随机性：机会与期望

应坚刚

2019 年 6 月 24 日

目　录

前言 **1**

参考文献 **2**

第二章　前世 **3**

2.1 确定与不确定 3

2.2 概率论简史 5

第三章　随机现象的数学化 **12**

3.1 数学的语言 12

3.2 样本空间与事件 13

3.3 关于概率的直觉 14

第四章　古典概率 **17**

4.1 等可能性与古典概率 17

4.2 样本空间的选取 18

4.3 经典古典概率问题 19

4.3.1 分赌注问题 19

4.3.2 掷硬币 20

4.3.3 生日问题 22

4.3.4 抽签与顺序 23

4.3.5 配对问题 24

4.4 反射原理 26

4.4.1 计票问题 27

4.4.2 首次回归时间 28

4.5 几何概率 29

4.5.1 坐车问题 29

4.5.2 约会问题 30

4.5.3 Buffon 问题 31

4.5.4 Bertrand 问题 33

4.6 无穷和 35

4.6.1 三人比赛 36

第五章　简单概率概念 38

5.1 概率空间 38

5.2 随机变量与分布 39

5.3 独立与乘法法则 40

5.3.1 重复Bernoulli 试验 42

5.4 条件概率 43

5.5 全概率公式 44

5.5.1 抽签与顺序无关 45

5.5.2 排位置 46

5.5.3 骰子游戏 46

5.5.4 赌徒输光问题 46

5.6 Bayes 公式 47

5.6.1 检测方法的有效性 48

5.6.2 Monty Hall 问题 48

第六章　数学期望与方差 51

6.1 分布的期望与方差 51

6.2 0,1-值随机变量 52
6.3 二项分布 52
6.4 等待成功 53
6.5 等待模式出现的时间 53
6.6 配对数 55
6.7 等待最高报价 56

第七章　母函数 **57**
7.1 首次通过时 58
7.2 种群灭绝概率 60

第八章　大数定律 **64**
8.1 Chebyshev 不等式 64
8.2 Bernoulli 大数定律 65
8.3 独立地重复 66
8.4 蒙特卡洛算法 67

第九章　预期与结果 **70**
9.1 样本平均与中位数 70
9.2 投资陷阱 70

第十章　随机商品与定价 **74**
10.1 公平 74
10.2 风险与风险溢价 75
10.3 对冲与无风险定价 76
10.4 期望与心理预期: 圣彼得堡悖论 79

第十一章　概率应用于推断 **82**

11.1 品茶女士与假设检验 83

11.2 对于推断的不同观点: 频率与Bayes 86

11.2.1 Bernoulli 的大数定律 86

11.2.2 Bayes 公式及其推断思想 87

11.3 频率学派与Bayes 学派的争议 90

前言

随机现象, 机会, 可能性等等, 无处不在，且与我们的生活息息相关. 它看似没有规律，但实际上隐含着深刻的规律. 概率论就是研究随机现象背后所蕴藏的规律的数学理论. 是分析数据的理论基础，它在现代社会中变得越来越重要，在数学中的地位也越来越高.

这是一本讲述随机现象与概率论的通识课程讲义, 重点是帮助学生了解随机性, 了解概率论的方法与历史, 让学生知道概率论有些什么用, 也让学生知道概率论不能做什么. 最后, 我们也会讲一点以概率为基础的统计推断.

讲义的定位是数学的科普, 我们的想法是让学生在问题中学习理解随机现象与概率, 理解问题本身, 理解解决问题的概率思想方法. 讲义中用到的数学知识并不高深, 基本上是初等数学, 以及初等的微积分, 但是这不等价于简单, 很多计算还是需要动脑子思考的, 所以只要有兴趣, 肯思考, 花点时间, 中学生到大学生, 无论文科理科, 都可以愉快地学习.

需要提醒读者的是, 相比其它的数学, 在学习概率论时, 直观起到关键的作用, 很多问题, 如果直观上明白了, 理论上就简单了. 著名的概率学家W. Feller 写的名为**概率论及其应用**的书中有这样一段话: *如果概率论对于生活是真实的, 那么每一个经验应该对应一个可以证明的命题.* 这个话特别有理, 也许读者还没有把一个经验转化为定理的能力, 但至少可以体会一下, 概率论对于生活是不是真实的.

应坚刚

2019/4/24

参考文献

[1] Feller, W., Probability Theory and its Application, Vol. I(1959: Third edition), Vol. II(1970), Wiley & Son

[2] Grimmett, G., Stirzaker, D., Probability and Random Processes, (3rd edition), Oxford University Press, 2001

[3] Grimmett, G., Stirzaker, D., One Thousand Exercises in Probability, Oxford University Press, 2001

[4] Ross, S., A First Course in Probability, 6th edition, Pearson Education 2002 (其中译本由机械工业出版社2006 年出版).

[5] Ross, S., 概率模型导论, 人民邮电出版社, 2015

[6] von Mises, R., Probability, Statistics and Truth, Dover Publications, Inc., 1957

[7] Salburg, D., The Lady Tasting Tea, Henry Holt and Company, New York, 2001

[8] 王梓坤, 概率论基础及其应用, 北京师范大学出版社, 北京, 2007

[9] 复旦大学, 概率论, 高等教育出版社, 北京, 1979

[10] 杨振明, 概率论, 科学出版社, 北京, 2001

第二章　前世

§2.1　确定与不确定

人每天都与各种现象打交道, 有的是确定的, 有的是不确定, 也称为随机的. 面对随机现象, 人的一生会有各种选择, 各种机会和可能性, 对人类来说充满诱惑.

提起随机, 大多数人可能首先想到一些常见的游戏, 例如打牌打麻将, 还有抽签买彩票等. 其实随机现象充斥生活的每个角落, 无时不刻会遇到, 只不过我们对此已经习以为常. 例如你是一个银行前台职员, 在一个普通工作日的早上, 你知道6 点多天差不多亮了, 当然夏天亮得早一点, 冬天亮得晚一点, 但不会有太大不同, 如果天晴, 太阳肯定会从东方升起, 这是确定的, 但是你可能对气温, 风力, 还有是否下雨会感到不确定, 要看看天气预报, 但天气预报也不一定每次都能预报准确, 这就是不确定性. 然后你吃好早饭, 在计划好的时间出门, 这是你可以控制的, 也是确定的, 接着去坐

图 2.1　掷硬币

公交车上班, 车上有没有座位? 会不会堵车? 什么时间到办公室? 这些都是不确定的. 上班之后, 有多少人来办事? 办什么业务? 中间有没有空闲? 这些也是不确定的. 再比如一位小学一年级的教师站在讲台后, 看着面貌不同性格迥异的学生, 遥想他们的未来, 从机会来看, 这些学生应该是差不多的, 但是最终的结果肯定大相径庭, 有的顺风顺水, 有的逆风奋斗, 有的英年早逝, 不一而足, 令人不由得感慨命运无常.

人类对确定的现象有一种天生的亲切感, 例如看着四季依次交替, 太阳东升西落, 即使是灾难, 心里有预期, 也会比较踏实. 但对于随机现象却会有一种陌生甚至畏惧的感觉, 特别是灾难, 例如台风, 暴雨, 洪水, 因为突如其来, 难以预测, 俗话说: “天有不测风云, 人有旦夕祸福,” 正是这种情感的写照. 还有些现象, 例如日全食, 古时人们对此很害怕, 称为天狗吞日, 以为它会伴随灾难降临, 殊不知后来人类认识到这是某种自然规律, 是可以预测的, 因此不再神秘和畏惧.

随机现象也是神秘的, 因为它无法被预测, 古时人们就把这种人无法掌控的东西当作是天的意志, 当对某事迟疑不决的时候, 就运用占卜, 求卦等手段来进行决策.

这里有个真实的故事, 发生在19 世纪, 一艘从英国开往美国的轮船在途中遭遇风暴, 船要沉没, 船长首先命令把所有的行李扔掉以减轻重量, 但是还不够, 为了大多数人的安全, 船长指定了一些乘客把他们放入救生艇自生自灭, 最后这些乘客都遇难了, 但船安全抵达美国. 这些一些乘客起诉船

长谋杀, 法官审理后认为船长有罪, 认为他为了大多数乘客而放弃一些乘客并没有错, 但是船长错在不应该私自决定放弃哪些乘客, 而应该交给上帝来决定, 也就是用抽签或者其他随机手段来决定.

然而, 随着时间推移, 人类对随机现象会形成有用的直觉, 即学会通过持续观察并总结经验来进行推断, 从而应对随机性. 例如上班的人需要估计早上该几点出门使得迟到的可能性小到可接受的程度. 关于天气预报, 民间有很多谚语, 例如: “一雾三晴，重雾三日必大风。”“满天乱飞云，雨雪下不停。”“喜鹊枝头叫，出门晴天报。”“风大夜无露，阴天夜无霜。”这些都是古人总结出来的, 不一定准确, 但有一定的参考价值.

对于充满不确定性的未来, 人类一直在探索, 致力于把不确定变成确定, 上千年来人类在科学研究方面取得了巨大的进展, 例如物理学化学的进展使得人类对头顶的星空和居住的地球了解更多, 生物学和医学方面的进展使得可以治愈部分以前束手无策且知名的疾病, 大大地提高了人的寿命. 但是仍然有许多充满了不确定性, 让人感觉无能为力.

概率论是探究不确定性的一种数学模型, 它用公理化界定不确定性, 然后用数学逻辑来推导这样的不确定性有什么性质, 除了其自身在数学上的价值, 概率论对某些随机现象的模拟可以说是非常真实的, 在很多情况下, 可反哺于解释随机现象中的规律, 对人类认识随机现象有极大的参考价值.

§2.2 概率论简史

关于概率论的历史, 要分两个时期讲. 如果只是谈人类对随机现象的兴趣和认识, 那么应该说是很早很早不可考的早期. 据考古发现, 远古时候人类就有占卜, 算卦等活动了, 所谓占卜, 就是用龟壳、蓍草、铜钱、竹签、纸牌或占星等手段和征兆来推断未来的吉凶祸福，除此之外, 民间也很早就有抽签掷骰子打骨牌等赌博游戏了. 这些都属于人类对随机现象的利用. 在这个时期, 人类对随机现象中可能性的大小并没有明确的记载, 但是显然不能排除人类已经对可能性的大小有感觉也有兴趣知道. 在日常言语中, 可能性与机会差不多是等义的, 可能性的大小在现在也称为概率或者几率. 据记载, 真正对概率这个数字进行描述和计算, 要归功于16 世纪中期意大利数学家G. Cardano, 卡尔达诺是个意大利数学家, 生于1501 年, 死于1575 年, 他最著名的工作是关于三次方程的解. 他喜欢赌博, 对于赌博的热情驱使他在晚年写了题为“机会的游戏”一书阐述了很多概率的问题与思想, 这本书实际上在他死后才出版. 他在此书中给出的概率是以“甲乙各自的有利场合数之比$r:s$”这种形式给出，而不是现在通常定义的概率: $\frac{r}{r+s}$. 他还讨论了独立重复, 大数定律等等的问题, 但只是有叙述, 没有严格的证明. 他的观点还比较粗糙, 处于探索概率的初期, 他的书中也还有很多的错误.

对于概率理论发展起到关键作用的应该是法国数学家Pierre de Fermat 与Blaise Pascal 在1654 年关于分奖金问题以及其

图 2.2 Pierre de Fermat, 1601-1665

它一些概率问题讨论的通信. Fermat 与Pascal 的探讨富含思想, 推理相当严格, 为后续概率的数学理论大厦放下了第一块基石.

差不多同时, 在1657 年, 荷兰著名科学家C. Huygens 出版了"关于机会游戏的推理"一书, 对当时关注的概率问题有系统的研究.

到18 世纪初期, 瑞士数学家Jakob Bernoulli 在1713 年, 去世后八年, 出版了《猜度术》, 其中讨论了很多复杂概率问题的计算, 最重要的是给出并严格证明了大数定律(law of large

图 2.3 Blaise Pascal, 1623-1662

numbers), 即在独立地重复一个随机试验n 次, 当n 很大时, 成功的频率趋向于成功的概率. 大数定律很直观, 很多人都或多或少地对此有感觉, Cardano 曾经叙述过这个思想, 但是Bernoulli 首次给出严格的证明.

这个时期另外一个关于概率论的著名著作是旅居英国的法国数学家de Moivre 在1718 年出版的《机会理论》一书, 此著作的主要贡献的给出并证明了最初形式的中心极限定理, 现在此定理通常称为de Moivre-Laplace 中心极限定理, 它可以被看作是大数定律的精细化.

另外我们还应该提到在18 世纪的中叶, 1763 年出版的《解决机会理论中一个问题的论文》, 作者是Thomas Bayes, 这篇论文是他去世后两年由他的朋友Richard Price 整理出版的, 论文是从概率论角度来回答哲学家David Hume 关于过去的经验对认识未来的影响之问题. 现在, 文中的公式称为Bayes

公式, 它是统计推断理论中的重要观点.
法国数学家Pierre-Simon Laplace 对概率论的发展也做出了重要贡献, 他分别在19 世纪初期的1812 年和1814 年出版了两本著作《概率的分析理论》与《概率的哲学探讨》, 给出古典概率的定义及概率应该服从的一些规则, 讨论概率在哲学, 科学, 社会科学, 法律审判中的应用, 也就是用于统计推断.
20 世纪是个伟大的世纪, 尤其对于概率论. 由于概率本身的发展以及概率论在自然科学和社会科学中的广泛应用, 人们清楚地知道, 概率论自身正在孕育着一个伟大的学科, 但是概率论还不是数学. 什么是数学? 数学是建立在公理体系上的学问, 正如Euclid 的几何学一样, 而概率还没有自己的公理体系. 著名德国数学家Hilbert 把概率论的公理化列在他于1900 年国际数学家大会上发布的著名23 问题之一, 第六个问题, 与Riemann 猜想, 连续统假设等并列.
谁解决了这个问题呢? 天才的苏联数学家A.N. Kolmogorov 于1933 年彻底解决了这个问题, 当然他也是站在别人的肩膀上, 1900 年, Lebesgue 发表了历史上最著名的博士论文, 构造出Lebesgue 测度, 建立了Lebesgue 积分. Lebesgue 测度的关键是可列可加性, Kolmogorov 借用测度的可列可加性, 建立起概率的公理体系. 自此, 概率论才真正成为数学的一个分支, 数学家们可以根据逻辑来推导概率的性质, 用于解释并启发解决生活以及其他学科中的随机现象.

图 2.4 Jakob Bernoulli

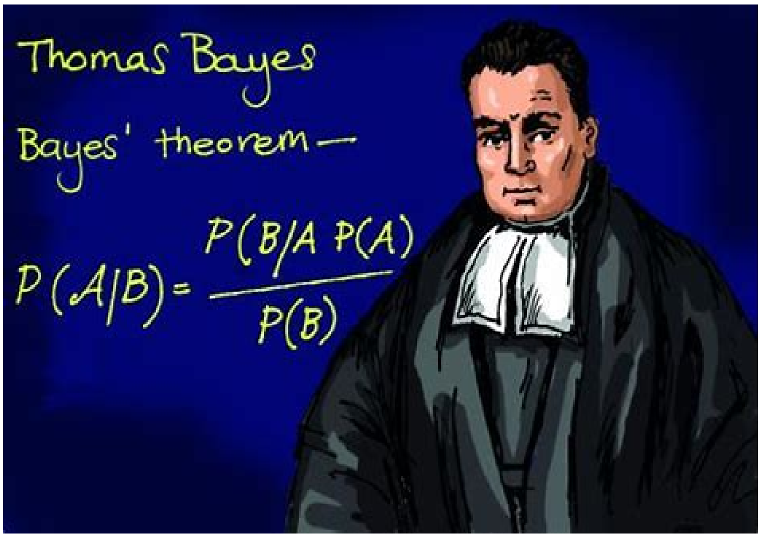


图 2.5 Thomas Bayes

图 2.6 Pierre-Simon Laplace

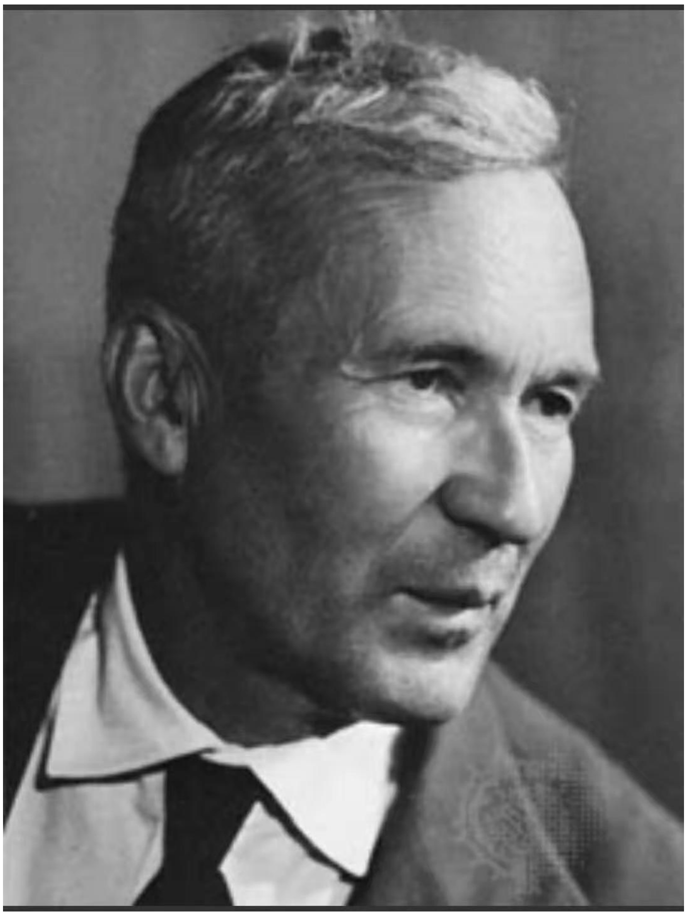


图 2.7 A.N. Kolmogorov

第三章　随机现象的数学化

§3.1　数学的语言

人类生活中有很多随机现象, 所以人们对机会或者可能性等术语是熟悉的, 受过适当教育的人甚至知道怎么来估算一些机会大小或者随机事件发生的可能性. 这些估算的方法源于直觉, 本讲义的大部分内容也是基于直觉, 只需要基于直觉, 我们就可以解决很多的经典概率问题. 因此在学习概率论的时候, 直觉是非常重要的, 甚至比理论还要重要.

尽管如此, 对于随机现象或者概率论的学习却相对比较晚, 通常要到高中后半阶段才会学习初等概率论或者统计. 而且因为很多的中学老师对概率的认识并不是特别透彻, 所以导致学生对随机现象以及概率通常是一知半解, 甚至有很多的误解. 本课程的目的是努力让进入大学但又没有机会系统学习概率论的学生们对概率论有个比较充分的认识, 在将来的工作中碰到随机现象时不再害怕, 碰到统计数据时知道怎么去解读.

要想用数学方法来研究随机现象, 第一个问题是要把随机现象数学化, 也就是要用数学的语言说话. 什么是数学的语言? 19 世纪后期所发展起来的集合与映射就是数学的语言. 我们假设读者已经在中学阶段学过数学的语言, 所以只是在下面简单重复一下.

一些特定的东西放在一起就组成一个集合, 集合中的东西称为集合的元素. 集合中元素是不能重复的, 也不排序. 符

号$a \in A$ 表示a 在集合A 中, 或者a 是A 的元素. 例如所有自然数的集合$\mathbf{N}$, 所有整数的集合$\mathbf{Z}$, 所有实数的集合$\mathbf{R}$, 所有中国人的集合, 所有女性的集合.

没有任何元素的集合称为空集, 记为$\emptyset$.

设有两个集合A, B, 如果A 中的任何元素也出现在B 中, 那么我们说A 是B 的子集, 或者A 包含于B, 或者B 包含A, 记为$A \subset B$, 也可以写$B \supset A$. 例如$\mathbf{N} \subset \mathbf{Z} \subset \mathbf{R}$. 中国人的集合和女性的集合是互不包含的.

空集是任何集合的子集, 当然A 自己也是A 的子集, 它们称为是A 的平凡子集, 其它子集是非平凡子集.

如果A 是B 的子集, B 也是A 的子集, 那么我们说两个集合是一样的, 记为$A = B$.

集合也可以运算得到新的集合. 集合有三种基本的运算, 一个是交, 两个集合A, B 的交是它们公共元素的全体, 记为$A \cap B$. 如果它们没有公共元素, 即$A \cap B = \emptyset$, 那么说它们不交, 或者互斥. 再一个是并, 两个集合A, B 的元素放在一起(重复的只放一个)组成的集合, 称为两个集合的并, 记为$A \cup B$. 交与并运算可以对多个集合进行. 还有一个运算是减, 把在集合A 中不属于集合B 的元素放在一起组成的集合, 称为A 减去B 的差, 记为$A \setminus B$, 即

$$A \setminus B = \{a \in A : a \notin B\}.$$

如果$B \subset A$, 那么$A \setminus B$ 也称为B 在A 中的补集或者余集, 记为$\complement_A(B)$, 如果A 是众所周知的, 那么也记为B^c.

集合X 到集合Y 的映射f 是指按照此规则, X 中任何一个

元素x 对应Y 中唯一的一个元素y, 记为

$$f : X \longrightarrow Y.$$

当Y 都是数集时, 这样的映射也称为X 上的函数. 例如在给定时间下, 身高是人类这个集合上的函数, 体重也是.

§3.2 样本空间与事件

实际生活中的随机现象是多种多样的, 先让我们从简单的开始. 最简单的随机现象就是掷硬币, 掷骰子, 抽签, 摸球, 抽卡片等可以任意重复的随机现象, 这样的随机现象通常称为随机试验, 因为它们像试验那样可以重复地做.

从哪里开始呢?

从我们关心的问题开始. 掷一个骰子, 出现偶数的概率是多少?

在这句话中, "掷一个骰子"是讲述什么样的随机试验,"出现偶数"这件事情有可能发生, 也有可能不发生, 我们称之为随机事件, 简称事件.

掷一个骰子, 可能出现1,2,3,4,5,6 之中的一个数字, 如果出现2,4,6 之一, 那么就是出现偶数这个事件发生了; 反之, 如果出现1,3,5 之一, 那么出现偶数这个事件没发生.

一般的随机试验也是这样, 它会有一些可能出现的结果, 然后我们关心其中一些结果出现的概率大小.

这样, 我们把一个随机试验的所有可能出现的结果放在一起组成一个集合, 称为该随机试验的样本空间, 通常用希腊字母Ω 表示.

“所有可能出现的结果”这句话并不是很严格, 例如掷一个骰子, 如上所言是有6 个可能结果

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

但是如果我只关心是不是掷出6, 那么在我看来只有两个结果: 6 与非6; 而对于只关心是不是掷出一个偶数的人来说, 他眼中也只有两个结果: 偶数与奇数. 所以所有可能出现的结果是主观的, 由观察者的视角决定. 对于我们来说, 标准是: 做一次随机试验, 样本空间中的所有可能结果必须**有且只有**一个出现. 例如掷骰子, 1, 偶数, 素数三个结果不能作为样本空间的所有可能结果, 因为当2 出现时, 它是偶数也是素数, 违反“有且仅有”的标准.

样本空间里面的元素亦称为基本事件.

