概率论与数理统计

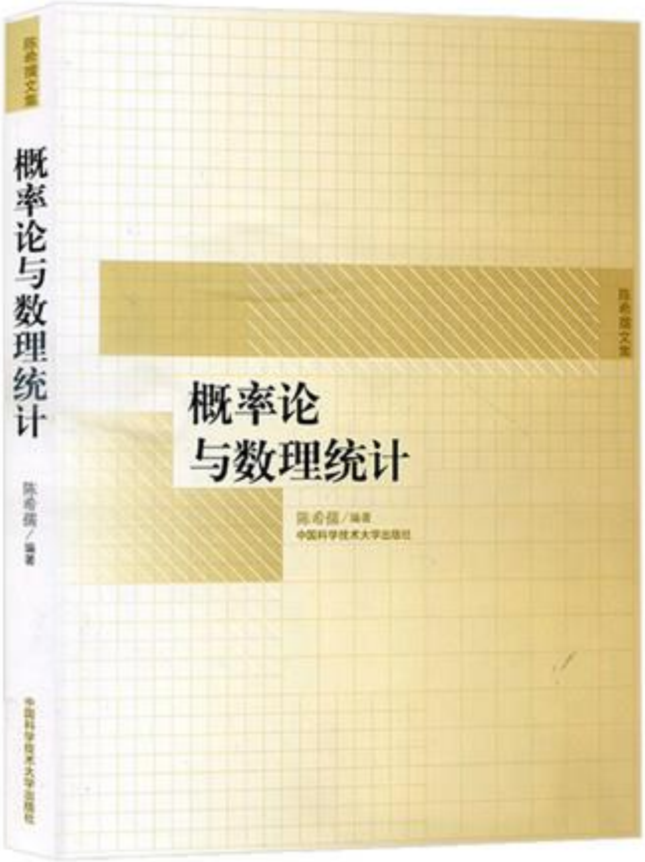
杨青

yangq@ustc.edu.cn

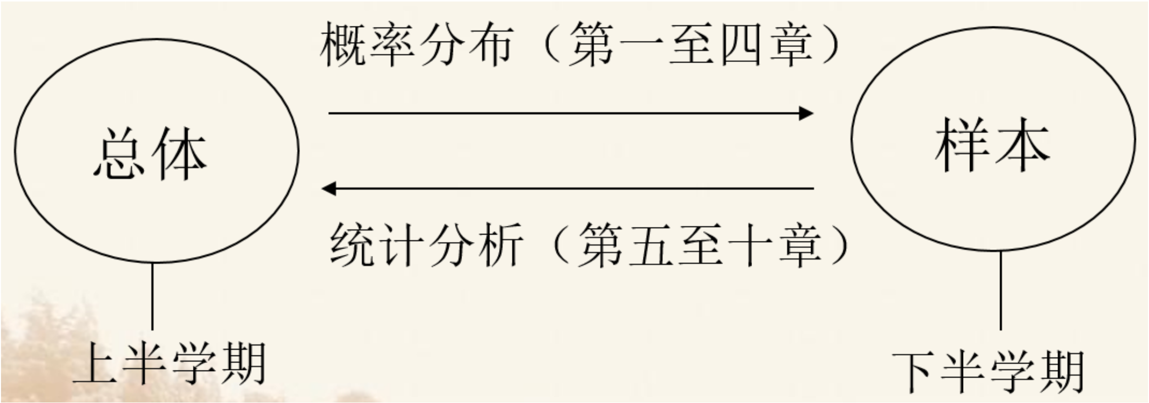
课程简介

教材

阅读材料



目　录


概率分布（第一至四章）
总体
样本
统计分析（第五至十章）
上半学期
下半学期

成绩构成

- 平时成绩
  - 每周一次作业
  - 附加分
- 期末成绩
  - 全校联考（闭卷考试，可带计算器）
- 助教：

李祥超(lxc1468401361@mail.ustc.edu.cn)

