

第 3 章

古典概率与概率论的公理化

形而上者谓之道，形而下者谓之器，化而裁之谓之变，推而行之谓之通。
——《易经·系辞》

自然界和人类社会中的现象本质上可分为确定性的和不确定性的（又称随机性的）两类。确定性的现象可以在某些条件下预言它是否发生：若发生则称之为必然事件 (certain event)，否则称之为不可能事件 (impossible event)。

□ 必然事件，如：

- (i) 一个标准大气压下，纯水在 100°C 沸腾；
- (ii) 物体在无外力作用下速度保持不变；
- (iii) 光线通过引力场将发生偏移。

□ 不可能事件，如：

- (i) 太阳从西方升起；
- (ii) 欧氏几何中的三角形两边之和小于第三边。

显然，必然事件的否定就是不可能事件，反之亦然。读者以往学过的数学、宏观物理学、初等化学等基本上都是研究这类确定性现象的，但宇宙万物间具有绝对确定性的现象少之又少，人类更多面对的是不确定性的现象，又称随机现象。

对于随机现象，它们在一定的条件下可能发生也可能不发生，在得知其发生与否之前，我们称之为随机事件 (random event)，简称事件，如：

- (1) 抛一枚均匀的硬币出现正面；
- (2) 一个均匀的骰子掷出奇数点；
- (3) 明年三月份交货的黄金期货价格为每盎司 765.50 美元；
- (4) 未来十年全球温度将持续上升；
- (5) 某特效药能治愈某人的胃癌等。

如何研究这些随机事件呢？传统的做法是通过多次的随机试验 (random trial) 来揭示隐藏在大量观测结果背后的规律。虽然每次试验的结果都不确定，而且少量试验也看不出什么规律，但随着试验次数的增加，那些隐藏着的“必然性”就会逐渐浮现出来。

例如, 对“抛一枚均匀的硬币出现正面”这一随机事件, 我们采用的随机试验是“在相同条件下抛该枚硬币”, 只要抛足够多次, 出现正面的次数与抛次之比就必然稳定在 $1/2$ 附近。

有人可能质疑: 既然在相同条件下抛硬币, 出现的结果应该是一样的, 哪里有随机性可言? 事实上, “在相同条件下”这一要求并不能完全达到, 由于技术上或能力上的局限, 总有一些人为不可控制的因素影响着试验的结果, 譬如地球引力的微小变化、气流的轻微扰动、抛硬币者的心理波动等, 况且抛硬币动作本身也不可能达到绝对精确的重复。

为了验证抛一枚均匀硬币足够多次以后出现正面的频率会呈现一定的规律性, 历史上有多位充满好奇心的学者, 如英国数学家奥古斯塔斯·德摩根 (Augustus de Morgan, 1806—1871), 法国博物学家、数学家、生物学家布丰 (Comte de Buffon, 1707—1788), 美国数学家威廉·费勒 (William Feller, 1906—1970), 英国统计学家卡尔·皮尔逊 (Karl Pearson, 1857—1936) 等, 都亲手做过抛硬币的随机试验 (表 3.1)。

表 3.1 历史上一些知名学者做过的抛硬币试验的结果

试验者	抛次	正面次数	正面频率
德摩根	2 048	1 061	0.5181
布丰	4 040	2 048	0.5069
费勒	10 000	4 979	0.4979
皮尔逊	12 000	6 019	0.5016
皮尔逊	24 000	12 012	0.5005

为了研究随机现象中的数量规律, 需要概率论这一数学分支。为了使问题能够得到形式化的描述, 概率论的研究要求。

(1) 随机试验 \mathcal{E} 所有可能的结果组成的集合 Ω 是已知的, 我们称之为基本事件集合。 Ω 中的任一元素 ω 称为一个基本事件 (elementary event) 或样本点 (sample point), 记作 $\{\omega\}$ 或 ω (在不引起歧义的情况下)。例如, 抛硬币试验的基本事件集合 $\Omega = \{\text{正面}, \text{反面}\}$, 其中 $\{\text{正面}\}$ 和 $\{\text{反面}\}$ 都是基本事件。

(2) 在相同条件下, 随机试验 \mathcal{E} 可以不断重复。对于那些无法重复的随机试验, 如“某特效药能治愈某人的胃癌”, 可以适当地修改条件, 把与此人病情、生理、生活规律、工作环境等相似的服用此药的其他胃癌患者也作为观察对象, 只要试验结果对适当修改的条件不太敏感, 研究者依然可以从这些“重复性试验”中寻找规律来预测该特效药能否治愈此人的胃癌。

