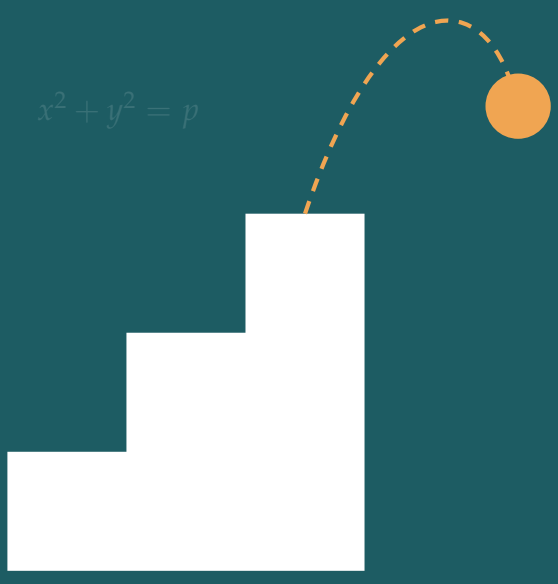
STEP 备试 · 大学数学

Beyond STEP

第 9 章

计数与机会



KEITH McNULTY

Beyond STEP

以历届 STEP 试题为门
走进大学数学

第 9 章

计数与机会

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

第 9 章：计数与机会

本册是 *Beyond STEP* 全书 13 章中的第 9 章。全书、各章的单行本、每一节的可列印工作纸，以及每一道 STEP 试题的详细解答，均可在 `beyondstep.keithmcnulty.org` 免费阅读及下载。

2026 年 11 月首次于网上发表。

本 PDF 内凡属作者本人创作的内容，包括正文、解答、插图、练习及答案，均按照 Creative Commons CC0 1.0 Universal 授权条款贡献予公有领域：你可以为任何目的复制、改编及分享，毋须征求许可。详见 `creativecommons.org/publicdomain/zero/1.0`。

本书讨论的 STEP 试题，版权属 Cambridge University Press & Assessment 所有。本书并无转载这些试题：每一题均附连结，通往官方来源的题目，即 STEP Support Programme 的试题资料库或 OCR 的历届试卷。

如有错漏，欢迎到 `github.com/keithmcnulty/step_book_mcnulty` 提交 issue 指正。

章

9

计数与机会

计数是最古老的数学，但要数得巧妙，至今仍是最困难的事情之一。

本章内容

9.1	直线分割平面	STEP II 1990
9.2	小型版的拉姆齐定理	STEP I 1989
9.3	为篱笆桩着色	STEP II 1995
9.4	把数列看成函数	STEP I 2016
9.5	凑零钱	STEP I 1997
9.6	奇怪的骰子	STEP I 1993
9.7	相等的幂和	STEP I 1992
9.8	没有人拿到自己的恤衫	STEP II 2016
9.9	生日问题	STEP II 2007
9.10	总能找赎	STEP I 2011
9.11	集齐一套	STEP II 1994
9.12	随机子集	STEP III 2019
9.13	做了手脚的骰子	STEP I 1999
A	练习答案	第 89 页

计数的艺术

组合数学是拓扑学的贫民窟。

— J. H. C. Whitehead，说于 1950 年代。今天已经没有人抱持这种看法。

人人都会数数。但若要问：用硬币凑出 £1 有多少种方法？安排十个人入座，使得没有人坐在自己的位子上，有多少种方法？用四种颜色为一幅地图着色，又有多少种方法？单纯的计数很快就行不通了。情况多得无法逐一列出，而且各种情况还会互相重叠。组合数学（combinatorics）是关于计数与排列的数学，它是一门艺术：把这类问题整理得井井有条，使我们不必逐一列举也能解决。

正如 Whitehead 那句话所反映的，在很长一段时间里，组合数学只被视为一堆巧妙技巧的集合，而不是一门严肃的学科。这种情况在二十世纪下半叶改变了，推动力一方面来自计算机科学（computer science），因为计数一个算法（algorithm）所需的步骤数目，本身就是一个组合问题；另一方面则来自人们发现它与代数、几何和概率（probability）之间有深刻的联系。今天，它是数学中最活跃的领域之一。2022 年，许埈珥（June Huh）获颁数学界的最高荣誉菲尔兹奖（Fields Medal），部分原因是他运用代数几何（algebraic geometry）的思想，证明了一个 1968 年提出的猜想，这个猜想所涉及的，正是 Section 9.3 中的色多项式（chromatic polynomial）。

计数为何重要

概率最简单的形式就是计数：当所有结果的可能性均等时，一个事件的概率等于有利结果的数目除以结果的总数。因此，每一条关于纸牌、骰子和彩票的问题，都是计数问题；本章后半部分取材自概率试卷，正是这个缘故。但计数的重要性远远不止于游戏。计算机科学家要计数一个算法所需的运算次数，以及它可能出错的方式。化学家要计数分子可能具有的结构。生物学家要计数一条 DNA 链可以折叠的方式。密码学家要计数攻击者必须尝试的密钥数目，以及（如 Section 9.9 所示）要有多少个，才会出现两个相同的情况。

你在大学会遇到甚么

大多数大学都会在一年级或二年级开设组合数学或离散数学（discrete mathematics）的课程，其主题正是本章的主题。你会学到基本的计数原理（乘法原理（product rule）、双射（bijection）、容斥原理（inclusion–exclusion）、鸽巢原理（pigeonhole principle）），以及怎

样建立和求解递推关系（recurrence relation）。你会遇到生成函数（generating function），它把数列变成函数，把计数变成代数。你会研究图（graph），从着色（colouring）到拉姆齐理论（Ramsey theory），还有偏序集（partially ordered set），例如子集所构成的格（lattice）。概率课程也会运用同样的工具，再加上期望值的线性性质（linearity of expectation），来分析随机结构。再往后，组合数学会与代数（对称函数、表示论）、数论（分拆、加性组合）以及计算机科学（算法分析、复杂度理论）连系起来。

贯穿本章的问题

三大问题

1. **怎样不逐一列举而完成计数？**有哪些结构（递推关系、双射、生成函数）能让你准确地计数庞大的集合？
2. **甚么时候某件事必然发生？**甚么时候单凭数量之大，就足以迫使某种规律、巧合或配对出现？
3. **计数与概率怎样互相帮助？**甚么时候最好用计数来求概率，甚么时候又最好用概率来求数目？

本章分为两半。前七节取材自纯数学试卷，先讲递推关系（直线分割平面、为篱笆桩着色）和鸽巢原理（拉姆齐定理），然后建立生成函数，并用它来计数硬币、设计奇特的骰子，以及把数字分成幂和相等的队伍。后六节是概率试卷中的经典计数问题：错排（derangement）与 $1/e$ 的出现、生日问题、选票问题及其反射原理（reflection principle）、优惠券收集问题（coupon collector）、随机子集，以及为何不可能把骰子灌铅，使得每个总和出现的可能性均等。

本章的思想

§	课题	思想
9.1	平面的区域	多加一个时，有甚么改变
9.2	拉姆齐理论	鸽巢原理
9.3	图着色	耦合递推关系；色多项式
9.4	生成函数	数列变成函数
9.5	分拆（partition）	通过系数来计数
9.6	Sicherman 骰子	多项式的因式分解
9.7	Thue–Morse	重根带来平衡的和
9.8	错排	容斥原理；$1/e$
9.9	巧合	补事件；平方根定律
9.10	选票问题	反射原理
9.11	优惠券收集问题	期望值的线性性质；$n \ln n$
9.12	随机子集	每次只看一个元素
9.13	灌铅骰子	以算术—几何平均不等式（AM–GM）证明不可能

怎样阅读本章

永远从小的情况入手。画三条直线，列出三张牌的六种次序，数一数凑出 5p 的方法。小的情况能给你数据去猜想，最后也可以用来检验你的公式。但在找到理由之前，不要轻信任何规律：Section 9.1 中有一个著名的数列 $1, 2, 4, 8, 16, \ldots$，它的下一项是 31。

9.1 直线分割平面

题目

STEP II 1990，第 4 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/90-s2-q4`

引子

用 n 刀直线把一个薄饼切开，各块不必大小相同，最多可以切成多少块？一刀得 2 块，两刀得 4 块，三刀得 7 块（不是 8 块：不妨试试），四刀得 11 块。数列 $1, 2, 4, 7, 11, 16, \ldots$ 称为*懒惰宴会承办人数列*（lazy caterer's sequence），而本题要求的正是它的公式。

这道题是一种贯穿整个组合数学（combinatorics）的技巧的第一课：不要试图一次过把所有东西数完，而应该问：多加一个对象时，有甚么改变？而答案原来是欧拉公式（Euler's formula）$V - E + F = 2$ 的一个特殊情况，你在 Section 7.9 讨论柏拉图立体时已经见过这条公式。

动笔之前

停一停，试一试 仔细画出 $n = 1, 2, 3, 4$ 的情况，并保持各直线处于一般位置（general position）（没有两条平行，没有三条共点）。数一数每种情况中线段和区域的数目。你能猜出当中的规律吗？

你应该会得到：

n	0	1	2	3	4
线段	0	1	4	9	16
区域	1	2	4	7	11

线段的数目看来是 n^2。区域的数目则依次增加 $1, 2, 3, 4$：相邻两项之差正是自然数。

逐步解题

阶段 1　数线段

> **想一想**　集中看其中一条直线。其余的直线中有多少条与它相交？它们把这条直线切成多少段？

每条直线都与其余 $n-1$ 条直线逐一相交（没有一条与它平行），而且交于 $n-1$ 个不同的点（没有三条直线交于同一点）。这 $n-1$ 个点把该直线切成 n 段：两端各有一段无限长的，中间有 $n-2$ 段有限长的。共有 n 条直线，所以共有

$$n \times n = n^2 \text{ 条线段.} \qquad \blacksquare$$

这是直接计数，无需使用归纳法（induction）：我们是逐条直线去数线段的。

阶段 2　加一条直线时会怎样？

> **想一想**　假设已画好 $n-1$ 条直线，现在加上第 n 条。它被其余直线切成 n 段。每一段对它所穿过的区域有甚么影响？

画出新直线时，它的 n 段各自穿过一个原有的区域，并把该区域一分为二（图 9.1.1）。不同的线段位于不同的区域内，因为新直线要由一个区域走到另一个区域，就必须越过一条原有的直线。因此，加上第 n 条直线后，区域的数目恰好增加 n：

$$R(n) = R(n-1) + n.$$

阶段 3　解递归关系

由 $R(0) = 1$（没有直线，只有一个区域）开始，

$$R(n) = 1 + (1 + 2 + \cdots + n) = 1 + \frac{n(n+1)}{2} = \frac{n^2 + n + 2}{2}.$$

若要写成归纳法证明：公式对 $n = 1$ 成立，因为 $\frac{1}{2}(1+1+2) = 2$；而若 $R(n-1) = \frac{1}{2}\big((n-1)^2 + (n-1) + 2\big) = \frac{1}{2}(n^2 - n + 2)$，则 $R(n) = \frac{1}{2}(n^2 - n + 2) + n = \frac{1}{2}(n^2 + n + 2)$。 $\blacksquare$

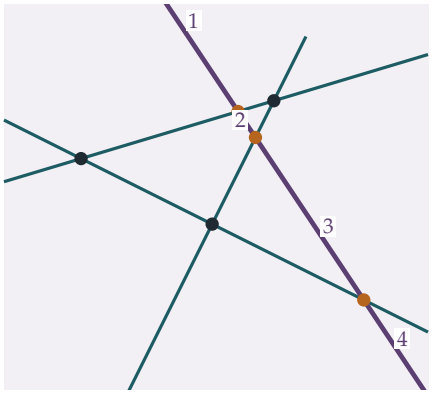


图 9.1.1 三条直线（蓝绿色）把平面分成 7 个区域。第四条直线（紫红色）与其余三条相交于琥珀色的点，这些点把它切成四段，编号为 1 至 4。每一段把一个区域一分为二，所以第四条直线增加了 4 个区域，合共 11 个。

回顾

- **每次只数一部分。**线段是逐条直线去数的；区域则是按每条新直线所增加的数目去数的。
- **把几何上的一步化为递归关系。**“每条新线段把一个区域一分为二”这项观察，化成了 $R(n) = R(n-1) + n$。
- **一般位置十分重要。**若有平行的直线，或有三条直线共点，区域的数目就会较少。试找出每项假设在哪里用过。

背后的数学

乔装的欧拉公式

以下是第二个证明，它不仅推导出答案，更解释了答案从何而来。把这些直线组成的图形看成一幅地图：交点是顶点（vertices），线段是边（edges），区域是面（faces）。图中有 $V = \binom{n}{2}$ 个顶点（每一对直线对应一个）和 $E = n^2$ 条边。无限长的线段只有一个端点，所以我们额外加上一个“在无穷远处”的顶点，让它们全部在那里相遇；这样，整幅图就成为画在球面上的地图（想像把平面包在一个球上，无穷远点位于球顶）。由 Section 7.9 的欧拉公式 $V - E + F = 2$ 可得

$$\left(\binom{n}{2} + 1\right) - n^2 + F = 2, \qquad \text{所以} \qquad F = n^2 - \frac{n(n-1)}{2} + 1 = \frac{n^2 + n + 2}{2}.$$

同一个数，却来自完全不同的方向。这些区域还可以整齐地分成两类：其中 $2n$ 个是无界的（每两条相邻的无限射线之间有一个），其余 $\frac{1}{2}(n^2 + n + 2) - 2n = \binom{n-1}{2}$ 个是有界的。

二项式系数与高维空间

这条公式可以写成

$$R(n) = \binom{n}{0} + \binom{n}{1} + \binom{n}{2},$$

而这并非巧合。把图形转动，使得没有一条直线是水平的。一个下方有界的区域必有一个最低点，而这个点必定是交点；每个交点又恰好是一个区域（即它正上方的那个）的最低点：这样就得到 $\binom{n}{2}$ 个区域，每一对直线对应一个。其余每个区域都向下无限延伸，因而会与一条画在所有交点下方很远处的水平线相遇；那 n 条直线把这条水平线切成 $n+1$ 个区间，每个这样的区域内各有一个。合共 $\binom{n}{2} + n + 1$ 个。在三维空间中，n 个处于一般位置的平面把空间切成

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3} = \frac{n^3 + 5n + 6}{6}$$

块（即蛋糕数（cake numbers）$1, 2, 4, 8, 15, 26, \ldots$），而这个规律在任何维数都成立。证明用的是同一个归纳法：新加的平面被其余平面切成若干区域，其排列方式就如平面上的直线所分成的区域，而每个这样的区域都把一块空间一分为二。

施泰纳与超平面排列

直线和平面所分割出的区域数目的公式，是施泰纳（Jakob Steiner）在 1826 年发现的。今天，它是*超平面排列*（hyperplane arrangements）理论的起点；这套理论计数任意一组直线、平面及其高维类比所形成的区域数目，包括并非处于一般位置的情况。Thomas Zaslavsky 在 1975 年证明了一条非凡的定理：利用一个记录了哪些超平面相交的多项式，就可以数出任何排列的区域数目。同样的想法也出现在机器学习之中：由“ReLU”单元构成的神经网络，会用超平面把输入空间切成多个区域，而区域的数目正好衡量它所能表示的函数有多复杂。

思维技巧

问一问：多加一个时，有甚么改变

要计数由 n 个对象组成的构形，可把它与 $n-1$ 个对象的构形比较。计数其中的改变，往往比计数整体容易得多。

这是把归纳法用作发现的工具，而不仅是证明的工具。它是组合数学中最常见的手法之一，能把几何问题化为可以求解的递归关系（recurrence relation）。

延伸练习

1 **圆。**证明：若 n 个圆中每两个都交于两点，且没有三个圆共点，则它们把平面分成 $n^2 - n + 2$ 个区域。

2 **平行线。**若 n 条直线中有 k 条互相平行（其余方面均处于一般位置），共有多少个区域？

3 **蛋糕数。**用归纳法证明：n 个处于一般位置的平面把空间分成 $\frac{1}{6}(n^3 + 5n + 6)$ 个区域。

4 **会失效的规律。**在圆周上取 n 个点，并把每一对点用弦连起来，其中没有三条弦在圆内交于同一点。当 $n = 1, 2, 3, 4, 5$ 时，分别得到 $1, 2, 4, 8, 16$ 个区域。数一数 $n = 6$ 时的区域数目。（提示：使用欧拉公式；每四个点给出一个交点。）证明区域的数目是 $1 + \binom{n}{2} + \binom{n}{4}$。

延伸阅读

- **Ronald Graham, Donald Knuth & Oren Patashnik, *Concrete Mathematics***（Addison-Wesley），第 1 章。开首讨论的正是这个问题，还有“薄饼”和“之字形”的变化题。
- **Miklós Bóna, *A Walk Through Combinatorics***（World Scientific）。一本平易近人的入门教材，设有归纳法、图和欧拉公式等章节。
- **Richard Guy, “The strong law of small numbers”**，*American Mathematical Monthly* 95 (1988)。收录了许多看似明显、其实不成立的规律，包括练习中的圆问题。

9.2 小型版的拉姆齐定理

题目

STEP I 1989，第 4 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/89-s1-q4`

引子

在任何一个六人派对上，要么有三个人彼此都认识，要么有三个人彼此都不认识。无论这六个人是谁，这一点都必定成立。换成五个人，就没有这个保证了。这个小小的事实是*拉姆齐定理*（*Ramsey's theorem*）的第一个情形，它道出了一条贯穿现代组合数学（combinatorics）的原则：*完全的无序是不可能的*。任何足够大的结构，无论怎样着色或排列，都必定包含一个高度有序的部分。

证明只需四行。但顺理成章的下一个问题（要多少人才能保证有四个互相认识的人或四个互不认识的人？五个呢?）却通向数学中一些最困难的未解难题。

动笔之前

想一想 那六个点之所以重要，只在于哪些点对以哪种颜色相连：几何只会令人分心。一共有多少条线段？每个点又连出多少条？

一共有 $\binom{6}{2} = 15$ 条线段，每个点连出 5 条。（一般位置（general position）这个条件只是确保任意三点都构成一个真正的三角形。这道题其实是关于六个顶点上的*完全图*（*complete graph*）K_6，它的边以两种颜色着色。）

逐步解题

阶段 1 五条线段，两种颜色

> **想一想** 从 A 连出五条线段，每条不是金色就是银色。关于它们的颜色，你可以肯定甚么?