朱希悦(zhuxiyue@mail.ustc.edu.cn)

不确定性

- 概率论和数理统计是以不确定性为研究对象的科学，为人们认识随机现象提供了重要的思维模式和解决问题的方法

人生处处都是不确定性

数据是人们认识和理解不确定性的媒介

概率论与数理统计的区别

- 概率论：刻画了不确定性的本质（随机现象的数量度量）

- 数理统计：试图通过现实观测数据去理解和运用这种本质（从样本推断总体）

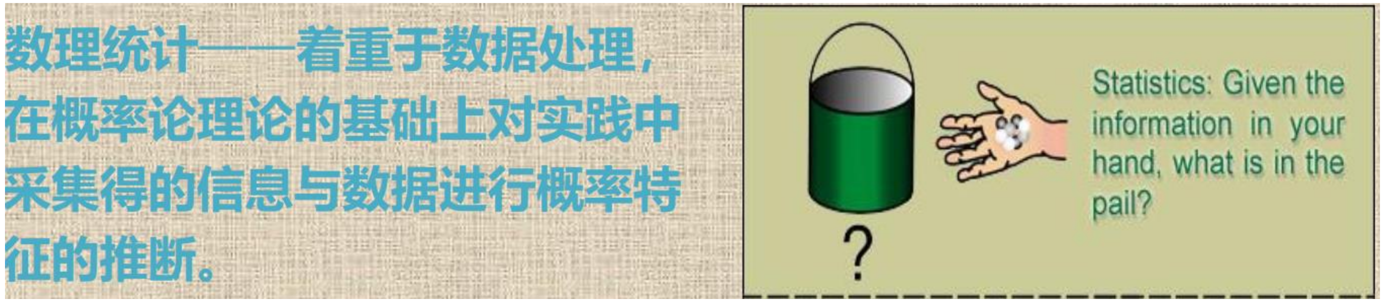


Diagram showing the difference between statistics and probability. (Image by MIT OpenCourseWare. Based on Gilbert, Norma. *Statistics.* W.B. Saunders Co., 1976.)

第一章： 事件及其概率

大纲：

- 随机试验和随机事件
  - 事件的关系及其运算
- 概率的定义及性质
- 条件概率
  - 全概率公式和 Bayes 公式
- 事件的独立性

1. 随机试验和随机事件

随机现象: 自然界中的一种客观现象, 当人们观测它时, 不能预先确定会出现哪种结果, 而仅仅知道是多种可能结果之一.

随机试验: 随机现象的实现和对它某个特征的观测.

- 随机试验中要求试验的结果至少 2 个
- 每次试验或观测得到其中的一个结果, 在试验或观测之前不能预知是哪个结果发生
- 一般还要求试验在相同条件下能够重复

如观测把硬币抛 4 次后正面向上的次数; 观测某地的温度变化; 某电话总机单位时间内转接的电话次数.

基本事件：随机试验中的每个单一结果，它犹如分子中的原子，在化学反应中不能再分，所以有基本两字.

如把硬币抛 3 次后有 8 种可能结果：正正正、正正反、正反正、反正正、正反反、反正反、反反正、反反反. 这 8 种可能结果的每一个都是基本事件.

样本空间(Sample Space)：随机试验中所有基本事件所构成的集合，通常用 Ω 或 S 表示. 样本空间中的元素，称为样本点，通常用 ω 等表示.

(1). 掷一枚骰子, 观察出现的点数. 则 $\Omega = \{1, 2, 3, 4, 5, 6\}$.

(2). 考察某一地区的年降雨量, 则 $\Omega = \{x | 0 \le x < T\}$, 这里 T 表示某个常数, 表示降雨量不会超过 T.

样本空间说明:

- 根据样本空间 Ω 的大小, 可以将样本空间分为三类: 有限样本空间 (仅含有有限个样本点)、可数无穷样本空间 (含有无穷且可数个样本点) 和不可数样本空间 (含有无穷且不可数个样本点) .

