
绪论

概率论与数理统计 · 复旦大学经济学院

ECON130001 2026 秋季 江嘉骏

先从你学过的说起

高中学过的概率

「 $\text{概率} = \text{有利结果数} \div \text{总结果数}$ 」，抛硬币、掷骰子、摸球。

大一学过的工具

微积分（积分——后半学期算连续型分布要用）、线性代数（矩阵——多维随机变量要用）。

★ 但高中那条公式有个前提没说

「 $\text{有利结果数} \div \text{总结果数}$ 」只在**所有结果等可能**时成立。那「明天下雨的概率是 70%」呢？这里的「总结果数」是多少？

为什么需要这门课

背景 两个人赌博，约定**先赢满 5 局的人拿走全部赌金**。打到甲 4 胜、乙 3 负时，出了事，赌局中断，再也打不下去了。钱怎么分？

转折 最自然的答案是**按已赢局数分，甲 4 份、乙 3 份**。这也正是当时的标准答案——写在**卢卡·帕乔利** 1494 年的《算术集成》里，而帕乔利是复式记账法的奠基人，欧洲当时最好的数学家之一。

可塔尔塔利亚指出一个要命的反例：如果**只打了一局就中断**，按这个规则赢的人拿走全部赌金——可他离 5 局还差 4 局。

问题 问题出在哪？——**分钱该看的不是已经发生了什么，而是还没发生的会怎样。可「还没发生的事」怎么算？**

来源

分赌本问题 (problem of points) 。帕乔利 1494 年《Summa de arithmetica》按已赢局数比例分配，塔尔塔利亚在 16 世纪中叶指出其反例；1654 年帕斯卡与费马的通信给出了正确解法，这场通信通常被视为概率论的起点。访问日期 2026-08-01。

本讲学习目标

- ① 学会概率论最早的那个思路：把**还没发生的事**也列出来再数
- ② 见识三个**直觉会失灵**的例子，明白为什么需要一套形式化的语言
- ③ 分清**概率、统计、计量经济学**各自回答什么问题

