

概率论与数理统计

杨青

yangq@ustc.edu.cn

课程简介

教材



阅读材料

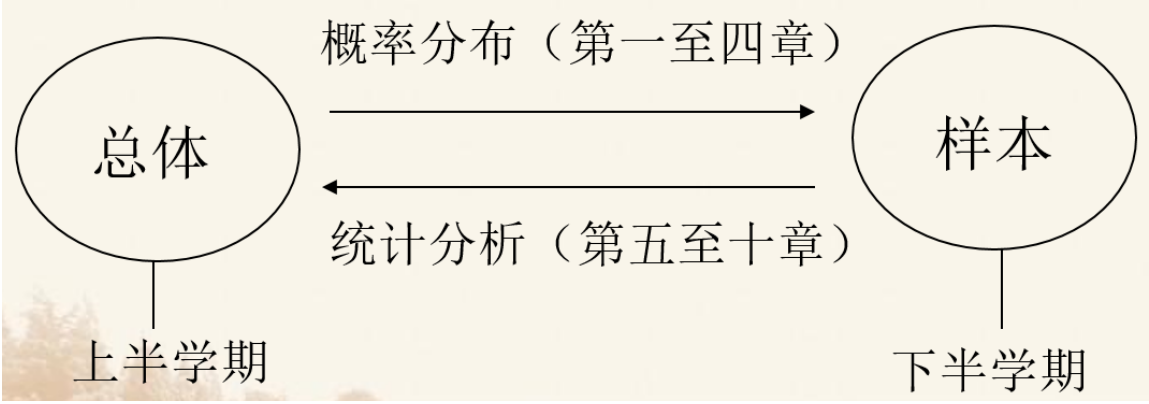


目 录

| | |
|--------------------------------|-----|
| 第一章 事件及其概率 | 1 |
| 1.1 概率论简史 | 1 |
| 1.2 随机试验和随机事件 | 3 |
| 1.3 概率的定义和性质 | 9 |
| 1.4 条件概率 | 19 |
| 1.5 独立性 | 29 |
| 1.6* 扩展进阶: 求概率的一些方法 | 33 |
| 1.7 扩展阅读 1: 贝叶斯公式和垃圾邮件识别 | 36 |
| 1.8 扩展阅读 2: 三门问题 | 38 |
| 本章总结 | 40 |
| 习题 | 41 |
| 第二章 随机变量及其分布 | 46 |
| 2.1 随机变量的概念 | 46 |
| 2.2 离散型随机变量的分布 | 47 |
| 2.3 连续型随机变量的分布 | 59 |
| 2.4 随机变量函数的分布 | 72 |
| 2.5 扩展阅读: 正态分布的由来 | 76 |
| 本章总结 | 80 |
| 习题 | 81 |
| 第三章 多维随机变量及其分布 | 88 |
| 3.1 多维随机变量及其分布 | 88 |
| 3.2 边缘 (际) 分布 | 95 |
| 3.3 条件分布 | 99 |
| 3.4 相互独立的随机变量 | 102 |
| 3.5 随机向量函数的分布 | 105 |
| 3.6 扩展阅读: 辛普森悖论 | 113 |
| 本章总结 | 115 |

| | |
|--------------------------|-----|
| 第四章 随机变量的数字特征和极限定理 | 123 |
| 4.1 数学期望和中位数 | 123 |
| 4.2 方差和矩 | 135 |
| 4.3 熵的基本概念 | 149 |
| 4.4 大数定律和中心极限定理 | 154 |
| 4.5 扩展阅读: 数学期望的计算 | 165 |
| 本章总结 | 170 |
| 习题 | 171 |
| 第五章 统计学基本概念 | 180 |
| 5.1 统计学发展简史 | 180 |
| 5.2 基本概念 | 183 |
| 5.3 抽样分布 | 191 |
| 5.4 扩展阅读 1: 民意调查 | 199 |
| 5.5 扩展阅读 2: 双盲对照试验 | 202 |
| 本章总结 | 203 |
| 习题 | 204 |
| 第六章 参数点估计 | 207 |
| 6.1 参数点估计的概念 | 207 |
| 6.2 矩估计法 | 208 |
| 6.3 最大似然估计 | 210 |
| 6.4 优良性准则 | 215 |
| 6.5 点估计量的大样本理论 | 225 |
| 6.6 扩展阅读: 德军坦克问题 | 227 |
| 本章总结 | 229 |
| 习题 | 230 |
| 第七章 区间估计 | 240 |
| 7.1 基本概念 | 240 |
| 7.2 枢轴变量法 | 243 |
| 7.3 大样本方法 | 248 |
| 7.4 自助法置信区间 | 253 |