如何用数学的方法形式地表示随机事件呢? 以“掷骰子出现奇数点”为例, 掷骰子的基本事件集合是所有可能出现的点数, 即 $\Omega = \{1, 2, 3, 4, 5, 6\}$, 其中出现奇数点的所有可能结果是 $A = \{1, 3, 5\}$ 。如果“掷骰子出现奇数点”这一随机事件发生了, 骰子的点数必定是集合 A 中的某一个。很自然地, 人们用集合 $A = \{1, 3, 5\}$ 来表示“掷骰子出现奇数点”这一随机事件, 用骰子实际掷出的点数是否属于 A 来判定该随机事件是否发生 (譬如, 骰子被掷出的点数是 2, 则随机事件“掷骰子出现奇数点”没有发生)。像 $A = \{1, 3, 5\}$ 这样由不少于两个基本事件构成的随机事件, 我们称之为复合事件 (composite event)。

任一随机事件都可用基本事件集合 Ω 的某个子集来表示, 于是, 集合论理所当然地成为概率论的数学基础。后文中, 凡提到随机事件, 都用集合来表示。

在 Ω 的所有子集中, Ω 自身和空集 \emptyset 是两个极端的例子。全集 Ω 包含了随机试验所有可能的结果, 不管试验结果如何, 事件 Ω 总是发生的, 显然 Ω 表示一个必然事件。而空集 \emptyset 不包含任何元素, 所以它用来表示不可能事件。

例 3.1 抛一枚均匀的硬币两次, 若用 H 表示正面 (head), T 表示反面 (tail), 则基本事件集合是 $\Omega = \{(T, T), (T, H), (H, T), (H, H)\}$, 基本事件的个数是 $|\Omega| = 4$ 。

Ω 共有 $2^{|\Omega|} = 16$ 个子集, 因此, 其所有子集构成的幂集合 (power set) $\mathcal{P}(\Omega)$ 含有 16 个元素。其中, 元素 $\{(T, T), (H, H)\}$ 表示复合事件 “两次抛出的结果相同”; $\{(H, T), (T, H), (H, H)\}$ 表示复合事件 “至少抛出一个正面”, 简记作 $\{HT, TH, HH\}$ 。

基本事件并不要求机会等同, 如抛一枚不均匀的硬币, 出现正面和反面的机会不等。再如, 图 3.1 所示的转盘的例子: 转盘的基本事件集合是 $\Omega = \{1, 2, 3\}$, 所有事件的集合是 2^{Ω} 。

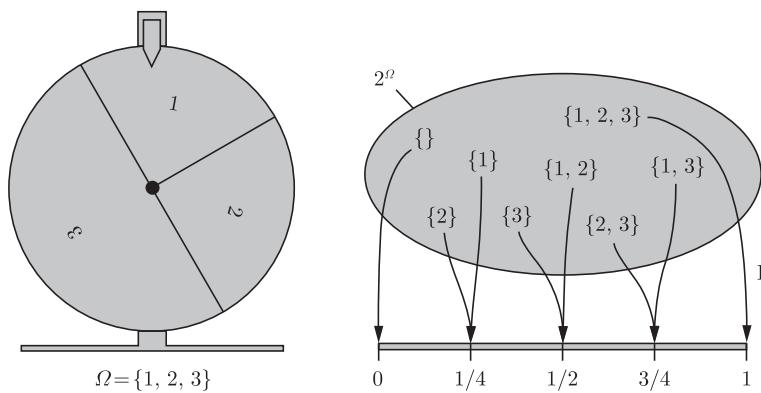


图 3.1 转盘的事件及其概率

在本质上, 一个随机事件是否发生是无法预测的。在试验之前, 我们不可能精确预知一个均匀骰子掷出的点数, 也就不可能预知 “掷出奇数点” 是否发生。需要澄清的是, 本质上无法预测和复杂得难以预测是截然不同的。下面两个例子有助于更好地理解随机性, 它们所谈论的都不是真正意义上的随机性, 虽然看上去很像。

※例 3.2 确定的动力学系统有时也可以表现出貌似的 “随机性”, 即混沌 (chaos), 它是复杂系统 (complex system) 的研究对象。混沌的本质是系统在经过长期演化后对初始条件变得敏感, 以至于 “差之毫厘, 失之千里” [64, 65]。

譬如, 在气象学方面, 理论上已经证明利用动力学模型精确预报两三周后的天气情况是不可能的, 一个形象的比喻是所谓的 “蝴蝶效应” (butterfly effect)——巴西一只蝴蝶偶尔扇动几下翅膀将引发数月后美国德克萨斯州的一场龙卷风。