概率论的起源：1654 年的一封信

帕斯卡把这个问题写信问了费马

两人 1654 年的这场通信，通常被视为概率论的起点。费马的回答是：**盯着已经赢了几局，方向就错了。**

该问的是另一个问题

如果继续打下去，各自赢的可能性是多少？ 甲还需再赢 1 局，乙还需再赢 2 局——所以**最多再打 2 局**必分胜负。

这 2 局的结果只有四种，等可能：

甲甲 甲乙 乙甲 乙乙

前三种甲都赢了，只有「乙乙」乙才赢。

答案

甲拿 $3/4$ ，乙拿 $1/4$ 。**把「没发生的事」也列进样本空间，问题就解了。**

最「简单」的难题：扔硬币

一枚公平硬币，正反面可能性相同

- 连扔两次，都是正面的可能性？
- 连扔三次，都是正面的可能性？
- 连扔三次，**恰有两次**正面的可能性？
- 连扔三次，**只有一次**正面的可能性？

再问一个

连扔**十次**，恰有五次正面的可能性是多少？

前四问都能靠列举硬算，第五问一列就是 1024 种。 **需要的不是耐心，是方法。**



看似最简单的随机现象

直觉失灵之一：同月同日生

先猜一个数

这个教室里，**至少有两个人生日相同**的可能性有多大？

再猜一个：一个班要有**多少人**，这件事才有一半的机会发生？

多数人的猜测

「一年 365 天，总得有一两百人吧。」

实际答案

23 人就超过 50%，**50 人**时高达 97%。

交互：生日悖论模拟器

怎么用

把人数拖到全班实际人数，看概率；再点「模拟 2000 次」，看实际比例。

23 个人为什么就够

要比的**不是人数和 365**，是**配对数**和 365。23 个人能配出 $\binom{23}{2} = 253$ 对。



交互演示：拖动班级人数，看同天生日的概率如何上升

https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/birthday_paradox.html

直觉失灵之二：换还是不换

美国电视节目「Let's Make a Deal」

- 三道门，一道后面是汽车，另两道是羊
- 你先选一道，**不打开**
- 主持人打开剩下两道中的一道，**是羊**
- 问：要不要把你选的门换成另一道没开的？

一种看法

剩两道门，一车一羊，各 $1/2$ ——换不换都一样。

正确结论

换。主持人知道车在哪，他的动作携带了信息。换了之后中奖概率**翻倍**：从 $1/3$ 变成 $2/3$ 。



选一道，主持人开一道羊，换不换？

交互：三门问题模拟器

怎么玩

选一扇门，主持人打开一扇有羊的门，再决定换还是不换。手玩几局看不出差别，点「模拟 1 万局」比较两种策略的胜率。

为什么该换

「换」相当于赌自己第一次选错了，而第一次选错的概率正是 $2/3$ 。



交互演示：手玩几局，再让机器模拟一万局对比换与不换

https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/monty_hall.html

直觉失灵之三：布丰投针

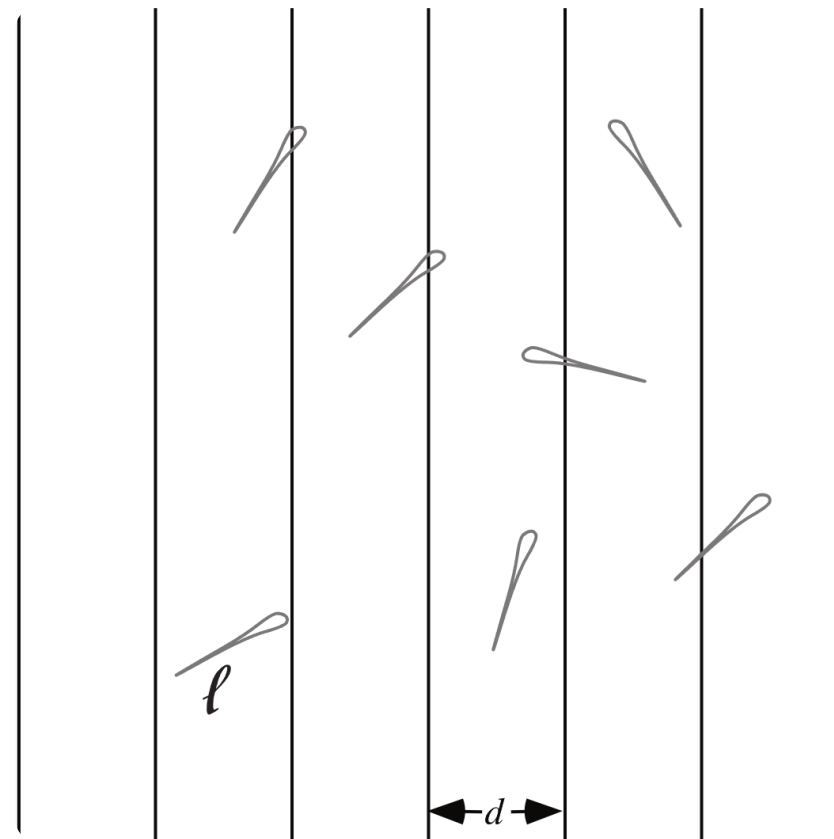
问题

一张纸上画满等距平行线，间距 d 。随机扔一根长度 L ($L < d$) 的针。

针与某条平行线相交的概率是多少？

困惑

- 这是个纯几何问题，为何与概率有关？
- 而且——答案里竟然出现了 π 。



针落在平行线上，相交与否

布丰投针：把几何与概率接起来

答案

$$P = \frac{2L}{\pi d}$$

证明思路

设针中点到最近平行线的距离为 x 、针与平行线的夹角为 θ ，相交条件是 $x \leq \frac{L}{2} \sin \theta$ 。在 (θ, x) 的矩形上算面积比即得。完整推导要用到二重积分与几何概型，见第 2 讲与教师笔记。