事件通常对应于若干基本事件组成的集合, 也就是样本空间的一个子集, 事件发生相当于其中一个基本事件发生. 例如掷骰子出现偶数对应于子集$\{2, 4, 6\}$, 出现素数对应于$\{2, 3, 5\}$. 因为至少有一个结果出现, 所以空集是不可能的, 也称为不可能事件. Ω 自身也是一个子集, 它一定是发生的, 所以称为必然事件. 不可能事件与必然事件是确定的, 其它事件是不确定的.

更有意思的是, 事件的运算也可以对应于集合的运算, 事件A 不发生, 那就是A^c, 即A 在样本空间中的补集; 事件A 与B 同时发生, 那就是$A \cap B$; 事件A 或者B 发生, 那就是$A \cup B$.

这样就可以用集合的语言来叙述随机现象中我们所关注的

问题, 与生活中的语言不同, 集合的语言非常严格与清晰, 不会引起歧义.

§3.3 关于概率的直觉

对于一个随机现象来说, 事件可能发生也可能不发生. 那我们是否可以谈论事件发生的可能性大小呢? 所谓可能性大小, 就是概率. 我们时常需要在面对随机现象时作决策, 这时, 概率就是一个非常重要的指标, 有很大的参考价值. 但要注意的是, 概率只是表达事件发生的可能性大小, 并不能明确预言事件会不会发生.

我们接着要说的概率是一个数学分支, 可是它的背景就是生活中熟悉的概率, 也就是说, 数学的概率应该符合大多数人对于生活中概率的认识和感觉, 把这样的东西抽象出来, 就是我们所要的概率.

我们对生活中的概率有什么样先验的认识和感觉呢? 这是很主观的, 难以完全说清楚. 作者先在这里列出几点, 大家可以思考一下是否认可:

1. 不可能事件的概率是0, 必然事件的概率是1, 其它事件的概率是处于0,1 之间;
2. 如果随机事件重复的话, 概率的大小说明会在其发生的频率中显示, 即概率大的事件发生得更经常;
3. 概率有可加性, 称为加法法则: 两个不同时发生的事件至少有一个发生的概率等于各自它们概率之和. 例如不输的概率等于胜的概率加和局的概率;

4. 概率还有乘法法则: 经常有直观上显然是不相干或者说相互独立的两个随机现象, 例如左手掷一个硬币, 右手掷一个骰子, 可以认为两者的结果是不相干的, 或者说是独立的. 独立的随机试验之下的事件同时发生的概率等于它们各自概率的积. 所以硬币正面朝上且骰子掷出6点的概率等于1/2 与1/6 的乘积, 是1/12.

第一点只是约定俗成, 第二点是感觉, 第三点概率有可加性是最本质的. 为什么概率有可加性? 物体的重量有可加性, 这很容易验证, 只要拿两块物体来秤一下重量, 然后再分别秤一下重量, 就可以验证了. 但概率不一样, 没有办法像秤重量那样秤一下概率.

如同承认几何原本上的五个公设就是Euclid 几何一样, 把第三点有限的可加性改为更强大的可列的可加性就是柯尔莫戈洛夫建立的概率论.

第四点是直观的, 它在古典概率模型中也是可以证明的. 我们暂时先使用这些直观性质来计算概率, 然后在后面作进一步解释.

习题: 设有 A, B, C 三个事件, 请用集合运算表示下列事件:

1. 仅有 A 出现;
2. A, B 都出现, C 没有出现;
3. 三个事件都出现;
4. 至少一个事件出现;
5. 至少两个事件出现;

6. 仅有一个事件出现;

7. 仅有两个事件出现;

8. 不多于两个事件出现.

第四章　古典概率

§4.1　等可能性与古典概率

通常用$\mathsf{P}(A)$ 表示一个事件A 发生的概率.

现在我们从简单的随机试验开始. 首先是掷硬币, 硬币有两面, 就称为正反面吧. 落地的时候究竟哪一面朝上是随机事件, 随便问一个人, 他都会说两面出现的可能性一样, 所以概率都是$1/2$, 其中2 是样本空间的元素个数.

同样, 掷一个骰子, 掷出任何一个数的可能性是一样的, 所以概率都是$1/6$, 其中6 是样本空间的元素个数.

仔细琢磨这两个例子, 这里有两个关键点, 一是任何结果的可能性一样, 所谓等可能性; 二是一个约定, 也就是说, 必然事件概率约定为1. 类似地, 只要样本空间中的基本事件是等可能的, 那么每个事件发生的可能性就是$1/|\Omega|$, 其中$|\cdot|$ 表示集合中的元素个数.

也许有人要问, 为什么硬币两面是等可能的? 这在数学上无法证明, 只能说是合乎自然且理想的假设. 合乎自然的意思是, 硬币的两面物理上几乎没有差别, 我们没有理由说哪一面出现的可能性大一点. 当然这只是一种简单化的假设, 如若考虑问题的角度不同, 完全可以作其它假设.

元素有限且每个结果等可能的样本空间称为古典概率空间, 其随机试验称为古典概率模型.

现在来看概率的可加性: 两个不会同时发生的事件A, B 至少有一个发生的概率是它们各自概率之和, 用数学语言说:

如果$A\cap B=\emptyset$, 那么

$$\mathsf{P}(A\cup B)=\mathsf{P}(A)+\mathsf{P}(B).$$

例如掷三枚硬币至少有两个正面朝上的概率等于恰好有两个正面朝上的概率与有三个正面朝上的概率之和.

还是要回到简单的掷骰子问题, 我们已经知道掷得每个点的概率都是1/6, 然后掷得1 或者2 的概率自然是2/6, 因为掷得1 或者2 这个事件对应子集$\{1,2\}$, 其中有两个元素; 掷得偶数的概率是3/6, 因为掷得偶数这个事件对应子集$\{2,4,6\}$, 其中有三个元素. 因此, 如果A 是样本空间为Ω 的古典概率模型中的一个事件所对应的子集, 那么概率理所当然应该是

$$\mathsf{P}(A)=\frac{|A|}{|\Omega|}.$$

这从数学上非常完美, 与之前所说的完全相容, 例如每个基本事件的概率都是

$$\frac{1}{|\Omega|},$$

不可能事件的概率是0, 必然事件的概率是1.

从定义可以看出, 计算概率的关键是计数, 或者说数元素个数, 不需要知道具体有些什么元素.

显然, 这样定义的概率有可加性, 因为当A,B 互斥时, 并集$A\cup B$ 中元素个数等于他们各自元素个数之和,

$$|A\cup B|=|A|+|B|,$$

这是元素个数的可加性, 推出概率的可加性

$$\mathsf{P}(A\cup B)=\mathsf{P}(A)+\mathsf{P}(B).$$

§4.2 样本空间的选取

对于一个随机试验来说, 样本空间可以有不同的取法, 因为人们对所有可能的结果理解不一.

例1. 掷一枚硬币三次. 如果依次记录其正反面, 则

$$\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\};$$

如果只记录正面出现的次数, 则

$$\Omega = \{0, 1, 2, 3\},$$

其中数字表示正面次数.

第一个样本空间是等可能的, 每个概率都是1/8; 第二个样本空间不是等可能的, 例如正面次数等于0,1,2,3 分别对应于第一个样本空间的子集

$$\{000\}, \{001, 010, 001\}, \{011, 101, 110\}, \{111\},$$

所以概率分别是1/8,3/8,3/8,1/8.

例2. 掷两枚骰子. 每个骰子6 个结果, 所以分别看两个骰子的数字, 有36 个等可能的结果

$$\Omega = \{(i, j) : 1 \le i \le 6, 1 \le j \le 6\}.$$

若只看两个骰子的点数之和, 那么样本空间是

$$\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\},$$

其中的概率依次为1/36,2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36.

样本空间可以按照观察者的关注角度选取, 只要一个随机试验存在一个等可能的样本空间, 我们就认为它是古典概率模型, 而且一般总是就取等可能的样本空间.

§4.3 经典古典概率问题

下面我们来讨论一些经典的概率问题, 目的是加深大家对于随机性及其概率的理解. 解决这些问题基本上只需要直觉, 不需要太多的数学工具, 因此大家在思考这些问题的时候, 首先尽量发挥想象力, 使用直觉; 其次在使用直觉解决问题时, 要能够判别使用了什么隐藏的假设或者工具, 这样才能真正理解随后的概率论严格化数学理论的展开.

§4.3.1 分赌注问题

分赌注问题是法国贵族赌徒de Méré 请教数学家B. Pascal 的, Pascal 写信给Fermat 讨论这个问题. 这个问题催生出概率这个学科, 所以说好的问题是无价的.

PROBLEM: 两个赌徒甲乙各出32 块钱作为赌注, 每赌一局胜者得一分先得3 分者赢得全部赌注. 现在甲得2 分, 乙得1 分时, 甲乙同意终止赌局, 问应该怎么分赌注才是公平的?

下面是Pascal 在给Fermat 的信中解释怎么解决这个问题.

我们假设第一个人得了两分, 另外一个人得了一分, 他们现在再玩一次, 如果第一个人赢, 那么他就拿走全部赌注, 如果另一个人赢了, 那么他们二比二平. 接着, 如果他们愿意终止, 那么他们每人拿一半赌注.

那么, 先生, 考虑到如果第一个人赢, 他取64元, 如果输, 他取32. 这时, 如果他们不想玩这一局, 就此终止, 那么第一个人会说: "对我来说，32 元是保证的，对于另外的32 元，也许可能是你的，也许可能是我的, 机会是相等的. 因此我们应该平分这32 元, 且给我另外的32 元." 那么他就拿48 元, 另一个人拿16 元.

由此我们可以看出, 按照3:1 的比例分配赌注是公平的.

现在我们来看, 如果继续赌局的话, 甲乙最终赢的赌注的概率比是多少? 甲已经有2 分, 乙有1 分, 继续赌局. 甲先得三分有两个途径: 1. 接着的第一局甲胜, 概率为1/2; 2. 接着的第一局乙胜, 但第二局甲胜, 概率为1/4. 因此甲最终赢的概率是3/4, 那么乙最终胜的概率是1/4, 他们两者最终赢得赌注的概率比就是3:1.

也就是说, 赌注按照赢的概率之比例来分配就是公平的.

§4.3.2 掷硬币

PROBLEM: 甲乙各掷n, $n+1$ 枚硬币. 问乙得到的正面数比甲多的概率是多少?

当$n=1$ 时, 总共三枚硬币, 8 个等可能结果,

$$\Omega=\{000,001,010,100,011,101,110,111\},$$

其中第一个符号表示甲的硬币, 后两个符号表示乙的两个硬币. "乙得到的正面数比甲多"这个事件对应于子集$\{001,010,011,111\}$, 其中4 个元素, 所以所求概率是4/8=1/2.

当$n=2$ 时, 总共5 枚硬币, 32 个等可能结果, 每个概率都是1/32,

$$\begin{aligned}\Omega=\{&00000,00001,00010,00011,00100,00101,00110,00111,\\&01000,01001,01010,01011,01100,01101,01110,01111,\\&10000,10001,10010,10011,10100,10101,10110,10111,\\&11000,11001,11010,11011,11100,11101,11110,11111\},\end{aligned}$$

其中前两个是甲的硬币, 后三个是乙的硬币. 怎么列出所有可能结果? 这里是利用二进制表示的方法从小到大列出所有结果. 这个事件对应于子集

$$\begin{aligned}\{&00001,00010,00011,00100,00101,00110,00111,\\&01011,01101,01110,01111,\\&10011,10101,10110,10111,\\&11111\},\end{aligned}$$

其中共有8 个元素, 所以所求概率是8/16=1/2.

因此我们猜测不论n 多大, 概率总是1/2. 但是当$n=3$ 时, 上面那种数个数的方法是行不通的.

用字母A 表示“乙得到的正面数比甲的正面数多”这个事件, 用X,Y 分别表示甲乙的正面个数. 那么

$$A=\{X<Y\}=\bigcup_{i<j}\{X=i,Y=j\}.$$

因此

$$\mathsf{P}(A)=\sum_{i=0}^{n}\sum_{j=i+1}^{n+1}\mathsf{P}(X=i)\mathsf{P}(Y=j),$$

其中用到甲乙两人硬币正面数之间是独立的这个假设.

概率$\mathsf{P}(X=i)$ 是指掷n 个硬币得到i 个正面数的概率. 掷n 个硬币这个随机试验的样本空间有2^n 个等可能个元素, 如上所示, 每个是n 位置放上0 或者1. $X=i$ 这个事件相当于n 位置中恰有i 个1, 因此有$\binom{n}{i}$ 个可能结果, 因此

$$\mathsf{P}(X=i)=\binom{n}{i}/2^n.$$

因此

$$\mathsf{P}(A)=\frac{1}{2^{2n+1}}\sum_{i=0}^{n}\sum_{j=i+1}^{n+1}\binom{n}{i}\binom{n+1}{j}.$$

由二项式定理,

$$\frac{1}{2^{2n+1}}\sum_{i=0}^{n}\sum_{j=0}^{n+1}\binom{n}{i}\binom{n+1}{j}=1.$$

现在我们要说其中组成$\mathsf{P}(A)$ 的与剩下的和一样

$$\sum_{i=0}^{n}\sum_{j=i}^{n+1}\binom{n}{i}\binom{n+1}{j}=\sum_{i=0}^{n}\sum_{j=0}^{i}\binom{n}{i}\binom{n+1}{j}.$$

这由组合的对称性$\binom{n}{i}=\binom{n}{n-i}$ 推出. 为什么呢? 看下面的矩阵

$$\begin{array}{ccccccc}
\binom{n}{0}\binom{n+1}{0} & \diagdown & \binom{n}{0}\binom{n+1}{1} & \binom{n}{0}\binom{n+1}{2} & \cdots & \cdots & \binom{n}{0}\binom{n+1}{n+1}\\
\binom{n}{1}\binom{n+1}{0} & \binom{n}{1}\binom{n+1}{1} & \diagdown & \binom{n}{1}\binom{n+1}{2} & \cdots & \cdots & \binom{n}{1}\binom{n+1}{n+1}\\
\binom{n}{2}\binom{n+1}{0} & \binom{n}{2}\binom{n+1}{1} & \binom{n}{2}\binom{n+1}{2} & \diagdown & \cdots & \cdots & \binom{n}{2}\binom{n+1}{n+1}\\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\
\binom{n}{n}\binom{n+1}{0} & \binom{n}{n}\binom{n+1}{1} & \binom{n}{n}\binom{n+1}{2} & \cdots & \cdots & \diagdown & \binom{n}{n}\binom{n+1}{n+1}
\end{array}$$

$\mathsf{P}(A)$ 中的组合数是矩阵中斜线上部的和, 即$i<j$ 的所有$\binom{n}{i}\binom{n+1}{j}$, 由组合公式

$$\binom{n}{i}\binom{n+1}{j}=\binom{n}{n-i}\binom{n+1}{n+1-j}.$$

因此$\{(i,j):0\le i\le n, i+1\le j\le n+1\}$ 与$\{(i,j):0\le i\le n, 0\le j\le i\}$ 通过映射$(i,j)\mapsto(n-i,n+1-j)$ 是一一对应的, 因此上面等式两边是一样. 推出$\mathsf{P}(A)=1/2$.

这个方法计算太复杂, 不够直观, 下面我们给一个直观的方法. 用字母B 表示“乙得到的反面数比甲的反面数多”这个事件. 因为正反面对称, 所以$\mathsf{P}(A)=\mathsf{P}(B)$. 下面我们证明$A$,$B$ 是对立事件, 即至少有一个发生, 但不会同时发生.

如果它们都不发生, 即乙得到的正面数不比甲的正面数多, 且乙得到的反面数也不比甲的反面数多, 那么乙的硬币总数不比甲的总数多, 矛盾.

如果它们同时发生, 即乙得到的正面比甲的正面数多, 且乙得到的反面数也比甲的多, 那么乙的硬币数至少比甲的硬币数要多两个, 矛盾.

所以$\mathsf{P}(A)+\mathsf{P}(B)=1$, 推出$\mathsf{P}(A)=1/2$.

§4.3.3 生日问题

PROBLEM: 设有n 个人, 问至少有两个人生日相同的概率是多少?

把 n 个球随机地放入 N 个盒子中, 每个球等可能地放入任何一个盒子. 那么每个球有 N 种放法, 因此样本空间 Ω 共有 N^n 种可能的放法, 每一种放法是等可能的. 现在设 $n\le N$, 考虑事件: 每个盒子至多只有一个球. 我们只需计

算这个事件有多少基本事件就可以了. 由乘法原理知道, 它应有

$$N(N-1)\cdots(N-n+1)$$

个, 即 N 个数中取 n 个的排列 $(N)_n$. 因此每个盒子至多只有一个球的概率为 $(N)_n/N^n$.

一年是 365 日, 一个人的生日假设是从 365 日中随机选一日, 不同人的生日也可以假设是独立的, 这样n 个人生日各不相同的概率为 $(365)_n/365^n$, 因此至少有两人生日相同的概率为

$$p_n = 1 - \frac{(365)_n}{365^n}.$$

比如

$$\begin{array}{cccccccccc} n = & 15, & 20, & 25, & 30, & 35, & 40, & 45, & 50, & 55, \\ p_n = & 0.25, & 0.41, & 0.57, & 0.71, & 0.81, & 0.89, & 0.94, & 0.97, & 0.99, \end{array}$$

其中几个数字颇令人惊讶. 例如55 个人就有99% 的概率有两人生日一样.

§4.3.4 抽签与顺序

抽签是生活中经常用的一个工具, 例如小时候生产队分桃子, 一堆一堆的桃子, 个数重量差不多, 但是各人对哪堆好意见不同, 那通常就用抽签方法解决, 这样公平. 一般来说, 抽签是依次一个个抽, 那么我们自然会问抽到好签的机会是否是顺序有关. 在做这个问题之前, 先想一想你直觉是什么.

PROBLEM: 假设是4 个签, 2 个写字母A, 2 个写字母B. 5 个人顺序不放回抽签, 问抽到1 字的机会是否与顺序有关?

把4 根签编号为1,2,3,4, 其中1,2 写字母A, 3,4 写字母B. 按照4 个人顺序不放回抽签得到的数字写样本空间, 应该是1,2,3,4 的所有排列

$$\begin{aligned}\Omega = \{&1234, 2134, 3214, 3124, 1324, 2314,\\ &1243, 2143, 3241, 3142, 1342, 2341,\\ &1423, 2413, 3421, 3412, 1432, 2431,\\ &4123, 4213, 4321, 4312, 4132, 4231\}.\end{aligned}$$

A_i 表示第i 个抽签者抽到A 这个事件. 那么A_i 就是第i 个位置是1 或者2 的那些基本事件, 所以

$$\begin{aligned}A_1 = \{&1243, 2134, 1324, 2314, 1243, 2143,\\ &1342, 2341, 1423, 2413, 1432, 2431\}\\ A_2 = \{&1234, 2134, 3214, 3124, 1243, 2143,\\ &3241, 3142, 4123, 4213, 4123, 4231\}\\ A_3 = \{&3214, 3124, 1324, 2314, 1423, 2413,\\ &3421, 3412, 4123, 4213, 4321, 4312\}\\ A_4 = \{&3241, 3142, 1342, 2341, 3421, 3412,\\ &1432, 2431, 4321, 4312, 4232, 4231\}.\end{aligned}$$

每个事件的元素都是12 个, 所以

$$\mathsf{P}(A_1) = \mathsf{P}(A_2) = \mathsf{P}(A_3) = \mathsf{P}(A_4) = \frac{1}{2}.$$

因此抽签得中的机会与顺序无关, 这其实也是经验或者直觉.

§4.3.5 配对问题

PROBLEM: n 对夫妇参加舞会, 舞会将n 位男女随机地配成n 对舞伴, 问没有一对夫妇配成舞伴的概率是多少?
用A 来表示问题中的这个事件, 当$n=1$ 时, $\mathsf{P}(A)=0$, $n=2$ 时, $\mathsf{P}(A)=1/2$, $n=3$ 时, $\mathsf{P}(A)=1/3$. 对一般的n, 用A_i 表示第i 对夫妇配对了这个事件. 那么

$$A = A_1^c \cap A_2^c \cap \cdots \cap A_n^c = \left(\bigcup_{i=1}^n A_i\right)^c.$$

因此我们只需要算

$$\mathsf{P}\left(\bigcup_{i=1}^n A_i\right).$$

这些事件可能会同时发生, 所以不能直接用可加性. 容斥定理是解决这样问题的一个方法.
容斥定理:

$$\begin{aligned}\mathsf{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n \mathsf{P}(A_i) \\ &- \sum_{i<j} \mathsf{P}(A_i \cap A_j) \\ &+ \sum_{i<j<k} \mathsf{P}(A_i \cap A_j \cap A_k) \\ &- \cdots \\ &+ (-1)^{n-1}\mathsf{P}(A_1 \cap A_2 \cap \cdots \cap A_n),\end{aligned}$$

其中等号右边的第m 项的符号是$(-1)^{m-1}$, 和是从n 个事件取m 个事件的所有组合的交的概率的和.

来看$n=2$ 时, 我们要证明$\mathsf{P}(A_1\cup A_2)=\mathsf{P}(A_1)+\mathsf{P}(A_2)-\mathsf{P}(A_1\cap A_2)$. 事实上, $A_1\cup A_2=A_1\cup(A_2\setminus A_1)$, 可加性推出

$$\mathsf{P}(A_1\cup A_2)=\mathsf{P}(A_1)+\mathsf{P}(A_2\setminus A_1).$$

然而, $A_2=(A_2\setminus A_1)\cup(A_2\cap A_1)$, 所以

$$\mathsf{P}(A_2\setminus A_1)=\mathsf{P}(A_2)-\mathsf{P}(A_1\cap A_2).$$

因此结论成立. 一般的情况可以应用数学归纳法.

容斥定理实际上是说怎么数多个集合并的元素个数. 例如两个集合的并的元素个数, 先数各自集合元素个数加起来, 那么相交部分其实数了两次, 所以要减去相交部分的元素个数.

现在我们可以来解决配对问题了. 配舞伴可以看成以下过程: 男士依数字顺序站好, 女士随机排列, 这样样本空间可以看作是$1,2,\cdots,n$ 的全排列, 所以$|\Omega|=n!$. A_i 表示i 对夫妇配对, 其他任意排列, 这样$|A_1|=(n-1)!$. 因此$\mathsf{P}(A_i)=1/n$. 同理

$$\mathsf{P}(A_i\cap A_j)=\frac{1}{n(n-1)}.$$

当$1<i_1<\cdots<i_m\le n$ 时,

$$\mathsf{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_m})=\frac{1}{n(n-1)\cdots(n-m+1)}.$$

应用容斥定理,

$$\begin{aligned}\mathsf{P}\left(\bigcup_{i=1}^{n}A_i\right)&=n\cdot\frac{1}{n}-\binom{n}{2}\frac{1}{n(n-1)}+\cdots+(-1)^{n-1}\frac{1}{n!}\\&=1-\frac{1}{2!}+\frac{1}{3!}-\cdots+\frac{(-1)^{n-1}}{n!}.\end{aligned}$$

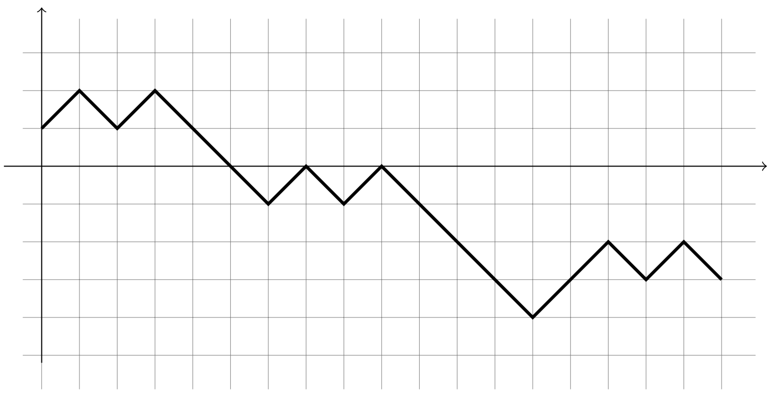
推出

$$\mathsf{P}(A) = 1 - \mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{m=0}^{n} \frac{(-1)^m}{m!}.$$

当n 趋于无穷时, 没有配对的概率的极限是e^{-1}.