- 样本空间的元素应该是相互不同的, 根据试验的不同目的, 样本空间应该予以不同的选择.

- 但是总的原则是样本空间应该尽可能详细, 即尽可能包含所有可能的结果.

看下面的例子

(1). 将一枚硬币抛三次，考察正反面出现的情况;

(2). 将一枚硬币抛三次，考察正面出现的次数。

这两个试验的目的不同，因此样本空间的选取也不同。

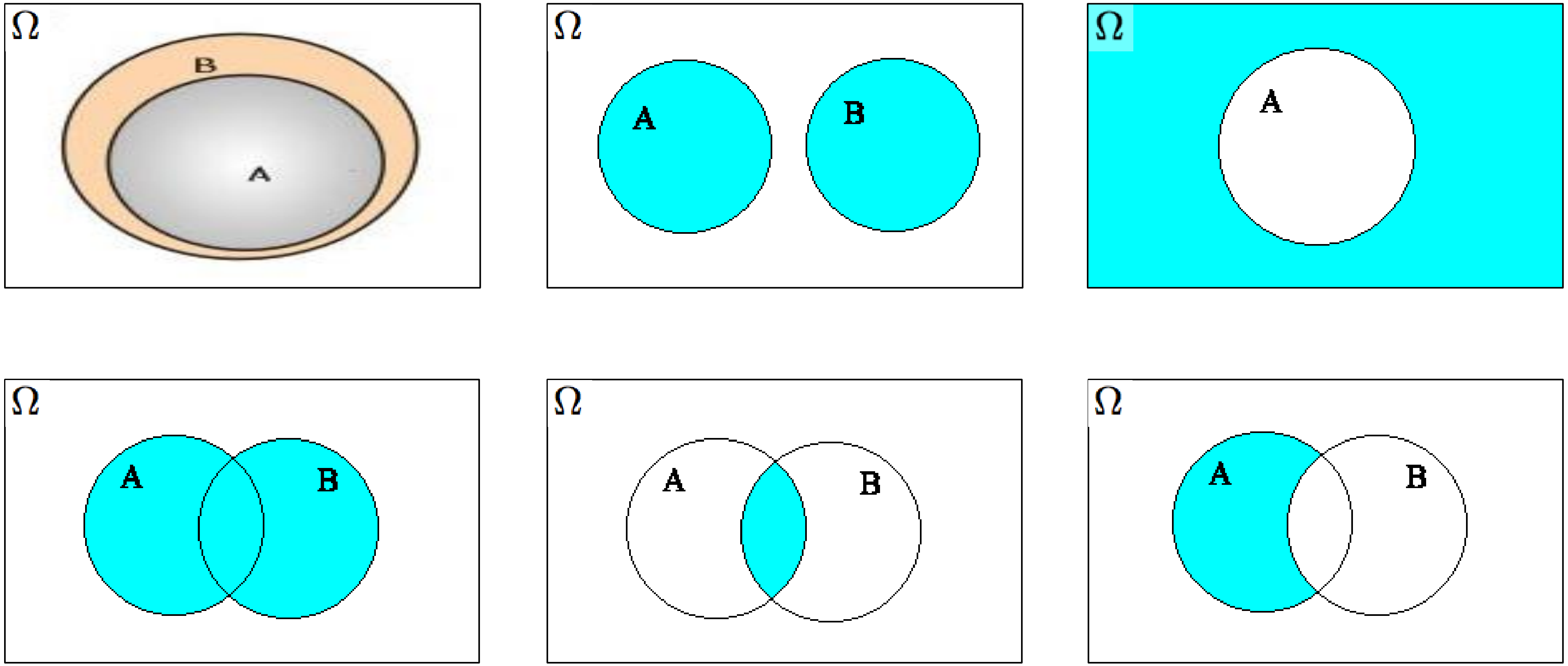
随机事件：简称事件 (Event)，在随机试验中我们所关心的可能出现的各种结果，它由一个或若干个基本事件组成.