以函数 $f(x) = 2x^2 - 1$ 的迭代为例, 考虑 $[-1, 1]$ 区间上的动力系统: 初值 x_0 设为 0.4 (实线) 和 0.4001 (虚线), 图 3.2 是 100 次迭代的结果的折线图。

$$\begin{aligned} x_1 &= f(x_0) \\ x_2 &= f(x_1) \\ &\vdots \\ x_{100} &= f(x_{99}) \end{aligned}$$

初始值的小扰动 0.0001 在刚开始迭代的一小段时间内并没有引起函数值太大的变化。然而, 在长期迭代后却引起了迭代结果貌似随机的变化, 它们看上去是非周期的、不规则的和无法预测的。但这种 “随机性” 与掷骰子、抛硬币等有着根本的区别: 混沌系统的短期表现 (即最初的几次迭代) 是可

知的，然而掷骰子在任何时候都是不能精确预测结果的。所以，混沌系统的“随机性”不是概率论意义上的随机性。

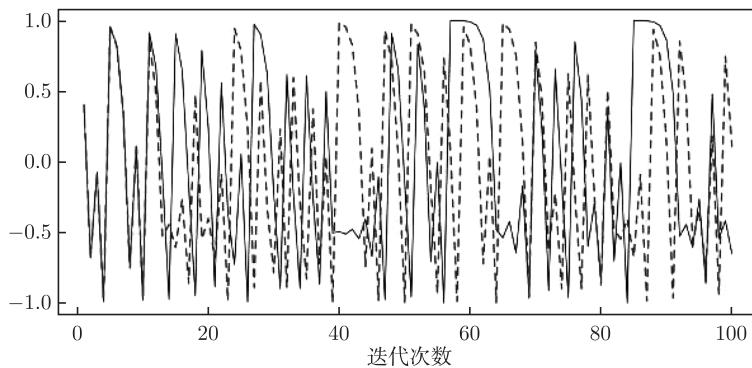


图 3.2 初值的小扰动在经过一定时间的演化后引发了混沌系统的“随机”变化

研究确定性的数学工具也可以用来研究不确定性。例如，前言中邮票所示的分形（具有自相似性的几何对象，没有随机性）与布朗运动（随机过程的一种，见 §8.2.3）相结合的分形布朗运动可用于描述山脉、云层、地形地貌等自然界的不规则形状（见第 392 页的图 8.9）。再如，热传导方程可用于解一些随机微分方程（见 §8.3.2）。

※例 3.3（阿诺尔德变换） 下述平面上的可逆周期映射被称为阿诺尔德变换。

$$(x, y)^\top \mapsto (2x + y, x + y)^\top \bmod 1$$

图 3.3 是俄罗斯数学家弗拉基米尔·阿诺尔德 (Vladimir Arnold, 1937—2010) 的头像经过阿诺尔德变换数次迭代后的结果。不难看出，阿诺尔德变换可以把一幅单位正方形的图像置乱，经过若干次迭代后，图像看上去是随机生成的，但是经过一定周期后原始图像又得以恢复——这一特点让阿诺尔

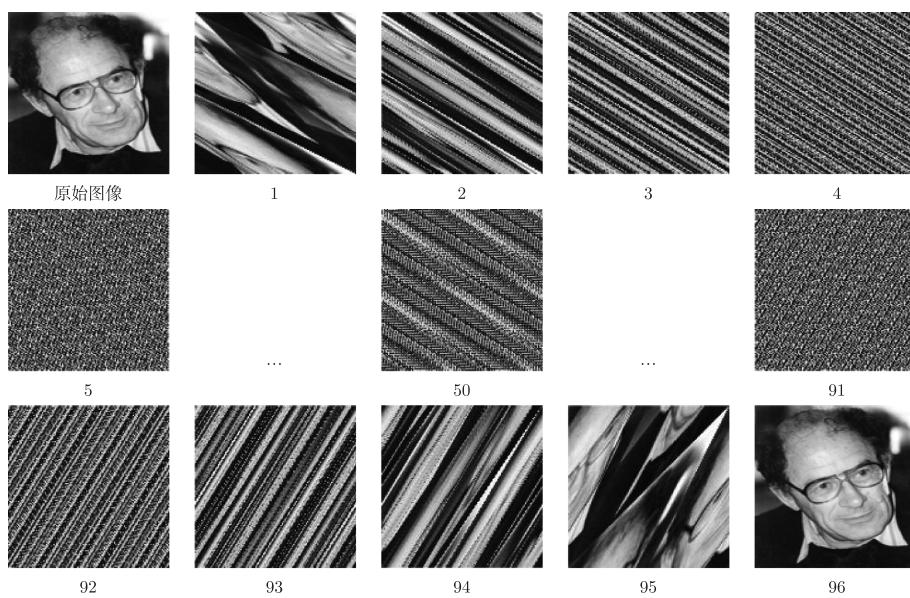


图 3.3 阿诺尔德变换

德变换常用于信息隐藏和图像加密。

在图 3.3 中, 原始图像经过 96 次迭代后又重新得到恢复, 在这一过程中, 该动力系统似乎表现出“随机性”, 其实不然——第 n 次迭代的结果总是第 $n-1$ 次迭代的结果按照固定的模式分割拼装而成。换句话说, 如果我们知道第 $n-1$ 次迭代的结果, 就可以准确无误地构造出第 n 次迭代的结果。