§4.4 反射原理

平面上的格点是指坐标都是整数的点, 把横轴看成时间轴, 纵轴看成状态轴. 一个格点轨道是指在格点上向右方移动, 每次向右上或者右下移动一格, 的轨道, 即在(i, j) 处的点等可能地移到$(i+1, j+1)$ 或者$(i+1, j-1)$.



格点轨道对应于简单对称的随机游动, 设甲乙两个人掷一枚硬币进行赌博, 正面甲赢一元, 反面乙赢一元. 用$x(n)$ 表示甲在时刻n 的钱数, 那么$\{(n, x(n)) : n \geq 0\}$ 是个格点轨道. 上图是掷硬币产生的一个轨道.

研究这样的随机游动与研究格点轨道是一样的. 固定起点, 时间长度为n 的格点轨道一共有2^n 条, 它们是等可能的.

PROBLEM: 从点(i, j) 出发, 是否可以到达点(m, n)? 如果可以到达, 总共有多少条?

从点(i, j) 到(m, n), 时间跨度为$m - i$, 假设在这段时间掷了r

个正面, s 个反面, 那么(i,j) 可达(m,n) 当且仅当存在非负整数r,s 使得$r+s=m-i$, 且$r-s=n-j$, 即

$$\begin{cases} 2r=(m-i)+(n-j); \\ 2s=(m-i)-(n-j). \end{cases}$$

因此方程有解的充分必要条件是$(m-i)$ 与$(n-j)$ 有相同的奇偶性且

$$|n-j|\le m-i.$$

当上述条件满足时, (i,j) 到达(m,n) 相当于掷$m-i$ 次硬币得

$$[(m-i)-(n-j)]/2$$

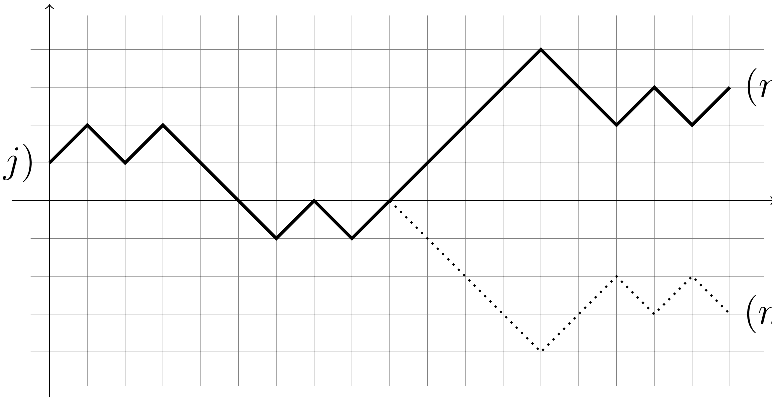
次正面, 用组合原理, 共有

$$\binom{m-i}{[(m-i)-(n-j)]/2}$$

种方法, 因此到达的轨道也是这么多.

PROBLEM: 设$j>0, n>0$, 问从(i,j) 到(m,n) 的轨道有多少是碰到时间轴的? 即轨道上有一个点的纵坐标是零.

用A 表示这样碰到时间轴的轨道全体, B 表示从(i,j) 到$(m,-n)$ (它是(m,n) 关于时间轴的对称点)的轨道全体. 建立A 到B 的一个变换: 任取A 的一条轨道, 把其上最后一个纵坐标等于0 的点之后的轨道关于时间轴反射下来, 之前保持不动, 得到新的轨道, 这条轨道的终点是$(m,-n)$, 所以属于B. 见下图



这个变换是A 到B 的一一对应(请读者证明之), 即单(不同的轨道的像不同)且满(B 中轨道必定是A 中轨道反射来的). 因此A 中的轨道总数为

$$|A| = |B| = \binom{m-i}{[(m-i)+n+j]/2}.$$

这个结论称为反射原理. 它的主要原因是轨道的对称性.

§4.4.1 计票问题

PROBLEM: 总数为$m+n$ 的投票人依次投票给甲乙两个候选人, 已知甲乙的票数分别为m,n 且$n>m$, 问在投票过程中乙一直领先的概率是多少? 这里一直领先是指乙票数一直多于甲.

这是个经典的问题, 相当于求$(0,0)$ 到达$(m+n,n-m)$ 的格点轨道中不碰到时间轴(起点除外)的格点轨道比例.

首先(0,0 到$(m+n,n-m)$ 的格点轨道总数是$\binom{n+m}{m}$.

其次, $(0,0)$ 到$(m+n,n-m)$ 的格点轨道中不碰到时间轴(起点除外)的格点轨道数, 等于$(1,1)$ 到$(m+n,n-m)$ 的格点轨道中不碰到时间轴的格点轨道数, 等于$(1,1)$ 到$(m+n,n-m)$ 的格点轨道总数, $\binom{m+n-1}{n-1}$, 减去其中碰到时间轴的轨道数.

后者可以用反射原理计算.

由反射原理, $(1,1)$ 到$(m+n,n-m)$ 碰到时间轴的格点轨道数, 等于$(1,1)$ 到$(m+n,m-n)$ 的格点轨道总数, 即等于$\binom{m+n-1}{m-1}$.

因此所求概率为

$$\frac{\binom{m+n-1}{n-1}-\binom{m+n-1}{m-1}}{\binom{n+m}{m}}=\frac{n-m}{n+m}.$$

§4.4.2 首次回归时间

PROBLEM: 从$(0,0)$ 到$(2n,0)$ 的轨道中途中不碰到时间轴的概率是多少? 注意$(0,0)$ 可达$(n,0)$ 的条件是n 是偶数. 或者问从0 出发的随机游动在$2n$ 时刻首次回到0 的概率是多少?

用A_n 表示这个事件. 因为要求途中不碰到时间轴, 所以轨道整体在时间轴上方或者下方, 数量是一样的. 那么我们来算位于时间轴上方的轨道集A_n^+, 有多大. 显然$(0,0)$ 到$(2n,0)$ 且途中在时间轴上的轨道数, 等于$(1,1)$ 到$(2n-1,1)$ 不碰到时间轴的轨道数, 对于总数减去碰到时间轴的轨道数. 由反射原理, 其中碰到时间轴的轨道数, 等于$(1,1)$ 到$(2n-1,-1)$ 的轨道数. 因此

$$|A_n^+|=\binom{2n-2}{n-1}-\binom{2n-2}{n-2}=\frac{1}{2(2n-1)}\binom{2n}{n}.$$

从而概率为

$$\mathsf{P}(A_n)=\frac{|A_n|}{2^{2n}}=\frac{1}{2n-1}\binom{2n}{n}\frac{1}{2^{2n}},$$

其中$\binom{2n}{n}2^{-2n}$ 是0 出发在$2n$ 时刻回到0 的概率.

巧妙地,

$$\frac{1}{2n-1}=\frac{2n}{2n-1}-1=\frac{2n\cdot 2n}{2n(2n-1)}-1.$$

这样

$$\begin{aligned}\mathsf{P}(A_n)&=\left(\frac{2n\cdot 2n}{2n(2n-1)}-1\right)\binom{2n}{n}2^{-2n}\\&=\binom{2(n-1)}{n-1}2^{-2(n-1)}-\binom{2n}{n}2^{-2n}.\end{aligned}$$

假设从0 出发, $x(0)=0$, 定义首次回归时间

$$T=\inf\{n\geq 1:x(n)=0\}.$$

由此

$$\begin{aligned}\mathsf{P}(T<\infty)&=\mathsf{P}(\bigcup_{n=1}^{\infty}A_n)\\&=\sum_{n=1}^{\infty}\mathsf{P}(A_n)\\&=1-\lim_{n}\binom{2n}{n}2^{-2n}=1,\end{aligned}$$

这说明以概率1, 随机游动在有限时间会回到0.

§4.5 几何概率

前面所说的古典概率模型是指仅有有限多个等可能发生的结果的随机试验, 还有一种直观的古典概率模型称为几何概率模型. 例如从$[0,1]$ 区间随机地取一个点, 记为X, 那么X有无限的可能, 但是X 落在$[1/2,2/3]$ 中的概率应该是此区间的长度$1/6$. 这是因为当我们说在区间内随机地取一个点

时, 也就意味着“均匀”或者“等可能”地取得每一个点. 可是点有无穷多，取得具体每个点的可能性必须是零，所以描述这种等可能的方法应该是“点落在区间I 中的概率与区间的长度成比例”, 而落在整个区间上的概率是1, 所以比例系数必须是整个区间长度的倒数.
设Ω 是空间(可以是1 维, 2 维, 也可以是高维)的一个有界区域, 从Ω 中随机取一个点, 记为X, 那么X 落在Ω 中的一个区域A 的概率为

$$\mathsf{P}(X \in A) = \frac{|A|}{|\Omega|},$$

其中$|\cdot|$ 表示区域的体积, 注意在3 维及以上的空间时, 称为体积, 在2 维空间时, 成为面积, 在1 维空间时, 称为长度. 在解决前面的古典概率问题时, 方法是数元素个数, 在解决几何概率问题时, 方法是计算长度, 面积, 或者体积.

§4.5.1 坐车问题

PROBLEM: 一个人经常加班, 下班时间比较随机, 下班后到地铁站坐地铁, 这里有两趟地铁, 一趟去自己父母家, 另一趟去岳父母家, 每趟地铁都是20 分钟一趟, 为了公平, 他决定哪趟先到就坐那趟, 但是几个月之后发现他还是去岳父母家更多一些, 为什么?
这个问题留给读者思考.

§4.5.2 约会问题

PROBLEM: 两人约在8:00-9:00 这个时间段在某咖啡店见面, 并约定先到之人最多等15 分钟, 问他们能够见面的概率是多少?

首先要把问题化成在某个区域中随机地取一个点. 用 x, y 分别表示甲乙两人抵达的时间, 由于随机性, 两人抵达的时间相当于在平面区域

$$\Omega = \{(x, y) : 8 \leq x, y \leq 9\}$$

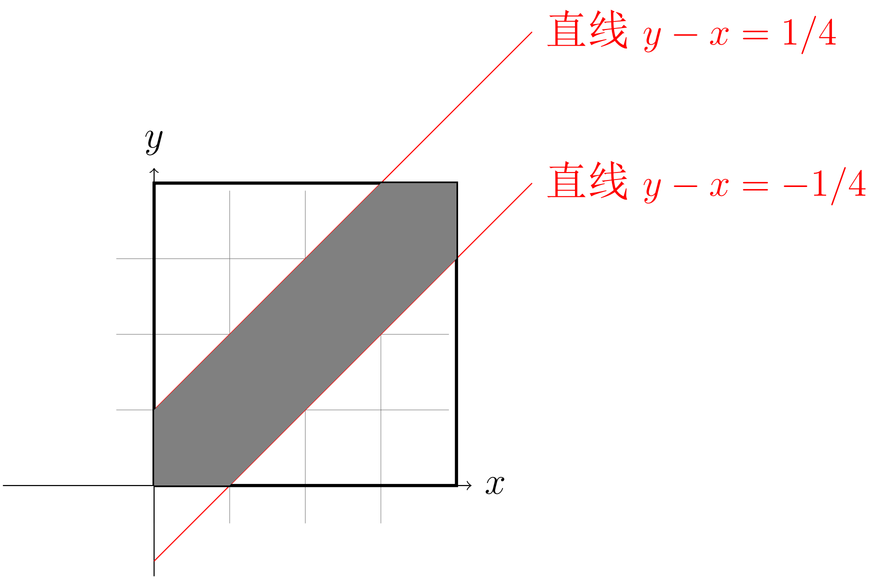
中随机地取一个点, 这等价于假设,

1. 他们两人在约定时间段的任何时间抵达的可能性都是一样的, 也就是等可能性;
2. 他们到达的事件是互相独立的.

而两人能见面这个事件相当于这个点落在区域

$$A = \{(x, y) \in \Omega : |x - y| \leq \frac{1}{4}\}$$

内. 它的概率就等于这个集合的面积与整个样本空间的面积的比, 即 7/16.

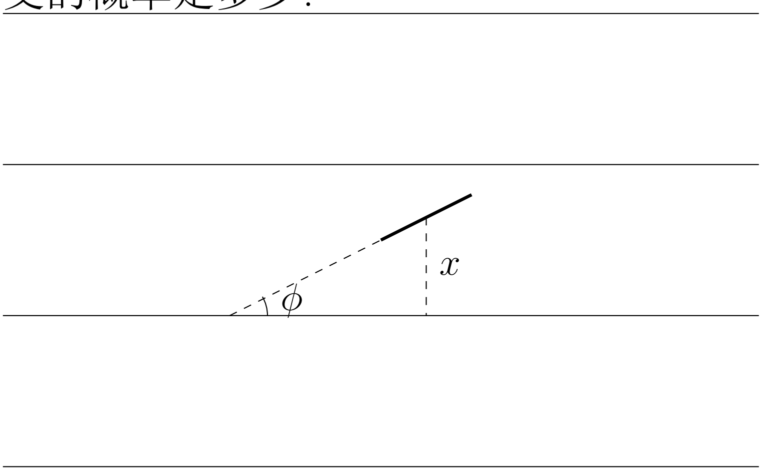


习题: 一个人试图扔一个一元硬币 (直径 3cm) 到 2 米外的一个边长为 4cm 的正方形台面上, 现在他的硬币已经落在台面上, 问硬币完全在台面里的概率是多少?

§4.5.3　Buffon 问题

下面的问题称为Buffon 投针问题.

PROBLEM: 向一个画着等距离平行线的平面上投针, 平行线间的距离为 l, 针的长度为 a, $l > a$, 问此针与平行线相交的概率是多少?



设针的中点与最近的平行线的距离是 x, 针 (或其延长线) 与平行线的夹角为 ϕ, 我们假设 (x, ϕ) 是在区域

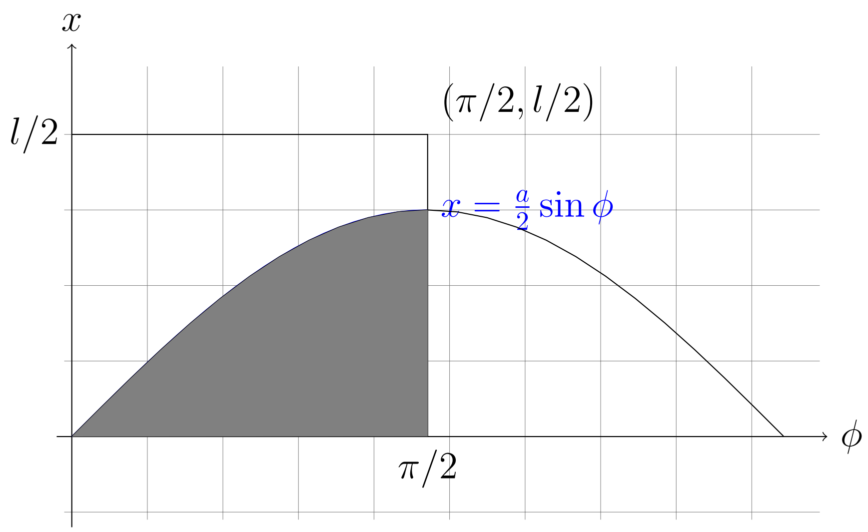
$$[0, \frac{l}{2}] \times [0, \frac{\pi}{2}]$$

上随机取的一个点, 这等于假设

1. x, ϕ 在各自范围内是等可能的;
2. x 与ϕ 是独立的.

事件 A: 针与平行线相交发生当且仅当 (x, ϕ) 落在区域 $\{(x, \phi) : x \leq \frac{a}{2} \sin \phi\}$ 因此概率是面积的比

$$\mathbb{P}(A) = \frac{\int_0^{\frac{\pi}{2}} \frac{a}{2} \sin \phi \mathrm{d}\phi}{\frac{l}{2} \cdot \frac{\pi}{2}} = \frac{2a}{\pi l}.$$

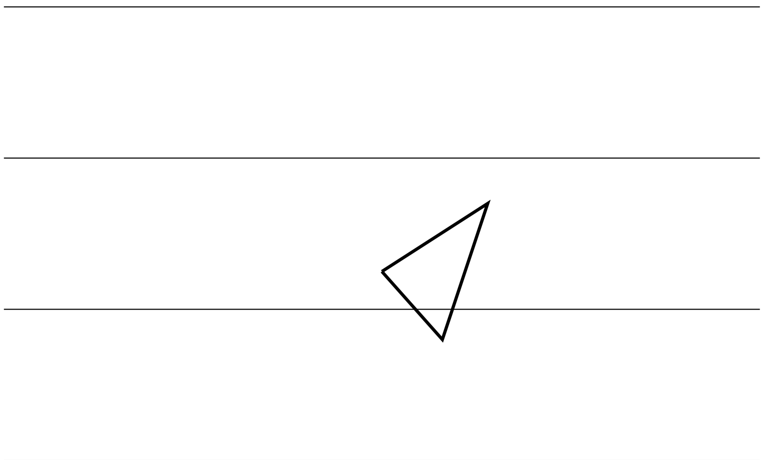


在这个解答中，我们要假设角度ϕ 和距离x 是独立的，这其实不是那么容易让人信服. 一般地，我们可以对两个看上去没有什么关系的随机试验假设独立，但这里，这两个量来自同一个随机试验，独立性不是那么显然.

Buffon 问题还有一个漂亮的解答，不必使用独立性假设，简述如下. 设 $f(x)$ 是长为 x 的针与平行线相交的概率，利用概率的可加性推出

$$f(x+y) = f(x) + f(y).$$

因此存在常数 k 使得对 $x \le l$ 有 $f(x) = kx$，下面我们要计算 k.



先看一个三角形abc 扔在平面上与平行线相交的概率, 其中 a, b, c 是三角形三条边及其长度. 用A, B, C 分别表示a, b, c 与平行线相交的这个事件, 那么$A \cup B \cup C$ 就表示三角形与平行线相交这个事件. 由容斥定理,

$$\begin{aligned}\mathsf{P}(A \cup B \cup C) =& \mathsf{P}(A) + \mathsf{P}(B) + \mathsf{P}(C) \\ &- \mathsf{P}(A \cap B) - \mathsf{P}(B \cap C) - \mathsf{P}(C \cap A) \\ &+ \mathsf{P}(A \cap B \cap C),\end{aligned}$$

事实上, 一个三角形与平行线相交当且仅当三角形的两条边和平行线相交, 所以$\mathsf{P}(A \cap B \cap C) = 0$, 而

$$\begin{aligned}\mathsf{P}(A) &= \mathsf{P}(A \cap B) + \mathsf{P}(A \cap C) \\ \mathsf{P}(B) &= \mathsf{P}(B \cap A) + \mathsf{P}(B \cap C) \\ \mathsf{P}(C) &= \mathsf{P}(C \cap A) + \mathsf{P}(C \cap B).\end{aligned}$$

由此推出三角形与平行线相交的概率为

$$\begin{aligned}\mathsf{P}(A \cup B \cup C) &= \frac{1}{2}(\mathsf{P}(A) + \mathsf{P}(B) + \mathsf{P}(C)) \\ &= \frac{1}{2}(f(a) + f(b) + f(c))\end{aligned}$$

$$= \frac{1}{2}k(a+b+c),$$

因此这个概率是周长的 $k/2$ 倍.
同理, 任何一个直径不超过 l 的凸多边形随机扔在平面上与平行线相交的概率也是周长的 $k/2$ 倍, 取极限得一个直径不超过 l 的凸图形与平行线相交的概率同样如此. 而一个直径为 l 的圆扔在平面上必然与平行线相交, 即有

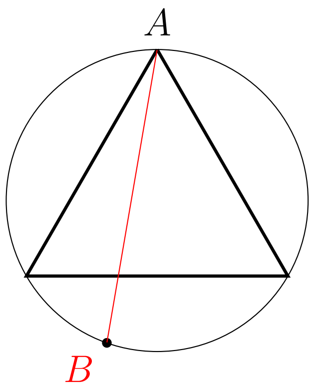
$$k/2 = \frac{1}{l\pi}.$$

注意一个图形的直径是指其中最远两个点的距离.

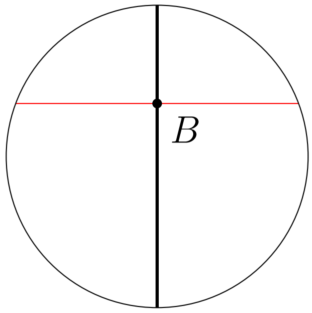
§4.5.4 Bertrand 问题

PROBLEM: 在一个圆周上随机地任取一根弦, 问其长度大于圆的内接等边三角形边长的概率是多少?
这个问题在历史上曾称为 Bertrand 悖论, 因为问题有不同的答案. 但随着理论的完善, 人们认识到导致不同答案的原因是问题中随机性的含义不清楚, 或者说因为对随机性有不同的理解. 所以后来称之为 Bertrand 奇论, 它是帮助理解概率论中随机性的很好的例子.
在这个问题里, 随机性至少有三种理解:

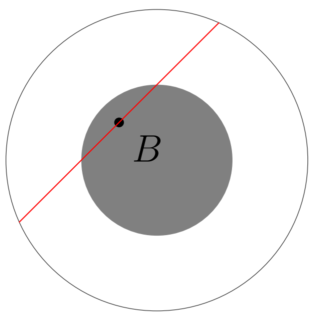
(1) 先在圆周上取定一点 A, 然后再在圆周上随机地取一个点 B, 连接 A 与 B 成为弦;



(2) 先取定一条直径, 然后在直径上随机地取一个点 B 作一条过此点与直径垂直的弦;



(3) 以圆内的任何点作为中点的弦是唯一决定的, 因此以这个对应, 随机地取一条弦就等同于随机地在圆内取一个点 B.



说这三种随机性不同是说如果按 (1) 随机地取弦 AB, 那么它与圆心的距离就不可能是在半径上等可能的, 它的中点也不会等可能地出现在圆内. 关于此, 在后面会有进一步解

释.

一条弦长大于圆的内接等边三角形边长当且仅当弦的中点与圆心的距离小于 $\frac{1}{2}$ 的半径长, 那么在 (1) 的情况下, 点 B 必须落在点 A 对面的 $\frac{1}{3}$ 圆弧上, 因此概率为 $\frac{1}{3}$, 在 (2) 的情况下, 点 B 必须与圆心距离不小于 $\frac{1}{2}$ 半径长, 因此概率为 $\frac{1}{2}$, 在 (3) 的情况下, 点 B 必须落在半径为原来圆半径长的 $\frac{1}{2}$ 的圆内, 因此概率为两圆面积的比 $\frac{1}{4}$.

§4.6 无穷和

从上面所涉及的概率计算可以看出, 可加性与独立性是非常重要的两个手段. 下面我们问一个简单的问题: 重复不断地掷一枚硬币, 我们是不是一定可以在有限次得到正面? 问题似乎真的很简单, 掷一枚硬币, 一般来说, 不需要几次, 我们就可以看到正面的, 但是我们能保证几次就一定能看到正面吗? 10 次可以吗? 100 次可以吗? 什么叫一定? 理论上可以证明吗?