随机事件常用大写英文字母 A, B, C, D 等表示. 如果用语言表达, 则要用花括号括起来.

必然事件 (Ω)： 在试验中一定会发生的事件；
不可能事件 (ϕ)：在试验中不可能发生的事件.

- 样本空间本身称为必然事件, 因其在随机试验中必然会发生 .习惯上, 人们将必然事件发生的概率设置为 1.
- 空集称为不可能事件, 由于其不包含任何样本点, 故在随机试验中不可能发生 .习惯上, 人们将不可能事件发生的概率设置为 0.
- 但是发生概率为 1 的事件未必是必然事件, 发生概率为 0 的事件未必是不可能事件.

事件的关系及其运算



运算法则

- 集合的运算法则适用于事件的运算

$$A \cup A = A,$$

$$A \cap A = A,$$

$$A \cup BC = (A \cup B)(A \cup C),$$

$$A \cap (B \cup C) = AB \cup AC,$$

$$(A \cup B)(C \cup D) = AC \cup BC \cup AD \cup BD$$

- De Morgan德摩根对偶法则:

$$\left(\bigcup_{k=1}^{n} A_k\right)^c = \bigcap_{k=1}^{n} A_k^c,$$

$$\left(\bigcap_{k=1}^{n} A_k\right)^c = \bigcup_{k=1}^{n} A_k^c.$$

例子

设 A, B, C 是三个事件，试表示下列事件

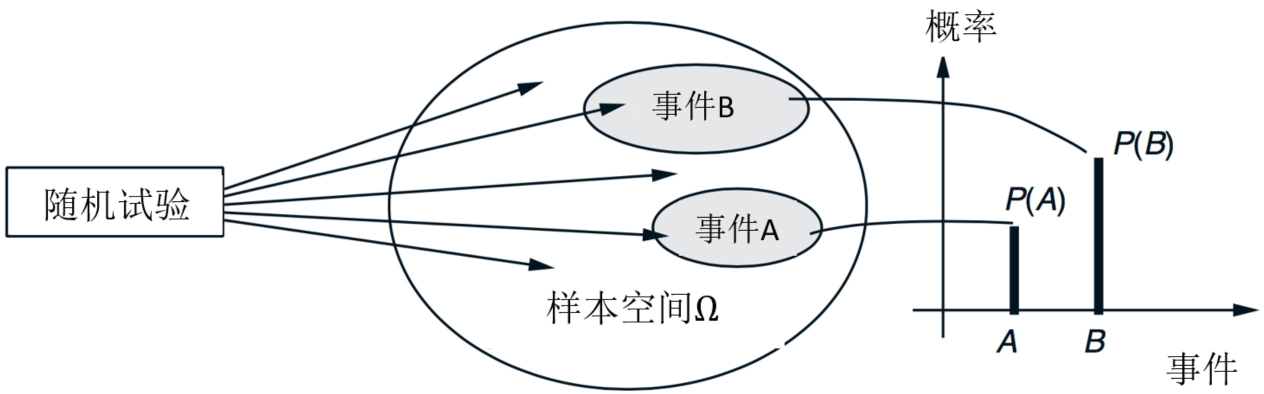
1. 事件 A, B 发生而 C 不发生;
2. 事件 A, B, C 不同时发生;
3. 事件 A, B, C 中至多有一个发生;
4. 事件 A, B, C 中至少发生两个;
5. 事件 A, B, C 中恰好发生两个;

- 概率的定义（初等描述）：

直观地讲, 概率是随机事件发生可能性大小的数字表征, 取值于区间 [0,1] .

换句话说, 概率是事件的函数.

性质：1) $P(\Omega)=1$, $P(\Phi)=0$. 2) $0\leq P(A)\leq 1$.