第 3 章的关键概念

知识的积累如同织网, 只有把知识点连接起来才能形成牢不可破、牵一而动的体系, 否则零散孤立的知识很容易被遗忘。我们用有向图来描述内容对象之间的关系, 这里对象可以是概念、方法或结果, 其中椭圆节点 (特别地, 圆形节点是拉普拉斯重点论述过的) 是最为关键或重要的 (图 3.4)。节点之间的关系是:

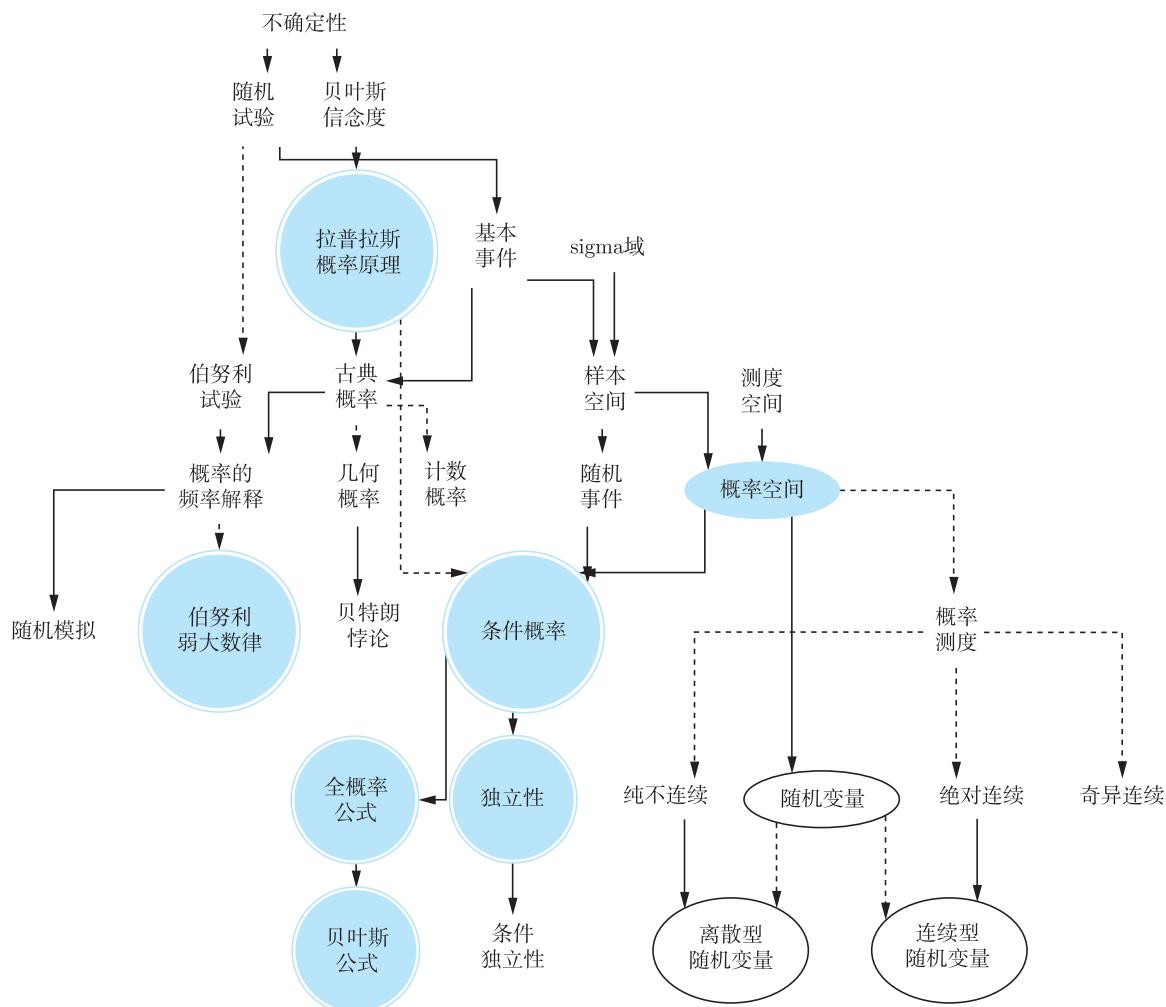


图 3.4 第 3 章的知识图谱

□ $a \rightarrow b$ 表示 “ b 是 a 的一个子类或一部分” (即, b is a kind/part of a);