| | | |
|------------|--------------------|------------|
| 7.5 | 置信限 | 257 |
| 7.6 | 扩展阅读: “足球赛会杀人”的真假 | 258 |
| | 本章总结 | 260 |
| | 习题 | 261 |
| 第八章 | 假设检验 | 266 |
| 8.1 | 问题的提法和基本概念 | 266 |
| 8.2 | 正态总体参数检验 | 273 |
| 8.3 | 比例 p 的检验 | 289 |
| 8.4 | 似然比检验 | 294 |
| 8.5 | p 值 | 299 |
| 8.6 | 扩展阅读: 多重假设检验 | 302 |
| | 本章总结 | 304 |
| | 习题 | 305 |
| 第九章 | 非参数假设检验 | 315 |
| 9.1 | 拟合优度检验 | 315 |
| 9.2 | 威尔科克森秩和检验 | 326 |
| 9.3 | 符号检验 | 329 |
| 9.4 | 其他非参数检验概述 | 331 |
| 9.5 | 扩展阅读: 正态性检验 | 333 |
| | 本章总结 | 337 |
| | 习题 | 338 |
| 第十章 | 相关分析和回归分析 | 341 |
| 10.1 | 相关分析 | 342 |
| 10.2 | 回归分析 | 348 |
| 10.3 | 多元回归中自变量的选择和模型诊断简述 | 366 |
| 10.4 | 扩展阅读: 相关与因果 | 369 |
| 10.5 | 附录 | 371 |
| | 本章总结 | 374 |
| | 习题 | 375 |



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)



- 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）

概率论——从数学模型进行理论推导，从同类现象中找出规律性。



- 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）

数理统计——着重于数据处理，在概率论理论的基础上对实践中采集得的信息与数据进行概率特征的推断。

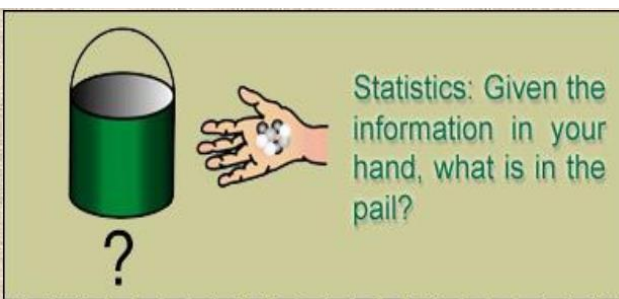


Diagram showing the difference between statistics and probability. (Image by MIT OpenCourseWare. Based on Gilbert, Norma. *Statistics*. W.B. Saunders Co., 1976.)

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 $\Omega = \{x | 0 \leq x < T\}$, 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。