概率空间

如何求出事件 A 的概率（记为 $P(A)$）?

(1) 古典概型:

- 称一个随机试验为古典概型, 如果

第一 (有限性) 试验结果只有有限个 (记为 n) ,
第二 (等可能性) 每个基本事件发生的可能性相同.

为计算事件 A 的概率, 设 A 中包含 m 个基本事件, 则事件 A 的概率为

$$P(A)=\frac{m}{n}=\frac{|A|}{|\Omega|}$$

记号: 为方便起见, 以 $|B|$ 记事件 B 中基本事件的个数.

- 计算古典概率, 主要用到排列组合的知识.

- 从 n 个不同的元素中，有放回地取出 r 个元素组成的可重复排列的种数为 n^r 种。从 n 个不同的元素中，不放回地取出 r 个元素组成的不重复排列的种数为 $n(n-1)\cdots(n-r+1) = P_n^r$.

- 从 n 个不同的元素中，不放回地取 r 个组成的组合，种数为

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!} = C_n^r$$

例子

在运用排列组合公式时，要清楚次序问题.

甲乙丙丁四人进行乒乓球双打练习，两人一对地结为对打的双方，有多少种不同的结对方式?

“组合”是一种“有编号的分组模式”，或者说，按照组合模式计算出的分组方式数目中，已经天然地把组的不同编号方式数目计算在内了.

欲将 6 个人分为 3 组，每组 2 人，分别从事 3 项不同工作，求分配方式数.

↑Example

↓Example

要把 7 人分为 3 个小组, 执行同一种任务, 其中一个组 3 人, 另两个组各 2 人, 求分组方式数.

↑Example
↓Example

- 为了适应这种分为多个“不同的”组的问题需求, 人们总结出“多组组合模式”.

多组组合模式：有 n 个不同元素, 要把它们分为 k 个不同的组, 使得各组依次有 $n_1, n_2, \cdots, n_k$ 个元素, 其中 $n_1+n_2+\cdots+n_k=n$, 则一共有

$$\frac{n!}{n_1!\cdot n_2!\cdot \ldots \cdot n_k!}$$

种不同分法.

不尽相异元素的排列模式 有 n 个元素, 属于 k 个不同的类, 同类元素之间不可辨认, 各类元素分别有 n_1, n_2, $\cdots$, n_k 个, 其中 $n_1+n_2+\cdots+n_k=n$, 要把它们排成一列, 则一共有

$$\frac{n!}{n_1!\cdot n_2!\cdots n_k!}$$

种不同排法.

例子

一批产品有 N 个，其中废品有 M 个。现从中随机取出 n 个，在以下两种情形下，分别求“其中恰好有 m 个废品”这一事件的概率。
(1) 有放回地选取；　(2) 不放回地选取

n 个男生, m 个女生排成一排 $(m \le n+1)$. 求事件 $A = \{$任意两个女孩不相邻$\}$ 的概率。

盒子模型

r 个不同的球任意放入编号为 1 至 n 的 n 个盒子，每球入各盒均等可能，求下列事件的概率

(1) A={指定的 r 个盒子各含一个球}

(2) B={每盒至多有一球}

(3) C={某指定盒中恰有 m 个球}

- 若r个球相同？

- 生日问题

求 r 个人中没有两个人生日相同的概率．若把 r 个人看作上面问题中的 r 个球，而把一年的 365 看作为盒子，则 $n = 365$，这时事件 B 的概率即为所求概率。例如当 $r = 40$ 时，$P(B) = 0.109$，这个概率已经相当小；而当 $r = 50$ 时, $P(B) = 0.03$。进一步当 $r = 55$ 时，$P(B)$ 之值只有 0.01，这实在是出乎意料地小。

(2) 几何概型:

- 古典概型中有两个限制, 如果去掉有限性, 保留基本事件的等可能性, 就称为几何概型 （样本点无限多）.