- $a \rightarrow b$ 表示“对象 b 的定义基于对象 a ”，或者“对象 b 由对象 a 诱导出”(即, b is defined/induced by a)，其中 a, b 之间没有类的包含关系。

3.1 古典概率模型

如果随机试验 \mathcal{E} 的每个基本事件 $\omega \in \Omega$ 发生的机会都是等同的，我们称之为古典概率问题，并把解决此类问题的概率模型称为古典概率模型。卡尔达诺在他的著作《论赌博游戏》(1663) 中，着重考虑了掷骰子的概率问题。以掷两个均匀骰子为例，卡尔达诺明确意识到所有可能的结果是 36 个有序对 (i, j) ，其中 $i, j = 1, 2, \dots, 6$ ，而不是 21 个无序对，见图 3.5(a)。并且，卡尔达诺认为每个有序对出现的机会都是等同的，即都是 $1/36$ 。两个骰子的点数之和可能是 $2, 3, \dots, 12$ ，相应的概率分别为 $1/36, 1/18, \dots, 1/36$ ，见图 3.5(b)。点数之和为 7 的概率是 $1/6$ ，最有可能发生。

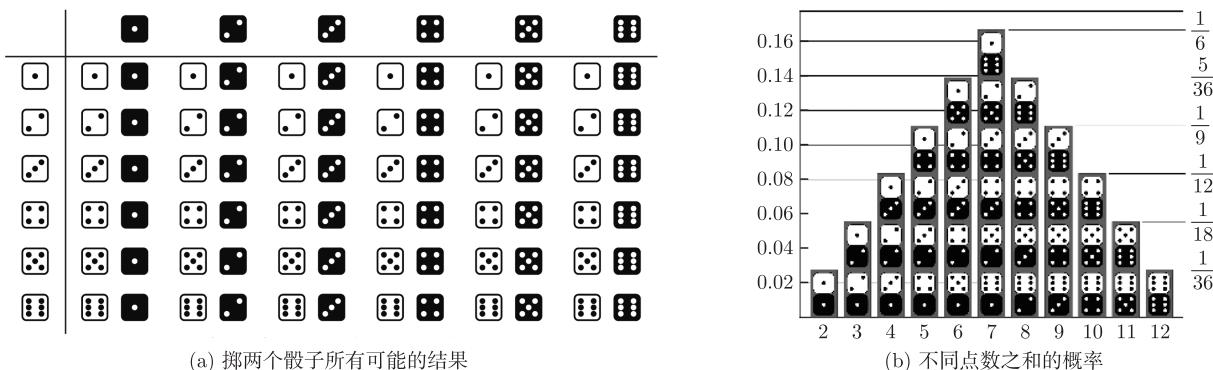


图 3.5 掷两个均匀的骰子所得点数之和的概率分析

科学之父、意大利数学家、物理学家、天文学家、哲学家伽利略·伽利莱 (Galileo Galilei, 1564—1642) (图 3.6) 曾被赌徒求教连续掷三次骰子，点数之和为 9 和为 10 的哪个概率大。赌徒们认为它们相等，因为点数之和为 9 的组合情况与点数之和为 10 的组合情况一样多，分别是

$$\begin{aligned} &(1, 2, 6), (1, 3, 5), (1, 4, 4), (2, 2, 5), (2, 3, 4), (3, 3, 3) \\ &(1, 3, 6), (1, 4, 5), (2, 2, 6), (2, 3, 5), (2, 4, 4), (3, 3, 4) \end{aligned}$$

考虑三个骰子的顺序，共有 $6^3 = 216$ 个基本事件。伽利略发现点数组合 $(3, 3, 3)$ 的情形只有 $\square\square\square$ 一种排列，而 $(2, 2, 5)$ 则有 $\square\square\square, \square\square\square, \square\square\square$ 三种排列。伽利略得到点数之和为 9 和 10 的分别有 25 和 27 种排列情形，于是其概率分别是

$$\begin{aligned} \frac{25}{6^3} &\approx 11.6\% \\ \frac{27}{6^3} &\approx 12.5\% \end{aligned}$$