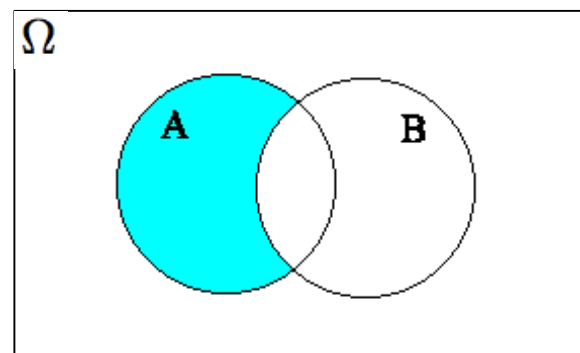
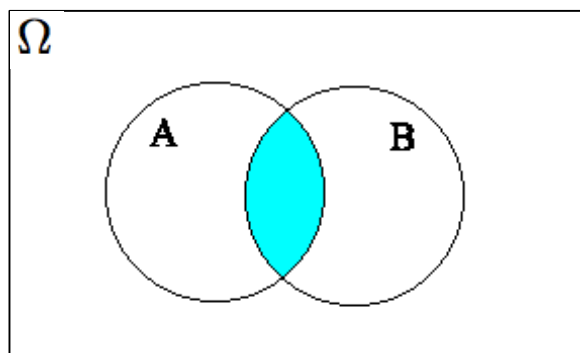
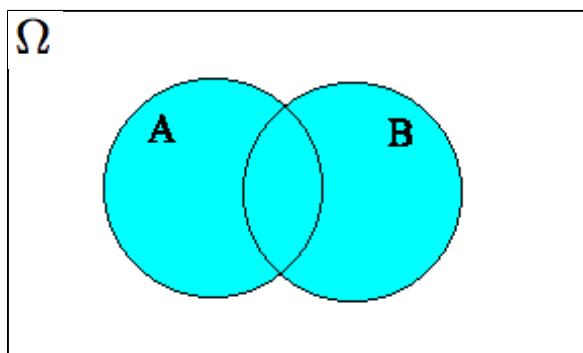
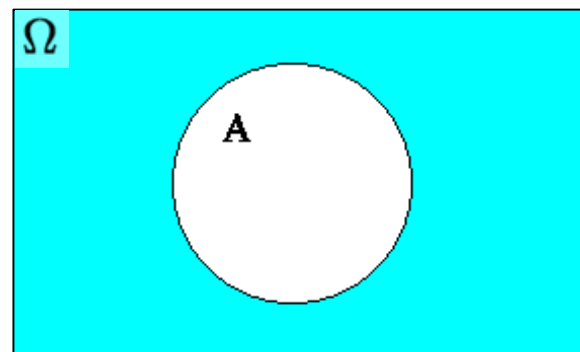
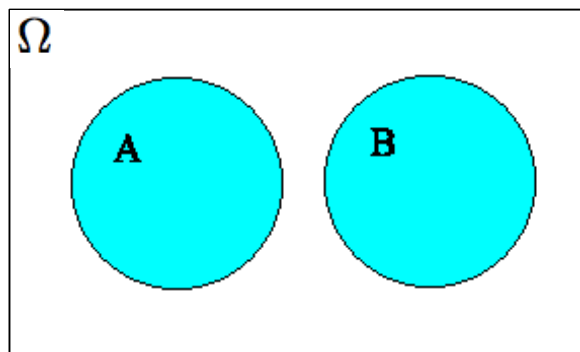
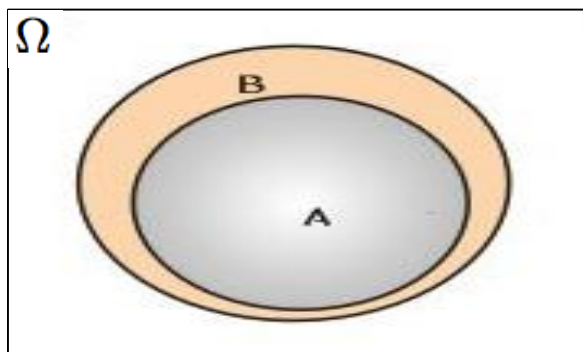
随机事件常用大写英文字母 A, B, C, D 等表示. 如果用语言表达, 则要用花括号括起来.

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



- 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:

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

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

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；

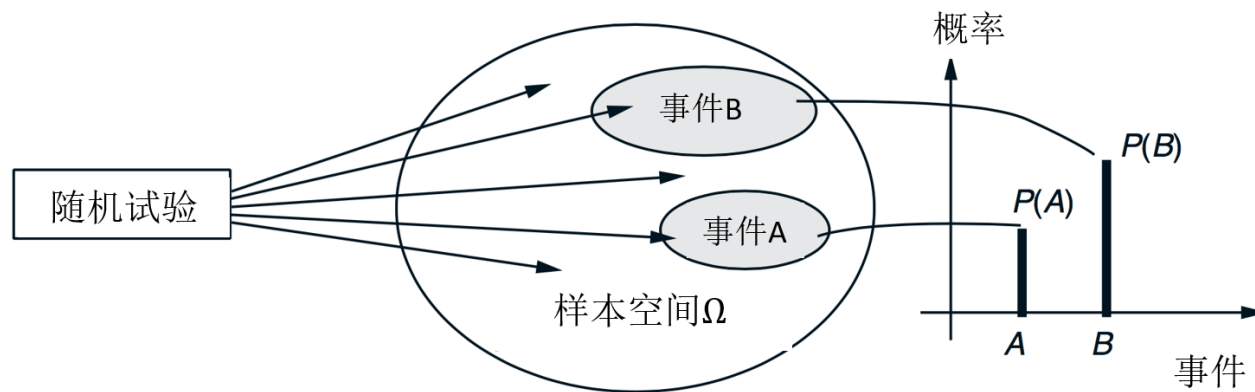
2. 概率的定义及性质

- 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) $0 \leq P(A) \leq 1$.



如何求出事件 A 的概率
(记为 $P(A)$)?

概率空间

2.1. 古典概型与几何概型

(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

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



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, \dots, n_k 个元素, 其中 $n_1 + n_2 + \dots + n_k = n$, 则一共有

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

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, \dots, n_k 个, 其中 $n_1 + n_2 + \dots + n_k = n$, 要把它们排成一行, 则一共有

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

种不同排法.

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 ($m \leq n + 1$). 求事件 $A = \{\text{任意两个女孩不相邻}\}$ 的概率。

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) $A = \{\text{指定的 } r \text{ 个盒子各含一个球}\}$
- (2) $B = \{\text{每盒至多有一球}\}$
- (3) $C = \{\text{某指定盒中恰有 } m \text{ 个球}\}$

• 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

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

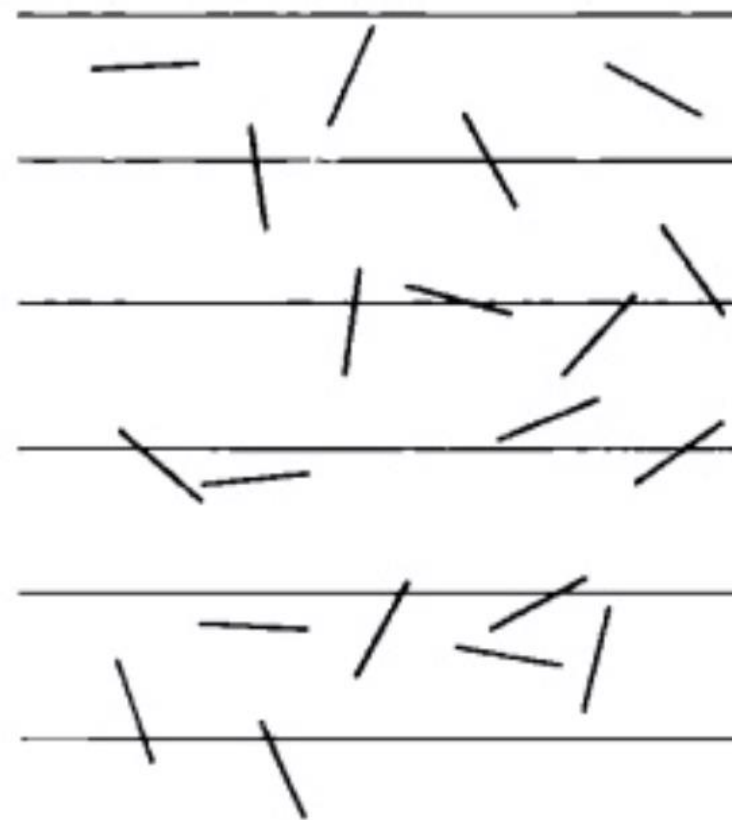
为事件 A 的几何概率。这里等可能性体现在“落在区域 A 的概率与区域 A 的测度成正比并且与其形状位置无关。”

这里 $m(\Omega)$ 表示 Ω 的“大小”。

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 $t(t \leq T)$ 再离去。试求事件 $A = \{\text{甲乙将会面}\}$ 的概率。