设 Ω 是欧氏空间中确定的集合，满足条件 $0 < m(\Omega) < +\infty$。对 Ω 中的任何可测子集 A，称

$$P(A) = \frac{m(A)}{m(\Omega)}$$

为事件 A 的几何概率。这里等可能性体现在“落在区域 A 的概率与区域 A 的测度成正比并且与其形状位置无关。"

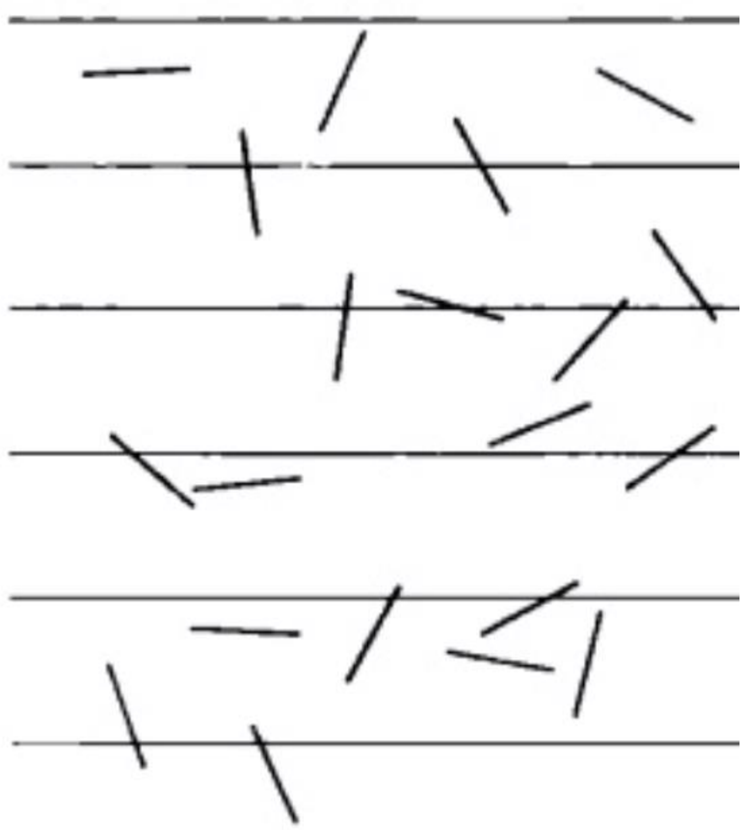
这里 $m(\Omega)$ 表示 Ω 的“大小”.