具体情况如下，

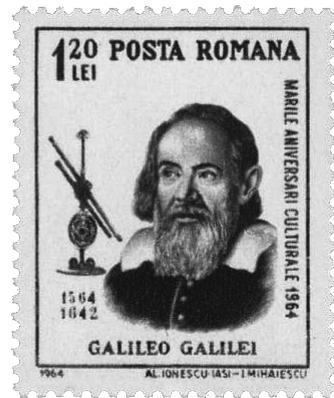


图 3.6 伽利略

点数之和为 9		点数之和为 10	
组合	数目	组合	数目
1 2 6	6	1 3 6	6
1 3 5	6	1 4 5	6
1 4 4	3	2 2 6	3
2 2 5	3	2 3 5	6
2 3 4	6	2 4 4	3
3 3 3	1	3 3 4	3
总计	25	总计	27

卡尔达诺和伽利略都选对了基本事件集合，这是一个了不起的认识，要知道两百年后法国知名数学家让·勒朗·达朗贝尔 (Jean le Rond d'Alembert, 1717—1783) (图 3.7) 也曾在例 3.4 上犯错：抛一枚均匀的硬币两次，出现正面的次数可能是 0, 1 或 2，达朗贝尔认为出现这三个结果的机会等同。果真如此吗？

例 3.4 连抛两次均匀的硬币，根据例 3.1，出现 0, 1 和 2 次正面的随机事件用 Ω 的子集分别表示为

$$\begin{aligned} A_0 &= \{(T, T)\} \\ A_1 &= \{(T, H), (H, T)\} \\ A_2 &= \{(H, H)\} \end{aligned}$$



图 3.7 达朗贝尔

于是，事件 A_0 , A_1 , A_2 出现的机会分别是 $1/4$, $1/2$ 和 $1/4$ 。

 在这个例子中假定硬币是均匀的，所以我们有理由认为每个基本事件出现的机会都是等同的。为刻画随机事件 A 出现的机会，即 A 的概率 $P(A)$ ，我们用构成随机事件 A 的基本事件占整个基本事件集合 Ω 的几成来描述。显然，

$$\begin{aligned} 0 &\leq P(A) \leq 1, \text{ 其中 } A \subseteq \Omega \\ P(\emptyset) &= 0 \\ P(\Omega) &= 1 \end{aligned}$$

除了硬币、骰子、转盘，在描述概率问题时，我们还常用到这样的道具——若干球和一些带标号的罐子。球的状态有以下两种情形。

(1) 无标号或不可分辨的 (indistinguishable)：即假定这些球的大小、颜色、质地等物理特性都一样，无法对它们进行区分。

(2) 可分辨的 (distinguishable)：即可以通过球的颜色或标号对它们进行区分。有时颜色和标号都要用到，譬如约定“ n 个可分辨的黑球”意味着 n 个黑球的标号分别为 $1, 2, \dots, n$ 。

拉普拉斯在《概率的哲学随笔》和《概率的分析理论》中喜欢用罐子和球的道具，来形象地描述一个概率问题，罐子-球模型有助于对概率的抽象思维*。

* 很多伟大的思想来自形象思维。例如，广义相对论 (general relativity) 用升降机来解释“等效原理” (equivalence principle)，即重力场与以适当加速度运动的参照系是等价的。物理学家经常先有直观的想法，然后再用严谨的数学语言来表达它，虽然有时这份表达也非易事。

例 3.5 随机地把两个球放入两个罐子，基本事件集合因罐子和球是否可辨而异。表 3.2 分别讨论球和罐子可分辨和不可辨时的基本事件集合 Ω 。

表 3.2 例 3.5 的基本事件集合

球	罐子	
	可辨	不可辨
可辨	$\Omega = \{(ab -), (a b), (b a), (- ab)\}$	$\Omega = \{[ab -], [a b]\}$
不可辨	$\Omega = \{(* * -), (* *), (- * *)\}$	$\Omega = \{[* * -], [* *]\}$

- 随机选取 n 个人，其生日的情况相当于把 n 个可辨的球随机地放入 $N = 365$ 个可辨的罐子（对应于一年的 365 天）。
- 反复掷骰子 n 次，（不计次序）所掷的点数情况相当于把 n 个不可辨的球随机地放入 $N = 6$ 个可辨的罐子（对应于骰子的六个点数）。

例 3.6 把 n 个球随机地放入标号为 $1, 2, \dots, N$ 的 N 个罐子。

- 若球可分辨，每个球有 N 个选择，共有 N^n 个可能的结果。
- 若球不可分辨，原问题相当于从 $N+n-1$ 个位置中选出 $N-1$ 个位置放隔板，其余位置放球。所有可能结果的总数是 C_{N+n-1}^n ，也是不定方程 $x_1 + x_2 + \dots + x_N = n$ 的非负整数解的个数（另一求解方法见例 3.16）。