用X 表示第一次得到正面时所掷的次数, 例如$X = n$ 表示第n 次掷得正面, 而且是第一次看到正面. 那么前$n-1$ 次都是反面, 而第n 次是正面, 则

$$\mathsf{P}(X = n) = \frac{1}{2^n}.$$

再例如$X > n$ 表示前n 次掷得的都是反面, 因此$\mathsf{P}(X > n) = 1/2^n$. 不论$n$ 多大, 前n 次掷得反面的概率都不是零, 虽然可能非常非常小. 所以理论上讲, 没有人能保证多少次就

一定可以得到正面, 也就是说从概率角度讲, X 不是有界的. 那有限次内看到正面吗? 有限次内看到正面是事件$X < \infty$. 怎么计算这个事件的概率呢? X 是有限的, 即$X=1$, 或者$X=2$, 或者$X=3$, 一直下去, 没有尽头. 数学上

$$\{X<\infty\}=\{X=1\}\cup\{X=2\}\cup\{X=3\}\cup\cdots\cdots,$$

右边有无穷个互斥的事件. 上面的可加性可能会自然地驱使很多读者在计算左边的概率时把右边各事件的概率加起来.

慢, 这超越了前面所说的可加性了, 那个可加性只是对有限多个事件成立, 称为有限可加性. 有限可加性不能自动地推广到无限的情况.

如果我们停止在有限可加性, 那么上面这个概率就无法计算了. 要再进一步, 我们必须要承认这样的无限可加性, 也称为可列可加性. 这样得

$$\begin{aligned}\mathsf{P}(X<\infty)&=\mathsf{P}(X=1)+\mathsf{P}(X=2)+\mathsf{P}(X=3)+\cdots\\&=\frac{1}{2}+\frac{1}{2^2}+\frac{1}{2^3}+\cdots=1,\end{aligned}$$

理论上证明了我们100%(概率1)可以肯定有限次内可以看到正面出现, 100% 可以肯定也称为几乎可以肯定. 反过来, $\mathsf{P}(X=\infty)=0$, 也就是说有限次内看不到正面的概率是零

不可能事件是空集, 概率一定是零, 反之, 概率零事件不一定是空集, 不一定是不可能事件, 从概率来看不可能, 不是绝对不可能. 也称为几乎不可能.

§4.6.1 三人比赛

PROBLEM: A, B, C 三人下棋, 规则如下: 两人下, 赢者与第三人下, 一直到其中一人连赢两局为胜. 三人的水平相当, A 与 B 先下. 问各人最终胜的概率是多少?
用 ACBACBB 表示这样的一次比赛: A 赢, C 赢, B 赢, A 赢, C 赢, B 连赢. 最终的结果是 B 胜, 共 7 局, 这样的结果出现的概率应该是 $\frac{1}{2^7}$. 那么样本空间 Ω 是这样的排列全体, 是一个无限集.
A 最终胜出这个事件包含样本空间里如下的序列:

(1) 首局 A 赢: AA, ACBAA, ACBACBAA, ACBACBACBAA, $\cdots$, 它们发生的概率依次为 $\frac{1}{2^2}, \frac{1}{2^5}, \frac{1}{2^8}, \cdots$;

(2) 首局 A 输: BCAA, BCABCAA, BCABCABCAA, $\cdots$, 它们发生的概率依次为 $\frac{1}{2^4}, \frac{1}{2^7}, \frac{1}{2^{10}}, \cdots$.

因此 A 最终胜出的概率为

$$\frac{\frac{1}{2^2}}{1-\frac{1}{2^3}} + \frac{\frac{1}{2^4}}{1-\frac{1}{2^3}} = \frac{5}{14}.$$

由对称性知 B 胜出的概率与 A 胜出的概率是一样的, 而类似的计算得 C 胜出的概率为 $\frac{4}{14}$, 三者概率和是 1, 也就是说比赛终有一个人会赢.
习题: 从标准的 52 张扑克牌中取5张牌, 求下列情况各有多少不同选择:

1. 至少有一个对子; —item full house(三张数字一样再加一个对子);

2. 正好是个顺子;

3. 正好是个同花顺子;

4. 四种花色都出现.

习题: 设有 m 个持 50 元钱的人和 n 个持 100 元钱的人在一个窗口排队买票, $m \geq n$, 票价是 50 元且窗口开始没有零钱. 求所有买票的人都不需要等待找钱的概率.

第五章　简单概率概念

§5.1　概率空间

在继续下去之前, 我们需要把概率的基本概念梳理一下, 如概率空间, 条件概率, 独立性, 随机变量, 分布, 期望, 方差, 等等, 还有几个基本的公式, 如全概率公式, Bayes 公式等.
一个随机试验, 有样本空间Ω, 一个事件A 是指Ω 的一个子集, 事件发生的概率是一个0,1 之间的数字, 用$\mathsf{P}(A)$ 表示. 用花写的字母$\mathscr{F}$ 表示事件的全体, 我们说概率是事件上的函数, 满足

1. 可列可加性: 如果事件$A_1, A_2, \cdots, A_n, \cdots$ 两两互斥, 则
$$\mathsf{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n\geq 1} \mathsf{P}(A_n),$$
2. 正规性$\mathsf{P}(\Omega) = 1$, 必然事件的概率是1.

这是Kolmogorov 建立的概率公理体系, 三元组$(\Omega, \mathscr{F}, \mathsf{P})$ 被称为概率空间.
从上面的定义可以推出

1. 不可能事件的概率是0: $\mathsf{P}(\emptyset) = 0$.
2. 可加性: 如果事件A, B 互斥, 则
$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B).$$
3. 单调性: 如果事件$A \subset B$, 那么$\mathsf{P}(A) \leq \mathsf{P}(B)$.

Kolmogorov 的体系像是一个壳, 壳外面附有一套完整而强大的理论, 只要你的概率模型能套入这个壳, 那么就可以使用这一套强大的理论. 前面所述的古典概率模型和几何概率模型可以套入这个壳.

Kolmogorov 的体系对概率来说是一个很好的壳, 能够很好地符合概率的直观, 但不能说不符合Kolmogorov 这个体系的就不是概率, 只能说不是Kolmogorov 意义下的概率.

PROBLEM: 在Kolmogorov 的意义下, 是否可以做到等可能地取一个自然数?

§5.2 随机变量与分布

前面看到, 在随机试验中, 我们经常用数值来表示结果, 即考虑样本空间上的函数, 这样的函数称为随机变量(注意: 随机变量是个函数), 例如掷硬币, 正面用1 表示, 反面用0 表示, 就是随机变量; 实际上, 给定事件A, 我们总可以定义一个随机变量, 记为1_A, 它像一个小旗, 在A 发生的时候往上挥舞, 等于1, 否则往下挥舞, 等于0. 这样的随机变量称为Bernoullli 随机变量, 也称为事件A 的指标. 这样的随机变量是最基本的, 其它随机变量是这样的随机变量的线性组合之后取极限. 掷n 个硬币, 正面个数也是随机变量.

取有限多个值, 例如掷n 个硬币的正面个数, 或者可列多个值, 例如等待成功出现的时间, 称为简单或者离散型随机变量. 把取值的概率写出来就是随机变量的分布. 我们只需要把所有取值的概率写出来就可以了, 例如掷n 个硬币得

到正面的个数为X, 则

$$\mathsf{P}(X=i)=\binom{n}{i}/2^n,\ 0\le i\le n.$$

分布的表示方法很多，简单的情况可用图或者表来表示, 这比较直观易懂,第一排表示随机变量的取值，下面对应的数表示取此值的概率, 例如, 掷3 个硬币, 正面次数的分布

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1/8 & 3/8 & 3/8 & 1/8 \end{pmatrix}.$$

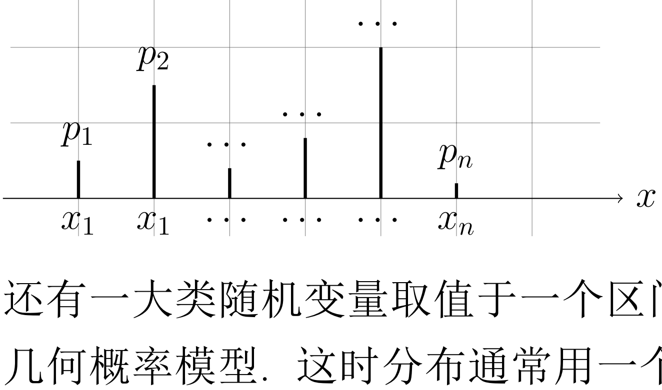
一般地, 一个如下形式的图表被称为是一个分布：

$$\begin{pmatrix} x_1 & x_2 & \cdots & \cdots & x_n \\ p_1 & p_2 & \cdots & \cdots & p_n \end{pmatrix},$$

其中 x_1、x_2、$\cdots$、x_n 是实数，p_1、p_2、$\cdots$、p_n 是非负数, 作为概率值, [1]其总和为 1，即成立

$$p_1+p_2+\cdots+p_n=1.$$

所以分布表示总数为1 的量怎么分布在一些点上.



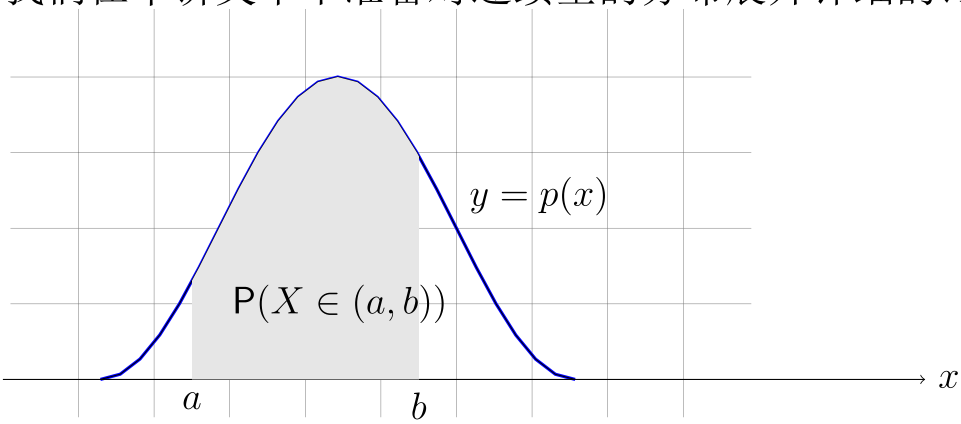
还有一大类随机变量取值于一个区间或者一个区域, 例如几何概率模型. 这时分布通常用一个密度函数来表示. 所谓密度函数是一个非负函数$p(x)$, 介于该函数与横轴之间的

[1]一般要求是正数，因为 0 概率值在分布中所在的这一列总可以删去.

区域面积等于1. 它可以用来描述分布, 随机变量X 落在区间(a,b) 的概率为密度函数在(a,b) 上的积分

$$\mathsf{P}(X \in (a,b)) = \int_a^b p(x)dx.$$

我们在本讲义中不准备对连续型的分布展开详细的讨论.



通常, 对于一个随机试验来说, 尽管随机变量的结果是不可预知的, 但分布是清楚的. 例如买彩票, 中奖的概率是清晰的, 公开的, 但中奖号码是不可预知的. 这也是随机现象的本质.

分布也是一个生活中常用的词, 人口分布, 收入分布, 年龄分布, 商品品牌的市场占有率分布, 等等. 当然概率其实就是一种分布, 体现随机律怎么分布在样本空间上, 随机变量的分布体现随机变量怎么分布在实数上.

§5.3 独立与乘法法则

独立性在概率论计算中是最重要的两个性质之一, 其实在前面的计算中, 我们实际上已经在使用独立性了. 例如分赌注问题中假设每局赌博是独立的, 在约会问题中假设两

人到达时间是独立的, 在蒲丰问题中假设针到平行线的距离和角度是独立的, 等等. 那些独立性的使用似乎是自然的, 但实际上我们不能这样肆无忌惮地使用独立性, 至少我们需要知道什么是独立性, 什么情况下我们可以使用独立性, 这些使用是不是合理. 否则的话, 数学的严密性就会变成一个笑话.

两个事件A,B 独立, 是指下式成立

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B).$$

如果一个随机试验可以分割为两个随机试验E_1 与E_2, 且它们的结果互相不影响, 则我们就可以假设它们是独立的, 也就是假设试验E_1 中的事件A_1 与E_2 中的事件A_2 是独立的, 因此

$$\mathsf{P}(A_{\cap} A_2) = \mathsf{P}(A_1)\mathsf{P}(A_2),$$

它被称为概率的乘法法则.

计算概率的两个基本法则是前面说过的可加性或者称为加法法则, 与现在所说的乘法法则. 加法法则与乘法非常直观自然, 经常会在不知不觉中使用.

两个随机试验的独立通常是合理假设, 不可以证明, 就如同假设一个硬币是圆形一样. 例如我们直观地知道, 掷一个硬币, 再掷一个硬币, 因为它们不会互相影响, 所以它们可以被合理地认为是独立的, 因此我们假设它们是独立的; 即使是同时掷多个硬币, 也可以假设是独立的, 因为相互的影响可以忽略; 摸一个球, 放回去, 再摸一个球, 可以说是独立的; 摸一个球, 不放回, 再摸一个球, 应该不独立, 因为互相

有影响, 第一次摸到的球不会被第二次摸到了.

在一个不可分割的随机试验中的事件是否独立是很难直观判断的, 例如从一副牌(52张, 没有大小王)中摸一张牌, A 是摸得红桃, B 是摸得10, 它们是否独立? 这个需要用定义来判断. $A\cap B$ 是摸得红桃10, 所以

$$\mathsf{P}(A\cap B)=\frac{1}{52}=\frac{1}{4}\cdot\frac{1}{13}=\mathsf{P}(A)\mathsf{P}(B),$$

因此这两个事件是独立的.

两个随机变量独立是类似的, 即一个随机变量取值如何与另一个是无关的, 独立的. 精确地说, 随机变量X,Y 独立, 是指对任何$x,y\in\mathbf{R}$, 事件$\{X\le x\}$ 与$\{Y\le y\}$ 独立.

同样, 独立随机试验各自的随机变量是独立的. 例如甲乙两人各掷n 枚硬币, 他们得到的正面个数是互相独立的. 但同一个随机试验中的随机变量之间的独立性并不直观, 回忆前面的Buffon 问题, 假设针的角度与针中点到平行线的距离之间独立也不是非常自然.

§5.3.1 重复Bernoulli 试验

只关心两个结果的随机试验通常称为Bernoulli 试验, 一个结果称为成功, 概率是$0<p<1$, 另一个结果称为失败, 概率是q, 当然$q=1-p$. 掷硬币是Bernoulli 试验; 掷骰子, 如果掷出6 算是成功, 否则算失败, 也是Bernoulli 试验, 成功概率是$1/6$. 通常用一个随机变量ξ 来表示成功的指标, 即成功的话, $\xi=1$, 失败的话, $\xi=0$. 只取0 与1 两个值的随机变量称为Bernoulli 随机变量.

独立地重复Bernoulli 试验是我们经常碰到的随机试验, 例

如重复掷硬币或者骰子等, 生活也是日复一日地在重复, 每天都在经历成功或者失败.

独立地重复一个成功概率为p 的随机试验, 用ξ_k 表示第k 次成功的指标.

成功次数: 用X 表示重复n 次Bernoulli 试验中成功的次数, 那么X 的取值是0 到n 间的整数. 成功次数为k 这个事件$\{X = k\}$ 的概率怎么计算?

把 n 次试验看成具有 n 个标号的位置, 其中每个位置都有两种可能: 成功或者失败, 分别标记为 1 及 0. "成功次数为k" 的事件 $X = k$ 可以看成是从 n 个位置里拿出 k 个位置标记为 1, 而其它标记为 0, 这样的选择共有 $\binom{n}{k}$ 种. 因为独立性，每种标记发生的概率是 $p^k q^{n-k}$. 由概率的可加性, 成功次数为 k 的概率为

$$\mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k},$$

其中 $0 \le k \le n$. 因为 $\binom{n}{0} = \binom{n}{n} = 1$, 可用下图表表示 X 的分布：

$$\begin{pmatrix} 0 & 1 & 2 & \cdots & k & \cdots & n \\ q^n & \binom{n}{1} p q^{n-1} & \binom{n}{2} p^2 q^{n-2} & \cdots & \binom{n}{k} p^k q^{n-k} & \cdots & p^n \end{pmatrix}$$

从而可以从这个角度证明二项式定理

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \mathbb{P}(X = k) = 1.$$

这是这个分布称为二项分布的理由. 对任何的n, n 之内至少有一次成功的概率

$$\mathsf{P}(X > 0) = 1 - \mathsf{P}(X = 0) = 1 - (1 - p)^n.$$

显然, 只要$p > 0$,

$$\lim_n \mathsf{P}(X > 0) = 1.$$

直观地说，做一件事情，不管成功概率多小, 只要执着地努力, 重复的次数足够多, 就有很大可能会成功. 如同俗话所说：失败是成功之母. 反过来说，如果不断地重复，小概率的坏事也终有可能发生. 例如开车，开一次车发生事故的概率 p 很小, 但是因为每天都在开车, 长期下去还是很可能会发生事故的. 所以，不仅每次开车都要格外小心，减小事故概率 p, 而且要尽可能地减少开车次数 n，这就能使发生事故的概率尽量减小.

等待成功的时间: 另外一个有意思的随机变量是等待成功的时间, 用X 表示首次成功时随机试验的次数, 则$X = k$ 表示前$k-1$ 次试验是失败, 而第k 次试验是成功. 由独立性

$$\mathsf{P}(X = k) = q^{k-1}p,\ \ k = 1, 2, 3, \cdots.$$

有限次之内成功的概率

$$\mathsf{P}(X < \infty) = \sum_{k=1}^{\infty} q^{k-1}p = 1.$$

PROBLEM: 打乒乓球比赛, 对于弱者来说, 比赛的规则是一局定胜负好还是三局两胜好?

§5.4 条件概率

设有两个事件A, B, 且$\mathsf{P}(A) > 0$, A 发生的条件下B 发生的概率称为条件概率

$$\mathsf{P}(B|A) = \frac{\mathsf{P}(B \cap A)}{\mathsf{P}(A)}.$$

在考虑条件概率时, 事件A 发生后, 样本空间改变了, A^c 被排除了. 概率$\mathsf{P}(B)$ 与条件概率$\mathsf{P}(B|A)$ 一般不一样, 当且仅当两者独立时一样.

设一个袋子里有两个球, 一白一黑, 甲摸一个球, 不放回, 乙再摸剩下的球, A, B 分别表示甲乙摸到白球. 显然

$$\mathsf{P}(A) = \mathsf{P}(B) = 1/2.$$

如果考虑条件概率, 那么甲若摸走白球, 乙就不可能摸到白球, 所以$\mathsf{P}(B|A) = 0$; 甲若摸走黑球, 乙就肯定摸到白球, 所以$\mathsf{P}(B|A^c) = 1$.

注意不要被概率与条件概率迷惑, 概率是不变的, 条件改变导致的概率变化是条件概率.

习题: 有标号为 $1, 2, \cdots, m$ 的 m 张卡片, 随机地一张一张抽取, 已知第 k 张是前 k 张中最大的, 求它是 m 的概率.

PROBLEM: 怎么通过掷一枚硬币得到1/3 这个概率?

PROBLEM: 设一个随机试验E 有两个事件$A \subset B$, A 发生的概率是$\mathsf{P}(A)$, B 发生的概率是$\mathsf{P}(B)$. 怎么得到条件概率$\mathsf{P}(A|B)$?

§5.5 全概率公式

设随机试验的结果可以分成 n 种情况, 即设样本空间 Ω 可分成 n 个两两不同时发生 (两两不相交) 的事件 $\Omega_1, \cdots, \Omega_n$:

$$\Omega = \Omega_1 \cup \cdots \cup \Omega_n.$$

做一个随机试验，必定是这 n 种情况之一发生. 一个事件

发生自然也可以看成是在不同情况下分别发生，即

$$A = (A \cap \Omega_1) \cup \cdots \cup (A \cap \Omega_n).$$

因此，由概率的可加性和乘法公式得,

$$\begin{aligned}\mathsf{P}(A) &= \mathsf{P}(A \cap \Omega_1) + \cdots + \mathsf{P}(A \cap \Omega_n) \\ &= \mathsf{P}(A|\Omega_1)\mathsf{P}(\Omega_1) + \cdots + \mathsf{P}(A|\Omega_n)\mathsf{P}(\Omega_n).\end{aligned}$$

因此有下面的

全概率公式：

$$\mathsf{P}(A) = \sum_{k=1}^{n} \mathsf{P}(A|\Omega_k)\mathsf{P}(\Omega_k).$$

因为

$$\sum_{k=1}^{n} \mathsf{P}(\Omega_k) = 1,$$

所以全概率公式实际上是说概率 $\mathsf{P}(A)$ 是条件概率 $\mathsf{P}(A|\Omega_k)$, $1 \le k \le n$, 的加权平均, 而条件概率 $\mathsf{P}(A|\Omega_k)$ 的权重为 $\mathsf{P}(\Omega_k)$, $k = 1, \cdots, n$.

全概率公式是计算概率的一种重要方法. 现在我们用它来回答抽签时抽到好签的概率是否与抽签顺序有关的问题.

§5.5.1 抽签与顺序无关

PROBLEM: 在抽签的时候，抽到好签的概率与抽签的顺序有没有关系?

前面其实有过讨论, 这里我们应用全概率公式来重新讨论. 抽签有两种：放回与不放回. 放回的情况很简单，因为每次的结果是相互独立的，所以当然与顺序无关. 对不放回的情况，我们用前面不放回摸球的例子来说明. 一个袋子中有

3 个白球 2 个黑球, 5 个人依次不放回地摸球，我们来验证他们每个人摸到白球的概率都是 3/5.

A 表示第一个人摸到白球, B 表示第二个人摸到白球. 显然 $\mathsf{P}(A)=3/5$. 第二个人摸球的时候有两种情况: 一种是 A 发生, 即第一个人摸去白球, 这时袋子中剩有 2 白 2 黑; 另一种是 A 没发生, 第一个人摸去黑球, 这时袋子中剩有 3 白 1 黑. 在第一种情况下条件概率为 $\mathsf{P}(B|A)=2/4$, 而在第二种情况下条件概率为 $\mathsf{P}(B|\overline{A})=3/4$. 由于上述两种情况发生的概率分别是 3/5 与 2/5, 直观地可以看出 B 发生的概率应该是两个条件概率的下述加权平均:

$$\mathsf{P}(B)=\frac{2}{4}\cdot\frac{3}{5}+\frac{3}{4}\cdot\frac{2}{5}=\frac{12}{20}=\frac{3}{5},$$

其值与 $\mathsf{P}(A)$ 相等.