反过来用：拿实验测 π

大量重复投针，统计相交频率 $f = \frac{\text{相交次数}}{\text{总次数}}$ ，反推

$$\pi \approx \frac{2L}{d \cdot f}$$

这是**蒙特卡洛方法**的思想源头——用随机模拟去算一个确定的数。

这里埋了两个后面的核心

- **几何概型**：样本空间是几何图形时，用面积算概率——**第 2 讲**
- **大数定律**：频率为什么会稳定到概率上——**第 8 讲**

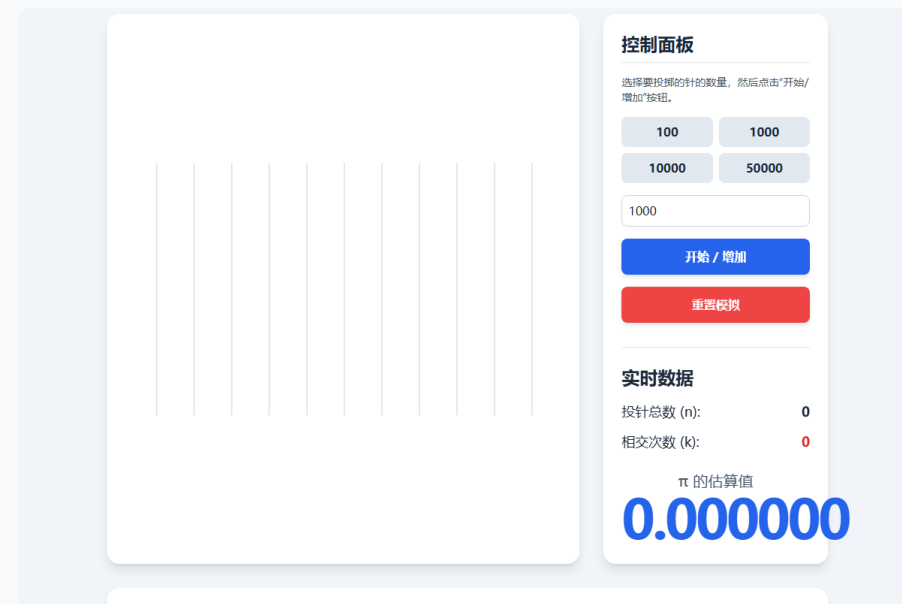
交互：布丰投针

怎么用

点「连续投针」让它一直投，右边的估计值会随投针数一起变。

几十次和几千次不是一回事

投几十次时估计值上下乱跳，投到几千次才稳定下来贴住 π 。**这就是大数定律的样子**，第 8 讲会把它写成定理。



交互演示：连续投针，观察估计值向 π 收敛

https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/buffon_needle.html