其中至少有三条颜色相同。假如金色的至多两条，银色的也至多两条，那么线段至多只有四条。这就是鸽巢原理（*pigeonhole principle*）：把五件物件放进两个盒子，必有一个盒子里至少有三件。不妨设（有需要时把两种颜色的名称对调）AB、AC 和 AD 全是金色。

阶段 2　看看另一端

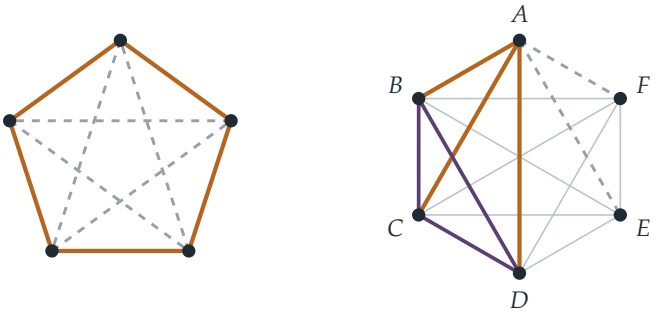
> **想一想** 现在看看三角形 BCD。如果它有任何一条边是金色，会怎样？如果一条都不是呢?

如果三角形 BCD 有任何一条边是金色，比如 BC，那么 ABC 的三条边 AB、AC 和 BC 全是金色。如果 BCD 没有一条边是金色，那么 BC、CD 和 DB 全是银色，BCD 就是一个银色三角形。无论哪种情况，都有一个单色三角形（图 9.2.1，右）。 ■

阶段 3　五个点并不足够

> **想一想** 若只有五个点，每个点只连出四条线段。怎样着色，才能使每个点连出的线段恰好每种颜色各两条?

把五个点放成一个正五边形（如果想要它们处于一般位置，可以把它们稍微提离平面）。把五边形的五条边涂成金色，五条对角线涂成银色（图 9.2.1，左）。金色的边构成一个五圈，当中没有三角形；银色的边构成一个五角星，它同样是一个五圈，当中也没有三角形。所以并没有单色三角形。 ■



五个点：没有单色三角形　　六个点：看看 BCD

图 9.2.1　左：五个点，五边形为金色（实线），五角星为银色（虚线），没有单色三角形。右：六个点时，从 A 连出的五条线段中有三条颜色相同，比如连到 B、C、D 的三条都是金色。这时三角形 BCD（紫红色）决定了结果：只要有一条边是金色，就与 A 构成一个金色三角形；没有金色的边，BCD 就是一个银色三角形。

回顾

- **撇除无关的东西。**空间中的点变成了一个着色的图；重要的只有边的颜色。
- **鸽巢原理。**五条边、两种颜色，迫使其中三条同色。
- **对合适的对象分情况讨论。**一旦找到三条金色的边，它们另一端的三角形就决定了一切。
- **紧确性（sharpness）。**五边形的例子说明“六”这个数不能再降低。一条好的定理，总会附带一个例子，说明它已是最佳的结果。

背后的数学

拉姆齐数

拉姆齐数（*Ramsey number*）$R(s,t)$ 是满足以下条件的最小的 N：把 K_N 的边以金、银两色着色，无论怎样着色，都必定包含一个金色的 K_s（即 s 个点，它们之间的所有边都是金色）或一个银色的 K_t。这道题已证明了 $R(3,3)=6$：六个点必定可行，而五边形说明五个点并不可行。

弗兰克·拉姆齐（Frank Ramsey）在 1928 年证明了这些数必定存在：对任意的 s 和 t，总有某个 N 符合要求。证明用的是归纳法（induction），把上面的论证加以推广，并得出

$R(s,t) \le R(s-1,t) + R(s,t-1)$。但要找出确切的数值却异常困难：

$$R(3,3) = 6, \qquad R(4,4) = 18, \qquad 43 \le R(5,5) \le 46.$$

$R(4,4)$ 的值是在 1955 年求得的。尽管已动用电脑作大规模搜索，$R(5,5)$ 至今仍然未知；上界 46 到 2024 年才获证明。电脑无法直接检查 K_{43} 的每一种着色：一共有 2^{903} 种，远远多于宇宙中原子的数目。

艾狄胥与外星人

保罗·艾狄胥（Paul Erdős）对拉姆齐理论（Ramsey theory）发展的贡献无人能及，他很喜欢讲这个故事。假设外星人入侵地球，并威胁说除非我们告诉他们 $R(5,5)$ 的值，否则就毁灭地球。我们应当调动所有电脑和数学家，这样或许能把它找出来。但假如他们问的是 $R(6,6)$，我们就应该设法消灭外星人。艾狄胥还在 1947 年证明了 $R(k,k) > 2^{k/2}$，所用的方法令人惊叹：他证明了，在少于 $2^{k/2}$ 个点上随机着色的图，有正的概率不包含任何单色的 K_k，所以好的着色必定存在，尽管没有人知道怎样把它构造出来。这种概率方法（*probabilistic method*）如今是组合数学中最强大的工具之一。

乱中有序

拉姆齐定理有很多近亲，它们表达的都是同一个想法：足够大的结构无法避开秩序。范德瓦尔登定理（*Van der Waerden's theorem*）说，如果把所有整数以有限种颜色着色，就存在任意长的单色等差数列（arithmetic progression）。艾狄胥—塞凯赖什定理（*Erdős–Szekeres theorem*）说，任何由 n^2+1 个相异的数组成的数列，都包含一个长度为 $n+1$ 的递增或递减子数列（subsequence）。而拉姆齐本人感兴趣的是一个逻辑问题：他是在一篇关于判定逻辑命题真假的论文中，把这条定理作为引理（lemma）来证明的。

思维技巧

鸽巢原理

如果把多于 kn 件物件放进 n 个盒子，则必有一个盒子里有多于 k 件物件。

这条原理显而易见，但要用得好却是一门艺术：功夫在于选定甚么是物件、甚么是盒子。在这道题中，物件是某一点上的五条边，盒子是两种颜色。在大学里，鸽巢原理的论证会出现在数论（以分数逼近无理数）、分析（波尔查诺–魏尔斯特拉斯定理（Bolzano–Weierstrass theorem））之中，更遍及整个组合数学。

延伸练习

1 **至少两个。**证明在 K_6 的任何双色着色中，其实至少有两个单色三角形。（提示：数一数“双色角”的数目，即在同一个顶点上的一对不同颜色的边。若一个点有 g 条金色的边和 $5-g$ 条银色的边，它就是 $g(5-g) \leq 6$ 个这样的角的顶点。每个非单色的三角形恰好包含两个这样的角。）

2 **三种颜色。**如果把 K_{17} 的边以三种颜色着色，证明其中有一个单色三角形。（提示：有一个点至少有六条同色的边。）

3 **三或四。**证明 $R(3,4) \leq 10$：在 K_{10} 的任何金、银着色中，都有一个金色三角形或一个银色的 K_4。（提示：一个点要么至少有 4 条金色的边，要么至少有 6 条银色的边。）事实上 $R(3,4) = 9$。

4 **单调子数列。**就 $n = 2$ 的情形证明艾狄胥–塞凯赖什定理：任何由 5 个相异的数组成的数列，都包含一个长度为 3 的递增或递减子数列。找出一个由 4 个数组成、两者皆无的数列。

延伸阅读

- **Miklós Bóna, *A Walk Through Combinatorics***（World Scientific），组合数学入门书，可读其中关于鸽巢原理和拉姆齐理论的章节。
- **Ronald Graham & Joel Spencer, “Ramsey theory”**，*Scientific American*（1990 年 7 月）。由两位顶尖研究者撰写的科普介绍。
- **Noga Alon & Joel Spencer, *The Probabilistic Method***（Wiley）。讲述艾狄胥这套方法的标准著作；第一章不难读懂。
- **Paul Hoffman, *The Man Who Loved Only Numbers***（Fourth Estate）。艾狄胥的传记。

9.3 为篱笆桩着色

题目

STEP II 1995，第 2 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/95-s2-q2`

引子

为排成一列的柱子髹色很容易：第一根有 3 种选择，其后每根各有 2 种（只要不与前一根同色即可），共有 $3 \times 2^{n-1}$ 种髹法。把这一列弯成一个圆圈，问题却变得出奇地棘手，因为最后一根柱子既要与相邻的前一根不同色，又要与第一根不同色，而这究竟排除了一种颜色还是没有排除任何颜色，取决于前面所有柱子的髹法。

题目用一个经典技巧绕过这个难处：按最后一根柱子的颜色把计数分成几种情况，再找出各种情况之间如何互相转化。答案 $2^n + 2(-1)^n$ 是*色多项式*（chromatic polynomial）的第一个例子，而色多项式正是四色定理（four colour theorem）背后的数学的一部分。

动笔之前

停一停，试一试 数一数为围成一圈的 3 根柱子髹色有多少种方法，4 根又有多少种。$3 \times 2^{n-1}$ 有没有多算？多算了多少？

围成一圈的三根柱子两两相邻，所以需要三种不同的颜色：共 $3! = 6$ 种方法。围成一圈的四根柱子：18 种方法（你可以在最后验证这一点）。与排成一列时的 $3 \times 2^2 = 12$ 和 $3 \times 2^3 = 24$ 比较：圆圈少了的，正是最后一根柱子与第一根同色的那些着色。

逐步解题

阶段 1　两类着色

想一想 考虑 $n+1$ 根柱子的一种着色（colouring），其中第一根和最后一根都是红色。把最后一根柱子拿走。余下的首 n 根柱子的着色有甚么特点?

第 n 根柱子与第 $n+1$ 根相邻，而后者是红色，所以第 n 根不是红色。拿走第 $n+1$ 根后，剩下的是 n 根柱子的一种着色，其中第一根是红色而最后一根不是红色：这正是 s_n 种着色之一。反过来，s_n 种着色中的任何一种都恰好有一种方法延伸，就是把第 $n+1$ 根柱子髹成红色。所以两个集合完全对应，$r_{n+1} = s_n$。

想一想 $r_n + s_n$ 是甚么?

每一种第一根柱子为红色的着色，其最后一根柱子不是红色就是非红色，所以 $r_n + s_n$ 数的是 n 根柱子中第一根为红色的全部着色。第一根之后，每根柱子有 2 种选择，所以 $r_n + s_n = 2^{n-1}$。因此

$$r_{n+1} + r_n = s_n + r_n = 2^{n-1}. \qquad \blacksquare$$

阶段 2　用数学归纳法（induction）证明公式

当 $n = 1$ 时只有一根柱子，髹成红色，所以 $r_1 = 1$，而公式给出 $\frac{1}{3}(1+2) = 1$。假设公式对 r_n 成立。那么

$$r_{n+1} = 2^{n-1} - r_n = 2^{n-1} - \frac{2^{n-1} + 2(-1)^{n-1}}{3} = \frac{2 \cdot 2^{n-1} - 2(-1)^{n-1}}{3} = \frac{2^n + 2(-1)^n}{3},$$

这正是 r_{n+1} 的公式。 $\blacksquare$

停一停，试一试 这条公式从何而来?递归关系（recurrence relation）$r_{n+1} + r_n = 2^{n-1}$ 就像一条微分方程（differential equation）：解是一个“特解”（particular solution）$c \cdot 2^n$ 加上一个满足 $r_{n+1} + r_n = 0$ 的“余函数”（complementary function）$A(-1)^n$。求 c 和 A。

阶段 3　圆圈

想一想 在 n 根柱子围成的圆圈中，固定第 1 根柱子的颜色，例如红色。现在在第 n 根与第 1 根柱子之间把圆圈剪开，拉直成一列。哪些直列着色对应于合乎要求的圆圈着色?