第三个人再摸, 记他摸到白球的事件为 C, 其概率是多少呢? 前面两个人摸球会产生四种情况:

$$\Omega_1=A\cap B,\ \Omega_2=A\cap\overline{B},\ \Omega_3=\overline{A}\cap B,\ \Omega_4=\overline{A}\cap\overline{B}.$$

用乘法公式, 这四种情况发生的概率分别是

$$\frac{3}{5}\cdot\frac{2}{4}=\frac{6}{20},\ \frac{3}{5}\cdot\frac{2}{4}=\frac{6}{20},\ \frac{2}{5}\cdot\frac{3}{4}=\frac{6}{20},\ \frac{2}{5}\cdot\frac{1}{4}=\frac{2}{20}.$$

例如

$$\mathsf{P}(\Omega_2)=\mathsf{P}(A\cap\overline{B})=\mathsf{P}(A)\mathsf{P}(\overline{B}|A)=\frac{3}{5}\cdot\frac{2}{4}=\frac{6}{20},$$

其余同理. 在这四种情况下, 袋子中剩下的黑白球的个数分别是：1 白 2 黑, 2 白 1 黑, 2 白 1 黑, 3 白 0 黑, 因此事件

C 的条件概率分别是 1/3、2/3、2/3、1. 再应用全概率公式, 就推出

$$\begin{aligned}\mathsf{P}(C) &= \sum_{k=1}^{4} \mathsf{P}(C|\Omega_k)\mathsf{P}(\Omega_k) \\ &= \frac{1}{3}\cdot\frac{6}{20} + \frac{2}{3}\cdot\frac{6}{20} + \frac{2}{3}\cdot\frac{6}{20} + 1\cdot\frac{2}{20} \\ &= \frac{36}{60} = \frac{3}{5},\end{aligned}$$

其值仍与 $\mathsf{P}(A)$ 相等.

类似地, 第四、第五个人摸到白球的概率仍然是 3/5, 也就是说, 摸到白球的可能性是与摸球的顺序无关的.

§5.5.2 排位置

PROBLEM: 有n 编号的人与n 个编号的位置, 第一个人随机地选一个位置坐下, 然后其他人依次选位, 每个人优先选择与其相同编号的位置, 如果该位置被占, 那么随机选个位置坐下. 问编号n 的人坐在编号为n 的概率是多少?

当$n=2$, 答案显然是1/2.

当$n=3$, 第一个人3 个选择, 如果选1, 那么第三个人肯定选到3; 如果选2, 那么第二个人的选择是1, 这时第三人坐3, 或3, 第三人不能坐3. 因此第三人坐位置3 的概率是

$$\frac{1}{3}\cdot 1 + \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 0 = \frac{1}{2}.$$

§5.5.3 骰子游戏

PROBLEM: 美洲有一种骰子游戏, 赌徒掷两个骰子. 如果掷出的点数之和是 7 或 11, 他就赢了. 如果掷出 2, 3 或 12, 他就输了. 如果掷出其他点数和, 记下这个数, 再掷骰子一

直到掷出这个数或者 7 为止, 如果是这个数, 则赢, 如果是7, 则输, 问赌徒赢的概率是多少?

§5.5.4 赌徒输光问题

PROBLEM: 赌徒甲乙掷硬币或者其它方式赌博, 每局输赢的概率各半, 每次输赢都是一元钱, 赌徒甲乙分别带有赌注总数a,b, 输光结束, 问赌徒甲先输光的概率是多少?

用A 表示甲先输光这个事件, X 表示甲现在掌握的赌注, 我们要算$\mathsf{P}(A|X=a)$, 我们应用全概率公式来推得$\mathsf{P}(A|X=x)$, $0\le x\le a+b$. 显然$X=0$, $X=a+b$ 这两种情况分别是甲输光和乙输光的情况, 比赛结束, 即有边界条件: $\mathsf{P}(A|X=0)=1$, $\mathsf{P}(A|X=a+b)=0$. 在$0<x<a+b$ 时, 比赛要继续, 且这局的结果是赢一元或输一元, 那么

$$\mathsf{P}(A|X=x)=\mathsf{P}(A|X=x-1)/2+\mathsf{P}(A|X=x+1)/2.$$

用$f(x)$ 表示$\mathsf{P}(A|X=x)$, 得

$$f(x)=\frac{f(x-1)+f(x+1)}{2}.$$

$f(x)$ 是等差的, 即$f(x)=cx+d$. 再利用边界条件, $f(0)=1$, $f(a+b)=0$, 推出$d=1$, $c=-1/(a+b)$, 因此

$$\mathsf{P}(A|X=x)=\frac{a+b-x}{a+b}.$$

特别地

$$\mathsf{P}(A|X=a)=\frac{b}{a+b}=\frac{1}{a/b+1}.$$

当b 相比于a, 即赌徒乙带的赌注远远超过甲时, 甲输光的概率几乎是1.

如果甲乙水平不同, 甲赢的概率是p, 输的概率是$q=1-p$, 那么上面的递推方是

$$f(x)=qf(x-1)+pf(x+1).$$

因为$p+q=1$, 所以有以下递推公式

$$f(x+1)-f(x)=\frac{q}{p}(f(x)-f(x-1)).$$

习题: 设有 A,B 两个盒子, 分别有三个白球和三个黑球, 每次从两盒子中任取一个球交换, 求 n 次后盒子中的球的颜色仍然是相同的概率.

§5.6 Bayes 公式

$\mathsf{P}(A)$ 表示A 发生的概率, 如果事件B 发生了, 就是有了新的信息, 在这个信息再看A 发生的概率

$$\mathsf{P}(A|B)=\frac{\mathsf{P}(A\cap B)}{\mathsf{P}(B)}=\frac{\mathsf{P}(B|A)\mathsf{P}(A)}{\mathsf{P}(B)}.$$

这个公式给出条件概率$\mathsf{P}(A|B)$ 与$\mathsf{P}(B|A)$ 的关系, 称为Bayes 公式.

§5.6.1 检测方法的有效性

PROBLEM: 对于某种致命疾病而言, 现有的检测方法昂贵而复杂, 某公司开发了一种成本低廉且快速的医学检测方法,对有这种疾病的人检测时, 90% 有阳性反应. 没有这种疾病的人检测时, 也有 5% 的人有阳性反应. 已知该疾病的患病率大约是5%, 问该检测方法准确率高吗?

上面两个指标听起来会感觉这个方法不错, 但这两个指标并不反映真正的准确率. 检测方法的准确率应该是指当一个人检测有阳性反应时, 他的确有这种疾病的概率的大小. 用 A 表示一个人患有此疾病的事件, A 的概率是患病率, B 表示他的检测有阳性反应的事件. 那么我们要计算条件概率 $\mathsf{P}(A|B)$. 由条件, $\mathsf{P}(A) = 0.05$, $\mathsf{P}(B|A) = 0.9$, $\mathsf{P}(B|A^c) = 0.05$, 由全概率公式

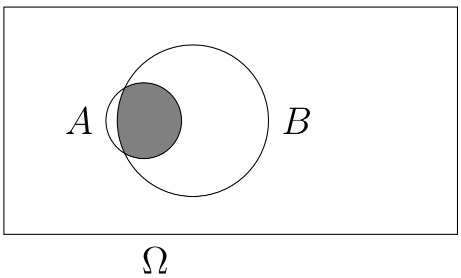
$$\mathsf{P}(B) = \mathsf{P}(B|A)\mathsf{P}(A)+\mathsf{P}(B|A^c)\mathsf{P}(A^c) = 0.9\times0.05+0.05\times0.95 = 0.0925,$$

再由 Bayes 公式,

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(B|A)\mathsf{P}(A)}{\mathsf{P}(B)} = \frac{0.9\times0.05}{0.0925} = \frac{18}{37}.$$

比较 $\mathsf{P}(B|A) = 0.9$, $\mathsf{P}(A|B)$, 不到1/2, 不是一个很高的数字. 其实这个概率与患病率有很大关系. 如果 $\mathsf{P}(A) = 0.02$, 那么 $\mathsf{P}(A|B) \sim 29\%$.

为什么问题中$\mathsf{P}(B|A)$ 很大, 但关键的指标$\mathsf{P}(A|B)$ 不大, 其实这容易理解, $\mathsf{P}(B|A)$ 是$A\cap B$ 在A 中的比例, 而$\mathsf{P}(A|B)$ 是$A\cap B$ 在B 中的比例. 见图



§5.6.2 Monty Hall 问题

在美国的某刊物有一个栏目名为: ask Marilyn, 回答读者提出的各种问题, 其中有一个后来被称为是 Monty Hall 的问题. 有一个由 Monty 主持的电视游戏栏目是这样的: Monty 让参与人 Voila 在三个完全一样的大门 A, B, C 中任选一个, 三门后分别有两只羊与一部汽车, Monty 知道门后是什么 (这点非常重要). 当 Voila 选定 (比如) A 后, Monty 打开一个放有羊的门 (比如) B, 然后告诉 Voila 可以再选择.

PROBLEM: Voila 是选择换与不换有没有差别?

Marilyn 的答案是换. 但答案在当时引起了很大的争议. 因为很多人认为在 Monty 打开门之后, 剩下的两个门有汽车的概率是一样的, 所以换不换没有区别. 还有许多人编了电脑程序来模拟.

实际上, 争议也是因为对问题的理解不同, 依我们理解, 这是一个条件概率问题, 就是计算在 Monty 打开 B 门的条件下, Voila 选择换而得到汽车的概率. 要计算这个概率, 我们需要假设

(1) Monty 总是打开后面是羊的门;

(2) 如果他可以选择, 他总是随机地选择.

在这两个假设下, 有下面的情况会出现:

(1) A 门后是汽车, 这时 Monty 打开 B 门的概率是 $\frac{1}{2}$.

(2) B 门后是汽车, 这时 Monty 打开 B 门的概率是 0.

(3) C 门后是汽车, 这时 Monty 打开 B 门的概率是 1.

这样, 由 Bayes 公式推出在 Monty 打开 B 门的条件下, Voila 选择换而得到汽车的概率为

$$\frac{\frac{1}{3}}{\frac{1}{3}\cdot\frac{1}{2}+0+\frac{1}{3}\cdot 1}=\frac{2}{3}.$$

如果直接从概率空间入手, 考虑换门, 则样本空间为

$$\{AB, AC, BA, BC, CA, CB\},$$

(其中 AB 表示从 A 换为 B). 假设车在 C 门, 在主持人打开一个羊的门之后, 样本空间变成为 $\{AC, BC, CA\}$ 或者 $\{AC, BC, CB\}$, 无论那种情况换而得车的概率都是 2/3.
也可以这样看问题, 用 S 表示主持人打开一个有羊的门, K 表示换门而得车这个事件. 因为主持人总是 (可以) 打开有羊的门, 还是假设车在 C 门, 故

$$\begin{aligned}\mathsf{P}(K|S) &= \mathsf{P}(S\cap K)\\ &= \frac{1}{3}(\mathsf{P}(S\cap K|A)+\mathsf{P}(S\cap K|B)+\mathsf{P}(S\cap K|C)) = 2/3,\end{aligned}$$

其中条件概率表示 Voila 的选择.
另外一个简单的解法如下 [2]. 也就是说 Voila 可以这样来考虑, 设 H 是 Voila 第一次的选择是车的事件, K 是 Voila 后来选择换门而得到车的事件, 那么在主持人打开的门是羊的条件下, K 等于 H^c, 也就是说, 如果 Voila 一开始选择正确的话, 换了就肯定不是车了, 如果他一开始选择的是羊, 那么在主持人开了羊的门之后, 他换门后肯定是车, 因此

$$\mathsf{P}(K|S) = 1 - \mathsf{P}(H) = 2/3.$$

[2]中山大学任佳刚教授提供

尽管有这样明确的解答, 但仍然有许多人置疑这个答案. 下面的解释也许可以说明一些问题: 主持人实际上是以两个门来换 Voila 的 A 门. 网上可以找到许多这个问题的Monte-Carlo 模拟.

第六章　数学期望与方差

§6.1　分布的期望与方差

随机变量有分布, 分布就有中心, 称为数学期望, 简称为期望或者平均. 设随机变量X 的分布为

$$\mathsf{P}(X = x_i) = p(x_i),\ 1 \le i \le n.$$

X 的期望$\mathsf{E}[X]$ 定义为

$$\mathsf{E}[X] = \sum_{i=1}^{n} x_i p(x_i),$$

即x_i 的概率权平均. 实际上X 可以取可列无穷多个值, 这是上面的和变成无穷级数. 随机变量分布的期望类似于物体的重心.

数学期望是随机变量组成的集合到实数的一个映射, 它有下面的性质, 称为线性性质:

1. $\mathsf{E}[cX] = c\mathsf{E}[X]$;
2. $\mathsf{E}[X+Y] = \mathsf{E}[X] + \mathsf{E}[Y]$.

如果X, Y 独立, 那么

$$\mathsf{E}[XY] = \mathsf{E}[X]\mathsf{E}[Y].$$

从定义看出, 一个非负随机变量的期望也是非负的, 因此期望还有单调性: 如果$X \ge Y$, 那么$\mathsf{E}[X] \ge \mathsf{E}[Y]$.

随机变量的方差定义为

$$D[X] = \mathsf{E}[(X - \mathsf{E}[X])^2],$$

它衡量X 的随机程度, 或者说波动程度. 特别地, 当$D[X]=0$ 时, X 是个常数, 没有随机性.
将定义的右边展开,

$$D[X]=\mathsf{E}[X^2-2X\cdot\mathsf{E}[X]+(\mathsf{E}[X])^2]=\mathsf{E}[X^2]-(\mathsf{E}[X])^2,$$

这个计算公式可能更加方便, 其中

$$\mathsf{E}[X^2]=\sum_{i=1}^{n}x_i^2p(x_i).$$

方差的两个性质:

1. $D[cX]=c^2D[X]$;

2. 如果X,Y 独立, 那么

$$D[X+Y]=D[X]+D[Y].$$

§6.2 0,1-值随机变量

取值0,1 的随机变量是Bernoulli 随机变量, 对于一个事件A, 我们总是用1_A 表示它的指标, 它是Bernoulli 的, 反过来, Bernoulli 随机变量ξ 也一定是事件$\{\xi=1\}$ 的指标. 这时, $\mathsf{E}[\xi]=p$, 其中$p=\mathsf{P}(\xi=1)$. 另外$\mathsf{E}[\xi^2]=p$, 所以$D[X]=\mathsf{E}[\xi^2]-(\mathsf{E}[\xi])^2=p-p^2=pq$, 其中$q=1-p$. 可以看出$D[X]$ 在$p=1/2$ 时最大, 即随机性最大.

§6.3 二项分布

PROBLEM: 求二项分布的随机变量X 的期望与方差.

设独立地重复一个成功概率为p 的随机试验n 次, 成功的次数为X, 那么

$$X = \xi_1 + \xi_2 + \cdots + \xi_n,$$

其中ξ_i 是第i 次试验成功的标志, 且$\xi_1, \cdots, \xi_n$ 是互相独立的. 利用期望与方差的性质得

$$\begin{aligned} \mathsf{E}[X] &= np, \\ D[X] &= D[\xi_1] + \cdots + D[\xi_n] = npq. \end{aligned}$$

期望等于试验次数与成功概率的乘积, 这正是我们的直观感觉.

PROBLEM: 设袋子里有a 个白球与b 个黑球, 随机地摸出n 个球, 用X 表示其中的白球数, 求X 的期望与方差.

§6.4 等待成功

PROBLEM: 重复成功概率为p 的Bernoulli 试验, 首次成功的时刻X, 称为等待首次成功出现的时间, 求其期望与方差. 其分布是

$$\mathsf{P}(X = k) = q^{k-1}p, \ k \geq 1.$$

期望

$$\mathsf{E}[X] = \sum_{k \geq 1} k q^{k-1} p = \frac{1}{p},$$

因为

$$\sum_{k \geq 1} k q^{k-1} = \left(\sum_{k \geq 0} q^k \right)' = \frac{1}{(1-q)^2} = \frac{1}{p^2}.$$

即平均等待成功的时间与成功概率成反比, 也是直观的.

要计算方差, 需要先计算二阶矩$\mathsf{E}[X^2]$,

$$\begin{aligned}\mathsf{E}[X^2] &= \sum_{k\geq 1} k^2 q^{k-1} p \\ &= pq\sum_{k\geq 1} k(k-1)q^{k-2} + p\sum_{k\geq 1} kq^{k-1} \\ &= pq\left(\frac{1}{1-q}\right)^2 + \frac{1}{p} \\ &= \frac{2pq}{p^3} + \frac{1}{p} = \frac{2-p}{p^2},\end{aligned}$$

因此

$$D[X] = \mathsf{E}[X^2] - (\mathsf{E}[X])^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

PROBLEM: 求等待第r 次成功时刻的期望与方差.

PROBLEM: 重复掷一个骰子, 求等待所有数字出现的时间的期望与方差.

§6.5 等待模式出现的时间

重复地掷一个硬币, 依次记录掷出的结果, 1, 0 分别表示正反面, 那么得到一个01 序列, 一个有限长度的01 序列称为是一个模式, 例如(1), (10), (101), (1111), (11101001) 等.

PROBLEM: 任意给定一个模式, 它是不是一定在有限时间内出现?

答案是肯定的. 对于一个给定的长度为k 的模式, 把它看作成功, 成功的概率是$p = 1/2^k > 0$. 然后把每k 次掷硬币当作一个随机试验, 独立地重复, 前面已经说过, 有限时间内一定会成功, 而且期望等待时间是$1/p = 2^k$.

你也许不明白这意味着什么. 一个模式可以很长, 例如一本书可以编译成为一个模式, 一个图书馆里所有的书也可以编译成为一个模式, 所以这结果告诉我们, 只要不断重复地掷硬币, 你肯定会看到这个图书馆的所有书出现.

PROBLEM: 求给定模式的平均等待时间.

用ξ 表示给定模式首次出现的时间. 首先要注意, 如果把模式中的01 互换得到的模式与原模式是对称的, 平均等待时间也是一样的. 因此为我们总是考虑以1 开始的模式; 其次, 上面这样分组考察模式出现方式不一定是模式首次出现的时间. 例如模式(101), 在下面的亲测的掷硬币序列中

$$001, 110, 011, 010, 111, 001, 100, 110, 101, 111, 000, 001, 100, 110, 111, 100 \cdots,$$

$\xi = 11$, 二若按3 个一组观察, 那么模式出现的时间是第9 组, 第27 次. 所以我们需要有新的方法.

期望也可以应用全概率公式, 分不同的情况求条件下的期望, 然后平均

$$\mathsf{E}[\xi] = \mathsf{E}[\xi|\Omega_1]\mathsf{P}(\Omega_1) + \cdots + \mathsf{E}[\xi|\Omega_n]\mathsf{P}(\Omega_n),$$

其中$\mathsf{E}[\xi|B]$ 表示事件A 发生的条件下ξ 的期望.

例如ξ 是等待模式(1), 那么, 在1 出现时, 结束, $\xi = 1$, $\mathsf{E}[\xi|1] = 1$, 而在0 出现时, 我们显然要重新等, $\mathsf{E}[\xi|0] = 1 + \mathsf{E}[\xi]$. 因此

$$\mathsf{E}[\xi] = \mathsf{E}[\xi|1]/2 + \mathsf{E}[\xi|0]/2 = 1/2 + 1/2(\mathsf{E}[\xi] + 1),$$

得$\mathsf{E}[\xi] = 2$.

再ξ 是等待模式(10), 先掷一次, 如果是1, 模式有望, 需继续等待; 如果是0, 模式无望, 重新开始等待. $\mathsf{E}[\xi] = \mathsf{E}[\xi|1]/2 +$

$\mathsf{E}[\xi|0]/2$, 因为$\mathsf{E}[\xi|0]=1+\mathsf{E}[X]$, 所以$\mathsf{E}[\xi]=1+\mathsf{E}[\xi|1]$. 再掷一次,

$$\begin{aligned}\mathsf{E}[\xi|1]&=\mathsf{E}[\xi|11]/2+\mathsf{E}[\xi|10]/2;\\ \mathsf{E}[\xi|11]&=1+\mathsf{E}[\xi|1],\ \mathsf{E}[\xi|10]=2;\\ \mathsf{E}[\xi|1]&=3,\ \mathsf{E}[\xi]=4.\end{aligned}$$

再设ξ 是等待模式(101), 反复利用上面的思想, 得方程

$$\begin{aligned}\mathsf{E}[\xi]&=\frac{1}{2}(\mathsf{E}(\xi|1)+\mathsf{E}(\xi|0))=\frac{1}{2}(\mathsf{E}(\xi|1)+1+\mathsf{E}[\xi]),\\ \mathsf{E}(\xi|1)&=\frac{1}{2}(\mathsf{E}(\xi|10)+\mathsf{E}(\xi|11))=\frac{1}{2}(\mathsf{E}(\xi|10)+1+\mathsf{E}(\xi|1)),\\ \mathsf{E}(\xi|10)&=\frac{1}{2}(\mathsf{E}(\xi|101)+\mathsf{E}(\xi|100))=\frac{1}{2}(3+3+\mathsf{E}[\xi]).\end{aligned}$$

解方程得 $\mathsf{E}[\xi]=10$.

PROBLEM: 求模式(1111) 的平均等待时间? 求模式连续n个1 的平均等待时间?

PROBLEM: 从这几个答案, 你能看出什么规律吗? 对一般的模式等待问题有没有猜测?

§6.6 配对数

在配对问题中, 用X 表示配对数, ξ_i 表示第i 对夫妇成功配对的指标, 那么$\mathsf{E}[\xi_i]=\mathsf{P}(\xi=1)=1/n$ 且

$$X=\xi_1+\cdots+\xi_n.$$

因此

$$\mathsf{E}[X]=\mathsf{E}[\xi_1]+\cdots+\mathsf{E}[\xi_n]=1.$$

即平均配对数总是1.

为了计算配对数的方差, 需计算二阶矩

$$\mathbb{E}[X^2]=\sum_{i=1}^n \mathsf{E}[\xi_i^2]+\sum_{i\neq j}\mathsf{E}[\xi_i\xi_j].$$

因为ξ_i 是Bernoulli 的, 故$\xi_i\xi_j$ 也是Bernoulli 的. 因此$\mathsf{E}[\xi_i^2]=1/n$, 而当$i\neq j$ 时,

$$\begin{aligned}\mathsf{E}[\xi_i\xi_j]&=\mathsf{P}(\xi_i=1,\xi_j=1)\\&=\mathsf{P}(\xi_j=1|\xi_i=1)\mathsf{P}(\xi_i=1)\\&=\frac{1}{n-1}\cdot\frac{1}{n}.\end{aligned}$$

这样推出

$$\begin{aligned}\mathsf{E}[X^2]&=n\cdot\frac{1}{n}+n(n-1)\frac{1}{n(n-1)}=2,\\D[X]&=\mathsf{E}[X^2]-(\mathsf{E}[X])^2=1.\end{aligned}$$

方差也是恒等于1.

§6.7 等待最高报价

PROBLEM: 在网上发布卖一个自行车, 第一个报价为X_1, 然后我决定卖给下一个报价高于第一个报价的人, 问期望等待时间是多久?

假设报价依次为

$$X_1,X_2,\cdots,X_n,\cdots,$$

可以假设它们是独立同分布的随机变量. 定义

$$N:=\inf\{n:\ X_n>X_1\},$$

即首次超过第一个报价的编号. 那么我们要求$\mathsf{E}[N]$.

