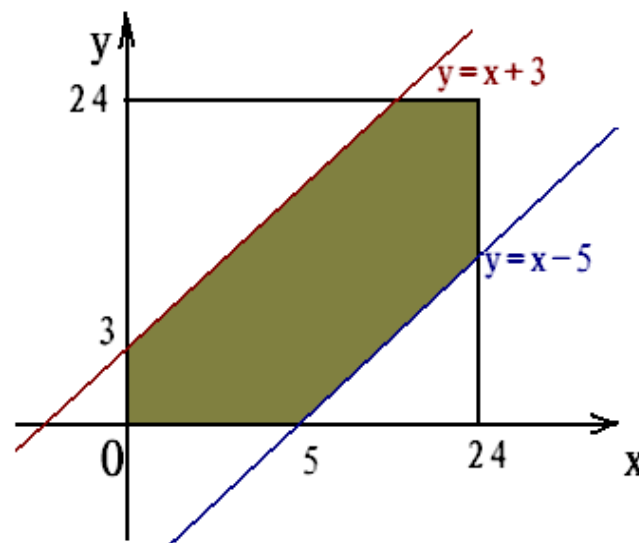
- ◆ 甲、乙两艘轮船都要停靠同一个泊位，他们可能在某一天的任意时刻到达，设甲乙两艘轮船停靠泊位的时间分别为3小时和5小时。求有一艘轮船停靠泊位时必须等待一段时间的概率。

- ◆ 设甲、乙到达时刻分别为x点、y点

$$\Omega = \{(x, y) | 0 \leq x \leq 24, 0 \leq y \leq 24\}$$

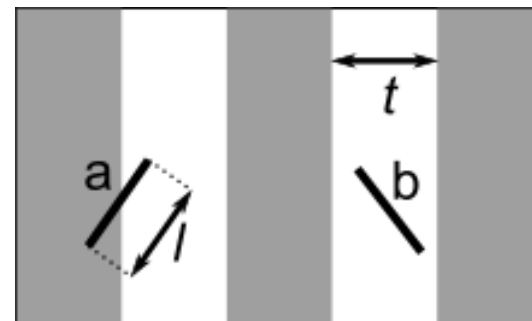
$$A = \{(x, y) | (x, y) \in \Omega, \text{且} y - x < 3, x - y < 5\}$$

- ◆ 故 $P(A) = \frac{24 \times 24 - \frac{1}{2} \times 21 \times 21 - \frac{1}{2} \times 19 \times 19}{24 \times 24} = \frac{175}{576}$



Buffon投针实验

- ◆ 18世纪，布丰提出以下问题：设我们有一个以平行且等
- ◆ 距（间距为 a ）木纹铺成的地板（如右图），现在随意抛
- ◆ 一支长度 b 比木纹之间距离小的针，求针和其中一条木纹
- ◆ 相交的概率。



试验者	时间	针长	投掷次数	相交次数	π 的近似值
Wolf	1850	0.8	5000	2532	3.1596
Smith	1855	0.6	3204	1218	3.1554
De Morgan	1860	1.0	600	382	3.137
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	0.83	3408	1808	3.1415929
Reina	1925	0.5419	2520	859	3.1795

Buffon投针实验

- ◆ 以 x 表示针投到平面上时针的中点 M 到最近的一条平行直线的距离， φ 表示该针与该平行线的夹角，那么
- ◆ 针在平面上的位置可以由 (x, φ) 完全决定。
- ◆ 样本空间为

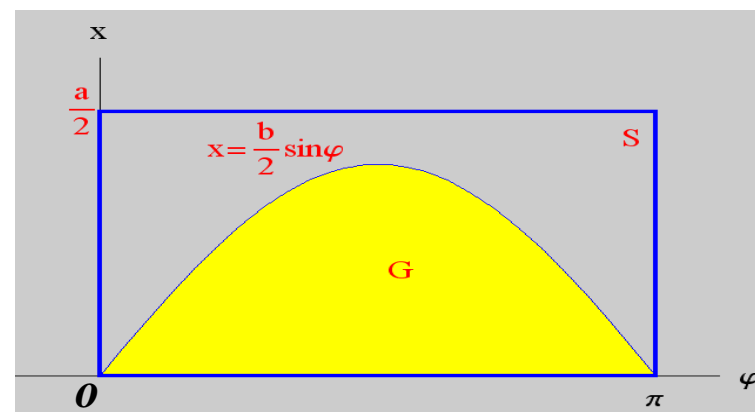
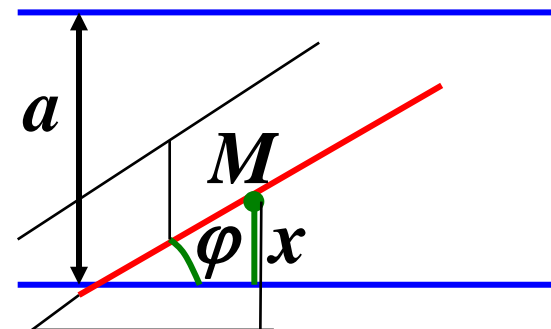
$$S = \{(x, \varphi) \mid 0 \leq x \leq \frac{a}{2}, 0 \leq \varphi \leq \pi\}$$

- ◆ 事件 $A = \{\text{针与某一条平行线相交}\}$ 发生的充要条件

- ◆ 为 $0 \leq x \leq \frac{b}{2} \sin \varphi, 0 \leq \varphi \leq \pi.$

$$P(A) = \frac{\mu(G)}{\mu(S)} = \frac{G \text{ 的面积}}{S \text{ 的面积}} = \frac{\int_0^\pi \frac{b}{2} \sin \varphi d\varphi}{\frac{a}{2} \times \pi}$$

$$= \frac{b}{\frac{a}{2} \times \pi} = \frac{2b}{a\pi}.$$



- ◆ **Dataguru（炼数成金）**是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。
- ◆ 关于逆向收费式网络的详情，请看我们的培训网站 <http://edu.dataguru.cn>



Thanks

FAQ时间