拉直之后，圆圈变成由红色开始的一列 n 根柱子，其中最后一根还必须与第 1 根不同色：最后一根不是红色。这些正好就是那 s_n 种着色。第 1 根柱子可以是 3 种颜色中的任何一种，所以圆圈着色的数目是

$$3s_n = 3(2^{n-1} - r_n) = 3 \cdot 2^{n-1} - 2^{n-1} - 2(-1)^{n-1} = 2^n + 2(-1)^n. \qquad \blacksquare$$

验算：$n=3$ 给出 $8-2=6$，而 $n=4$ 给出 $16+2=18$。

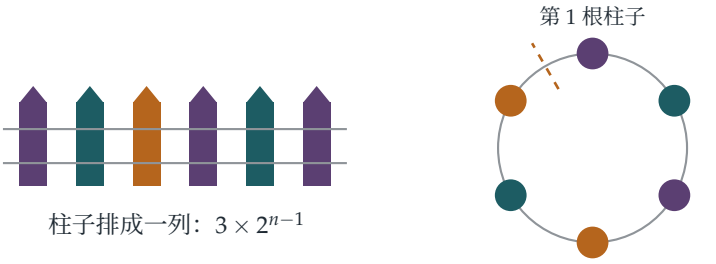


图 9.3.1 左：排成一列时，第一根之后的每根柱子只须与前一根不同色。右：围成一圈时，最后一根柱子还必须与第 1 根柱子不同色。沿虚线把圆圈剪开，圆圈着色就变成首尾两根柱子不同色的直列着色：即题目中所数的 s_n 种着色之一。

回顾

- **分成互相转化的情况。**记下最后一根柱子是否红色，便得到两个数列 r_n 和 s_n，它们由简单的递归关系联系起来。
- **双射（bijection）。**$r_{n+1}=s_n$ 是靠把着色一一配对来证明的，而不是靠计算。
- **剪开圆圈。**一个循环的问题变成了一个对两端附加条件的直线问题。

背后的数学

色多项式

柱子和它们之间的相邻关系构成一个*图*（graph）：排成一列时是一条路径（path），围成一圈时是一个圈（cycle）。*正常着色*（proper colouring）给相邻的顶点（vertex）不同的颜色。用 k 种颜色的正常着色数目是 k 的一个多项式，称为该图的*色多项式*，由伯克霍夫（George Birkhoff）于 1912 年引入。用 k 种颜色重复题目中的论证，可得

$$\text{路径：}k(k-1)^{n-1},\qquad \text{圈：}(k-1)^n+(-1)^n(k-1).$$

取 $k=3$，圈的公式就是 $2^n+2(-1)^n$。伯克霍夫希望通过证明每幅地图所对应的图的色多项式在 $k=4$ 时为正数，从而证明四色定理（每幅地图都可以用四种颜色着色，使相邻的国家颜色不同）。这个方法始终未能成功；定理最终在 1976 年由阿佩尔（Kenneth Appel）和哈肯（Wolfgang Haken）借助大量电脑运算证明。色多项式至今仍是一个活跃的研究领域，并与统计物理（statistical physics）有出人意表的联系：在那里它们以描述磁性的*波茨模型*（Potts model）的形式出现。

用矩阵数出同一个答案

还有另一种方法可以看出答案，用的是第 7 章的线性代数（linear algebra）。设 $\mathbf{M}$ 为对角线上是 0、其余位置是 1 的 3×3 矩阵：当颜色 j 可以跟在颜色 i 之后时，第 (i,j) 项为 1。那么 $\mathbf{M}^{n-1}$ 的第 (i,j) 项数的是由颜色 i 开始、以颜色 j 结束的直列着色，所以 r_n 是 $\mathbf{M}^{n-1}$ 的一个对角线项。圆圈着色是经过 n 步后回到起点的一个序列，所以圆圈着色的数目就是 $\mathbf{M}^n$ 的迹（trace）。由于 $\mathbf{M} = \mathbf{J} - \mathbf{I}$，其中 $\mathbf{J}$ 的每一项都是 1，$\mathbf{M}$ 的特征值（eigenvalue）是 $3-1=2$（特征向量（eigenvector）为 $(1,1,1)$）和 -1（两重），而 $\mathbf{M}^n$ 的迹是各特征值的 n 次方之和：

$$\operatorname{tr} \mathbf{M}^n = 2^n + (-1)^n + (-1)^n = 2^n + 2(-1)^n.$$

这种*转移矩阵*（transfer matrix）方法在组合数学（combinatorics）和统计物理中随处可见，用来数算逐步构造出来的排列。

思维技巧

建立互相耦合的递归关系

当一个数列单靠自己无法满足简单的递归关系时，不妨多记录一点资料（这里是最后一根柱子的颜色），再找出联系各种情况的递归关系。

这就是*状态*（state）的概念：只记下继续下去所必须知道的资料，不多也不少。它是电脑科学中动态规划（dynamic programming）的基础，也是概率中马可夫链（Markov chains）（Section 12.16）和转移矩阵方法的基础。

延伸练习

1 **k 种颜色。**用 k 种颜色重做这道题。证明 $r_{n+1} = s_n$ 仍然成立，并且 $r_{n+1} + r_n = (k-1)^{n-1}$，求 r_n，并由此推导出圈的色多项式 $(k-1)^n + (-1)^n(k-1)$。

2 **没有两根红柱相邻。**把排成一列的 n 根柱子髹成红色或白色，而且没有两根相邻的柱子同为红色，共有多少种方法？（你应该会得到斐波那契数（Fibonacci numbers）。）

3 **轮子。**加一根位于中心的柱子，与圆圈上的每根柱子都相邻（一个“轮子”）。用 4 种颜色为它着色有多少种方法？（*提示*：中心用去一种颜色，圆圈必须使用其余三种。）

4 **项链。**如果圆圈的两种着色在其中一种可由另一种旋转而得时视为相同，那么当 $n = 4$ 时有多少种不同的着色？$n = 5$ 时呢？（这是*伯恩赛德引理*（Burnside's lemma）的开端。）

延伸阅读

- **Robin Wilson, *Introduction to Graph Theory***（Pearson）。标准的图论入门教材；涵盖着色和四色问题。
- **Robin Wilson, *Four Colours Suffice***（Penguin）。四色定理的故事，为一般读者而写。
- **Herbert Wilf, *generatingfunctionology***（可在网上免费阅读）。讲述递归关系及其解法，包括转移矩阵方法。

9.4 把数列看成函数

题目

STEP I 2016，第 8 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/16-s1-q8`

引子

用计算机算一算 $1/89$：0.011235955…。开首的数字是 $0, 1, 1, 2, 3, 5$，正是斐波那契数（Fibonacci numbers）。这是巧合吗？不是，而这道题目正好解释了原因。

这个想法是组合数学（combinatorics）中最强大的想法之一。取一个无穷数列，把各项用作幂级数（power series）的系数，从而把整个数列包装成单一个函数。这样，关于数列的事实就变成关于函数的事实，你便可以用代数和微积分来处理。Herbert Wilf 有一句令人难忘的话：“生成函数（generating function）是一条晾衣绳，我们把一列数字挂在上面展示。”

动笔之前

想一想 哪个数列的生成函数是 $\dfrac{1}{1-x}$？$\dfrac{1}{1-2x}$ 又如何？如果把一个数列的生成函数乘以 x，数列会有甚么变化？

等比级数（geometric series）$\frac{1}{1-x} = 1 + x + x^2 + \cdots$ 是 $1, 1, 1, \ldots$ 的生成函数，而 $\frac{1}{1-2x} = 1 + 2x + 4x^2 + \cdots$ 则是 2 的各次幂的生成函数。乘以 x 会把每一项向后移一位：$x \cdot \frac{1}{1-x} = x + x^2 + \cdots$ 对应的数列是 $0, 1, 1, 1, \ldots$。

逐步解题

阶段 1　自然数与三角形数

想一想　用二项级数（binomial series）展开 $(1-x)^{-2}$。x^n 的系数是甚么？

由二项级数可得

$$(1-x)^{-2} = 1 + 2x + 3x^2 + 4x^3 + \cdots = \sum_{n\ge 0} (n+1)x^n,$$

所以 $x(1-x)^{-2} = x + 2x^2 + 3x^3 + \cdots = \sum nx^n$：这就是 $u_n = n$ 的生成函数。

同样地，$(1-x)^{-3}$ 中 x^n 的系数是 $\frac{(-3)(-4)\cdots(-3-n+1)}{n!}(-1)^n = \frac{(n+1)(n+2)}{2}$，所以

$$x(1-x)^{-3} = \sum_{n\ge 1} \frac{n(n+1)}{2}\, x^n = x + 3x^2 + 6x^3 + 10x^4 + \cdots,$$

这就是三角形数（triangle numbers）$u_n = \frac{1}{2}n(n+1)$ 的生成函数。

想一想　怎样用 n 和 $\frac{1}{2}n(n+1)$ 来表示 n^2？

由于 $n^2 = 2\cdot\frac{1}{2}n(n+1) - n$，而生成函数可以像数列一样相加，

$$\sum n^2 x^n = \frac{2x}{(1-x)^3} - \frac{x}{(1-x)^2} = \frac{2x - x(1-x)}{(1-x)^3} = \frac{x(1+x)}{(1-x)^3}. \qquad \blacksquare$$

阶段 2　递推关系变成方程

想一想　把 $u_n = ku_{n-1}$ 乘以 x^n，然后对所有 $n \ge 1$ 求和。两边用 f 表示，分别是甚么？

左边是 $\sum_{n\ge 1} u_n x^n = \mathrm{f}(x) - u_0 = \mathrm{f}(x) - a$。右边是 $kx\sum_{n\ge 1} u_{n-1}x^{n-1} = kx\,\mathrm{f}(x)$。所以 $\mathrm{f}(x) = a + kx\,\mathrm{f}(x)$，于是

$$\mathrm{f}(x) = \frac{a}{1-kx}. \qquad \blacksquare$$

当然，$u_n = ak^n$；把 $\frac{a}{1-kx}$ 展开成等比级数，便可确认这一点。

阶段 3　斐波那契数

对 $u_n = u_{n-1} + u_{n-2}$（其中 $n \ge 2$）作同样的处理：

$$\sum_{n\ge 2} u_n x^n = x\sum_{n\ge 2} u_{n-1}x^{n-1} + x^2 \sum_{n\ge 2} u_{n-2}x^{n-2},$$

即

$$\mathrm{f}(x) - u_0 - u_1 x = x\big(\mathrm{f}(x) - u_0\big) + x^2\,\mathrm{f}(x).$$

代入 $u_0 = 0$ 和 $u_1 = 1$，上式即 $\mathrm{f}(x) - x = x\,\mathrm{f}(x) + x^2\,\mathrm{f}(x)$，所以

$$\mathrm{f}(x) = \frac{x}{1 - x - x^2}. \qquad \blacksquare$$

阶段 4 **解释** $1/89$

> **停一停，试一试** 暂且不理收敛（convergence）的问题（这里没有问题：级数在 $|x| < 0.618$ 时收敛），把 $x = \frac{1}{10}$ 代入 f。你得到甚么？

$$\frac{1}{10}F_1 + \frac{1}{100}F_2 + \frac{1}{1000}F_3 + \cdots = \mathrm{f}\Big(\frac{1}{10}\Big) = \frac{1/10}{1 - 1/10 - 1/100} = \frac{10}{89}.$$

除以 10，得 $\frac{1}{89} = \sum F_n/10^{n+1}$。斐波那契数依次放在相继的小数位上，当它们大于 9 时便会互相重叠（图 9.4.1）。$1/89$ 开首的数字显示出斐波那契数；其后的数字则因进位而变得模糊。

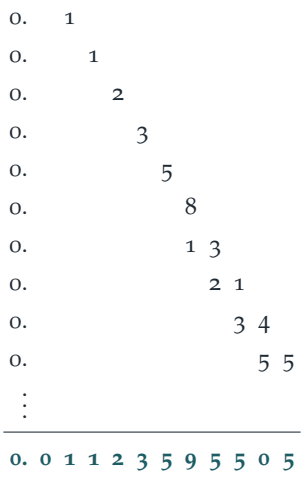


图 9.4.1 把 $F_n/10^{n+1}$ 对 $n = 1, 2, 3, \ldots$ 相加。每个斐波那契数的末位都比前一个向右多移一个小数位。一旦它们有两位数字，便会互相重叠，而进位会改变第七位起的数字：$\frac{1}{89} = 0.01123595505\ldots$。

回顾

- **把数列变成函数。**整个无穷数列变成了一个有理函数（rational function）。

- **递推关系变成方程。**乘以 x^n 再求和，便把规则 $u_n = u_{n-1} + u_{n-2}$ 变成一条关于 f 的方程，可以用代数方法解出。
- **从旧的造出新的。**把生成函数相加、平移和求导，便能得出新的生成函数，无须从头做起。

背后的数学

从生成函数到公式

生成函数不只是把数列编码：它还可以把数列解出来。分母可以因式分解为 $1 - x - x^2 = (1 - \varphi x)(1 - \psi x)$，其中 $\varphi = \frac{1+\sqrt{5}}{2}$ 和 $\psi = \frac{1-\sqrt{5}}{2}$ 是 $t^2 = t + 1$ 的根。由部分分式（partial fractions）可得

$$\frac{x}{1 - x - x^2} = \frac{1}{\sqrt{5}}\left(\frac{1}{1 - \varphi x} - \frac{1}{1 - \psi x}\right),$$

再把每个分式展开成等比级数，

$$F_n = \frac{\varphi^n - \psi^n}{\sqrt{5}}.$$

这就是*比内公式*（*Binet's formula*）。由于 $|\psi| < 1$，第二项非常小，所以 F_n 是最接近 $\varphi^n/\sqrt{5}$ 的整数，而相邻两个斐波那契数之比趋向黄金比（golden ratio）（这是 Section 2.2 的主题）。同一方法可以解任何常系数线性递推关系（linear recurrence with constant coefficients）；练习中有一个来自 STEP I 2004 的例子。

生成函数为何如此有用

真正的威力在相乘时才显现出来。乘积 $f(x)g(x)$ 中 x^n 的系数是 $\sum_k u_k v_{n-k}$，即对所有把 n 拆成两部分的方法求和。要数算由两个部分组成的东西，这正是我们所需要的。接下来的两节会用它来数算找换零钱的方法（Section 9.5），以及找出点数古怪、但总和的分布与普通骰子相同的骰子（Section 9.6）。在概率中，同一想法给出*概率生成函数*（*probability generating function*），它把独立随机变量之和变成乘积。

棣莫弗、欧拉与拉普拉斯

棣莫弗（Abraham de Moivre）在 1720 年代用生成函数来解线性递推关系，所用的方法基本上就是上面的方法。欧拉（Leonhard Euler）在 1748 年用它们来研究整数的分拆（partitions），而拉普拉斯（Pierre-Simon Laplace）则以它们作为其著作 *Théorie analytique des probabilités*（1812）的基础。到了二十世纪，生成函数成为组合数学的核心，而解析方法（把生成函数看成复变量的函数来研究）如今能为复杂得无法精确求解的计数问题给出精确的估计。

思维技巧

转换表示方式

一个关于数列的问题，在把数列变成函数、矩阵或图形之后，可能会变得容易。

生成函数把递推关系翻译成代数，把计数翻译成乘法。在大学里你会遇到许多这类翻译：傅里叶级数（Fourier series）把函数变成数列，拉普拉斯变换（Laplace transforms）把微分方程变成代数，而矩阵则把几何变成算术。

延伸练习

1 **每次两位数字。**试做 STEP II 2022，第 3 题（载于 OCR 提供下载的 2022 年试卷，`www.ocr.org.uk/Images/696587-2022-past-papers.zip`）。它解释了为何 $\frac{1}{89}$ 的小数展开式以斐波那契数开首，并要求找出一个分数，其小数每次两位数字地显示这个规律。

2 **解递推关系。**试做 STEP I 2004，第 7 题（`step.maths.org/questions/04-s1-q7`）。它用生成函数和部分分式解一个线性递推关系，然后把这个方法倒过来做。

3 **逐一数数。**证明 $\frac{1}{9801} = 0.00\,01\,02\,03\,04\ldots$，其中每两位数字为一组，逐一向上数。这个规律最先在哪里出错？为甚么？

4 **平方和。**解释为何 $\frac{1}{1-x}f(x)$ 是部分和 $u_0 + u_1 + \cdots + u_n$ 的生成函数。利用这一点和第 (i) 部分的答案，求 $1^2 + 2^2 + \cdots + n^2$ 的公式。

延伸阅读

- **Herbert Wilf, *generatingfunctionology***（A K Peters；网上免费阅读）。经典的入门书。读完本节后，最适合接着读第 1 章。
- **Ronald Graham, Donald Knuth & Oren Patashnik, *Concrete Mathematics***（Addison-Wesley），第 7 章。
- **Philippe Flajolet & Robert Sedgewick, *Analytic Combinatorics***（Cambridge University Press；网上免费阅读）。远远超出大学一年级的程度，但开首数页展示了这门学问的去向。

9.5 凑零钱

题目

STEP I 1997，第 1 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/97-s1-q1`

引子

这看来是全书最容易的一题：把所有可能逐一列出便是。可是列举很快就会失控。用英国人口袋里的硬币，凑出 20p 有 41 种方法，凑出 50p 有 451 种，凑出 £1 更有 4563 种。这道题真正要求的是一套*系统*，也就是一种整理计数的方法，使得既无遗漏，亦无重复。

这套系统背后是一段优美的数学。每种硬币对应一个等比级数（geometric series），凑出 n 便士的方法数就是这些级数之积的一个系数（coefficient）；把同一个想法一并用于所有正整数，便得到欧拉（Euler）的*分拆*（partition）理论，那是数论中最深刻的课题之一。

动笔之前

想一想 只用 2p 和 1p 硬币，凑出 n 便士有多少种方法？为甚么答案只取决于你用了多少枚 2p 硬币？

一旦选定 2p 硬币的数目，其余的就必须全是 1p 硬币，所以 2p 硬币的每个可能数目 $0, 1, \ldots, \lfloor n/2 \rfloor$ 都恰好对应一种方法。这样共有 $\lfloor n/2 \rfloor + 1$ 种方法，记作 $a(n)$。例如 $a(10) = 6$，$a(5) = 3$。