第七章　母函数

先介绍一下有用的母函数. 一个数列$\{a_n : n \geq 0\}$ 可以定义一个幂级数

$$f(x) := \sum_{n\geq 0} a_n x^n,$$

称为是数列的母函数. 稍微有点高数的知识就知道, 当数列有界的时候, 级数收敛域包含区间$|x| < 1$, 这时数列由母函数唯一决定. 数列及其母函数的例子如下

$$\begin{aligned} \frac{1}{1-x} &= 1 + x + x^2 + \cdots + x^n + \cdots , \\ \frac{1}{(1-x)^2} &= 1 + 2x + 3x^2 + \cdots + (n+1)x^n + \cdots , \\ \log(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} + \cdots + \frac{(-1)^{n-1}x^n}{n} + \cdots . \end{aligned}$$

如果某个随机变量τ 取值为非负整数, 那么

$$\mathsf{E}[x^\tau] = \mathsf{P}(\tau = 0) + x\mathsf{P}(\tau = 1) + x^2\mathsf{P}(\tau = 2) + \cdots + x^n\mathsf{P}(\tau = n) + \cdots ,$$

即分布律作为数列$\mathsf{P}(\tau = n)$ 的母函数是$\mathsf{E}[x^\tau]$, 也称为τ 的母函数. 算母函数和算分布列是等价的, 而有时候算母函数比较简单, 原因是两个独立随机变量τ, σ 的和的母函数等于它们各自母函数的乘积

$$\mathsf{E}[x^{\tau+\sigma}] = \mathsf{E}[x^\tau]\mathsf{E}[x^\sigma].$$

注意我们容许τ 取正无穷为值, 但是

$$\mathsf{P}(\tau < \infty) = \sum_{n\geq 0} \mathsf{P}(\tau = n) = f(1).$$

如果$\mathsf{P}(\tau < \infty) = 1$, 则$\tau$ 的期望为

$$\mathsf{E}[\tau] = f'(1).$$

§7.1 首次通过时

接着讨论随机游动. 甲乙两人赌博, 每局甲胜的概率是p, 输赢一元钱. 不妨设一开始赌资为0, 用$x(n)$ 表示甲在n 局之后的钱数, 这是简单随机游动, 在$p = 1/2$ 是称为是对称的. 随机游动是概率论中最重要的主题, 也是永恒的主题, 实际上可以证明, 任何(可分空间上的)概率问题都可以转化为对称简单随机游动中的问题.

PROBLEM: 设a 是整数, 求首次通过时

$$T_a = \inf\{n > 0 : x(n) = a\}$$

的母函数及随机游动在有限时间内会通过a 的概率$\mathsf{P}(T_a < \infty)$.

先设$a > 0$, 令

$$f_a(x) := \mathbb{E}[x^{T_a}],$$

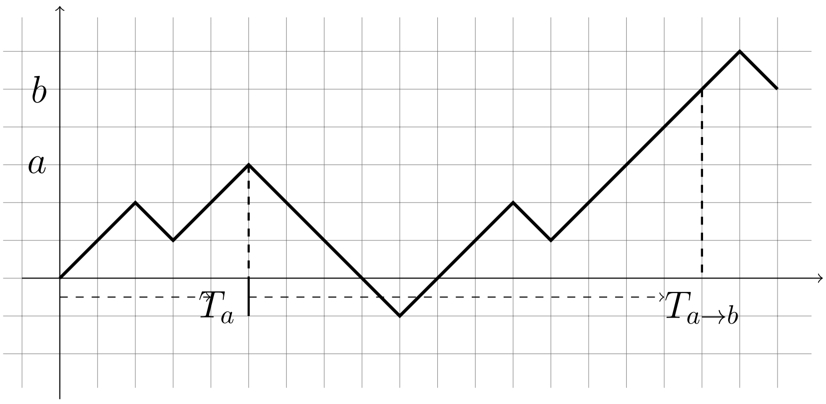
称为从0 出发首次通过a 时的母函数.

1. 如果$a < b$, 那么从a 出发首次通过b 的母函数等于从0 出发首次通过$b - a$ 的母函数.

2. 如果$b > a > 0$, 那么

$$f_b(x) = f_{b-a}(x) \cdot f_a(x).$$

第一条性质是平移不变性. 第二条性质是由于独立性. 从0出发首次达到b 分成为首次到达a 的时间T_a 加上, 重新从时间0 开始计, 从a 出发首次抵达b 的时间$T_{a\to b}$. 随机游动的后一段和前一段是独立的, 即T_a 之后怎么走与T_a 之前无关. 好好想一想, 是不是这样?



因此

$$\begin{aligned} f_b(x) &= \mathsf{E}[x^{T_b}] = \mathsf{E}[x^{T_a+T_{a\to b}}] \\ &= \mathsf{E}[x^{T_a}]\mathsf{E}[x^{T_{a\to b}}] = f_a(x)f_{b-a}(x). \end{aligned}$$

由此推出

$$f_a(x) = (f_1(x))^a.$$

我们只需算T_1 的母函数就可以了.

应用全概率公式, 让随机游动走一步, 它往右走一步, 概率是p, 这时就碰到1 了, 所以$T_1 = 1$; 或者往左走一步, 概率是q, 这时随机游动位于-1 处, 那么原先的T_1 在现在看来就是1 加上从-1 出发到1 的首次通过时, 它和从0 到2 的首次通过时同分布. 因此

$$\begin{aligned} f_1(x) &= \mathsf{E}[x^{T_1}] = xp + xq\mathbb{E}[x^{T_2}] \\ &= xp + xq(f_1(x))^2. \end{aligned}$$

令$y=f_1(x)$, 我们得到方程

$$xqy^2-y+xp=0.$$

因为$0<f_1(x)<1$, 解出

$$f_1(x)=\frac{1-\sqrt{1-4x^2pq}}{2xq}.$$

由对偶性得

$$f_{-1}(x)=\mathsf{E}[x^{T_{-1}}]=\frac{1-\sqrt{1-4x^2pq}}{2xp}.$$

由此推出, 0 点出发在有限步内到达1 的概率为

$$\mathsf{P}^0(T_1<+\infty)=f_1(1)=\frac{1-|p-q|}{2q}=\begin{cases}1, & p\geq q;\\ \frac{p}{q}, & p<q.\end{cases}$$

即当随机移动偏向右时, 它必定在有限时间内到达1. 这时计算其期望值h

$$\mathsf{E}[T_1]=f_1'(1)=\begin{cases}+\infty, & p=q;\\ \frac{1}{p-q}, & p>q.\end{cases}$$

再来计算0 出发首次回归0 的时间τ 的母函数, 同样让随机游动走一步, 它或者到1, 概率为p, 然后从1 出发首次抵达0, 分布等同于从0 出发首次通过-1; 或者到-1, 概率为q, 然后从-1 出发首次抵达0, 其分布等同于从0 出发首次通过1. 性,

$$\begin{aligned}f(x)=\mathsf{E}[x^{\tau}]&=xp\mathsf{E}[x^{T_{-1}}]+xq\mathsf{E}[x^{T_1}]\\&=xpf_{-1}(x)+xqf_1(x)=1-\sqrt{1-4x^2pq}.\end{aligned}$$

因此0 出发在有限步内返回0 的概率

$$\mathsf{P}(\tau<\infty)=1-\sqrt{1-4pq}=1-|p-q|=\begin{cases}2q, & p>q\\ 1, & p=q\\ 2p, & p<q,\end{cases}$$

注意只有在对称简单随机游动时, 此概率为1.
当$p=1/2$ 时, 首次回归时τ 的期望是无穷, 事实上,

$$\mathsf{E}[\tau]=f'(1)=(1-\sqrt{1-x^2})'_{x=1}=\left(\frac{x}{\sqrt{1-x^2}}\right)_{x=1}=\infty.$$

§7.2 种群灭绝概率

现在我们考虑一个简单的种群繁衍模型, 设种群的每个个体繁衍后代的个数ξ, 独立于其他个体的繁衍, 且服从以下分布

$$\mathsf{P}(\xi=n)=p_n,\ n\geq 0,$$

即后代n 个的概率是p_n, 假设$p_1<1$, 否则后代总是一个, 属于平凡的情况.
设ξ 的母函数为

$$f(x)=\mathsf{E}[x^\xi]=p_0+p_1x+p_2x^2+\cdots+p_nx^n+\cdots,$$

因为母函数的各阶导数都是非负的, 所以母函数是凸的. 另外$f(1)=1$, $f(0)=p_0$, $f'(1)$ 是ξ 的期望.
设$X_0=1$, X_n 是第$n-1$ 代的所有个体X_{n-1} 繁衍的后代总数, 即

$$X_n=\sum_{i=1}^{X_{n-1}}\xi_i^{(n-1)},$$

其中$\xi_i^{(n-1)}$ 是第$n-1$ 代的第i 个个体繁衍的后代数, 它们是独立的且分布与ξ 相同. 这个模型称为Gordon-Watson 模型. 显然, 如果存在一个n 使得$X_n=0$, 那么种群就灭绝了.

PROBLEM: 问种群灭绝的概率是多少?

种群灭绝这个事件是

$$A=\bigcup_{n\geq 1}\{X_n=0\},$$

而若第n 代没有后代的话, 其后肯定也没有后代了, 所以$X_n=0$ 推出$X_{n+1}=0$, 因此

$$\mathsf{P}(A)=\lim_n \mathsf{P}(X_n=0),$$

其中$\mathsf{P}(X_n=0)$ 关于n 递增, 极限存在, 就是灭绝概率, 记为a.

怎么计算$\mathsf{P}(X_n=0)$? 如果我们知道X_n 的母函数, $\mathsf{E}[x^{X_n}]$, 它的常数项, 即母函数在$x=0$ 时的值就是$\mathsf{P}(X_n=0)$. 记$f_n(x)=\mathsf{E}[x^{X_n}]$. 那么

$$\begin{aligned}
f_n(x) &= \mathsf{E}\left[x^{\sum_{i=1}^{X_{n-1}}\xi_i^{(n-1)}}\right] \\
&= \sum_k \mathsf{E}\left[x^{\sum_{i=1}^{k}\xi_i^{(n-1)}}; X_{n-1}=k\right] \\
&= \sum_k \mathsf{E}\left[x^{\sum_{i=1}^{k}\xi_i^{(n-1)}}\right]\mathsf{P}(X_{n-1}=k) \\
&= \sum_k (\mathsf{E}[x^{\xi}])^k \mathsf{P}(X_{n-1}=k) \\
&= \sum_k (f(x))^k \mathsf{P}(X_{n-1}=k) \\
&= f_{n-1}(f(x)) = \cdots = f^n(x),
\end{aligned}$$

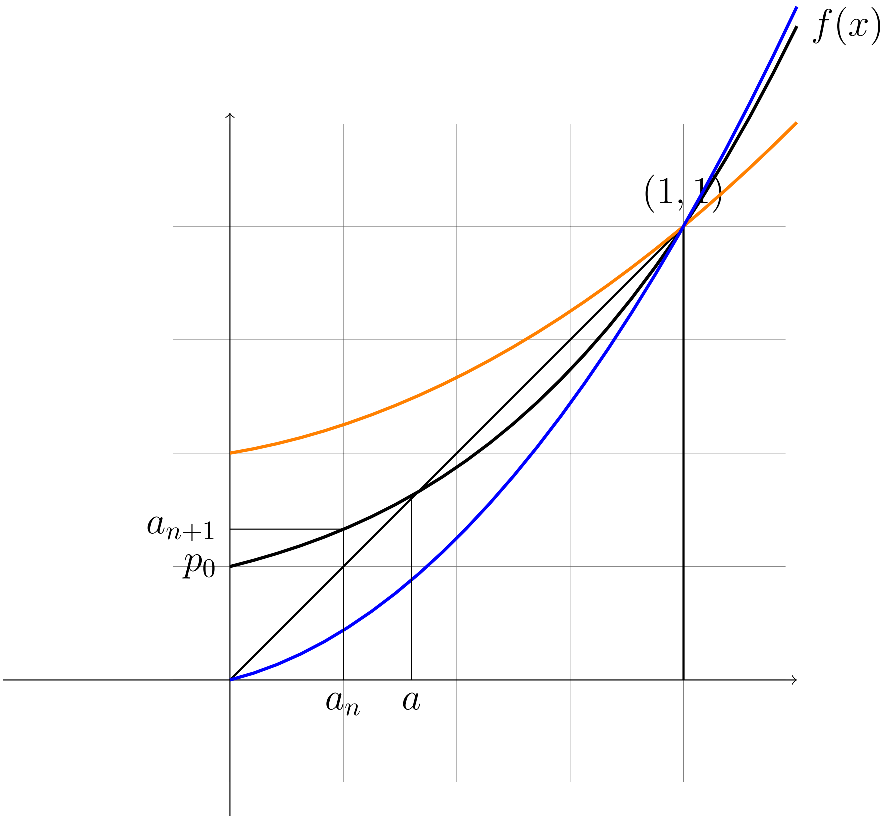
其中f^n 表示n 重复合函数, 即

$$f^2(x) = f(f(x)), f^3(x) = f(f^2(x)), \cdots .$$

看上去不错, $\mathsf{P}(X_n = 0) = f_n(0) = f^n(0)$, 记它为$a_n$, 那么有递推

$$a_{n+1} = f(a_n).$$

因为a_n 递增, 极限就是灭绝概率a, 肯定满足$a = f(a)$, 即灭绝概率是方程$x = f(x)$ 的根. 因为$f(x)$ 是凸的, 此方程最多两个根, 一个根是1.



第一种情况: $f'(1) \leq 1$, 即个体期望后代数不超过1. 这时方程有唯一根1, 所以$a = 1$, 种群必定灭绝.

第二种情况: $f'(1) > 1$, 即个体期望后代数超过1. 因为$f(0) \geq 0$, 所以肯定有两个根. 那么a 是哪个根?

实际上, a 一定是小的根, 即$a<1$. 理由如下: 因为$a_{n+1}\geq a_n$, 所以(a_n,a_{n+1}) 在直线$y=x$ 的上方. 而当x 处于小的根与1 之间时, $f(x)<x$, 因此a_n 肯定都处于小的根之前, 极限只能是小的根.

这时分两种情况: 一是$p_0=0$, 即每个个体至少有一个后代的. 那么小的根是0, 灭绝的概率是0, 种群几乎肯定不会灭绝. 而是$p_0>0$, 那么$0<a<1$, 灭绝概率在0, 1 之间, 种群灭绝依然是可能的.

第八章　大数定律

对于一个随机现象来说, 结果是不可预测的, 结果发生的概率是可以知道的. 例如掷一次硬币得正面的概率比掷骰子得点6 的概率要大不少, 但这不能保证正面会出现或者6 点不会出现, 结果依然是随机的. 这样的话, 概率除了告诉我们可能性大一点或者小一点之外, 概率的数值究竟有什么意义呢?

这个数值会在多次试验时体现出直观的意义, 也就是说成功的频率, 成功次数与试验次数之比, 逼近成功概率. 这个规律是如此自然直观, 我认为历史上肯定有很多人意识到. 第一个叙述这个规律的是意大利数学家Gerolamo Cardano (1501-1575), 因为他也是一个赌徒, 因为对诸多的概率问题有所讨论并写在他的著作The book of Games of Chance. 对于频率与概率的问题, 他大概是这么说的: when the probability for an event is p then by a large number n of repetitions the number of times it will occur does not lie far from the value np. 如果一个事件的概率是p, 那么重复次数n 很大之后, 它发生的次数不会远离np.

但是真正精确地阐述并证明这个规律的是150 年之后的瑞士数学家Jakob Bernoulli, 发表在他死后八年1713 年出版的The art of conjecture 一书中. 他自己显然非常看重这个规律, 称之为Golden Theorem, 认为它在价值上超过了书中其它内容. 这个规律现在通常称为大数定律.

在本讲义中给出的证明不同于Bernoulli 原始的证明, 证明

的关键是Chebyshev 不等式, 注意前面我们几乎都是计算, 没有证明, 包括期望很多性质, 如果读者真的对概率论有兴趣, 这些性质都是需要严格证明的.

§8.1 Chebyshev 不等式

设X 是一个随机变量, x 是个正实数, 则有
Chebyshev 不等式:

$$\mathsf{P}(|X| \geq x) \leq \frac{\mathsf{E}[X^2]}{x^2}.$$

证明的关键是应用期望的单调性. 因为

$$X^2 \geq X^2 1_{\{|X| \geq x\}} \geq x^2 1_{\{|X| \geq x\}},$$

所以

$$\mathsf{E}[X^2] \geq \mathsf{E}[x^2 1_{\{|X| \geq x\}}] = x^2 \mathsf{P}(|X| \geq x),$$

因此推出所要证明的不等式.

§8.2 Bernoulli 大数定律

回到独立地重复一个成功概率为p 的Bernoulli 试验n, X 是成功的次数, 那么X/n 是成功的频率, 它是一个随机变量, 期望为$\mathsf{E}[X/n] = np/n = p$, 方差为

$$D[X/n] = D[X]/n^2 = npq/n^2 = pq/n.$$

应用Chebyshev 不等式, 对任何的$\varepsilon > 0$ 有

$$\mathsf{P}\left(\left|\frac{X}{n} - p\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathsf{E}\left[\left(\frac{X}{n} - p\right)^2\right]$$

$$= \frac{D[X/n]}{\varepsilon^2} = \frac{pq}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

得到下面的结论

Bernoulli 大数定律:

$$\mathsf{P}\left(\left|\frac{X}{n} - p\right| > \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}.$$

仔细地阅读上面的结论, 它是说, 不管ε 多小, 频率与概率的差比ε 大的可能性不超过$(4n\varepsilon^2)^{-1}$, 它当试验次数n 趋于无穷时趋于零. 这也就是大数定律的本质所在, 即概率的数值意义会在许多次试验时体现出来, 或者说概率这个事件的抽象属性会通过频率表现出来.

通过频率来估计概率也就是统计的基本方法.

PROBLEM: 一个大水塘, 不把水抽干, 怎么来估计水塘里大概有多少鱼?

设水塘里共有n 条鱼, 先从水塘里打捞出1000 条鱼, 涂上一点不会掉落的红色, 然后放回水塘. 那么水塘里有红色的鱼的比例是$1000/n$. 过几天再从水塘里打捞出1000 条鱼, 查看有多少是红色的, 例如150 条, 那么$\frac{150}{1000}$ 就是水塘中有红色的鱼的比例的一个估计, 即

$$\frac{1000}{n} \approx \frac{150}{1000},$$

因此

$$n \approx \frac{1000^2}{150} \approx 6667.$$

Bernoulli 大数定律的推广形式如下: 一个随机试验中有个随机变量X, 独立地重复随机试验, 观察此随机变量, 得到

独立同分布的随机序列

$$X_1, X_2, X_3, \cdots, X_n, \cdots.$$

用S_n 表示前n 项的和, 那么S_n/n 就是前n 项的平均, 称为样本平均.

大数定律: 当n 趋于无穷时, 样本平均趋于平均(即期望):

$$\frac{S_n}{n} \longrightarrow \mathsf{E}[X].$$

上面所说的收敛实际上是依概率收敛, 概率论中的随机变量序列有很多种不同的收敛, 另外一种重要的收敛是几乎处处收敛, 也就是说, 在几乎所有的情况下, 都是收敛的. 几乎处处收敛比依概率收敛要强大, 在方差有限的条件下, 实际上有S_n/n 几乎处处收敛于$\mathsf{E}[X]$, 被称为强大数定律. 非概率专业的读者可能无法区分这两种收敛的不同, 可以忽略.

例如掷一个骰子, 骰子点数X 的数学期望为

$$\mathsf{E}[X] = \frac{1+2+3+4+5+6}{6} = 3.5,$$

那么掷n 次骰子, 骰子点数的平均数差不多是3.5.

§8.3 独立地重复

概率是赋予事件的一个数，它表达该事件有多大的可能性发生. 在等可能的假设下，掷硬币得正面的概率是 1/2, 掷骰子得 6 的概率是 1/6. 但是如果没有等可能假设，这些概率是很难知道的. 例如, 假设投掷的一个硬币由一个金属薄片和一个木头薄片黏贴组成，这样的硬币显然不再是等可

能的，掷这个硬币金属面朝上的概率 p 和木头片朝上的概率 q 依然是存在的，且 $p+q=1$, 但理论上没有人能够算出来p 与q 究竟是多少. 这时候因试验可以任意多次重复，仍可以用频率或者经验概率来进行估计，这正是大数定律所断言的.

大数定律成立的关键是独立地重复, 如果一个随机现象不能独立地重复, 概率的数值意义也无从谈起. 现实世界中更多的随机现象是不能独立重复的, 例如明天的天气, 明天股市变化, 明天身体情况, 等等, 因此概率论可以适用的范围不大. 尽管如此, 有些随机现象也许可以近似地独立重复, 就像统计中的数据.

现在换个问题. 一个具体的人，例如 20 岁的小明，能够健康活到 70 岁的可能性有多大? 因为不能预测，健康活到 70 岁是一个随机事件. 它有没有一个概率呢？因为小明这个人是不能重复的，大数定律失效. 这时即使说概率，也只是一种心理预期，无法加以检验. 这是作为数学的概率论无法解决的问题，称之为主观概率，因为它只是反映说话者的主观判断.

尽管小明问自己能健康活到 70 岁的可能性多大在数学上是没有意义的，但我们仍然可以做一点有意义的事情，就是把其他人看成是小明的某种重复，从而求得这个事件的“频率”，也就是求健康活到 70 岁的人数在总人数中的比例. 这个比例是统计数据，对具体的个人而言是没有意义的，但是它仍然有其统计意义，在某些场合是非常有用的，例如保险公司可以据此来计算人寿保险的保险费率.

从这里可以明白，生活中经常说的话，如“学某某专业好找工作”或者“锻炼让人长寿”等等，都是统计意义上的断言，对具体个人没有什么特别的意义.
注: 本节应用上海市高中数学教材

§8.4 蒙特卡洛算法

大数定律还给了我们一种新的算法，称为蒙特卡洛算法. 假设我们要算平面上一个不规则区域 D (例如某个国家在地图上所占的区域) 的面积，这样的图形用通常的面积计算方法是不容易求得的.
怎么应用蒙特卡洛算法呢？看个简单的例子，取一个正方形, 记为 Ω, 把所求的区域 D 放在此正方形中, 如下图所示然后把这个包含区域 D 的正方形放置在墙上，向它随意投掷 200 次飞镖，记录到飞镖落在 D 中的次数有 82 次.
因为是随机投飞镖，所以飞镖落在 D 中的概率等于面积比

$$p = \frac{|D|}{|\Omega|},$$

其中 $|D|$ 与 $|\Omega|$ 分别表示两个区域的面积 (这是另一种古典概率：称为几何概率模型). 现在飞镖落在 D 中的频率为

$$\hat{p} = \frac{82}{200} = \frac{41}{100}.$$

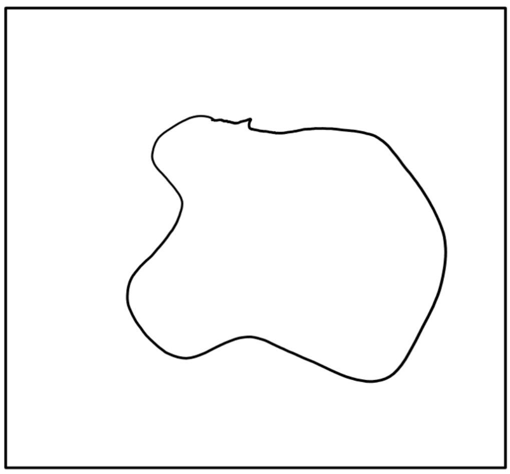
按大数定律，应有 $\hat{p} \approx p$, 因此

$$|D| \approx \frac{41}{100}|\Omega|,$$

其中正方形 Ω 的面积已知或者是容易计算的.