拉普拉斯在《概率的哲学随笔》中总共给出了十条概率演算的一般原理，其中原理一是概率的定义，“概率是有利情形的个数与所有可能情形的个数之比。”对“有利”的判定既有主观的一面，也有客观的一面。主观的一面容易理解，就是个体的认知（如，明天下雨的概率），它可以因人而异。客观的一面整理成集合论的语言，就是下面的性质。

性质 3.1 对于一个古典概率问题，如果基本事件集合 Ω 是有限的，不妨设其元素个数为 n ，记作 $|\Omega| = n$ 。因为每一基本事件 $\{\omega\}$ 发生的机会等同，所以 $\{\omega\}$ 的概率，记作 $P(\{\omega\})$ ，满足以下性质。

$$P(\{\omega\}) = \frac{1}{n}$$

$$\sum_{\omega \in \Omega} P(\{\omega\}) = 1$$

随机事件 $A \subseteq \Omega$ 的概率是“一个分数，分子是所有有利场合的数目，分母是所有可能场合的数目”（拉普拉斯，《概率的分析理论》）。

$$P(A) = \frac{|A|}{|\Omega|} = \frac{n_A}{n} = \frac{A \text{ 中基本事件的个数}}{\text{所有基本事件的个数}} \quad (3.1)$$

其中， $|A|$ 和 n_A 表示集合 A 中元素的个数，有时候也记作 $\#(A)$ 。

因为有关随机试验 \mathcal{E} 的每个随机事件都应表示为基本事件集合 Ω 的某个子集的形式，按照严格的写法，基本事件应该记为 $\{\omega\} \subseteq \Omega$ 。在不引起歧义的情况下，基本事件 $\{\omega\}$ 有时也记作 $\omega \in \Omega$ 。

例 3.7 在例 3.5 中，当球和罐子都可辨而且被选机会都是 $1/2$ 时，每个基本事件发生的机会都是 $1/4$ 。当球不可辨而罐子可辨且被选机会都是 $1/2$ 时，基本事件 $(*|*)$ 出现的机会是 $1/2$ ，其他两个基本事件出现的机会都是 $1/4$ 。

古典概率要求基本事件发生的机会等同，限制了概率论的研究对象（例如，图 3.1）。即便如此，在古典概率的研究范畴里，依然有很多问题难以解决或无法解决（具体实例见第 55、56 页的例 3.27）。

和例 3.28)，请读者体会古典概率的局限性。后文将不断突破古典概率的狭隘，从公理化的角度抽象地看待随机事件和概率。

对于古典概率，拉普拉斯解释道，“机会的理论就是把同类的所有事件化归为一定数量的等可能情况（这里所说的等可能性，就是对于属同一类的事件，人们以同等程度不能判定哪个事件会发生），并且确定在哪些情况下被考虑的事件可能发生。可能发生的次数与总的次数之比就是对该事件出现的概率的一个描述。”^[13]

同时，拉普拉斯还清楚地认识到，“当某一事件在所有试验中都发生，可能事件就变成必然事件，它的概率等于 1。在这种情况下，可能性与确定性变得可比了，尽管它们之间有着本质的不同……。”^[66]

解决古典概率问题有时需要很高的技巧，正如拉普拉斯阐述原理二时所说，“假定各种情形都是等可能的，否则就要先确定它们各自的概率，对它们的精确估计正是机会理论中的难点所在。于是，概率就是所有有利情况的概率之和。”

在例 3.1 中，拉普拉斯认为事件 $A = \text{“至少有一次出现正面”}$ 的有利情形是 $\{HH, HT, TH\}$ ，所以 $P(A) = 3/4$ 。或者，按照达朗贝尔的思路，有利的情形是两次正面 $A_2 = \{HH\}$ 和一次正面 $A_1 = \{HT, TH\}$ ，二者概率之和为 $1/4 + 1/2 = 3/4$ 。

※例 3.8 (连续正面问题) 抛一枚均匀的硬币 $n \geq 3$ 次，计算事件 $H_3 = \text{“至少连续出现 3 次正面”}$ 的概率 $P_n(H_3)$ 。

解 基本事件集合 Ω_n 是 H 和 T 构成的长度为 n 的序列的全体，所以 $|\Omega_n| = 2^n$ ，进而每个基本事件的概率为 2^{-n} 。事件“头五次抛硬币的结果是 THTTH”是以 THTTH 开头的所有长度为 n 的序列的集合，简记作 THTTH，它的概率是 2^{-5} 。当 n 很大时，通过列举 H_3 的元素来计算概率 $P_n(H_3)$ 是相当烦琐的。例如， $n = 20$ 时 $|\Omega_n| = 2^{20} = 1048576$ ，我们有 $|H_3| = 825259$ 。