逐步解题

阶段 1　十便士

想一想　按面值最大的硬币来整理计数。你可以用多少枚 10p 硬币？定下来之后，又可以用多少枚 5p 硬币？

- 一枚 10p 硬币：1 种方法。
- 不用 10p，用两枚 5p：1 种方法。
- 不用 10p，用一枚 5p：余下的 5p 由 2p 和 1p 凑成，有 $a(5) = 3$ 种方法。
- 不用 10p 也不用 5p：10p 全由 2p 和 1p 凑成，有 $a(10) = 6$ 种方法。

总数是 $1 + 1 + 3 + 6 = 11$。 ■

阶段 2　逐层建立

想一想　设 $b(n)$ 为用 5p、2p 和 1p 硬币凑出 n 便士的方法数。怎样用函数 a 表示 $b(n)$？

选定 5p 硬币的数目 k；余下的 $n - 5k$ 便士由 2p 和 1p 硬币凑成。所以

$$b(n) = a(n) + a(n-5) + a(n-10) + \cdots .$$

由此得

n	0	5	10	15	20
$a(n)$	1	3	6	8	11
$b(n) = a(n) + b(n-5)$	1	4	10	18	29

同样地，设 $c(n)$ 为连 10p 硬币也可使用时的方法数；则 $c(n) = b(n) + b(n-10) + \cdots$，所以 $c(10) = 10 + 1 = 11$（与第 1 阶段的结果一致），而 $c(20) = b(20) + b(10) + b(0) = 29 + 10 + 1 = 40$。

阶段 3　二十便士

最后轮到 20p 硬币：我们要么用它（一种方法），要么不用（$c(20)$ 种方法）。所以共有

$$40 + 1 = 41 \text{ 种方法.} \qquad \blacksquare$$

停一停，试一试　留意当中的结构：每加入一种新硬币就多一层，而每一层都是把上一层按新硬币的面值为间隔取值再相加。你能看出电脑会怎样为 £1 甚至 £100 进行这项计算吗？

回顾

- **按最大的部分来整理。**先固定面值最大的硬币的数目，计数便有条不紊。
- **建立表格。**每加入一种新硬币，就多出一行，由上一行计算出来。这就是*动态规划*（dynamic programming），是电脑计算这类数目的标准方法。
- **与已知的情况核对。**表格重现了 10p 的 11 种方法。

背后的数学

把硬币看成等比级数

以下这个想法把问题化为代数。所用 5p 硬币的数目可以是 $0, 1, 2, \ldots$，分别贡献 $0, 5, 10, \ldots$ 便士。把这一点记录成级数

$$1 + x^5 + x^{10} + x^{15} + \cdots = \frac{1}{1 - x^5}.$$

对每种硬币都这样做，然后相乘：

$$\frac{1}{(1-x)(1-x^2)(1-x^5)(1-x^{10})(1-x^{20})} = \sum_n w(n)x^n.$$

把各括号乘开时，每一项 x^n 都来自在每个括号中各选一项，也就是为每种硬币各选一个数目，而总值为 n。所以 x^n 的系数 $w(n)$ 正好就是凑出 n 便士的方法数。第 2 阶段的表格，其实就是把这个乘积逐个括号乘开。这正是 Section 9.4 中母函数（generating function）的想法，而乘法正是它大显威力之处：母函数之积所计算的，就是各项选择的组合数目。

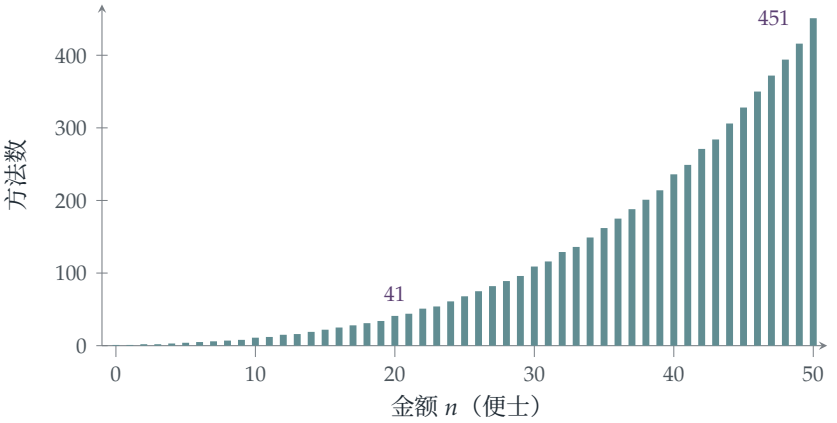


图 9.5.1 用 1p、2p、5p、10p、20p 和 50p 硬币凑出 n 便士的方法数。在 5 和 10 的倍数处出现的跳升，显示新硬币首次可用的位置。有六种硬币时，方法数的增长大致与 n^5 相若：一般而言，增长有如一个多项式，其次数比硬币种数少一。

分拆

假如每个正整数都可以作为硬币的面值，又会怎样？这时，不计次序而把 n 写成正整数之和的方法数，就是*分拆数*（partition number）$p(n)$。例如 $p(4)=5$，因为 $4=3+1=2+2=2+1+1=1+1+1+1$。欧拉在 1748 年看出

$$\sum_{n\ge 0} p(n)x^n = \frac{1}{(1-x)(1-x^2)(1-x^3)(1-x^4)\cdots},$$

并用它发现了一些非凡的恒等式（identity）。其中之一是：把 n 写成*互不相同*的部分之和的方法数，等于把它写成*奇数*部分之和的方法数。（用母函数来证明，只需一行代数：请在练习中一试。）

哈代、拉马努金与 $p(200)$

分拆数增长得很快：$p(10)=42$，$p(100)=190\,569\,292$。1918 年，哈代（G. H. Hardy）与拉马努金（Srinivasa Ramanujan）为 $p(n)$ 找到一条令人惊叹的公式，那是一个无穷级数（infinite series），其首项为

$$p(n) \approx \frac{1}{4n\sqrt{3}}\, e^{\pi\sqrt{2n/3}}.$$

为了检验这条公式，Percy MacMahon 利用欧拉的恒等式，徒手算出 $p(200)=3\,972\,999\,029\,388$。哈代与拉马努金的级数只取首几项，所得结果与正确答案相差不足一。拉马努金还发现 $p(5n+4)$ 必定可被 5 整除，$p(7n+5)$ 必定可被 7 整除，而 $p(11n+6)$ 必定可被 11 整除。至于为何再没有其他这类规律，则要到二十一世纪才得到解释。

思维技巧

藉系数来计数

当一个对象由若干独立的选择构成时，每项选择就是一个因式，而总大小为 n 的对象的数目，就是乘积中 x^n 的系数。

这把计数化为代数。在大学里，它是计数组合学（enumerative combinatorics）的基础，而概率母函数（probability generating function）也正是这样处理独立随机变量（independent random variables）之和的。

延伸练习

1 **两种六。**试做 STEP II 2012，第 1 题（`step.maths.org/questions/12-s2-q1`）。该题利用等比级数之积来计数。做完之后，请描述一个以这些系数为答案的硬币计数问题。

2 **互异与奇数。**证明 $(1+x)(1+x^2)(1+x^3)\cdots$ 是分拆成互不相同部分的母函数。利用 $1+x^k=\frac{1-x^{2k}}{1-x^k}$ 证明它等于 $\frac{1}{(1-x)(1-x^3)(1-x^5)\cdots}$，并由此推出欧拉定理。就 $n=6$ 验证这个结果。

3 **一条公式。**证明只用 1p、2p 和 5p 硬币凑出 n 便士的方法数，是最接近 $\frac{(n+4)^2}{20}$ 的整数。（先用较小的 n 验证。）

4 **次序有别。**把 1p 和 2p 硬币逐枚投入投币口来凑出 n 便士，若次序有别，共有多少种方法？（这个数列（sequence）你以前见过。）

延伸阅读

- **George Andrews & Kimmo Eriksson, *Integer Partitions***（Cambridge University Press）。一本为大学本科生而写的入门书，篇幅短小，浅白易读。
- **波利亚（George Pólya），“On picture-writing”**，*American Mathematical Monthly* 63 (1956)。波利亚以兑换一美元的问题（共有 292 种方法）来讲解母函数的经典文章。
- **Robert Kanigel, *The Man Who Knew Infinity***（Abacus）。拉马努金的生平，以及他与哈代的合作。

9.6 奇怪的骰子

题目

STEP I 1993，第 1 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/93-s1-q1`

引子

两颗六面骰子，各面标上正整数，除了平常的标法之外，还有没有别的标法，可以令由 2 到 12 的每个点数和出现的次数，都与普通骰子完全相同？1978 年，Martin Gardner 在《科学人》（*Scientific American*）杂志上宣布，这样的标法恰好只有一种：一颗骰子标上 $1,2,2,3,3,4$，另一颗标上 $1,3,4,5,6,8$。这对骰子由 George Sicherman 上校发现，现在市面上有售，称为 *Sicherman* 骰子。用它们来玩一局大富翁（Monopoly），与用普通骰子玩根本无从分辨。

本题是一道热身题：从点数和的列表，推断两颗骰子各面标了甚么数字。要有系统地解决这个问题，以至证明 Sicherman 骰子是唯一的一对奇异骰子，方法都是把每颗骰子化为一个多项式（polynomial）。

动笔之前

想一想 假设第一颗骰子有 p_i 个面显示 i，第二颗有 q_j 个面显示 j，那么掷出点数和 t 共有多少种方法？

方法的数目是 $\sum_i p_i q_{t-i}$：先选定第一颗骰子的点数 i，第二颗就必须显示 $t-i$。就题目中的第一对骰子而言，点数和 6 可以来自 $1+5$（有 1×1 种方法）、$2+4$（2×1 种）、$3+3$（2×2 种），此外再无其他可能，因为第一颗骰子没有 4 或 5，而第二颗没有 0。所以共有 $1+2+4=7$ 种方法。 ■

逐步解题

阶段 1　两端的情况

把两颗骰子称为 A 和 B，并以 a_i、b_i 分别表示两颗骰子上显示 i 的面数。

> **想一想**　点数和不可能是 2，也不可能是 12。这对 1 和 6 的分布有甚么启示？掷出 3 却确实有两种方法。哪一颗骰子上有 1？

掷不出点数和 2，即 $a_1 b_1 = 0$：最多只有一颗骰子上有 1。点数和 3 只可能是 $1+2$ 或 $2+1$，而且有两种方法，所以其中一颗骰子确实有 1；把它称为 A，于是 $b_1 = 0$。那么

$$\text{点数和}3: \quad a_1 b_2 = 2, \qquad \text{所以} (a_1, b_2) = (2,1) \text{ 或} (1,2).$$

同理，掷不出点数和 12，即 $a_6 b_6 = 0$。

阶段 2　逐步向上推

> **想一想**　写出点数和为 4 的方程，记住 $b_1 = 0$。在两种情况下，这个方程分别说明了甚么？

点数和 4 是 $1+3$ 或 $2+2$（由于 $b_1 = 0$，$3+1$ 不可能出现），所以 $a_1 b_3 + a_2 b_2 = 1$。

- **情况 (a)**，$a_1 = 2$，$b_2 = 1$：$2b_3 + a_2 = 1$，所以 $b_3 = 0$ 及 $a_2 = 1$。
- **情况 (b)**，$a_1 = 1$，$b_2 = 2$：$b_3 + 2a_2 = 1$，所以 $a_2 = 0$ 及 $b_3 = 1$。

点数和 5：$a_1 b_4 + a_2 b_3 + a_3 b_2 = 1$。

- 情况 (a)：$2b_4 + 0 + a_3 = 1$，所以 $b_4 = 0$，$a_3 = 1$。至此 A 有 $1,1,2,3$，B 有 2；B 其余五个面都是 5 或 6。
- 情况 (b)：$b_4 + 0 + 2a_3 = 1$，所以 $a_3 = 0$，$b_4 = 1$。至此 A 有 1，B 有 $2,2,3,4$；A 其余五个面取自 $4,5,6$，B 其余两个面取自 $5,6$。

阶段 3　情况 (a) 不成立

> **想一想**　在情况 (a) 中，写出点数和为 7 的方程，并利用 $b_5 + b_6 = 5$。

点数和 7：$a_1 b_6 + a_2 b_5 + a_3 b_4 + a_4 b_3 + a_5 b_2 = 2b_6 + b_5 + 0 + 0 + a_5 = 3$。代入 $b_6 = 5 - b_5$，得 $10 - b_5 + a_5 = 3$，所以 $b_5 = 7 + a_5 \geq 7$。但 B 只有六个面。情况 (a) 不可能成立。

阶段 4　情况 (b) 成立

想一想 在情况 (b) 中，利用点数和 6 和 11，以及 $a_6 b_6 = 0$ 这个事实。

点数和 6：$a_1 b_5 + a_4 b_2 = b_5 + 2a_4 = 4$（其余各个乘积都含有 $a_2 = a_3 = 0$ 或 $b_1 = 0$）。由于 $b_5 \le 2$，而且 $b_5 = 4 - 2a_4$ 是偶数，所以要么 $b_5 = 0, a_4 = 2$，要么 $b_5 = 2, a_4 = 1$。

- 若 $b_5 = 0$，则 B 是 $2,2,3,4,6,6$，于是 $a_6 = 0$，而 A 是 $1,4,4,5,5,5$。但这样点数和 7 就有 $a_1 b_6 + a_4 b_3 + a_5 b_2 = 2 + 2 + 6 = 10$ 种方法，而不是 3 种。
- 若 $b_5 = 2$，则 B 是 $2,2,3,4,5,5$，其中 $b_6 = 0$，而 A 是 $1,4$ 再加上四个取自 5 和 6 的面。点数和 11：$a_6 b_5 = 2a_6 = 6$，所以 $a_6 = 3$ 及 $a_5 = 1$。

唯一的可能是

$$A:\ 1,4,5,6,6,6, \qquad B:\ 2,2,3,4,5,5.$$

最后，我们必须验证这对骰子确实给出整个列表，因为我们只用了其中一部分点数和。把乘积展开（或利用下文的多项式），即可确认每一项都吻合。 ■

回顾

- **从两端入手。**最小和最大的点数和只牵涉一两个乘积，所以提供的信息最有力。
- **分情况讨论，逐一排除。**每个点数和都把可能性收窄一些，直到只剩下一种情况。
- **检查充分性（sufficiency）。**推论说明了骰子*必须*是甚么样子；最后的验证则说明这对骰子确实可行。

背后的数学

把骰子看成多项式

把每颗骰子写成一个多项式：每个显示 i 的面对应一项 x^i。普通骰子是 $x + x^2 + x^3 + x^4 + x^5 + x^6$，而题目中的两颗骰子是

$$A(x) = x + x^4 + x^5 + 3x^6, \qquad B(x) = 2x^2 + x^3 + x^4 + 2x^5.$$

相乘时，每一对面贡献 $x^i x^j = x^{i+j}$，所以 $A(x)B(x)$ 中 x^t 的系数（coefficient），就是掷出点数和 t 的方法数目。这又是 Section 9.4 中母函数（generating function）的想法：点数和的列表其实就是乘积 $A(x)B(x)$，而我们逐个情况的推论，正是把它因式分解（factorise）的一种方法。

Sicherman 骰子

两颗普通骰子的乘积是 $(x+x^2+\cdots+x^6)^2$。任何一对点数和分布相同的骰子，其多项式 $P(x)$、$Q(x)$ 必须满足 $P(x)Q(x)=(x+\cdots+x^6)^2$。现在作因式分解：

$$x+x^2+\cdots+x^6=x(1+x)(1+x+x^2)(1-x+x^2),$$

而这些因式不能再进一步分解（成为整数系数的多项式）。所以 P 和 Q 都必须由因式 x、$1+x$、$1+x+x^2$、$1-x+x^2$ 组成，而每个因式合共出现两次。以下三个条件把选择收窄：