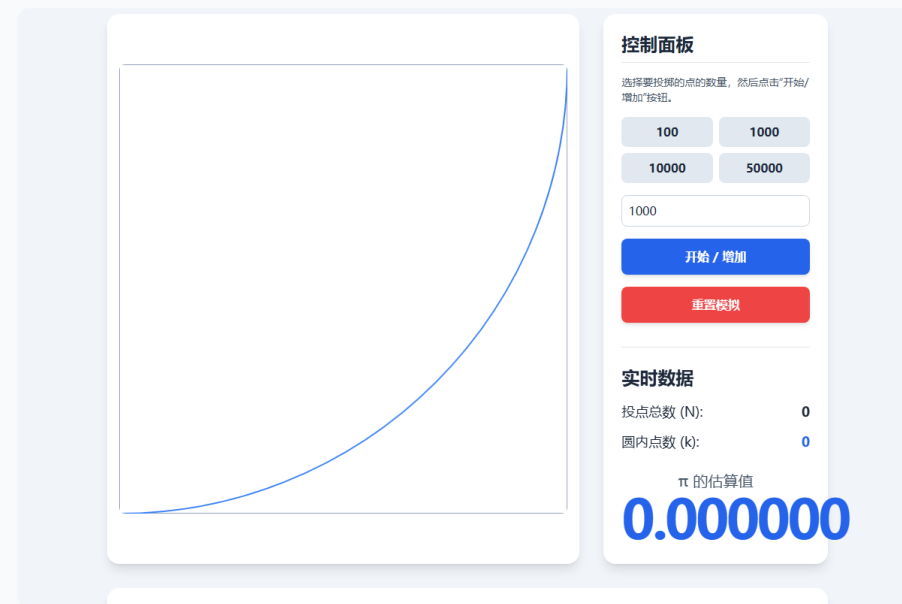
交互：蒙特卡洛模拟 π

怎么用

点「撒点」，在正方形里随机撒点，看落进圆内的比例。

和投针是同一个办法

撒点、数比例、反推面积——和上一页一样，都是**拿随机试验的频率去估一个确定的数**。换了个壳，思路没变。



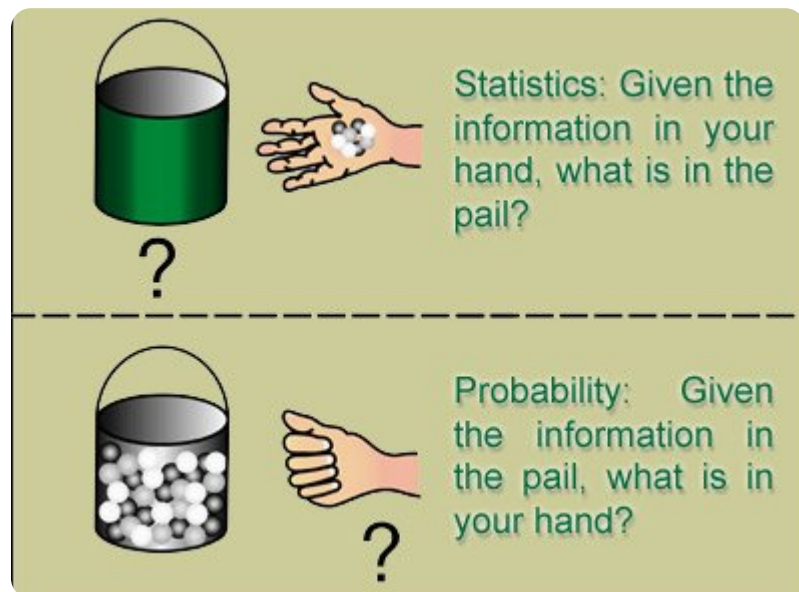
交互演示：正方形内随机撒点，用落入圆内的比例估算 π

https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/Monte_Carlo_pi.html

小结：三个例子说明了什么

- **直觉不可靠。**生日问题低估、三门问题误判、投针问题连方向都想不到。而且错的不是外行——帕乔利、当时的数学家、后来质疑三门问题的博士们，都错过。
- **错的往往不是计算，是「问题到底在问什么」没界定清楚。**「有人和我同天生日」与「任意两人同天生日」是两件事；「主持人随机开门」与「主持人知情开门」是两件事。
- 所以这门课的第一件事，是**造一套不会有歧义的语言**——这就是第 1 讲。