例子

甲乙两人约定在 $[0,T]$ 时段内去某地会面，规定先到者等候一段时间 $t(t \le T)$ 再离去。试求事件 A={甲乙将会面} 的概率。

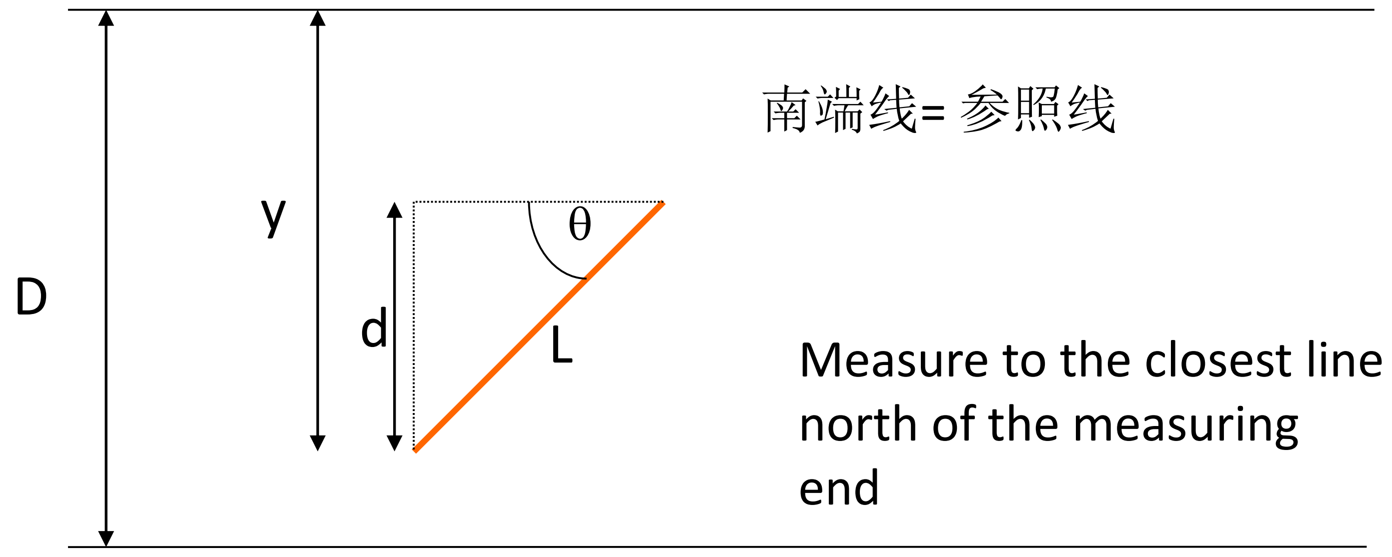
例子： Buffon投针

假设在平面上画上间距为D的平行线，现在随机投掷长度为L的的针(L≤D)，求针与其中某条平行线相交的概率。



Buffon投针的计算：几何方法

由于针的长度小于线的间距，所以必有一条平行线与针相邻但不相交。我们将这个线作为参照线（南端的线）。

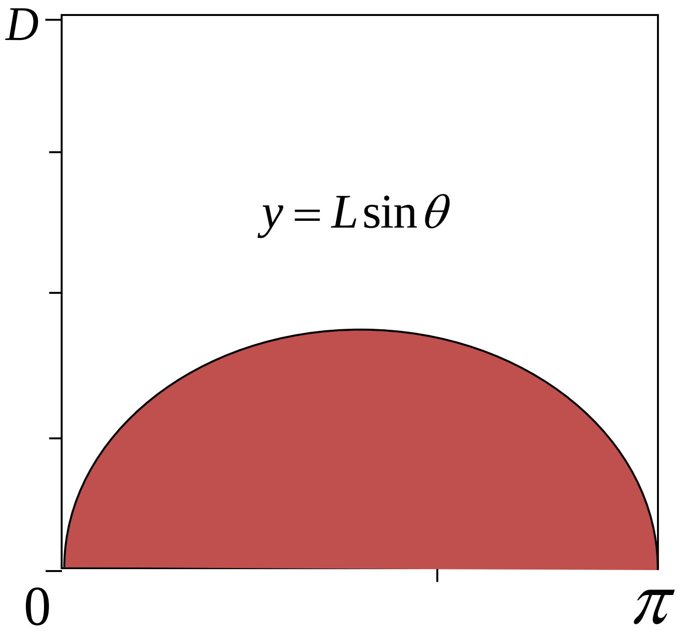


现在我们有：

$$d = L\sin\theta \qquad 0 \le \theta \le \pi \qquad 0 \le y \le D$$

关系式　　独立均匀分布　　独立均匀分布

如果 $y \le L\sin\theta$ ，则表示出现了相交。



左图是y与θ的联合分布图，其中红色区域表示$y < L\sin\theta$，即能够相交，而白色区域则表示没有相交。最后只需要计算出红色区域所占整体区域的比例即可。

$$红色区域 = L\int_0^{\pi}\sin\theta\, d\theta$$

$$= L(-\cos\pi + \cos 0)$$

$$= 2L$$

$$总区域 = D\pi$$

$$\Pr(相交) = \frac{红区域}{总区域} = \frac{2L}{D\pi}$$