- 每颗骰子有 6 个面，所以 $P(1)=Q(1)=6$；各因式在 $x=1$ 时的值分别是 1、2、3、1，所以每颗骰子各得一个 $1+x$ 和一个 $1+x+x^2$；
- 没有任何面显示 0，所以 $P(0)=Q(0)=0$：每颗骰子各得一个因式 x；
- 余下唯一的自由，是两个 $1-x+x^2$ 如何分配。每颗各得一个，就是普通骰子。两个都归同一颗骰子，则得

$$x(1+x)(1+x+x^2)=x+2x^2+2x^3+x^4,$$
$$x(1+x)(1+x+x^2)(1-x+x^2)^2=x+x^3+x^4+x^5+x^6+x^8.$$

这就是 Sicherman 骰子，1, 2, 2, 3, 3, 4 和 1, 3, 4, 5, 6, 8（图 9.6.1），而以上论证亦证明了它们是唯一的另一种可能。关键在于：多项式与整数一样，可以唯一地分解为不可约（irreducible）因式的乘积——数论与组合数学（combinatorics）在此相遇。

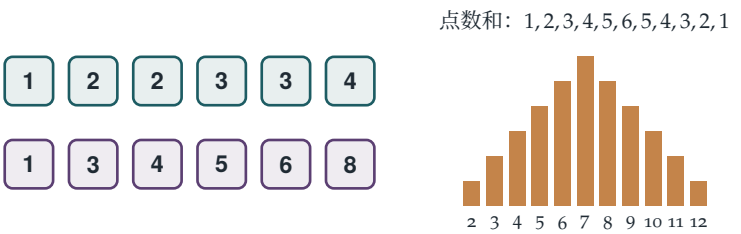


图 9.6.1　Sicherman 骰子。两颗一起掷，由 2 到 12 的每个点数和出现的频数，都与两颗普通骰子完全相同（右）：掷出 2 有一种方法，掷出 3 有两种方法，如此类推，直至掷出 7 有六种方法。

概率母函数

在概率论中，同一个想法以*概率母函数*（probability generating function）的形式出现：取值 $0,1,2,\ldots$ 的随机变量（random variable）X 有 $\mathrm{G}(t)=\mathrm{E}(t^X)=\sum \mathrm{P}(X=k)t^k$，而对于独立（independent）的 X 和 Y，$\mathrm{E}(t^{X+Y})=\mathrm{E}(t^X)\mathrm{E}(t^Y)$。独立随机变量之和，对应于母函数之积。STEP III 2003 Q14（见练习）就是用这个方法找出 Sicherman 骰子，

而 Section 9.13 则用一个相关的论证，说明无论怎样在两颗骰子上做手脚，都不能令所有点数和出现的机会均等。

思维技巧

把计数问题编码成多项式

当物件以数值相加的方式组合时，把每个物件表示成 x 的多项式。组合变成相乘，而有关组合的问题就变成有关因式分解的问题。

在这里，一个关于骰子的问题化成了多项式的唯一分解（unique factorisation）。在大学里，这类编码随处可见：在组合数学中，在概率论中，也在代数中——在代数里，多项式本身就是研究对象，它们构成一个环（ring），其算术与整数的算术相似。

延伸练习

1 **四面体骰子。**试做 STEP III 2003, 第 14 题（`step.maths.org/questions/03-s3-q14`），该题就四面骰子提出 Sicherman 的问题。（提示：$x + x^2 + x^3 + x^4 = x(1 + x)(1 + x^2)$。）

2 **验证列表。**展开 $(x + x^4 + x^5 + 3x^6)(2x^2 + x^3 + x^4 + 2x^5)$，并与题目中的列表对照。

3 **三颗骰子。**是否存在三颗六面骰子，各面标上正整数，不全是标准骰子，而其点数和的分布与三颗标准骰子相同？（提示：$(x + \cdots + x^6)^3$ 的因式现在包括三个 $1 - x + x^2$。）

4 **一枚硬币与一个八面体。**一颗标上 $1, \ldots, 8$ 的八面骰子，和一颗标上 $1, 2$ 的两面“骰子”（即一枚硬币）：能否用另一种方式重新标号，而点数和的分布保持不变？

延伸阅读

- **Martin Gardner, *Penrose Tiles to Trapdoor Ciphers*** (Mathematical Association of America)。收录了关于 Sicherman 骰子的那篇专栏文章。
- **Joseph Gallian & David Rusin, “Cyclotomic polynomials and nonstandard dice”**, *Discrete Mathematics* 27 (1979)。一般理论，适用于任意面数的骰子。
- **Herbert Wilf, *generatingfunctionology*** （可于网上免费阅览），第 1 章。

9.7 相等的幂和

题目

STEP I 1992，第 5 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/92-s1-q5`

引子

把 1 至 16 这些数分成两队，每队八个：

$$\{1,4,6,7,10,11,13,16\} \quad 和 \quad \{2,3,5,8,9,12,14,15\}.$$

两队的和都是 68。两队的平方和都是 748。两队的立方和都是 9248。很难相信这不是纯粹的巧合，但它其实是一个无穷无尽的规律的第三步：只要用正确的方法把 1 至 32 分开，连四次方之和也会相等。

这个分法由 Eugène Prouhet 于 1851 年发现，而它背后的规则，即 *Thue–Morse 数列*（Thue–Morse sequence），出现的地方多得惊人：数论、计算理论、国际象棋，以至最公平的轮流方式。本题通过一个多项式（polynomial）及其导数（derivative）把它发掘出来。

动笔之前

想一想 若多项式 $p(x) = \sum c_k x^k$ 满足 $p(1) = 0$，这对各系数（coefficient）有甚么含意？若 $p'(1) = 0$ 又如何？

$p(1) = \sum c_k$，所以 $p(1) = 0$ 表示各系数之和为零。而 $p'(x) = \sum k c_k x^{k-1}$，所以 $p'(1) = \sum k c_k$。若所有系数都是 ± 1，这两个条件就是说：$+1$ 与 -1 所在的位置数目相同，而且位置之和也相等。

逐步解题

阶段 1　一个三重根

想一想　把 $1-x^2$ 和 $1-x^4$ 因式分解。当中共出现多少个因式 $1-x$?

由于 $1-x^2=(1-x)(1+x)$ 及 $1-x^4=(1-x)(1+x)(1+x^2)$，

$$p_0(x)=(1-x)^3(1+x)^2(1+x^2),$$

所以 $(1-x)^3$ 是一个因式（factor）。 ■

阶段 2　在 1 处的导数

想一想　若 $(1-x)^3$ 整除 p_0，那么 $(1-x)$ 的多少次方整除 p_0'? 接着 $p_1=xp_0'$ 和 p_1' 又如何?

写成 $p_0(x)=(1-x)^3q(x)$。根据积法则（product rule），$p_0'(x)=-3(1-x)^2q(x)+(1-x)^3q'(x)$，它有因式 $(1-x)^2$。所以 $p_0'(1)=0$。于是 $p_1(x)=xp_0'(x)$ 同样有因式 $(1-x)^2$，而用相同的论证可知 p_1' 有因式 $(1-x)$，所以 $p_1'(1)=0$。 ■

阶段 3　读出各系数

展开可得 $(1-x)(1-x^2)=1-x-x^2+x^3$，然后

$$p_0(x)=1-x-x^2+x^3-x^4+x^5+x^6-x^7.$$

想一想　以系数 c_k 表示，$p_0'(1)$ 和 $p_1'(1)$ 分别是甚么?

由于 $p_1(x)=xp_0'(x)=\sum kc_kx^k$，我们有 $p_1'(1)=\sum k^2c_k$。系数在 $k=0,3,5,6$ 时为 $+1$，在 $k=1,2,4,7$ 时为 -1。所以 $p_0'(1)=0$ 和 $p_1'(1)=0$ 即是说

$$0+3+5+6=1+2+4+7\quad(=14),$$
$$0^2+3^2+5^2+6^2=1^2+2^2+4^2+7^2\quad(=70),$$

略去 0 之后，这正是所需的恒等式。 ■

阶段 4　再多一个因式

> **停一停，试一试** 要令立方和也平衡，我们需要第四个因式 $(1-x)$。应该用哪一个多项式取代 p_0？

取 $\mathrm{P}(x)=(1-x)(1-x^2)(1-x^4)(1-x^8)$，它有因式 $(1-x)^4$。它的系数来自 $\mathrm{p}_0(x)-x^8\mathrm{p}_0(x)$：先是 0 至 7 的正负号规律，然后在 8 至 15 重复同一规律，但正负号全部相反：

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
c_k	+	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+

把第 2 阶段的论证再用一次，可知对于 $r=0,1,2,3$ 都有 $\sum c_k k^r=0$。取 $+$ 号的位置 $\{0,3,5,6,9,10,12,15\}$ 和 $-$ 号的位置 $\{1,2,4,7,8,11,13,14\}$。当 $r=1,2,3$ 时，来自 0 的一项为零，所以可以略去。因此以下的次序

$$a_1,\dots,a_8=1,2,4,7,8,11,13,14,\qquad a_9,\dots,a_{15}=3,5,6,9,10,12,15$$

符合要求：两半的和都是 60，平方和都是 620，立方和都是 7200。 ■

回顾

- **导数能侦测重根（repeated root）。**因式 $(1-x)^m$ 会迫使首 $m-1$ 阶导数在 $x=1$ 处全部为零。
- **先乘以 x 再求导**把 $\sum kc_k$ 变成 $\sum k^2c_k$，如此类推便得到更高次方。
- **倍增。**每多一个因式 $(1-x^{2^m})$，长度便增加一倍，同时多一个次方达到平衡。

背后的数学

Thue–Morse 数列

$(1-x)(1-x^2)(1-x^4)(1-x^8)\cdots$ 各系数的正负号构成 *Thue–Morse* 数列。以 0 代表 $+$、1 代表 $-$，写出来就是：

$$0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1\,0\,1\,1\,0\ \dots$$

它有三种等价的描述。

- **倍增：**由 0 开始，不断把手上已有的部分的补（complement，即把 0 与 1 互换）接在后面：$0\to 01\to 0110\to 01101001\to\cdots$。
- **二进制数字：**把 k 写成二进制（binary）时，若当中 1 的个数是偶数，第 k 项便是 0；若是奇数，便是 1。（这就是 x^k 的正负号为 $(-1)^{1\text{的个数，取自}k}$ 的原因：要从乘积中得出 x^k，就要选取那些在二进制中组成 k 的因子 x^{2^m}，而每一个都带来一个负号。）

► **多项式：** $\prod_m (1 - x^{2^m})$ 的各系数，一如题目所示。

按 Thue–Morse 数列把 $0, 1, \ldots, 2^N - 1$ 分开，所得的两个集合对于每个 $r < N$ 都有相等的 r 次方之和（图 9.7.1）。这就是 *Prouhet–Tarry–Escott 问题*（Prouhet–Tarry–Escott problem）：找出两个不同的整数集合，使它们的 r 次方之和在 $r = 1, \ldots, k$ 时全部相等。Prouhet 的解用了 2^{k+1} 个数；究竟最少需要多少个数，是一个著名的未解决问题。

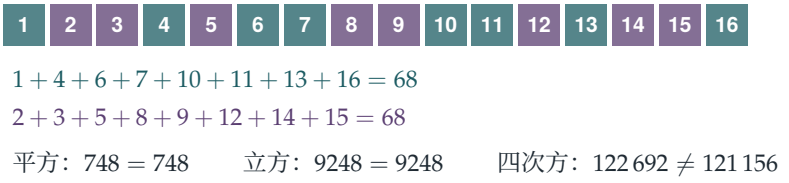


图 9.7.1　按 Thue–Morse 数列分开的 1 至 16。两个集合的和相等、平方和相等、立方和也相等，但四次方之和并不相等；要做到这一点，便需要 1 至 32。由左至右读出颜色，得到的规律是蓝绿–紫–紫–蓝绿、紫–蓝绿–蓝绿–紫、...。

最公平的轮流方式

两个小孩轮流从一堆糖果中挑选，最好的先被选走。若他们以 $ABABAB\ldots$ 的次序交替，那么每一对糖果中较好的一颗总是归 A，这是一种系统性的优势。$ABBA$ 的次序较为公平，而 $ABBABAAB$ 更加公平：这正是 Thue–Morse 数列，而次方和平衡恰恰就是个中原因。若糖果的价值由其位置的一个光滑函数给出，把该函数展开成多项式，便可看到 Thue–Morse 次序会把它的首几项抵销。同样的想法曾被提议用于足球的互射十二码（足球规则制定者自 2017 年起试行的“$ABBA$”次序），而网球决胜局的发球次序亦采用这个方式。

Thue、Morse 与 Euwe

这个数列曾多次被人发现。Prouhet 在 1851 年用过它。Axel Thue 在 1912 年研究没有重复规律的数列时重新发现了它：它永远不会连续三次出现同一段（即它是“无立方”（cube-free）的）。Marston Morse 在 1921 年研究曲面上的测地线（geodesics）时再次发现它。而国际象棋世界冠军尤伟（Max Euwe）在 1929 年用它证明了，当时生效的重复局面和棋规则并不能防止对局无限延续下去。时至今日，它是*自动数列*（automatic sequences）理论中的核心例子；自动数列是指由简单机器读取 n 的二进制数字而计算出来的数列。

思维技巧

把组合性质编码成多项式

次方和是否平衡，关键在于 $+$ 号和 $-$ 号落在哪些位置。以这些正负号作为系数的多项式，可把这个性质转化为一个关于根的命题。

本题从头到尾都没有要求你去搜索：$x = 1$ 处的一个重根已经包办了所有工作。利用代数来构造组合（combinatorial）对象，而不是靠反复试验去找，是大学组合数学中一再出现的想法。

延伸练习

1 **平移。**试做 STEP II 2025, 第 8 题（载于 OCR 提供下载的 2025 年试卷，`www.ocr.org.uk/Images/750486-2025-past-papers.zip`）。该题说明怎样把一个令和与平方和平衡的分法，通过平移其中一列再合并，变成一个更大而且连立方和也平衡的分法。

2 **平方的平方。**（同样出自 STEP II, 2025。）集合 $A = \{1,3,4,5,9,11\}$ 和 $B = \{2,6,7,8,10\}$ 的和相等，平方和也相等。对于每个正整数 n，把 $\{n^2, (n+1)^2, \ldots, (n+11)^2\}$ 分成两个和相等的集合。（提示：把 0 加入第二个集合。）

3 **平移不变性 (translation invariance)。**证明：若两个数集的 r 次方之和在 $r = 0, 1, \ldots, k$ 时全部相等，那么把每个元素都加上同一个数 c 所得的两个集合亦然。为甚么这表示题目中的集合 $\{1, \ldots, 15\}$ 同样可以换成任意 16 个连续整数?