概率、统计、计量：三件不同的事



方向相反的一对

概率论：给定条件 → 预测未知

已知硬币是公平的，问扔十次出现五次正面的可能性。**模型在手，推数据。**

统计学：给定数据 → 归纳规律

扔了十次出现九次正面，问这硬币还公平吗。**数据在手，推模型。**

计量经济学：从数据里找变量之间的关系

并且要回答：这个关系是真的，还是碰巧？

接下来

到这里，你大概知道这门课在做什么了。

但还有一个问题没回答：这些十七世纪赌桌上的东西，和你要学的金融、要用的 AI，到底有什么关系？

概率论在 AI 时代的「文艺复兴」

先问一句

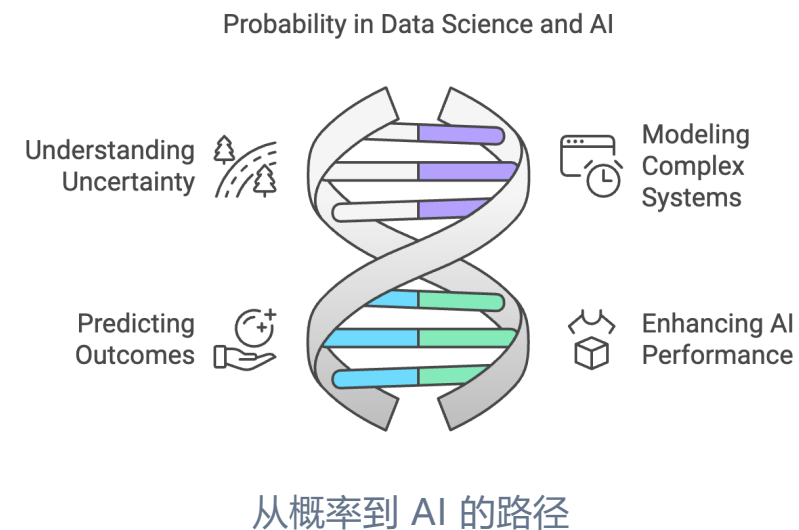
人工智能为什么算「智能」？

一个回答

因为它的核心能力之一，是在**充满不确定性**的世界里做出决策和预测。

而概率论正是量化不确定性的数学语言

说概率论是现代人工智能的基石，并不夸张。



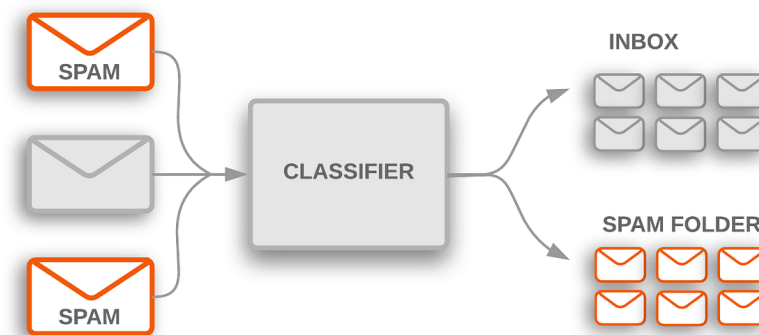
案例一：AI 如何识别垃圾邮件

朴素贝叶斯分类器的三步

1. **学习**：统计「中奖」「发票」这些词在正常邮件与垃圾邮件里各自出现的频率
2. **推理**：收到新邮件时，用**贝叶斯公式**算
 $P(\text{垃圾} \mid \text{出现「发票」})$ vs $P(\text{正常} \mid \text{出现「发票」})$
3. **决策**：哪个大就归哪类

这就是第 3 讲的内容

看起来很现代的自动分类，底座是**条件概率**。



训练 → 推理 → 决策

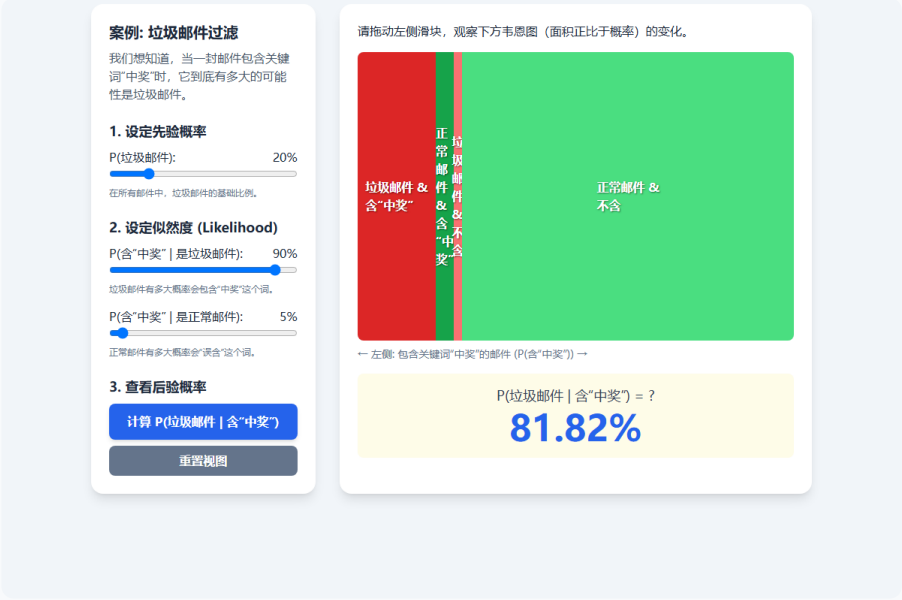
交互：垃圾邮件过滤器

怎么用

改几个词的出现频率，或者调整「垃圾邮件占比」，看判定结果怎么跟着变。

它没有在「读懂」邮件

分类器只是在算「出现这些词的前提下，这封信是垃圾邮件的概率」。今天只演示，第 3 讲我们自己把这个概率算出来。



交互演示：调整词频与先验，看分类器如何做判断

https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/bayes_spam.html

案例二：ChatGPT 如何「创作」

语言模型的核心任务

给定一段文字，预测**下一个最可能出现的词**。

例如输入「今夜的月色」

模型算出的不是一个词，而是一个**概率分布**：

下一个词	概率
真美	0.30
格外	0.20
那么	0.15
.....	