事件 H_3 可用图 3.8 所示的剪枝二叉树递归地构造出来：树中的每个节点代表一个事件，这些事件两两交集为空，其并集合就是 H_3 。树的高度为 $n-2$ ，根节点（即第 1 层节点）表示一开始就连抛三个正面，第 k 层节点的全体表示事件“从第 k 抛开始至少连续出现三次正面”。用 T_k 表示第 k 层节点的个数，则

$$T_0 = 0 \quad T_1 = T_2 = 1 \quad T_k = T_{k-1} + T_{k-2} + T_{k-3}$$

序列 $T_k, k = 1, 2, \dots$ 被称为 3-步斐波那契数列 (Tribonacci)，其中每一项都是前三项之和，它是“斐波那契数列”（图 3.9）的推广。该序列的前几项是 0, 1, 1, 2, 4, 7, 13, 24, 44, 81, 149, 274, 504, 927, 1705, 3136, 5768, 10609, 19513, 35890, …，它们恰是图 3.8 中树的各层节点数。显然，所求的概率是

$$P_n(H_3) = \sum_{k=0}^{n-2} 2^{-2-k} T_k = \frac{1}{4} \sum_{k=0}^{n-2} \left(\frac{1}{2}\right)^k T_k$$

不难验证 3-步斐波那契数列具有如下性质：

$$\sum_{k=0}^{\infty} x^k T_k = \frac{x}{1 - x - x^2 - x^3}$$

利用该性质轻而易举地就能论证当 $n \rightarrow \infty$ 时，

$$P_n(H_3) \rightarrow \frac{1}{4} \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k T_k = \frac{1}{4} \times \frac{2^{-1}}{1 - 2^{-1} - 2^{-2} - 2^{-3}} = 1$$

即抛的次数越多，事件 H_3 发生的机会就越大，这几乎是显而易见的！

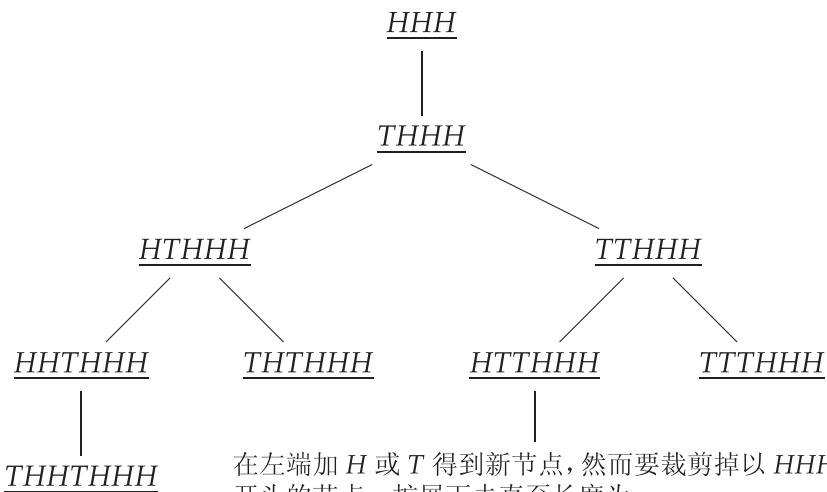
图 3.8 用剪枝二叉树表示事件 H_3 

图 3.9 斐波那契数列 (Fibonacci sequence): 每一项是前两项之和

值得注意的是，古典概率模型无法处理由无穷次重复操作构成的随机试验，如抛一枚硬币无穷次。即便能形式地刻画基本事件，但囿于对概率的认识，古典概率模型也无法计算“至少连续出现 $t < \infty$ 次正面”的概率。数学上必须以测度论^[67] 为基础，才能真正建立起概率论的大厦。

回顾数学史，新工具、新方法往往带来更高程度的抽象，甚至改变数学分支的命运。举个例子，因为有了解析几何 (analytic geometry)，古希腊数学家对圆锥曲线的研究终被历史尘封。然而，具体成果可以被遗忘，思想的延续性却不应被割裂。没有那些对圆锥曲线的系统研究，也不会诞生解析几何。类似地，如果不经历古典概率并深刻体会到它的局限性，新工具的引入也就成了无稽之谈。

借古喻今，也许百年之后，今天的很多数学成果会被扫进历史的故纸堆，但思想不会。我们追溯前人的数学思想，因为有思想性的数学才是有趣的数学。

3.1.1 计数概率

基本事件集合 Ω 有限的这类古典概率问题，实质上就是用排列组合的方法确定 Ω 和所关心的事件 A 的势 (cardinality)，也称作“基数”，所需的技巧也仅限于计数。很多不同应用背景的概率问题都可以化归到罐子-球模型加以讨论^[68]，这些简单的道具使得很多貌似不同、本质上同类的问题“原形