上面例子中所示的思想就是蒙特卡洛算法，也称为随机化算法. 该算法的优点是实现简单，但缺点是难以估计和控制误差, 因为大数定律告诉我们的是, 在试验次数很大时, 误差大于某个数的概率变得很小, 不能保证肯定地变得很小. 法国数学家蒲丰在 1777 年曾经设计过投针实验来计算圆周率 π 的近似值，相关阅读留作习题.

注: 本节引用上海市高中数学教材.



$D \subset \Omega$

第九章　预期与结果

对于一个随机变量来说, 结果是指它的实际取值, 预期是指它的期望, 从定义看预期是结果的加权平均, 应该会很好地反映结果, 也就是说, 预期好通常应该结果也好, 但实际上往往不是这样.

§9.1　样本平均与中位数

在一个小区里统计居民财富, 当然最好的办法是看每一户居民的财富, 但领导通常是看平均财富, 因为平均财富数字简单, 大概能够说明这个小区居民的财富情况了.
通常来说这是对的, 但也很可能出问题, 例如这小区99% 的户均财富在10 万以下, 但有1% 的居民很富有, 财富超过10 亿, 那么平均就差不多是1000 万了, 会严重地扭曲小区居民财富的真实情况. 也就是说财富差距很大的时候, 平均不一定是个好的指标, 预期不能反映真实结果.
这时, 另外一个指标也很重要, 那就是把居民财富按顺序排列之后位于中间的数字, 这个数字称为中位数. 在这个特例中, 那就是10 万元, 它差不多反映大部分人的真实情况, 但是它没有反映出富裕群体的情况. 因此通过简单的统计数字不一定能看到真实情况, 我们应该千万小心.
实际上, 平均和中位两个指标各有各的用处. 例如在一个公司的劳资双方谈判时, 公司的老板说员工的平均工资已经达到5 万年薪, 已经很高了, 但工会方说大部分工人的年

薪才是1 万多, 太低了. 这就是双方关注点不同, 公司的拥有者只关心他每年要支付的工资的总数, 所以关心平均工资, 而工人则关心自己拿到多少钱, 不关心老板总的付出多少钱, 因此劳资双方很难谈得拢.

§9.2 投资陷阱

在这一节中, 我们介绍一个投资的模型, 看看从这个模型能得到什么.

假设我有一笔钱, 放在银行拿利息比较稳定, 设每期的利率为正常数$r > 0$, 也就是说a 元钱一期之后变成$a(1+r)$ 元; 另一种方法是投资, 收益是随机的, 每期的投资收益率为随机变量$X > -1$, 即a 元钱一期之后变成$a(1+X)$ 元钱.

由于投资收益是随机的, 也就是有风险, 那必然有风险溢价, 即期望收益率$\mathsf{E}[X]$ 应该高于利率, 这是说市场总会在期望上给予愿意冒险者一定的奖励, 否则这个投资就没有足够的吸引力. 例如现在, 银行年利率为3%, 那投资的期望收益率应该3% 以上才会有吸引力, 负责大多数愿意选择把钱存在银行, 市场就不可能活跃.

用S_0 表示初始投资额, 这是常数. 用S_n 表示n 期投资之后的财富, 那么S_n 可以如下表示

$$S_n = S_0(1+X_1)(1+X_2)\cdots(1+X_n),$$

其中$X_1, X_2, \cdots, X_n$ 为第$1, 2, \cdots, n$ 期的投资收益率. 假设投资项目稳定, 它们是独立的且与X 同分布的.

利用期望的性质

$$\mathsf{E}[S_n] = S_0(\mathsf{E}[1+X])^n = S_0(1+\mathsf{E}[X])^n,$$

因为$\mathsf{E}[X] > r > 0$, 所以期望收益$\mathsf{E}[S_n]$ 趋于无穷. 因此从预期看, 投资总是很乐观的. 事实是这样吗? 答案是不一定, 至少从数学上看有一个很深的陷阱. 什么是陷阱? 陷阱是指期望很乐观的投资却几乎总是失败的, 亏损的, 即对于几乎所有$\omega \in \Omega$, $S_n(\omega)$ 趋于0.
投资陷阱的存在性: 存在这样的随机收益率X 使得

1. $\mathsf{E}[S_n] \longrightarrow +\infty$;

2. $S_n \longrightarrow 0$.

因为S_n 可以写成为

$$S_n = S_0 \exp\left(\sum_{i=1}^{n} \log(1+X_i)\right).$$

由强大数定律, 当

$$\mathsf{E}[\log(1+X)] < 0$$

时, $\sum_{i=1}^n \log(1+X_i)$ 趋于负无穷, 这时S_n 趋于零.
因此当

$$\begin{aligned} &\mathsf{E}[1+X] > 1,\ (\text{等价地},\ \log \mathsf{E}[1+X] > 0),\ \text{且} \\ &\mathsf{E}[\log(1+X)] < 0 \end{aligned}$$

时, 投资陷阱呈现. 称之为陷阱条件.

这实际上很大程度上是因为对数函数的凹性. 因为对数的凹性, 由Jensen 不等式知

$$\mathsf{E}[\log(1+X)] \le \log(1+\mathsf{E}[X]),$$

这样适当选取X 的分布, 右边是正的, 但左边可能会负的, 见下例.

假设投资有亏损的可能, 即X 可能大于0 (盈利) 也可能小于0 (亏损). 最简单地, 我们设X 取两个值, 一个是$b > 0$ 一个是$-1 < a < 0$, 它的分布为

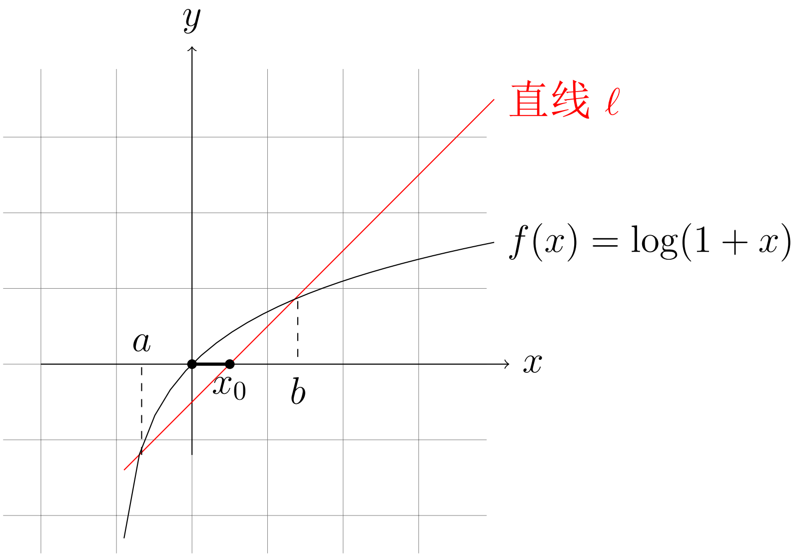
$$\begin{pmatrix} a & b \\ 1-p & p \end{pmatrix},$$

则$\log(1+X)$ 的分布为

$$\begin{pmatrix} \log(1+a) & \log(1+b) \\ 1-p & p \end{pmatrix}.$$

那么陷阱条件等价于

$$\begin{aligned} &\mathsf{E}[X] = pb + (1-p)a > 0, \\ &\mathsf{E}[\log(1+X)] = p\log(1+b) + (1-p)\log(1+a) < 0. \end{aligned}$$



考察函数$f(x) = \log(1+x)$, 通过两个点$(a, \log(1+a))$ 与$(b, \log(1+b))$ 作直线ℓ, 方程为

$$\frac{y - \log(1+b)}{x - b} = \frac{\log(1+b) - \log(1+a)}{b - a}.$$

这条直线在区间(a, b) 上位于在$y = \log x$ 之下.
对于$p \in (0,1)$, $pb + (1-p)a \in (a, b)$, 它在直线ℓ 上对应的点恰是$\mathsf{E}[\log(1+X)]$, 实际上,

$$\begin{aligned} y &= \log(1+b) + (pb + (1-p)a - b)\frac{\log(1+b) - \log(1+a)}{b - a} \\ &= p\log(1+b) + (1-p)\log(1+a) \\ &= \mathsf{E}[\log(1+X)]. \end{aligned}$$

直观地可以看到, 在$x > 0$ 时, $f(x) > 0$, 但直线上有一段却小于0, 这就是我们要找的陷阱区. 现在我们就把这一段区间先找出来. 通过解方程得到直线与$y = 0$ 的交点的x-坐标

$$x_0 = b - \frac{(b-a)\log(1+b)}{\log(1+b) - \log(1+a)} \in (0, b).$$

现在我们只需要找到p 使得$pb+(1-p)a \in (0,x_0)$ 就可以了. 实际上, 解方程

$$0 < pb + (1-p)a < x_0$$

得

$$0 < \frac{-a}{b-a} < p < \frac{x_0 - a}{b-a} < 1,$$

这时陷阱条件满足: $\mathsf{E}[1+X] > 1$ 但$\mathsf{E}[\log(1+X)] < 0$, 投资陷阱现象出现.

投资陷阱存在的结果是, 大多数投资是失败的, 但是投资成功的人会非常赚钱, 财富向少数人聚集, 这样造成贫富差距会越来越大.

第十章　随机商品与定价

现代社会是商品社会, 不仅有普通商品, 如食品, 消费品, 有服务品, 还有金融商品. 典型的金融商品是彩票, 银行存贷款, 股票, 保险, 期货, 衍生证券等等. 金融商品有个特点, 它的价值通常的随机的, 例如我们花10 块钱买一张彩票, 彩票没中奖的话一钱不值, 中奖的话可能价值千万; 再例如买财产保险, 如果家里一直平安, 它似乎不值钱, 但是一旦家里遭遇火灾, 保险就会体现价值; 股票更是如此, 它的价值一直在波动, 如果你能低价时买入高价时卖出, 那就可以赚钱.
随机商品最早的雏形就是赌博, 现在它虽然改头换面, 称为投资, 或称为避险, 但是本质上还是赌博, 赌你对未来的预测是否准确.
随机商品为什么会出现? 这是一个经济学问题. 随机商品出现的原因主要是人们对风险偏好的不同, 或者说人与人之间不仅需要交换通常的生活品, 还需要交换风险, 或者说交换不确定性. 不喜欢风险的人拿钱买保险, 降低风险, 把风险转移给那些喜欢风险的人.
通常商品的价值通常是通过成本或者需求来计算, 那么随机商品的价值怎么计算呢?

§10.1　公平

在一个典型的赌博中, 输赢是随机的, 输的人按照事先约定的规则付钱给赢的人, 因此对于赌博中一个固定的个体来

说, 他的所得是随机变量, 记为X, 正表示赢, 负表示输.

众所周知, 赌博的一个基本准则是公平. 什么是公平?

例如两个人赌博, 如果每次输赢的可能性一样, 那么每次的赌注是一样的. 如果每次的输赢可能性不同, 那么赌注就不同. 例如掷硬币, 每次输赢一块钱是公平的; 如果甲乙掷骰子, 甲掷出6 才算赢, 其他算输, 那么每次输赢都是一块钱就不公平了, 怎么才算公平呢? 乙的赌注是甲的5 倍时才是公平的, 即输赢数的比例应该等于赢输的概率比.

精确地说, 假如甲赢和输的概率分别是p 和$1-p$, 赌注分别为x 与y, 那么甲如果赢, 得y, 如果输, 付x, 这样的随机变量X 的分布为

$$\mathsf{P}(X=y)=p,\ \mathsf{P}(X=-x)=1-p,$$

则当

$$\frac{y}{x}=\frac{1-p}{p}$$

时才是公平的, 实际上, 这就是说

$$\mathsf{E}[X]=py-(1-p)x=0.$$

因此, 公平就是输赢数的期望等于0.

由大数定律, 当期望等于0 时, 重复n 局赌局之后, 每一局的平均所得

$$\frac{X_1+X_2+\cdots+X_n}{n}$$

接近于零. 这也是公平性的体现.

设随机商品的价值是X, 那么按照上一节的公平准则, 我们需要付出的价格等于X 的期望$\mathsf{E}[X]$. X 的结果无法计算, 但

是期望很多情况下是可以计算的, 即使无法计算, 我们也可以利用在大数定律原理下的Monte Carlo 方法计算.

§10.2 风险与风险溢价

通常随机商品的价值是随机的, 会在未来某个时刻确定. 市场上各种随机商品的目的是不一样的, 例如:

1. 彩票的价值是0 或者一个确定的钱数, 赌场中玩一次角子机的价值是它可能吐出的钱数, 这这些场合, 商品的目的是提供不确定性, 它的价值是非负的;
2. 投资是指投入一笔资金, 以收取随机的收益, 付出的是资金的时间价值;
3. 汽车保险的价值是支付你下一年的车损, 目的是消除车损的不确定性, 价值是非负的;
4. 期货合约是在敲定时间与以敲定价格购买某种货品, 必须履行, 它的价值是货品的到期实际价格减去敲定价格的差, 它的目的是消除货品价格波动的风险;
5. 期权合约是以敲定价格购买某种股票的权利, 和期货不一定, 期权可以放弃, 所以它的价值是非负的.

我们说用数学期望来给随机商品定价是一个基本准则, 但是实际上不一定如此. 例如彩票的售价肯定高于彩票价值的期望的, 为什么? 因为卖彩票的结构有运行费用, 专门的彩票, 例如福利彩票和体育彩票还要拿出一部分作为福利

或者资助体育事业. 但是还是有很多人会去买, 因为这是一个投入很低的发大财的机会, 尽管机会很小. 保险是付钱把未来的损失转移给保险公司, 保险公司也一样要要收取公司的运行费用, 还要因为承担风险而收取费用. 投资也是如此, 投资是有风险的, 所以期望收益通常要高于银行存款利率, 作为对投资者愿意冒风险的奖励, 称为风险溢价.
随机商品的定价是有风险的, 以保险为例, 保险公司承担很高的经营风险, 如突如其来的飓风海啸等灾难引起的损失对保险公司可能都是致命的. 保险公司怎么降低风险呢? 一个办法是提高保费, 除此之外是扩大规模. 例如, 设n 个人购买了相同的车险, 价格是车损的期望m, 假设车损为独立同分布的随机变量$X_1, X_2, \cdots, X_n$, 那么平均车损为S_n/n, 其方差为$D[X_1]/n$. 因此平均车损的随机性随着n 的增加而减少, 也就是说保险公司减少风险的一个有效办法是增加n.

§10.3 对冲与无风险定价

一般原始随机商品的风险可以想办法减少, 但是永远无法消除. 然而对于一类随机商品来说, 无风险的定价是可能的, 这就是衍生证券. 衍生证券是指其价值的随机性是来自于其它风险资产的价值的那些随机商品, 例如期货合约价值依赖于货品的价格变动, 期权合约价值依赖于股票的价格变动. 现在我们构造一个简单的金融市场来解释这种定价.
简单的金融市场有两种金融产品, 一个是存贷款利率都为r

的银行服务, 称为无风险资产; 另一个是随机收益率X 的股票, 称为风险资产. 设X 只取两个值$(-1 <)a < b$, 其中b 代表涨, a 代表跌, 概率分别为p 与$q = 1 - p$.

假设市场无套利, 也就是说市场上没有无风险的赚钱机会. 这时必有

$$a < r < b.$$

事实上, 如果$r \le a$, 那么钱太便宜, 我可以从银行贷款买股票, 等下个时刻卖掉股票还钱, 不管股市涨跌, 我一定不会亏, 而且当股票涨的时候, 我可以赚钱, 没有风险. 反过来, 如果$b \le r$, 那么股市太低迷, 我卖空股票, 把钱存银行, 等下个时刻, 我从银行把钱取出来买来股票还回去, 不管股市涨跌, 我一定不会亏, 而且当股票跌的时候, 我可以赚钱, 没有风险. 因此这两种情况都说明市场有套利机会.

什么是期权呢? 设股票现在的价格是S, 到下个时刻可能涨至$S(1 + b)$ 也可能跌至$S(1 + a)$. 最简单的欧式买入期权是这样一份合约, 投资者有在下个时刻以敲定的价格K 购买一份股票的权利, 显然K 必须满足

$$S(1 + a) < K < S(1 + b),$$

当股票上涨时, 投资者执行这个权力, 从市场上以价格K 买一股, 然后以市场价$S(1 + b)$ 卖掉, 投资者得利$S(1 + b) - K$; 反之, 当股票下跌时, 股票的市场价比敲定价还要便宜, 执行期权是亏损的, 所以他应该选择放弃执行期权, 得利为0. 因此期权对投资者带来的价值等于$(S(1+X) - K)^+$, 其中a^+ 是这样一个记号, 当$a > 0$ 时, $a^+ = a$, 当$a \le 0$ 时, $a^+ = 0$.

一般的期权是X 的一个函数, 记为$V(X)$.
期权是个典型的随机商品, 但它的价值是非负的, 投资者必须付出相应的代价, 那么销售这个期权的机构应该怎么给期权定价呢? 通常的办法是使用大数定律, 定价就是价值的期望:

$$\mathsf{E}[V(X)] = pV(b) + qV(a).$$

前面说过, 这个定价是有风险的, 风险可以通过扩大销售的办法来降低, 但无法消除. 销售机构不想承担风险, 下面介绍使用对冲的方法来消除风险.
什么是对冲? 对冲是金融市场上规避风险的操作. 例如某公司半年后要使用外汇, 为了防止半年后外汇过度上涨, 公司先买一个该外汇的看涨期货, 那么半年后不管外汇市场怎么波动, 公司都不会有太大风险, 因为外汇涨的话期货赚钱, 外汇跌的话公司使用外汇便宜. 这样的操作就是对冲.
想象销售机构卖出了一份期权, 向投资者收取了金额x, 现在机构想要通过投资市场来对冲卖出的期权. 假设要购买y 份股票, 这里x, y 都是待定的. 投资的过程如下: 购买y 份股票所需资金yS, 机构手里有资金x, 剩下$x - yS$, 放入银行, 这个数字的正负分别代表存款和贷款. 到下个时刻, 机构的股票价值$yS(1+X)$, 银行存款$(x - yS)(1+r)$, 加起来是

$$yS(1+X) + (x - yS)(1+r).$$

冲抵期权的意思是投资得到的财富应该等于期权的价值

$$yS(1+X) + (x - yS)(1+r) = V(X).$$

这个是不是可能呢? 恰好可以, 事实上, 因为X 仅有两个状态: 涨或者跌, 所以上面的方程可转化为方程组

$$\begin{cases} yS(1+b)+(x-yS)(1+r)=V(b), \\ yS(1+a)+(x-yS)(1+r)=V(a). \end{cases}$$

这方程有两个未知数, 所以有唯一解

$$\begin{aligned} x &= \frac{1}{1+r}\left(\frac{r-a}{b-a}V(b)+\frac{b-r}{b-a}V(a)\right), \\ y &= \frac{V(b)-V(a)}{S(b-a)}. \end{aligned}$$

重复一遍, x 是期权的价格, y 是投资购买的股票份数, 可以成为对冲投资策略.

上面的公式称为期权定价公式. 这个公式粗略地解释了期权无风险定价的可行性. 投资者看好股票, 来买个期权, 自己无需设计投资策略而可以去度假, 机构拿了卖出期权的钱, 根据以上投资策略买好股票, 等投资者度假回来, 机构卖掉股票取出存在银行的钱正好是投资者买的期权的价值, 完美.

上面的公式有一个漂亮的解释. $V(b)$ 与$V(a)$ 的系数

$$p'=\frac{r-a}{b-a},\ q'=\frac{b-r}{b-a}$$

都是正的而且其和等于1. 这说明

$$\frac{r-a}{b-a}V(b)+\frac{b-r}{b-a}V(a)$$

是某概率(记为P')下$V(X)$ 的期望$\mathsf{E}'[V(X)]$. 前面的系数$1/(1+r)$ 是对应于利率的折现率, 即将来的钱在今天的价值要打个折扣. 所以x 是期权在某个概率下的期望之折现.

这个概率还有一个非常有趣的现象, 一般来说, 由于风险溢价, 股票的期望收益率应该高于银行利率, 但是, 在这个新的概率测度下, 股票市场是与银行相比在期望也没有优势, 即

$$\mathsf{E}'[S(1+X)] = S\left(\frac{r-a}{b-a}(1+b)+\frac{b-r}{b-a}(1+a)\right) = S(1+r),$$

这说明投资期望收益率与银行利率一样.
再注意几点

1. 因为X 只有两个状态, 所以方程有解;
2. 期权的定价公式与市场涨跌概率p 无关;
3. 无套利称为市场的有效性, 意味着信息对称;
4. 期权可以被对冲称为市场的完备性, 意味着市场活跃;
5. 这是市场的一个非常简单的抽象.

市场的有效性和市场的完备性可以通过一个简单的例子解释. 一个见多识广的城里人去深山未开发的地区玩, 被邀请在当地一个土著家吃饭, 看到他使用的碗是一个汉朝的古董, 市场上同样的东西可以拍卖到一百万. 他问土著这个碗卖多少钱. 这时有两种情况: 一, 土著完全不了解外面的世界, 说十块钱你就拿走吧. 这是说市场不有效, 土著没有充分了解信息. 二, 土著通过电视或者以前的顾客了解这个碗在市场上值很多钱, 但是因为没有有钱人接手, 这个碗一直卖不出去, 所以他泄气了, 说一千块钱你就拿走吧. 这是说市场不完备, 值钱的东西不一定能卖出去.

§10.4 期望与心理预期: 圣彼得堡悖论

概率论是研究与解释随机现象的数学理论, 从上面许多例子看, 概率论能够解释很多实际的概率问题, 所以大多数学者对概率论是非常满意的, 认为它就是真正的概率了.

但是, 还是有许多问题, 概率论的解释与人心理的感受不符, 特别是在概率很小的时候. 例如, 掷10 个硬币, 如果10 个都是正面, 拿10,000 元, 否则是零. 因为10 个都是正面的概率是千分之一, 所以这个游戏的期望是10 块钱, 也许10 块钱是小数, 那么放大100 倍, 你愿意花1000 块钱赌10 个硬币都是正面, 赢了拿100 万吗?

我想大多数人都不愿意, 因为大家心理感觉, 掷一次, 10 个硬币都是正面是几乎不可能的. 这大概就是数学期望和心理期望的差异吧, 概率论的思考是线性的, 人的心理是非线性的.

圣彼得堡悖论是个极端的例子, 更能清楚地说明期望与心理预期的差异. 它是Bernoulli 家族的Daniel Bernoulli, 于1738 年出版的, 但有记载说是他的堂兄Nicolaus Bernoulli 在1713 年9 月写给Daniel 的一封信中首先陈述的.

PROBLEM: 某赌场提供一种游戏, 掷一枚硬币, 直至掷出正面即停止, 如果这时掷的次数是n, 那么玩者得到2^n 块钱, 因为首次正面出现在第n 次投掷的概率是2^{-n}, 所以所得钱数X 的分布是

$$\begin{pmatrix} 2 & 2^2 & 2^3 & 2^4 & \cdots & 2^n & \cdots \\ 2^{-1} & 2^{-2} & 2^{-3} & 2^{-4} & \cdots & 2^{-n} & \cdots \end{pmatrix}.$$

这个游戏的价值期望为

$$\mathsf{E}[X] = 2 \cdot 2^{-1} + 2^2 \cdot 2^{-2} + 2^3 \cdot 2^{-3} + \cdots + 2^n \cdot 2^{-n} + \cdots = +\infty,$$

因此按照期望来确定价值的话, 这个游戏的价值是无穷.