然后从这个分布里**采样**一个词，再开始下一轮。

这就是第 4-5 讲的内容

所谓「创作」，是一场基于**随机变量及其概率分布**的文字接龙。



下一个词的概率分布

概统与经济金融：四类应用

金融资产定价

用概率分布描述股价、利率的不确定性。期望 $E[R]$ 定义**收益**，方差 $\text{Var}(R)$ 定义**风险**。

第 7 讲

保险精算与风险管理

以**大数定律**为基石，靠风险汇聚让保险公司稳健经营。

第 8 讲

计量经济学与因果推断

回归找关系，**假设检验**判断这关系是真的还是碰巧。

第 11–12 讲

金融科技与 AI

信用评分算 $P(\text{违约})$ ；算法交易按概率信号下单；智能投顾按风险偏好配组合。

贯穿全课

关于考核

成绩构成

- **期末考试 70%**
- **平时成绩 30%**：两次小测，**各占 15%**

作业

课后练习，**无需提交。**

小测两次，时间暂定

每次 **50 分钟**，各占总成绩 **15%**，**提前一周通知。**

小测	时间（暂定）	范围
小测 1	第 7 周	第 1–5 讲（第一、二章）
小测 2	第 12 周	第 6–8 讲（第三、四、五章）

课程信息

上课

- 时间: **周一 4-5 节、周四 1-2 节**
- 地点: **H4203**
- 课件: 课前放到 eLearning, 方便预习与记笔记

答疑

- Office hour: 课后或邮件预约
- 办公室: 经济学院大楼 1005

任课教师

- 江嘉骏 副教授
- 北京大学金融学博士、数学与应用数学学士
- 纽约大学 Stern 商学院访问学者
- 研究方向: 行为金融、消费金融、金融科技
- Email: jiangjj@fudan.edu.cn