4 **公平分糖果。**糖果的价值为 $1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{8}$。比较 A 在 $ABABABAB$ 和 $ABBABAAB$ 两种次序下所得的总值。

延伸阅读

- **Jean-Paul Allouche & Jeffrey Shallit, "The ubiquitous Prouhet–Thue–Morse sequence"**, 载于 *Sequences and Their Applications*（Springer, 1999）。综述这个数列在各处的出现；可在网上免费阅读。
- **Matt Parker, *Things to Make and Do in the Fourth Dimension***（Penguin）。内容包括等次方和及其他数字趣闻。
- **Joshua Cooper & Aaron Dutle, "Greedy Galois games"**, *American Mathematical Monthly* 120 (2013)。公平轮流与 Thue–Morse 数列。

9.8 没有人拿到自己的恤衫

题目

STEP II 2016，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/16-s2-q12`

引子

一支十一人的足球队匆忙更衣，各人随手抓起一件球衣就穿。没有任何人穿上自己球衣的机会有多大？再设想那是一支五十人的大军，甚至是一百万人的人群。人多了，出错的人也多了，机会是否随之上升？还是因为有更多人可能碰巧拿对，机会反而下降？出人意料的是，这个机会几乎完全不变：由五个人开始，它与 $1/e \approx 0.368$ 的差距不超过半个百分点。

这是组合概率（combinatorial probability）中最古老的问题，由 Pierre de Montmort 于 1708 年提出，称为 *problème des rencontres*（相遇问题）。本题以容斥原理（inclusion–exclusion principle）解决这个问题；容斥原理是计数互相重叠的事物的工具，而所得的答案，是 e 在数学中最令人意想不到的现身之一。

动笔之前

停一停，试一试 先试 $n = 3$：列出卡 $1, 2, 3$ 的全部六种次序，数一数其中有多少种是没有任何一张卡在自己位置上的。如果你有耐性，再试 $n = 4$（共有 24 种次序）。

当 $n = 3$ 时，只有 231 和 312 没有任何卡在自己的位置上，所以概率是 $\frac{2}{6} = \frac{1}{3}$。当 $n = 4$ 时，24 种次序中有 9 种符合条件，概率为 $\frac{3}{8} = 0.375$。没有任何一张卡留在原位的排列，称为错排（derangement）。

逐步解题

阶段 1　三个事件的容斥原理

想一想　怎样把 $A \cup B \cup C$ 写成两个事件的并集（union），使两个事件的公式可以派上用场？

把 $A \cup B$ 看成单一事件，对它和 C 应用公式：

$$\mathrm{P}(A \cup B \cup C) = \mathrm{P}(A \cup B) + \mathrm{P}(C) - \mathrm{P}((A \cup B) \cap C).$$

现在 $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$，再应用一次公式，得 $\mathrm{P}((A \cap C) \cup (B \cap C)) = \mathrm{P}(A \cap C) + \mathrm{P}(B \cap C) - \mathrm{P}(A \cap B \cap C)$。把 $\mathrm{P}(A \cup B)$ 也展开，再合并各项，便得出结果。 ■

四个事件时，规律依旧：把四个单一事件的概率相加，减去六个两两交集（intersection）的概率，加上四个三重交集的概率，再减去 $\mathrm{P}(A \cap B \cap C \cap D)$。

阶段 2　各个概率

想一想　如果卡 i 在位置 i，其余的卡有多少种排法？如果卡 i 和卡 j 都固定了呢？

共有 $n!$ 种机会均等的次序。如果卡 i 固定在位置 i，其余 $n-1$ 张卡可以任意排列：共 $(n-1)!$ 种排法。所以

$$\mathrm{P}(E_i) = \frac{(n-1)!}{n!} = \frac{1}{n}, \qquad \mathrm{P}(E_i \cap E_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)},$$
$$\mathrm{P}(E_i \cap E_j \cap E_k) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}.$$

一般而言，指定的 k 张卡全部在自己位置上的概率是 $\frac{(n-k)!}{n!}$。

阶段 3　加起来

想一想　在 $\mathrm{P}(E_1 \cup \cdots \cup E_n)$ 的容斥公式中，涉及 k 个事件交集的项有多少个？它们的总和是多少？

从这些事件中选出 k 个，有 $\binom{n}{k}$ 种方法，而每个交集的概率都是 $\frac{(n-k)!}{n!}$。所以各个 k 重项的总和是

$$\binom{n}{k}\frac{(n-k)!}{n!} = \frac{n!}{k!\,(n-k)!} \cdot \frac{(n-k)!}{n!} = \frac{1}{k!}.$$

除了 $k!$ 之外，全部约去。配上正负相间的符号，

$$\mathrm{P}(\text{最少有一张卡在自己的位置上}) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1}\frac{1}{n!}. \qquad \blacksquare$$

所以出现错排的概率是

$$d_n = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n\frac{1}{n!} = \sum_{k=0}^{n}\frac{(-1)^k}{k!}.$$

阶段 4 恰好一张

? **想一想** 要恰好有一张卡在原位，是哪一张卡？其余的卡又必须怎样？

先选定哪一张卡在自己的位置上（n 种选法）。卡 i 在原位的概率是 $\frac{1}{n}$；在这个条件下，其余 $n-1$ 张卡必须在它们之间形成一个错排，其概率为 d_{n-1}。所以

$$\mathrm{P}(\text{恰好一张}) = n \cdot \frac{1}{n} \cdot d_{n-1} = \sum_{k=0}^{n-1}\frac{(-1)^k}{k!}. \qquad \blacksquare$$

回顾

- **先多算，再修正。**把各个 $\mathrm{P}(E_i)$ 相加，会把有两张卡固定的排列算了两次；减去两两交集的项可以修正，但对三张卡固定的情况又减得太多；如此类推。
- **对称性令各项相等。**每个 k 重交集的概率都相同，所以公式的每一层都是一个数乘以一个二项式系数（binomial coefficient）。
- **化为较小的情况。**“恰好一张”变成了 $n-1$ 张卡的错排。

背后的数学

e 从何而来

级数 $\sum(-1)^k/k!$ 就是指数级数（exponential series）$e^x = \sum x^k/k!$ 在 $x=-1$ 时的值。所以当 $n \to \infty$ 时，

$$d_n \to e^{-1} \approx 0.3679,$$

而恰好有一张卡对上的概率也趋向同一个极限。收敛（convergence）的速度快得惊人（图 9.8.1）：在 $1/n!$ 一项停下来所产生的误差小于 $1/(n+1)!$，所以在 $n=7$ 时，答案已经与 $1/e$ 吻合至小数点后四位。事实上，n 件物件的错排数目 $n!\,d_n$ 恰好是最接近 $n!/e$ 的整数。

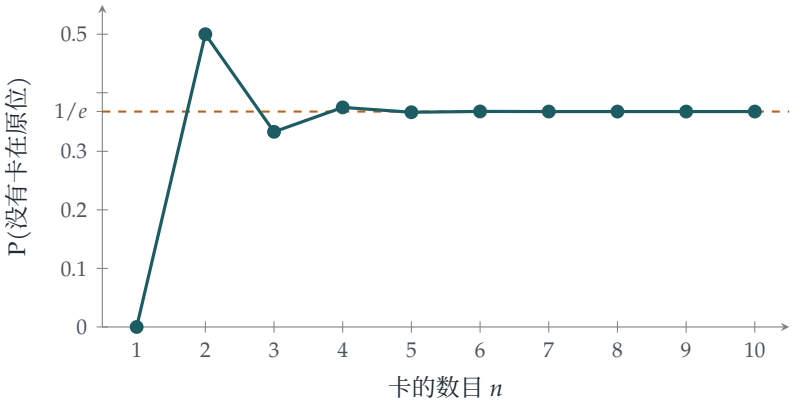


图 9.8.1　把 n 张卡随机洗匀后，没有任何一张卡留在自己位置上的概率。它在 $1/e$ 的上下交替出现，而由 $n = 6$ 开始，在这个比例尺下已经与 $1/e$ 无法分辨。

对上的数目服从泊松分布

结论还不止于此。设 N 为在自己位置上的卡的数目。那么对每个 k，都有 $\mathrm{P}(N = k) \to \frac{e^{-1}}{k!}$：在极限之下，对上的数目服从平均值为 1 的泊松分布（Poisson distribution）。平均值为甚么是 1，不用任何计算也看得出来：N 是 $E_1, \ldots, E_n$ 的指示变量（indicator）之和，每个的期望值（expectation）都是 $\frac{1}{n}$，而即使这些事件并非独立，期望值仍然可以相加。STEP III 2004 Q12（见练习）证明方差（variance）同样是 1。每当有许多机会很微、又近乎独立的事件时，泊松分布就会出现；错排是一个经典例子，也是第 12 章 Section 12.11 的主题。

Montmort 的 *treize* 游戏

Pierre Rémond de Montmort 是一位法国贵族，他在著作 *Essay d'analyse sur les jeux de hazard*（1708）中研究各种博彩游戏。在 *treize*（十三）这个游戏中，玩家派出十三张牌，同时口中数着“一、二、三……”，只要有一次派出的牌与所数的数字相同，便算胜出。Montmort 求出了胜出的概率，并在与尼古拉·伯努利（Nicolaus Bernoulli）的通信中，于 1713 年或之前确立了一般公式。欧拉（Euler）在 1753 年重新发现了这条公式。容斥原理其后由棣莫弗（Abraham de Moivre）加以形式化，而它的一般形式则由西尔维斯特（James Joseph Sylvester）在十九世纪建立。

思维技巧

容斥原理

要计数互相重叠的集合的并集，先把各集合的大小相加，减去两两重叠的部分，加上三重重叠的部分，如此类推。

当对称性令同一大小的重叠部分全部相等时，这个方法最见威力，本题正是如此。在大学里，它会在组合数学（combinatorics）、概率论和数论（number theory）中出现；在数论中，它用来计数不能被一系列质数中任何一个整除的整数，并由此引出欧拉函数（Euler's totient function）和默比乌斯函数（Möbius function）。

延伸练习

1 **球衣。**试做 STEP III 2004，第 12 题（`step.maths.org/questions/04-s3-q12`）。这道题用指示变量计数穿上自己球衣的球员人数，求出平均值和方差，并把结果与泊松分布比较。

2 **递推关系。**设 D_n 为 n 件物件的错排数目。考虑卡 1 去了哪个位置，以及原本在该位置的卡是否去了位置 1，证明 $D_n = (n-1)(D_{n-1} + D_{n-2})$。用 $D_3 = 2$、$D_4 = 9$ 验算。

3 **恰好 k 张。**证明恰好有 k 张卡在自己位置上的概率是 $\frac{1}{k!}d_{n-k}$，并且它趋向 $\frac{e^{-1}}{k!}$。

4 **互质的整数。**用容斥原理计数由 1 至 30 的整数中，不能被 2、3 和 5 任何一个整除的有多少个。你的答案与 $30\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\left(1-\frac{1}{5}\right)$ 有甚么关系？

延伸阅读

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press）。大学一年级教科书，涵盖容斥原理和配对问题。
- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1**（Wiley），第 IV 章。经典之作，载有配对问题的许多变化。
- **Miklós Bóna, *A Walk Through Combinatorics***（World Scientific），关于容斥原理的部分。

9.9 生日问题

题目

STEP II 2007，第 13 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/07-s2-q13`

引子

一个房间里要有多少人，才会令其中两人同一天生日的机会大于一半？大多数人会猜 180 左右，即 365 的一半。答案却是 23。在一班三十人的课堂里，有人同一天生日的概率（probability）是 70%；在五十人的派对上，则是 97%。

这个结果令人意外，是因为我们把两个不同的问题混为一谈：“有没有人与*我*同一天生日？”（这确实需要大约 253 人）和“有没有*任何两个人*同一天生日？”。23 个人可以组成 $\binom{23}{2} = 253$ 对，而每一对都是一次出现相同生日的机会。这道题把这个直觉变得精确，而同样的数学亦决定了保护你银行户口的密码技术有多安全。

动笔之前

想一想 为甚么计算*没有*两位宾客同一天生日的概率会比较容易？请宾客排成一行：第二位避开第一位生日的机会是多少？第三位同时避开前两位的机会又是多少？

“至少有两人相同”包含许多互相重叠的情况（两人相同、三人相同、两对不同的人各自相同……）。它的补事件（complement）“全部不同”却是单一而简洁的事件。第二位宾客避开第一位生日的概率是 $\frac{364}{365}$；在此条件下，第三位同时避开前两位的概率是 $\frac{363}{365}$；如此类推。

逐步解题

阶段 1　一个有用的近似

想一想　$y = e^{-x}$ 在 $x = 0$ 处的切线（tangent line）是甚么？

由于 $e^{-x} = 1 - x + \frac{x^2}{2} - \cdots$，当 x 很小时，我们有 $e^{-x} \approx 1 - x$。取 $x = r/n$，得

$$\frac{n-r}{n} = 1 - \frac{r}{n} \approx e^{-r/n}. \qquad \blacksquare$$

误差约为 $\frac{r^2}{2n^2}$，当 r 远小于 n 时，这个误差很小。

阶段 2　有人同一天生日的概率

k 个生日全部不同的概率是

$$\frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365-(k-1)}{365} = \prod_{r=0}^{k-1} \frac{365-r}{365}.$$

对每个因子使用上述近似（approximation），再利用指数函数相乘时指数相加这个性质，

$$\prod_{r=0}^{k-1} e^{-r/365} = \exp\Big(-\frac{0+1+\cdots+(k-1)}{365}\Big) = \exp\Big(-\frac{k(k-1)}{730}\Big).$$

因此，至少有一对人同一天生日的概率约为 $1 - e^{-k(k-1)/730}$。 $\blacksquare$

想一想　看看指数：$\frac{k(k-1)}{730} = \binom{k}{2}\Big/365$。用文字表达，这说明了甚么？

它是宾客的配对数目，乘以某一对人同一天生日的概率 $\frac{1}{365}$：也就是生日相同的配对数目的期望值（expected number）。这个近似告诉我们，*没有*任何配对相同的机会是 $e^{-(\text{相同配对的期望数目})}$，这又一次透露出泊松分布（Poisson distribution）的影子，正如 Section 9.8 所见。

阶段 3　二十三

我们需要 $1 - e^{-k(k-1)/730} \geq \frac{1}{2}$，即 $e^{-k(k-1)/730} \leq \frac{1}{2}$，亦即

$$\frac{k(k-1)}{730} \geq \ln 2 \approx \frac{253}{365}, \qquad \text{即} \qquad k(k-1) \geq 506.$$

由于 $22 \times 21 = 462$ 而 $23 \times 22 = 506$，最小的值是 $k = 23$。 $\blacksquare$

（23 位宾客时的精确概率是 0.507，22 位宾客时是 0.476，所以这个近似给出了正确答案。）

阶段 4　与主人同一天生日

想一想 对每位宾客来说，不与主人同一天生日的概率是多少？在这方面，各位宾客之间是否互相独立（independent）？

每位宾客各自独立地以概率 $\frac{364}{365} \approx e^{-1/365}$ 避开主人的生日。因此 k 位宾客中无人与主人同一天生日的概率是 $\left(\frac{364}{365}\right)^k \approx e^{-k/365}$，而我们需要

$$e^{-k/365} \le \tfrac{1}{2}, \qquad \text{即} \qquad k \ge 365 \ln 2 \approx 253.$$

所以必须有 253 位宾客。 ■

这个对比正是重点所在：要与某一个指定的人生日相同，需要 253 位宾客；但要有某一对人生日相同，只需 23 人，因为 23 个人可以组成 253 对。

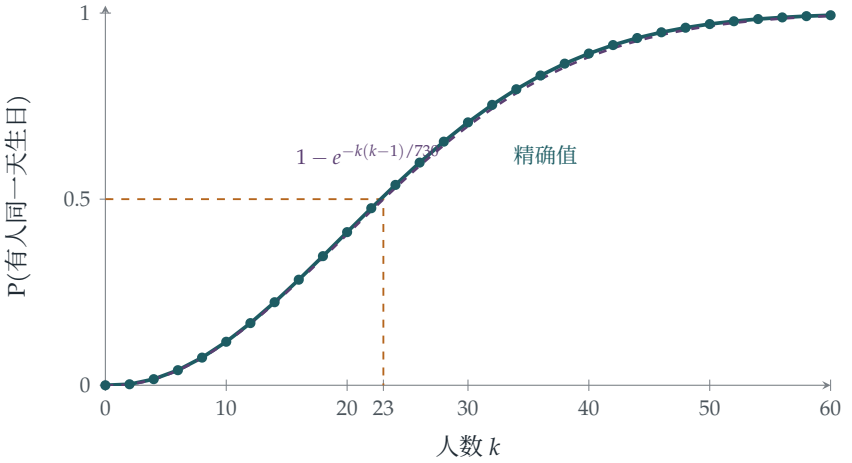


图 9.9.1 k 个人中至少有两人同一天生日的概率：精确值（蓝绿色）与近似值 $1 - e^{-k(k-1)/730}$（紫色虚线）。两者非常接近，并且都在 $k = 22$ 与 $k = 23$ 之间越过 $\frac{1}{2}$。

回顾

- **利用补事件。**“全部不同”只是一个乘积；“至少有一对相同”却是一团乱麻。
- **先线性化，再取指数。**把每个因子 $1 - x$ 换成 e^{-x}，乘积就变成了和。
- **数配对。**指数正是配对的数目乘以一对人相同的机会。$k(k-1)$ 中的平方，就是这个悖论的全部解释。

背后的数学

平方根定律

若把 365 天换成 n 个等可能的日子，同样的论证给出

$$\mathrm{P}(\text{有人相同，人数为}k) \approx 1 - e^{-k^2/(2n)},$$

所以只要 k 达到 $\sqrt{n}$ 左右，而不是 n，就很可能出现相同的情况。更准确地说，概率在 $k \approx 1.18\sqrt{n}$ 时达到 $\frac{1}{2}$，而在首次出现重复之前你需要遇到的人数，期望值约为 $\sqrt{\pi n/2}$（当 $n = 365$ 时，约为 24.6）。这条*平方根定律*（square-root law）正是巧合如此常见的原因：在任何一大批事件之中，总很可能有某一对在某方面吻合，尽管任何一个指定的吻合都不大可能发生。

生日攻击

电脑会计算文件的*杂凑值*（hash），以检查文件有否被窜改：那是一个指纹，可能的取值有（比方说）$n = 2^{128}$ 个。伪造者若想得到两份杂凑值相同的文件（一份无害的合约和一份欺诈的合约），并不需要试 2^{128} 份文件。根据生日问题的论证，为每份文件各自产生大约 $\sqrt{2^{128}} = 2^{64}$ 个变体，就很可能得到一对相同的杂凑值。这种*生日攻击*（birthday attack）正是现代杂凑函数采用 256 位元或以上的原因：有一半的位元因平方根定律而白白损失。类似的计算亦决定了随机识别码必须有多长，才能令两部电脑永远不会碰巧选中同一个。

谁最先提出?