再来看, 如果玩n 局, 每局所得记为$X_1, X_2, \cdots, X_n$, 它们是独立同分布的, 那么应用大数定律可以证明, 平均每一局所得

$$\frac{X_1 + X_2 + \cdots + X_n}{n}$$

也趋于无穷.

概率论的知识告诉我们, 只要是花有限的钱, 100, 1,000, 10,000, $\cdots$, 这个游戏肯定是值得的, 但是我们的内心会这么认为吗? 按我们的通常想法, 平均2 次正面就会出现1 次, 5 次之内正面不出现的概率就很小了, 10 次之内正面不出现的可能性几乎没有, 所以根据调查, 很少有人愿意出多于20 块钱玩这个游戏.

为什么期望的数学理论和心理预期会相差这么大? 这是概率界一直难以解释的问题. Daniel Bernoulli 用边际效用递减的想法来解释, 还有很多不同的解释, 但这些解释都不能从根本上解释为什么我们通常认为与实际符合得很好的概率论在这里与心理预期如此地不一致, 难道我们一代又一代的智者遗漏了什么吗?

第十一章　概率应用于推断

概率起源于各种游戏, 如赌博, 麻将, 纸牌等等, 也用于游戏, 用于推断这些游戏中的各种情况的可能性, 例如推断对手手里可能是什么牌. 应用概率的推断现在称为统计推断. 统计一词起源于国情调查，英文是statistics, 词根是国, 最早意为国情学. 一般来说，统计包括三个含义：统计工作、统计资料和统计科学. 统计工作、统计资料、统计科学三者之间的关系是：统计工作的成果是统计资料，统计资料和统计科学的基础是统计工作，统计科学既是统计工作经验的理论概括，又是指导统计工作的原理、原则和方法. 原始的统计工作即人们收集数据的原始形态已经有几千年的历史，而它作为一门科学，是从17世纪开始.

这里所说的统计是指基于概率论的统计, 是统计中统计科学的一个重要可以说是主要分支, 通常称为数理统计, 简称为统计. 统计推断是根据一些观察来推断一个随机现象或者一个数据体中的规律. 统计推断属于归纳推理, 即从特殊归纳到一般.

统计推断的通常表现形式是推断分布或者分布中的参数. 例如通过掷硬币来推断硬币是不是真的两面等可能, 通过分析以往的数据来推断之后几天的天气情况, 通过以往的数据来推断股市的涨跌, 等等. 统计推断的另一个重要表现形式是假设检验, 就是根据数据以及某个置信度来判断一个断言是否是真的, 例如检验硬币是不是等可能的, 检验三峡大坝对气候是否有影响, 检验空气质量对平均寿命的相

关性, 等等.

要注意的是, 统计推断具有随机性, 推断的结论依然是不确定的, 但这不是说推断是没有意义的. 关于这个问题, 统计学家Fisher 是这么说的: We may at once admit that any inference from the particular to the general must be attended by some degree of uncertainty, but this is not the same as to admit that such inference cannot be absolutely rigorous, for the nature and degree of uncertainty may itself be capable of rigorous expression. 翻译过来是说, 我们要承认任何从特别到一般的推断是有不确定性存在的, 但这不等于说这样的推断是不能绝对地严密的, 因为不确定的质与度本身是可以严密地表述的.

直到现在, 很多学者是不承认且极其反对归纳推理的, 认为归纳推理不严密, 可能因为滥用而导致严重的谬误, 但不可否认, 归纳推理推动了科学的发展, 是一种值得学习的推理方式, 同样不可否认, 像学者所担心的那样, 统计推断时常被严重滥用, 或者容易有意或者无意地被滥用. 甚至很多统计学者也这么认为.

因此在涉及写作和阅读一个统计推断时要特别小心, 作为一个统计学者, 你需要尽量准确地表述其中的不确定性的质与度, 分清楚其中的科学成分与统计成分. 而作为一个阅读者, 你需要尽量去理解结论所表述的不确定性的质与度, 理解其中的科学成分与统计成分. 这样就可能避免一些现在所说的“无意地被滥用”现象, 例如p-值困境.

.

§11.1 品茶女士与假设检验

在日常的生活中, 我们经常看到很多听起来很玄妙的争议, 例如连接放大器和音箱的线材对声音有没有影响, 用煤气烧的饭和用柴烧的饭在味道上有没有差别, 等等. 这种争议通常是各说各的, 没有一个明确的结论, 结果总是不欢而散. 原因是说话的人不知道怎么证明自己和驳斥别人.

1920 年代后期夏天的某个下午在剑桥大学的校园里就有这样一个争议, 一个真实但带有点八卦的故事. 一些绅士, 大学教授, 以及他们的夫人们坐在一起喝下午茶, 就是红茶和牛奶的混合体: 英国奶茶. 一位女士提出了一个观点, 把茶倒在牛奶中得到的奶茶比把牛奶倒在茶中得到的奶茶味道更好. 这个观点立刻引来那些具有科学精神的男士们的反击, 他们嘲笑说, 这能有什么区别呢? 从化学角度看, 都是茶和奶的混合体而已. 学者的争议不同于普通人的争议, 马上一位戴着厚眼镜带尖髯胡的男士就提出一个实验来测试, 让那位认为有区别的女生来品尝一系列的茶, 其中有些是茶倒入奶中, 另一些是奶倒入茶中. 想到这样的实验不难, 普通人也许也能想到, 难的是怎么专业地解释实验得到的结果.

PROBLEM: 怎么设计一个实验并按实验的结果来理由充足地判断这位说自己可以区分奶茶味道的女士所言是真是假?

科学的发现往往来自于偶然的小问题. 女士能不能区分这两种茶的味道这个问题不是什么太有意义的问题, 有意思

的是不是能找到一个有效的方法来判断这位女士说的是不是对的, 很快, 男士们就开始讨论怎么来做出判断. 但是, 稍微思考一下就知道, 科学研究中有同样的问题, 例如不同的肥料配方对于农业收成的影响, 什么样的植物成分对疾病有治疗效果, 等等. 因此, 小问题的讨论可以应用于大问题.
上面的品茶故事记载在于2001 出版的《品茶女士》一书中的一开始，书中提到的这位戴厚眼镜尖髯胡的男士是英国著名统计学家Sir Ranold Fisher. 有趣的是, 在1935 年, Fisher 出版了一本题为《实验设计》的书, 在此书的第二章的第一句话是这样的: A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. 他一开始就讲述了上述女士品茶故事中所提到的实验, 讲述他怎么设计实验且以什么理由确定这位女士是不是真的可以区分两种方法做出来的奶茶的味道. Fisher 实际上是利用这个实验来讲解他关于假设检验的思想.
具体地说, 在书中, Fisher 设计了这么一个现在大学生熟知并称为假设检验的实验, 即提出一个假设, 然后用实验的数据来看是不是有足够的信心推翻它. 他设计的实验如下: 准备8 杯茶, 4 杯是先倒奶, 然后茶倒入奶中, 另4 杯是先倒茶, 然后奶倒入茶中. 8 杯茶依一个完全随机的顺序递给女士品茶, 把这些信息告知女士, 然后要求女士通过品尝把4 杯茶倒入奶中的奶茶识别出来. 然后分析怎么按照这个女士可能给出的各种答案来推断女士是否真的可以区别茶的味道.

Fisher 设计的实验并不是要证明女士能够区分奶茶味道, 而是想要推翻下面的假设, 称为零假设(或者原假设):

$$H_0: \text{女士是随机地从8 杯茶中选择4 杯的.}$$

如果零假设成立, 那么从8 杯茶随机地挑出4 杯茶有$\binom{8}{4} = 70$ 中不同的等可能的选择. 用X 表示她选择正确的杯数, 那么X 的分布是

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{1}{70} & \frac{16}{70} & \frac{36}{70} & \frac{16}{70} & \frac{1}{70} \end{pmatrix}.$$

即挑出的4 杯茶完全正确的机会是1/70. 在这里, Fisher 提出一个显著性的概念, 它认为只要在零假设下发生的概率小于1/20 的事件就应该被认为是实际上不可能发生的. 因此, 如果它在某个实验中真的发生了, 那就说明假设不对, 所以应该可以安全地"拒绝"零假设了. 这也是显著性检验的意义.

在这个实验中, 也就是说, 如果最后的结果是品茶女士正确地挑出了4 杯茶倒入奶中的茶, 那么Fisher 的推理是这样的: 因为这个事件在零假设下发生的概率是1/70, 远远小于显著性的指标1/20, 所以就可以拒绝零假设, 也就是说拒绝女士是随机地选择的假设. 如果品茶女士准确地识别了3 杯, 1 杯错, 在零假设下, 这个发生的可能性是16/70, 远远超过1/20, 没有显著性, 也就是说, 这个事件的发生概率虽然不大, 但也并不让人觉得不可接受. 因此我们不能拒绝零假设. 这时我们会问, 能不能接受零假设呢?

Fisher 本人对此是谨慎的, 他认为实验可以否定零假设但不能肯定, 他写到: It should be noted that the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis. “零假设永远无法证明或者确立，但是可以通过实验手段被否定. 可以这么说，每个实验存在的目的是为了给事实提供一个机会去推翻零假设. ”

与零假设相反的假设称为备择假设, 零假设与备择假设在数学上是没有区别的, 但在实际应用时, 零假设一般是想要推翻的假设. 如果上面的零假设被实验结果拒绝, 那么接受备择假设, 就是说女士并不是随机地分辨奶茶的, 或者说对两种做法的奶茶有一定的识别能力. 另外, 零假设需是确切的, 即在零假设下, 所考察的随机现象的分布是确定已知的. Fisher 还是以女生品茶为例来说明, 我们可以把女士可以正确地识别茶的味道作为零假设, 这时分布是确定的, 即$\mathsf{P}(X = 4) = 1$, 完全识别的概率为1, 因此只要有一杯茶识别错误就可以推翻这个假设, 但是我们永远不能用有限次实验证明这个假设成立. 我们能不能把女士有一定的把握识别茶的味道当作零假设呢? 不能, 因为在这个假设下X的分布是不确定的.

Fisher 的这本书极大地推动了统计在科学研究中的应用, 在Fisher 之前, 统计通常意味着只是收集数据, 统计很少会产生新的知识, Fisher 的书教给我们怎么读数据以及怎么从数据里获得新知识的一个规范性的方法. 到现在, 统计是科学研究的重要工具, 统计也是我们生活不可或缺的部分, 用数据说话慢慢成为我们的习惯.

最后提一句, Fisher 这本书中并没有提到实验的最终结果, 因为这对于他的书来说不重要. 但是据当时在场的其他人说, 这位女士果然胸有成竹, 准确地识别出所有的茶, 不管是奶倒入茶中还是茶倒入奶中.

§11.2 对于推断的不同观点: 频率与Bayes

统计学界对于统计推断的认识还是有分歧的, 主要分为频率学派与bayes 学派. 频率学派建立在 Kolmogorov 的公理体系上, 认为概率(或者其它参数)是一个事件的属性, 也许不可知, 但是它是一个常数. 大样本(足够多的数据)是估计概率(或者其它参数) 的必要条件, 小样本情况下做的统计是没有实际意义的, 主要工具是大数定律; Bayes 学派认为, 没有所谓的固定不变的概率, 概率是一种以主观认识开始, 然后根据观察或者经验而不断修正的量, 这种认识特别符合现代人工智能的学习理论, 因此Bayes 学派从一个极端的少数派渐渐地大众化.

§11.2.1 Bernoulli 的大数定律

统计推断的第一个思想来自于Bernoulli 的大数定律, 假设一个随机试验中的一个事件A 发生的概率是p, 然后重复这个试验, 且用p_n 表达该事件在n 重复试验中实际发生的频率, 则对任何的$\varepsilon > 0$,

$$\lim_n \mathsf{P}(|p_n - p| > \varepsilon) = 0.$$

这里频率是个随机变量, 每次重复试验时可能都不同, 也依赖于重复次数, 概率是一个确定的数, 结论是说频率越来越

稳定, 概率是频率的极限. 确切地说, 把$|p_n - p|$ 看成误差, 那么误差大于某个确定数的可能性趋向于零.

也就是说, 我们可以透过频率这么一个可能观察的量, 或者可以从已知数据中获得的量, 来推断概率这么一个知道它存在但无法确定的量. 但可能性趋向于零, 不是说没有可能性, 所以大数定律的推断是无法彻底消除误差的, 误差大的可能性总是存在, 这个可能性当样本量n 不大或者当误差ε 要求很小的时候会比较大, 当样本量n 较大且精度要求不高时会比较小.

无论怎么样, 大数定律是统计推断最重要的方法与思想来源, 应用Bernoulli 思想来进行推断的统计学家通常称为频率派或者大样本派, 意指在样本量大的时候才有实际意义.

历史上有很多用投币来检验大数定律的例子. 在掷一枚硬币时，既可能出现正面，也可能出现反面，预先做出确定性的判断是不可能的. 但是假如硬币均匀，人们会相信出现正面的可能性与出现反面的可能性应该大体相等，即在大量试验中出现正面的频率应接近于 50%. 为了验证这一点，历史上曾有不少人做过实验，下面是用掷硬币来验证频率稳定性的著名例子.

实验者	掷硬币次数	出现正面次数	频率
蒲丰	4040	2048	0.5069
费勒	10000	4979	0. 4979
皮尔森	12000	6019	0.5016
皮尔森	24000	12012	0.5005

援引它不仅因为它的权威性，而且因为它简单，人人都可

以做. 从此表可以看到，当试验次数增大时，频率接近 1/2. 同时还可看到，说出现正面的概率是 1/2, 并不是意味着掷 2 次硬币必出现 1 次正面，掷 10 次必出现 5 次正面，而是在大量试验中，出现正面的频率值越来越稳定在 1/2 的附近. 频率在大数次的试验中稳定于某一常数 (概率) 这个事实非常重要. 正因为如此，在实际中可以把频率作为概率 (的估计值) 来应用. 频率也称为**经验概率**，计算它通常是为了估计概率.

§11.2.2 Bayes 公式及其推断思想

很多学者认为Bernoulli 大数定律给出的用频率推断概率并不是真正的推断, 因为大数的要求是先有概率, 然后断言当样本量充分大时, 以很高的可能性, 频率离概率很近, 所以这是从概率推断频率, 而我们真正需要的是以频率推断概率. 另外在通常的数据条件下, 样本量是固定的, 我们不知道它是否已经达到频率离概率很近所要求的可信度.

那么, 怎么才算是用频率推断概率呢? 想要达到这个目的, 首先要改变我们所认为的“概率是随机事件的一个属性”的观点. Bayes 认为, 随机现象的出现是由于人类的认识水平的局限性，简单地说，是由于人类的无知. 随着人类对世界的认识越来越多, 随机现象会越来越少, 例如古代人们肯定认为日食月食以及彗星的出现是随机现象, 是上帝的某种告诫, 但现在人们已经可以准确地预测这类天文现象, 因此它们不再是随机的.

让我们回顾Bayes 公式, 一个事件A 发生的概率$\mathsf{P}(A)$ 称为先

验概率, 当其它一个与之相关的事件B 发生后, 根据这个新获得的信息, 人们对于A 的可能性大小的认识就不同了, 变成条件概率$\mathsf{P}(A|B)$, 而

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(B|A)\mathsf{P}(A)}{\mathsf{P}(B|A)\mathsf{P}(A) + \mathsf{P}(B|A^c)\mathsf{P}(A^c)}.$$

这就是Bayes 公式, 它的基本思想是随机性随着新的信息的进入会改变, 这也就是Bayes 推断的思想: 用后续发生的事件B 来推断地修正预先对事件A 的随机性的认识. 为了这样的逆推断的以实施, 我们在一开始必须对A 的随机性有一个主观的假设.

例如, 我们通常认为一个硬币出现正面的概率是确定的, 尽管我们无法知道. 但Bayes 推断的观点认为不一定, 也许我们是从一个有100 枚质地均匀硬币和100 枚正面概率为2/3 的罐子中随机抽取的, 所以硬币以1/2 的可能性是公平的, 1/2 可能性是有偏的, 出现正面概率为2/3. 这是对于硬币正反面随机性的先验认识, 也就是我们对于硬币随机性的主观假设.

现在假设A 是说硬币公平的, B 是某人掷5 次该硬币得到2 次正面, 那么$\mathsf{P}(B|A)$ 与$\mathsf{P}(B|A^c$ 分别是硬币正面概率为1/2 与2/3 的情况下的一个二项分布, 根据Bayes 公式

$$\begin{aligned}
\mathsf{P}(A|B) &= \frac{\mathsf{P}(B|A)\mathsf{P}(A)}{\mathsf{P}(B)} \\
&= \frac{\binom{5}{2}(1/2)^5 \cdot 1/2}{\binom{5}{2}(1/2)^5 \cdot 1/2 + \binom{5}{2}(2/3)^2(1/3)^3 \cdot 1/2} \\
&= \frac{1}{1 + \frac{2^7}{3^5}} \approx 0.655.
\end{aligned}$$

这说明, 在利用获得的信息修正之后, 硬币是公平的概率是0.655, 不公平的概率为0.345, 这称为是后验概率. 如果你继续掷硬币, 这个概率可以继续被修正. 这个过程就是Bayes推断的全部想法: 用频率推断概率.

在这个例子中, 所做的假设, 硬币以相同的可能性可能公平或者不公平, 是有直观理由的, Bayes 的思想走得更远一点. 拿来一个不知道从哪个罐子取的硬币, 这时我需要主观地假设它的正面概率有一个先验的分布, 或者说, 假设概率是具有某个分布的随机变量.

这个想法看上去有点奇怪, 但是我们先不要多想, 顺着这条路继续走下去. 如果我对硬币一无所知, 那么假设它出现正面的概率是一个$(0,1)$ 上的均匀分布可能是个最好的选择, 即

$$\mathsf{P}(p \in (a,b)) = b-a,\ 1>b>a>0.$$

PROBLEM: 用这个硬币掷n 次得到m 次正面, 这时怎么推断第$n+1$ 次是正面的概率?

解: 用B 表示掷n 次得到m 次正面这个事件. 当$p=x$ 时, B 的概率是$\binom{n}{m}x^m(1-x)^{n-m}$, 由全概率公式

$$\mathsf{P}(B) = \int_0^1 \binom{n}{m} x^m (1-x)^{n-m} dx.$$

怎么严格地证明? 先离散化p, 我们可以认为p 是均匀地分布在下面k 个值上

$$\frac{1}{k}, \frac{2}{k}, \cdots, \frac{k-1}{k}, \frac{k}{k}.$$

那么全概率公式推出

$$\begin{aligned}\mathsf{P}(B) &= \sum_{j=1}^{k}\mathsf{P}(B|p=\frac{j}{k})\frac{1}{k}\\ &= \sum_{j=1}^{k}\frac{1}{k}\binom{n}{m}(j/k)^m(1-j/k)^{n-m}.\end{aligned}$$

让k 趋于无穷, $\mathsf{P}(B)$ 的极限便是上述积分.
为了计算$\mathsf{P}(B)$ 右边的积分, 利用分部积分公式得

$$\begin{aligned}&\int_0^1\binom{n}{m}x^m(1-x)^{n-m}dx\\ &=\int_0^1\binom{n}{m}(1-x)^{n-m}\frac{dx^{m+1}}{m+1}\\ &=\int_0^1\binom{n}{m}\frac{x^{m+1}}{m+1}(n-m)(1-x)^{n-m-1}dx\\ &=\int_0^1\binom{n}{m+1}x^{m+1}(1-x)^{n-m-1}dx,\end{aligned}$$

即积分与m 无关, 而m 可能的取值是0 到n, 所以

$$\mathsf{P}(B)=\frac{1}{n+1}.$$

因此, 在已知B 发生的条件下, 概率p 的后验密度函数是

$$\mathsf{P}(p\in dx|B)=(n+1)\binom{n}{m}x^m(1-x)^{n-m}dx.$$

用A_{n+1} 表示第$n+1$ 次掷硬币是正面这个事件. 由Bayes 公式, 以及上面的积分计算结果得

$$\begin{aligned}\mathsf{P}(A_{n+1}|B) &= \frac{\mathsf{P}(B\cap A_{n+1})}{\mathsf{P}(B)}\\ &= \frac{\int_0^1\mathsf{P}(B\cap A_{n+1}|p=x)dx}{\frac{1}{n+1}}\end{aligned}$$

$$
\begin{aligned}
&= (n+1)\int_0^1 \binom{n}{m} x^{m+1}(1-x)^{n-m}dx \\
&= (m+1)\int_0^1 \binom{n+1}{m+1} x^{m+1}(1-x)^{n+1-(m+1)}dx \\
&= \frac{m+1}{n+2}.
\end{aligned}
$$

实际上, 概率p 的后验期望与上述条件概率一样

$$
\mathsf{E}(p|B) = \frac{m+1}{n+2}.
$$

Bayes 推断在字面上看上去确实要比Bernoulli 大数定律导出的推断要更像从频率推断概率, 因为它像一个学习与认识的过程, 用发生的事情的信息一步步地来把握随机现象的奥秘, 因此它更适合于现在热门的机器学习和人工智能的年代.

§11.3 频率学派与Bayes 学派的争议

统计是一门源起于应用的学科, 通常需要满足应用的有效性以及逻辑的合理性. 从应用的有效性讲, 频率学派简单, 理论发展已经相当完备. 例如在上面的例子中, 对于频率学派来说, 不管硬币从哪里来, 频率m/n 就是概率p 的估计. 只是如上面所说, 推断的逻辑合理性有欠缺, 它是利用频率接近概率的原理来推断概率, 从逻辑上讲, 先有概率, 再有频率, 所以Bayes 学派说频率学派是概率推断频率不是所需要的从频率推断概率, Bayes 推断才是真正的统计推断.

但Bayes 学派也有问题, 首先把概率看成随机变量不符合Kolmogorov 的公理概率思想, 也就是说, 统计成为了一个可以用但没有

公理支撑的学科, 这对很多完美主义的学者来说是不可接受的; 其次, Bayes 推断要有一个开始, 需要确定先验分布. 概率的先验分布是根据实际感受选取的, 但经常是主观的, 不同的研究者很可能选取不同的先验分布, 最终的结果依赖于该分布的选取, 这与科学研究所追求的客观性矛盾. 这应该是Bayes 学派最受攻击的弱点; 最后, Bayes 估计的计算通常比频率学派的估计复杂和难算得多, 例如上面例子中的计算, 尽管最终的答案$(m+1)/(n+2)$ 在样本较大时与m/n 差不多, 但计算量大得多, 而且这个答案是在先验分布是最简单的均匀分布的假设下获得的. 另外计算的难度与先验分布的选取有关, 对于一些复杂的先验分布来说, 也许根本不可能显式地计算. 当然这个问题在实际应用上因现代的计算能力大幅度提高故不是太大的问题.

Bayes 学派也批评频率学派的估计对样本量有要求, 不能在小样本的情况下进行统计推断, 这限制了统计的应用, 因为很多统计问题是不可能收集到足够量的数据的, 例如地震, 飓风, 飞机失事等灾难. 而频率学派的回应是, 从本质上讲, 统计应该建立在大样本基础上, 小样本的统计是没有意义的, 它只会误导结论, 更容易被滥用. 顺便说一句, 作者本人的观点也是这样.

最后, 自然科学或者其他领域的研究中应用统计方法是非常常见的且已经有过度依赖的倾向. 我们应该对统计推断是否给我们新的知识持有怀疑态度. 统计推断的结论, 应该有其它相应学科的原理进行科学解释之后才是新知识, 否则它应该被打个问号.