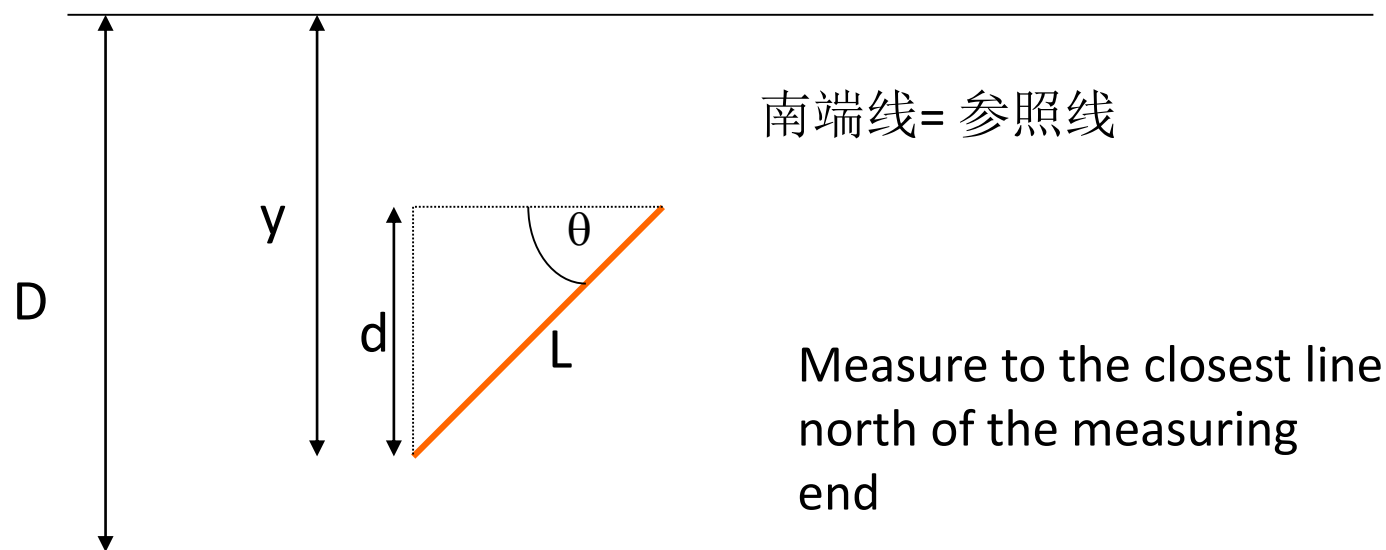
例子： Buffon投针

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 的针 ($L \leq D$)，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：

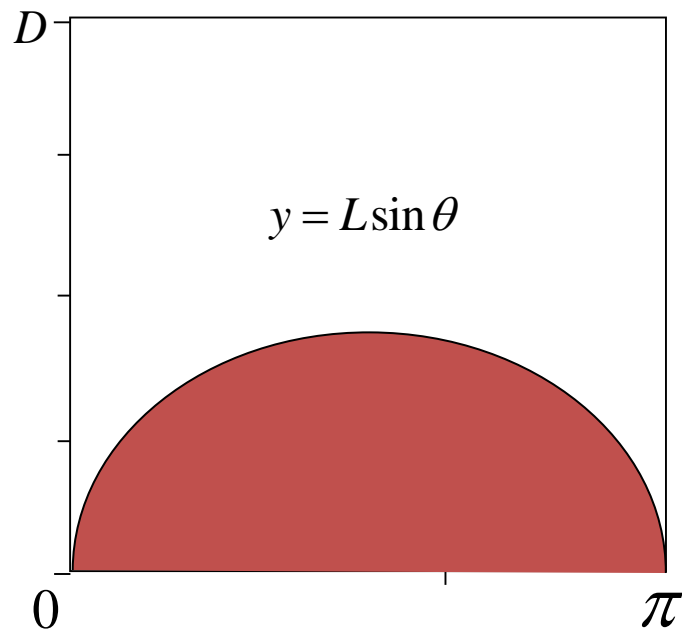
$$d = L \sin \theta \quad 0 \leq \theta \leq \pi \quad 0 \leq y \leq D$$

关系式

独立均匀分布

独立均匀分布

如果 $y \leq L \sin \theta$ ，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。

$$\begin{aligned} \text{红色区域} &= L \int_0^{\pi} \sin \theta \, d\theta \\ &= L(-\cos \pi + \cos 0) \\ &= 2L \end{aligned}$$

$$\text{总区域} = D\pi$$

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