教材与参考书

教材

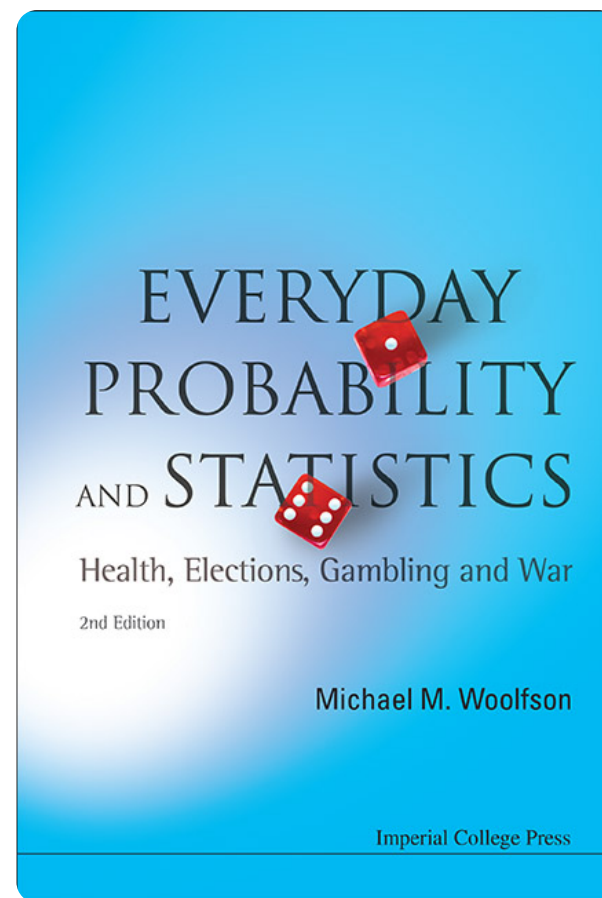
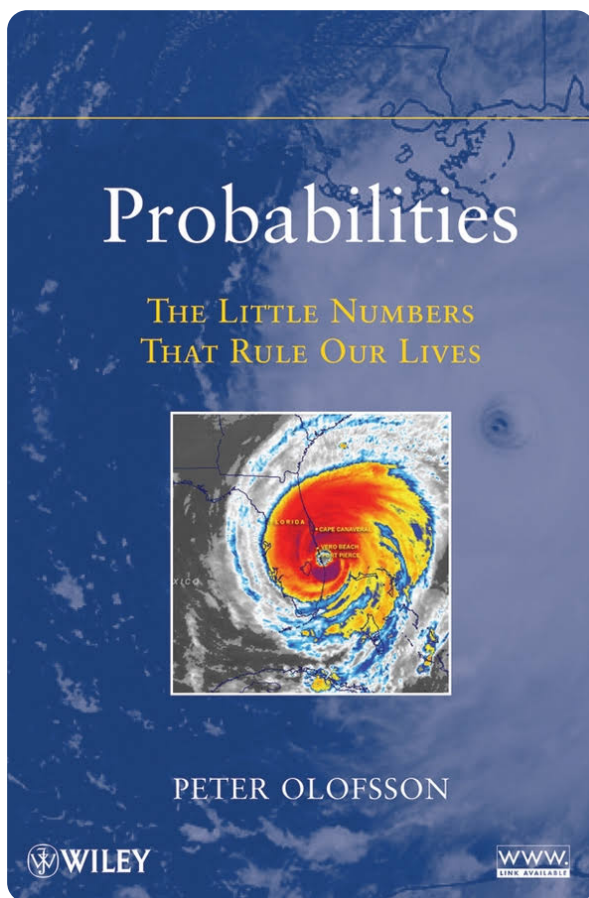
吴传生 《概率论与数理统计》（第 4 版）

参考书

- 袁荫棠 《概率论与数理统计》（修订版）
- 朱弘鑫 《概率论与数理统计》
- DeGroot & Schervish *Probability and Statistics*, 4th ed.

课外读物

对概率论里的趣味问题感兴趣，可以参考这两本。



课程安排

周	内容	教材	备注
1	绪论 · 随机事件与概率定义	第一章	
2	古典概型与几何概型	第一章	
3	条件概率与事件独立性	第一章	
4–5	离散型随机变量与分布函数	第二章	10.1、10.5 停课
6	连续型随机变量与密度函数	第二章	
7–9	多维随机变量及其分布	第三章	小测 1（时间暂定）
9–10	随机变量的数字特征	第四章	
11	大数定律与中心极限定理	第五章	
12–13	样本及抽样分布	第六章	小测 2（时间暂定）
13–15	参数估计	第七章	
15–16	假设检验	第八章	
16	复习与总结答疑	—	12.24 最后一次课

本讲总结

① 回到开场那个问题

分赌本该怎么分？——不看已经赢了几局，看**继续打下去会怎样**。把没发生的可能性——列出来，再数一数，答案就出来了：3 : 1。

这个动作——**把所有可能情况列成一个集合，再在上面计数**——就是这门课全部内容的原型。

② 核心结论

- 概率论：**由模型推数据**；统计学：**由数据推模型**。方向相反
- 直觉在概率问题上系统性地不可靠，而且错的常常不是计算，是**问题没界定清楚**
- 频率会稳定到概率上（布丰投针、蒙特卡洛）——这条要到第 8 讲才证

本讲总结（续）

③ 这学期你会用到的三条线

- **金融**：期望是收益，方差是风险，大数定律是保险的基石
- **AI**：贝叶斯公式做分类，概率分布做生成
- **计量**：假设检验回答「这个关系是真的吗」

④ 下一讲

今天所有的例子都在问同一件事：**某件事有多可能发生。**

可我们连「**某件事**」到底指什么都还没说清楚——「至少有两人同月同日生」是一件事，还是很多件事？

下一讲：先把「事件」这个词定义清楚。