生日问题通常归功于冯·米塞斯（Richard von Mises），他在 1939 年讨论过这个问题，不过它在此之前已经流传。据说 Harold Davenport 亦更早知道这个问题，只是觉得它太简单而没有发表。按 Warren Weaver 的说法，他在第二次世界大战期间喜欢在军官的晚宴上示范这个问题：他打赌在场二十二人之中有两人同一天生日，结果输了，直至一位女侍应透露她与其中一位军官同一天生日。

思维技巧

计算补事件；小心地近似

当“至少一个”很复杂时，就去求“一个也没有”的概率。当乘积难以处理时，把每个因子近似为指数函数，然后把指数相加。

这两个做法在大学里都是标准手法。第二个，即 $1 - x \approx e^{-x}$，是泊松近似理论的起点，也是渐近方法（asymptotic methods）的起点；后者用来估计那些复杂得

无法精确计算的量。

延伸练习

1 **首次重复。**试做 STEP II 2014，第 13 题（`step.maths.org/questions/14-s2-q13`）。这道题研究随机数产生器（random number generator）首次出现重复的值，其实是换了个面貌的生日问题，并把有关概率化成两个出人意表的恒等式（identities）。

2 **三人相同。**大约需要多少人，才会令某三人同一天生日的机会大于一半？（提示：同一天生日的三人组数目的期望值是 $\binom{k}{3}/365^2$。令它约等于 $\ln 2$。）真正的答案是 88。

3 **差之毫厘。**需要多少人，才会令两人的生日在同一天或相邻两天的机会达到一半？

4 **不均匀的生日。**现实中的生日并非均匀（uniform）分布。证明若各日的概率为 $p_1, \ldots, p_{365}$，则两个人同一天生日的机会是 $\sum p_i^2$，并且当所有 p_i 相等时这个值最小。（所以不均匀只会令相同的情况更容易出现。）

延伸阅读

- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1**（Wiley），第 II 章。生日问题及其各种近似。
- **Persi Diaconis & Frederick Mosteller, "Methods for studying coincidences"**，*Journal of the American Statistical Association* 84 (1989)。解释巧合为何常见；非常易读。
- **David Hand, *The Improbability Principle***（Bantam）。平方根定律，以及不大可能的事情为何时刻都在发生的其他原因。

9.10 总能找赎

题目

STEP I 2011，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/11-s1-q12`

引子

这个规律实在叫人难以抗拒：$\frac{m}{m+1}$、$\frac{m-1}{m+1}$、$\frac{m-2}{m+1}$。若有 n 个人拿着 £2 硬币，答案想必就是 $\frac{m-n+1}{m+1}$。事实正是如此，而最巧妙的证明（proof）用上了组合数学（combinatorics）中数一数二优雅的想法：*反射原理*（reflection principle）。它把图形的一部分沿一条直线反射，从而将一个困难的计数问题（从不缺零钱找赎的队伍有多少种）变成一个容易的计数问题。

同一个问题有许多不同的面貌：点票过程中胜出者从未落后的选举（1887 年的伯特兰投票问题（Bertrand's ballot problem））、永不破产的赌徒、正确配对的括号。当 $m = n$ 时，数出来的就是*卡塔兰数*（Catalan numbers），这大概是组合数学中除二项式系数（binomial coefficients）以外最常出现的数列（sequence）。

动笔之前

想一想 每位拿 £1 的顾客会给你一枚硬币；每位拿 £2 的顾客会取走一枚（作为找赎）。你甚么时候会无法继续？

留意你手上 £1 硬币的数目。每来一位 £1 顾客，数目加 1；每来一位 £2 顾客，数目减 1。你无法继续的情况，恰好就是一位 £2 顾客来到而你手上没有硬币：数目会变成 -1。所以你能成功，当且仅当（if and only if）在队伍中的每一刻，到目前为止的 £1 顾客人数都不少于到目前为止的 £2 顾客人数。

逐步解题

阶段 1　一枚 £2 硬币

当 $n = 1$ 时，只有 £2 顾客排第一才会失败。£2 顾客排在 $m+1$ 个位置中任何一个的可能性均等，所以

$$\mathrm{P}(\text{成功}) = 1 - \frac{1}{m+1} = \frac{m}{m+1}. \qquad \blacksquare$$

阶段 2　两枚 £2 硬币

> **想一想**　第一个人必须拿着 £1 硬币。第二个人可以出现甚么情况？在每种情况下，你余下面对的是甚么局面？

第一个人拿着 £1 硬币的概率（probability）是 $\frac{m}{m+2}$；否则你立即失败。这时你手上有一枚硬币，而余下的队伍有 $m-1$ 个 £1 的人和两个 £2 的人。

- 若第二个人拿着 £1 硬币（概率为 $\frac{m-1}{m+1}$），你手上便有两枚硬币，无论两枚 £2 硬币何时出现都足够找赎。成功是必然的。
- 若第二个人拿着 £2 硬币（概率为 $\frac{2}{m+1}$），你手上又变回没有硬币，余下 $m-1$ 个 £1 的人和一个 £2 的人。根据第 (i) 部分，以 $m-1$ 代替 m，你成功的概率是 $\frac{m-1}{m}$。

所以

$$\mathrm{P}(\text{成功}) = \frac{m}{m+2}\left(\frac{m-1}{m+1} + \frac{2}{m+1}\cdot\frac{m-1}{m}\right) = \frac{m}{m+2}\cdot\frac{m-1}{m+1}\cdot\frac{m+2}{m} = \frac{m-1}{m+1}. \qquad \blacksquare$$

> **停一停，试一试**　试用同样的方法处理 $n = 3$。这是行得通的，但情况会愈分愈多。有没有方法可以一次过数出所有成功的队伍？

阶段 3　把队伍看成路径

> **想一想**　以你手上的硬币数目对已服务的人数作图。一条队伍看起来是甚么样子？一条成功的队伍又是甚么样子？

一条队伍就是一条路径，由 m 步上升（£1）和 n 步下降（£2）组成，从高度 0 出发，终于高度 $m-n$。这样的路径共有 $\binom{m+n}{n}$ 条，每条出现的可能性均等。一条队伍成功，恰好就是它的路径从不低于 0。所以我们需要数出坏路径，即那些触及高度 -1 的路径。

阶段 4　反射

想一想 取一条坏路径。找出它第一次触及 -1 的时刻，把路径在这一刻之前的部分沿直线 $y=-1$ 反射。新的路径从哪里开始，又在哪里结束?

反射后的路径从 -2 出发（即 0 对直线 $y=-1$ 的镜像），终点与原来的路径相同，即 $m-n$（图 9.10.1）。反过来说，每条由 -2 到 $m-n$ 的路径都必定在某处穿过 $y=-1$（因为 $m \geq n$，它起点在这条线之下，终点在这条线之上）；把它第一次到达这条线之前的部分反射，便得回一条由 0 出发的坏路径。所以由 0 出发的坏路径，与由 -2 到 $m-n$ 的所有路径恰好一一对应。

一条以 $m+n$ 步由 -2 走到 $m-n$ 的路径共上升了 $m-n+2$，所以它有 $m+1$ 步上升和 $n-1$ 步下降：这样的路径有 $\binom{m+n}{n-1}$ 条。因此

$$\mathrm{P}(\text{失败}) = \frac{\binom{m+n}{n-1}}{\binom{m+n}{n}} = \frac{n!\,m!}{(n-1)!\,(m+1)!} = \frac{n}{m+1},$$

而

$$\mathrm{P}(\text{成功}) = 1 - \frac{n}{m+1} = \frac{m-n+1}{m+1}.$$

当 $n=3$ 时，这就是 $\frac{m-2}{m+1}$。 ■

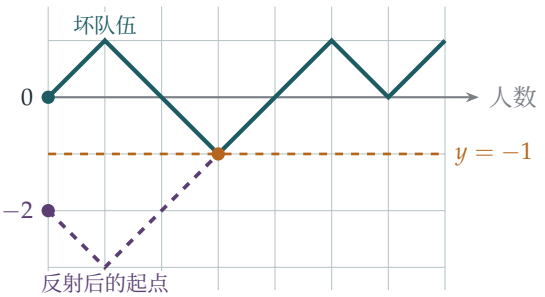


图 9.10.1 一条有 $m=4$ 枚 £1 硬币和 $n=3$ 枚 £2 硬币的队伍，画成一条路径（蓝绿色）。它在第三个人时首次到达 -1（琥珀色），这时你无法找赎。把路径在这一点之前的部分沿直线 $y=-1$ 反射，便得到一条由 -2 出发（紫色虚线）而终点相同的路径。坏队伍与由 -2 出发的路径之间的这种配对，就是反射原理。

回顾

- **找出合适的状态。**唯一要紧的是手上硬币的数目，这把队伍变成了一次游走。
- **以最初几步为条件。**第 (i) 和第 (ii) 部分的做法，是考虑最初几个人的情况，从而化约为较小的情形。
- **用双射（bijection）数出补集。**坏路径难以直接数出，但反射把它们与一个容易数算的集合配对起来。

背后的数学

卡塔兰数

当 $m = n$（£1 硬币刚好足够）时，成功队伍的数目是

$$\binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n},$$

即第 n 个卡塔兰数 C_n：$1, 1, 2, 5, 14, 42, 132, \ldots$。Richard Stanley 的著作 *Catalan Numbers* 列出了超过二百种由卡塔兰数计算数目的事物。其中包括：把 n 对括号正确排列的方法数目（“(())()”是允许的，“())(()”则不允许，理由正是上述的硬币论证，只要把“(”看成 £1、“)”看成 £2）、用互不相交的对角线把一个 $n+2$ 边形分割成三角形的方法数目，以及为 $n+1$ 个因子的乘积加上括号的方法数目。

随机游走

队伍就是随机游走（random walk），而反射原理是研究随机游走的基本工具之一。举例说，它可以证明由 0 出发的公平随机游走必定终会返回 0，也可以证明它在最初 $2n$ 步一直保持严格为正的概率是 $\frac{1}{2}\binom{2n}{n}2^{-2n} \approx \frac{1}{2\sqrt{\pi n}}$。在连续的形式下，把同样的反射论证应用于布朗运动（Brownian motion），便可得出其最大值的分布（distribution）。这是某些金融期权定价的基础，这类期权的回报取决于股价有否越过某个界限。

伯特兰投票问题

1887 年，约瑟夫·伯特兰（Joseph Bertrand）提出：在一场选举中，候选人 A 得 a 票，B 得 $b < a$ 票，那么 A 在整个点票过程中一直严格领先的概率是多少？答案是 $\frac{a-b}{a+b}$，同年由 Désiré André 证明，他的论证通常被视为反射原理（不过历史学家指出，André 实际的证明用的是一个稍有不同的双射）。我们的抽奖券问题是一个容许平手的变体：成功的意思是 £1 的数目从不低于 £2 的数目。其中一道练习会由我们的答案推导出伯特兰的答案。

思维技巧

双射计数

要数算一个集合，就找出它与一个你懂得数算的集合之间的一一对应（*one-to-one correspondence*）。

反射把坏队伍与由另一个起点出发、不受限制的路径配对起来。双射是组合数学中最令人满足的证明，因为它们解释了两个数为何相等，而不只是证明它们相等。在大学里，你会在组合数学中处处遇到双射证明，也会在研究随机游走和布朗运动时再次遇上反射原理。

延伸练习

1 **分情况处理三枚。**仿照第 (ii) 部分，以队伍中最初几个人为条件来证明第 (iii) 部分，看看反射原理省却了多少工夫。

2 **伯特兰投票问题。**在一场选举中，A 得 a 票，B 得 $b < a$ 票。证明 A 在整个点票过程中一直严格领先的概率是 $\frac{a-b}{a+b}$。（提示：第一票必须投给 A；此后，A 落后于 B 的票数绝不可多于…。）

3 **括号。**列出三对括号的 5 种正确排列，以及把五边形分割成三角形的 5 种方法。你能找出两者之间的一个双射吗？

4 **一张 £5 纸币。**一条队伍中有 m 个人拿着 £1 硬币，n 个人拿着 £5 纸币，而一张票的价钱是 £1。解释为何反射原理不再给出简单的答案。（这个推广困难得多。）

延伸阅读

- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1**（Wiley），第 III 章。有关随机游走和反射原理的经典论述。
- **Richard Stanley, *Catalan Numbers***（Cambridge University Press）。超过二百种诠释，并附练习。
- **Marc Renault, "Lost (and found) in translation: André's actual method and its application to the generalized ballot problem"**，*American Mathematical Monthly* 115 (2008)。反射原理的历史。

9.11 集齐一套

题目

STEP II 1994，第 13 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/94-s2-q13`

引子

一本足球贴纸簿要集齐 600 张贴纸，贴纸每次随机发售一张。要贴满整本贴纸簿，你预计要买多少张贴纸？最初几百张很快到手，因为几乎每一张都是新的。但贴纸簿愈贴愈满，新贴纸就愈来愈罕有，单是最后一张，平均就要买 600 次。总数大约是 4200 张，是贴纸簿大小的七倍。

这就是*集券问题*（coupon collector's problem）。它的答案结合了两个想法：等候一次成功所需的时间，以及调和级数（harmonic series）$1+\frac{1}{2}+\frac{1}{3}+\cdots$，后者的增长与 $\ln n$ 相若。你在 Section 2.3 见过调和级数发散（diverge）；在这里，它那缓慢的增长正正就是集齐一套收藏品所要付出的代价。

动笔之前

想一想 如果每包各自独立地以概率 p 开出一张新卡，那么直至开出第一张新卡为止（连这一包计算在内），你开包数目的期望值（expectation）是多少？

开包数目 X 服从*几何分布*（geometric distribution）：$\mathrm{P}(X=k)=(1-p)^{k-1}p$。它的期望值是 $\frac{1}{p}$。（有一个快捷的看法：$\mathrm{E}(X)=1+(1-p)\mathrm{E}(X)$，因为你先开一包，然后有 $1-p$ 的概率回到起点。）如果有三分之一的机会开出新卡，你预计要等三包。

逐步解题

阶段 1　等候一张新卡

手上已有 r 张不同的卡时，一包开出新卡的概率是 $\frac{n-r}{n}$，与之前各包独立。因此，直至开出下一张新卡为止，所需包数的期望值是

$$\frac{1}{(n-r)/n} = \frac{n}{n-r}. \qquad \blacksquare$$

阶段 2　比较和与积分

> **想一想**　在区间 $r \le x \le r+1$ 上，$\frac{1}{x}$ 介乎哪两个值之间？这对曲线下的面积有甚么启示？

由于 $\frac{1}{x}$ 是递减的，在 $[r, r+1]$ 上有 $\frac{1}{r+1} \le \frac{1}{x} \le \frac{1}{r}$。在长度为 1 的区间上积分，得

$$\frac{1}{r+1} \le \int_r^{r+1} \frac{\mathrm{d}x}{x} \le \frac{1}{r}. \qquad \blacksquare$$

用图来看（图 9.11.1）：曲线下的面积夹在高度为 $\frac{1}{r+1}$ 的长方形与高度为 $\frac{1}{r}$ 的长方形之间。把这些不等式由 $r = 1$ 加到 $r = n-1$，各个积分便连接成 $\int_1^n \frac{\mathrm{d}x}{x} = \ln n$，于是

$$\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \le \ln n \le 1 + \frac{1}{2} + \cdots + \frac{1}{n-1}. \qquad \blacksquare$$

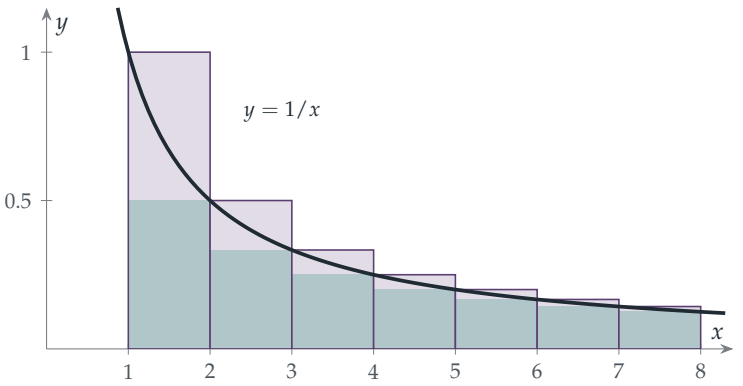


图 9.11.1　在每个区间 $[r, r+1]$ 上，$y = 1/x$ 之下的面积介乎蓝绿色长方形（高度 $\frac{1}{r+1}$）与较大的紫红色长方形（高度 $\frac{1}{r}$）之间。全部加起来，$\ln n$ 便介乎 $\frac{1}{2} + \cdots + \frac{1}{n}$ 与 $1 + \frac{1}{2} + \cdots + \frac{1}{n-1}$ 之间。曲线与紫红色长方形之间的空隙，加起来是一个有限的总和：欧拉常数（Euler's constant）$\gamma \approx 0.577$。

阶段 3　集齐全套

想一想 总包数是等候第一张卡、第二张新卡……各段等候时间之和。它的期望值是多少?

设 X_r 为由 r 张不同的卡增至 $r+1$ 张所需的包数。总数是 $X_0 + X_1 + \cdots + X_{n-1}$，而期望值可以相加，所以

$$\mathrm{E}(\text{总数}) = \sum_{r=0}^{n-1} \frac{n}{n-r} = n\Big(\frac{1}{n} + \frac{1}{n-1} + \cdots + 1\Big) = nH_n,$$

其中 $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ 是第 n 个调和数（harmonic number）。由第 2 步，$\ln n + \frac{1}{n} \le H_n \le \ln n + 1$，所以当 n 很大时 H_n 约为 $\ln n$，而我预计大约要吃掉

$$n \ln n \text{ 包.} \qquad \blacksquare$$

回顾

- **把漫长的等候分成多个阶段。**总时间拆成 n 段等候时间，每一段都属于简单的几何分布。
- **期望值可以相加。**我们从头到尾都不需要总数的分布，只需要各部分的期望值。
- **比较和与积分。**那幅图把一个我们无法求值的和化成一个对数，误差至多为 1。

背后的数学

这个估计有多准确?

这个比较还可以再精确一些。图 9.11.1 中曲线与上方长方形之间的阴影空隙，总面积是有限的，所以 $H_n - \ln n$ 趋向一个极限，

$$\gamma = \lim_{n\to\infty} (H_n - \ln n) = 0.5772156649\ldots,$$

即欧拉常数。（没有人知道 γ 是否有理数。）因此包数的期望值是 $n\ln n + \gamma n + \frac{1}{2} + \cdots$。以 600 张贴纸的贴纸簿为例，$600 \times (\ln 600 + 0.577) \approx 4185$，这就是引子中所引述的数字。

时间都花到哪里去了? 各个阶段极不平均。集齐全套的前一半只需大约 $n\ln 2 \approx 0.69n$ 包，而单是最后一张卡，平均就要 n 包。大部分工夫都花在最后几张卡上，这正是收藏者要互相交换重复卡的原因。

集中现象

期望值只是个开始：总数的变动有多大？总数的方差（variance）是 $\sum nr/(n-r)^2$，约为 $\frac{\pi^2}{6}n^2$（Section 2.4 中的和 $\sum 1/k^2 = \pi^2/6$ 在这里出现了）。所以标准差（standard deviation）约为 $1.28n$，当 n 很大时远小于平均值 $n\ln n$：总数集中在其平均值附近。艾狄胥（Paul Erdős）和雷尼（Alfréd Rényi）在 1961 年证明了一个更精确的结果：总数减去 $n\ln n$ 再除以 n，会有一个极限分布（limiting distribution），即耿贝尔分布（Gumbel distribution），这个分布也描述了许多随机量之中最大的一个（Section 12.25）。

无处不在的集券问题

这个问题可追溯至棣莫弗（Abraham de Moivre，1712 年）和拉普拉斯（Laplace）。如今它以许多不同的面貌出现。要检查一个系统的每个组件，需要做多少次随机测试？在一个网络上进行随机漫步（random walk），要多久才会走遍每个节点？电脑的杂凑表（hash table）要探查多少次？每一种情况的答案都涉及 $n\ln n$。在生物学中，同样的计算可用来估计要有多少个定序“读段”（reads）才能覆盖整个基因组。

思维技巧

期望值的线性性质（linearity of expectation）

和的期望值等于期望值之和，不论各部分是否独立。把一个复杂的随机量拆成简单的部分。

这是概率论中最有用的事实之一。在这里，它把集券问题简化为 n 段几何等候；在 Section 9.8 中，它不用任何计算就给出了配对数目的期望值。到了大学，同一个想法推动着算法分析，以及组合数学中的概率方法（probabilistic method）。

延伸练习

1 **半套。**证明集齐 $\frac{n}{2}$ 张不同的卡（n 为偶数）所需包数的期望值是 $n(H_n - H_{n/2})$，并且这个值约为 $n\ln 2$。

2 **不同的卡。**证明开了 k 包之后，你手上不同的卡的数目的期望值是 $n\left(1-(1-\frac{1}{n})^k\right)$。（提示：对每一张卡，求出你拥有它的概率，然后相加。）大约要开多少包，才能得到全套的 95%？

3 **两位收藏者。**两个小孩收集同一套卡，并共用所开的各包，因此对其中一人来说是重复的卡，对另一人来说可能是新的。证明集齐两套所需包数的期望值约为 $n(\ln n + \ln\ln n)$。（这题较难；先说服自己这个数目远小于 $2n\ln n$。）

4 **更精确的界。**改用梯形法则（trapezium rule）的图像而不用长方形，证明 $H_n - \ln n$ 是递减的，而 $H_n - \ln n - \frac{1}{2n}$ 是递增的。由较小的 n 值推导出 $0.5 < \gamma < 0.7$。

延伸阅读

- **Geoffrey Grimmett & David Stirzaker, *One Thousand Exercises in Probability***（Oxford University Press）。集券问题的许多变化。
- **Julian Havil, *Gamma: Exploring Euler's Constant***（Princeton University Press）。关于 γ 和调和级数的一切，写得深入浅出。
- **Michael Mitzenmacher & Eli Upfal, *Probability and Computing***（Cambridge University Press）。集券问题及其在电脑科学中的应用。

9.12 随机子集

题目

STEP III 2019，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 OCR 提供下载的 2019 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

`www.ocr.org.uk/Images/696590-2019-past-papers.zip`

引子

完全随机地选出 $\{1,2,\ldots,100\}$ 的两个子集（subset）。它们没有任何共同元素的可能性有多大？直觉也许会告诉你：“相当大：大多数成对的集合都不相同”。事实上，这个概率是 $\left(\frac{3}{4}\right)^{100}$，约为 3×10^{-13}。随机集合很大（大约占全部元素的一半），所以它们几乎总会重叠。

这道题很短，当中唯一的想法却很有威力：不要把集合当作整体来想，而要每次只想一个元素。一个随机子集不过是 n 次独立（independent）的掷硬币，每个元素掷一次；于是，凡是关于随机子集的问题，都变成关于这些掷硬币结果的问题。

动笔之前

? **想一想** 怎样用 n 条是非题的一串答案，来描述 $\{1,\ldots,n\}$ 的一个子集？

对每个元素 i，回答“i 在集合里吗？”这条问题。子集决定了答案，答案也决定了子集。每个元素有 2 个选择，所以一共有 2^n 个子集。均匀随机（uniformly at random）地选一个子集，就等于独立地掷一枚公平硬币来回答每条问题。

逐步解题

阶段 1 一个元素

按照掷硬币的描述，$\mathrm{P}(1 \in A_1) = \frac{1}{2}$。事实上，$2^n$ 个子集中恰好有一半包含 1：把每个不包含 1 的子集，与加入 1 之后所得的子集配成一对即可。

阶段 2　不相交的集合

> **想一想**　要使 $A_1 \cap A_2 = \varnothing$，每个元素 i 必须满足甚么条件？就单一个元素而言，这件事的概率是多少？为甚么可以把各个元素的概率相乘？

A_1 与 A_2 不相交（disjoint），当且仅当没有元素同时属于两者。对单一个元素 i 来说，有四种等可能的情况：两个集合都不在、只在 A_1、只在 A_2、两个都在（图 9.12.1 左）。只有最后一种是禁止的，所以 i 不同时属于两者的概率是 $\frac{3}{4}$。不同的元素由不同而且独立的掷硬币结果决定，所以

$$\mathrm{P}(A_1 \cap A_2 = \varnothing) = \left(\frac{3}{4}\right)^n. \qquad \blacksquare$$

对三个集合而言，元素 i 同时属于三者的概率是 $\frac{1}{8}$，所以 $\mathrm{P}(A_1 \cap A_2 \cap A_3 = \varnothing) = \left(\frac{7}{8}\right)^n$。一般而言，

$$\mathrm{P}(A_1 \cap \cdots \cap A_m = \varnothing) = \left(1 - \frac{1}{2^m}\right)^n. \qquad \blacksquare$$

阶段 3　子集

> **想一想**　要使 $A_1 \subseteq A_2$，一个元素的四种情况之中，哪一种是禁止的？

$A_1 \subseteq A_2$ 不成立，当且仅当有某个元素在 A_1 里而不在 A_2 里。同样是四种模式中有一种被禁止，所以

$$\mathrm{P}(A_1 \subseteq A_2) = \left(\frac{3}{4}\right)^n,$$

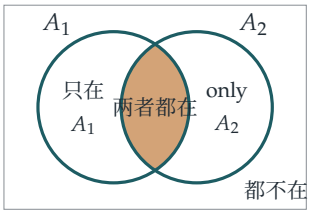
与不相交的情况相同。这并非巧合：$A_1 \subseteq A_2$ 当且仅当 A_1 与 A_2 的补集（complement）不相交，而随机集合的补集也是一个随机集合。

> **想一想**　要使 $A_1 \subseteq A_2 \subseteq A_3$，一个元素的 8 种归属模式（在或不在每个集合里）之中，哪些是容许的？

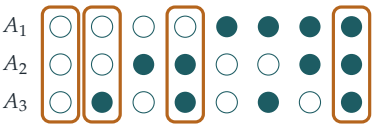
沿着 A_1, A_2, A_3 这条链（chain）走，一个元素一旦进入，就必须一直留在里面。所以容许的模式是“不在、不在、不在”、“不在、不在、在”、“不在、在、在”和“在、在、在”：8 种之中的 4 种（图 9.12.1 右）。因此 $\mathrm{P}(A_1 \subseteq A_2 \subseteq A_3) = \left(\frac{4}{8}\right)^n = \left(\frac{1}{2}\right)^n$。

对于由 m 个集合组成的链，一个容许的模式取决于该元素最先进入哪一个集合（$A_1, \ldots, A_m$ 其中之一，或者从不进入），所以 2^m 种模式之中有 $m+1$ 种是容许的：

$$\mathrm{P}(A_1 \subseteq A_2 \subseteq \cdots \subseteq A_m) = \left(\frac{m+1}{2^m}\right)^n. \qquad \blacksquare$$



每个元素：4 个等可能的位置



符合 $A_1 \subseteq A_2 \subseteq A_3$ 的模式：4 种（共 8 种）

图 9.12.1 左：S 的一个元素相对于 A_1 和 A_2 可以处于四个位置之一，每个位置的概率都是 $\frac{1}{4}$。不相交这个条件禁止“两者都在”（琥珀色）；$A_1 \subseteq A_2$ 则禁止“只在 A_1”。右：元素对 A_1、A_2、A_3 的八种归属模式（实心圆表示“在”）。链 $A_1 \subseteq A_2 \subseteq A_3$ 只容许加了方框的四种模式，即元素一旦进入便一直留在里面。

回顾

- **按元素分解。**每个集合层面的条件（不相交、子集、链），都是分别加在每个元素上的条件。
- **元素之间互相独立。**由于各个元素由独立的掷硬币结果决定，概率可以相乘。
- **数算模式。**每个答案都是（容许的模式数 ÷ 全部模式数）^{n}。

背后的数学

不用概率的计数

把每个概率乘以选择的总数，就得到一个计数。满足 $A_1 \subseteq A_2$ 的有序对 (A_1, A_2) 的数目是 $4^n \times \left(\frac{3}{4}\right)^n = 3^n$。你也可以直接数算这些有序对：先选定 A_2 的大小 k，然后以 $\binom{n}{k}$ 种方法选出 A_2 本身，再从它的 2^k 个子集中任选一个作为 A_1。所以

$$\sum_{k=0}^{n} \binom{n}{k} 2^k = 3^n,$$

这正是 $(1+2)^n$ 的二项式定理（binomial theorem）。用两种方法数算同一样东西，一次按元素、一次按大小，便证明了一条恒等式：这就是算两次（double counting）的方法，你在 Section 7.9 讨论柏拉图立体时已经见过。同样地，链 $A_1 \subseteq \cdots \subseteq A_m$ 的数目是 $(m+1)^n$。

子集的格

$\{1, \ldots, n\}$ 的所有子集按包含关系排序，形成一个称为布尔格（Boolean lattice）的结构，可以把它想像成一个 n 维立方体：每个子集是一个顶点，以它那一串是非题的答案作标

记，而沿着一条棱移动，就是加入或移走一个元素。链就是向上走的路径。有些深刻的结果描述了在给定条件下，子集族可以有多大。施佩纳定理（Sperner's theorem，1928 年）指出，在没有一个集合包含另一个集合的子集族之中，最大的一族有 $\binom{n}{\lfloor n/2 \rfloor}$ 个集合；艾狄胥—柯—拉多定理（Erdős–Ko–Rado theorem，1961 年）则说明，最多可以选出多少个 k 元集合，使得任意两个都有共同元素。两者都是组合数学（combinatorics）入门课程的标准内容，而把子集看成掷硬币的观点，给出了其中一些最漂亮的证明。

布尔的代数

布尔（George Boole）的 *The Laws of Thought*（《思维规律》，1854 年）把逻辑命题当作集合，把“与”、“或”、“非”当作交集、并集和补集。把一个集合编码成一串是非题的答案，正是电脑表示集合的方法：$\{1, \ldots, 64\}$ 的一个子集就是单独一个 64 位元的数，而交集就是“逐位元 AND”（bitwise AND）指令。本题的方法，即每次只看一个元素（一个位元），正是分析这类运算的方法。

思维技巧

每次只想一个元素

当对象是由每个元素各自的独立选择构成时，一个分别加在每个元素上的条件，其概率等于一个乘积。

这个方法把一个关于 2^n 个子集的问题，变成一个关于 4 种或 8 种模式的问题。到了大学，它会在组合数学（集合与多重集的生成函数（generating functions））、概率论（独立的坐标）和计算机科学（逐位元的推理）中再次出现。

延伸练习

1 **覆盖。**求 $\mathrm{P}(A_1 \cup A_2 = S)$ 和 $\mathrm{P}(A_1 \cup A_2 \cup \cdots \cup A_m = S)$。

2 **大小相同。**证明 $\mathrm{P}(|A_1| = |A_2|) = \binom{2n}{n}/4^n$。（提示：$|A_1| = |A_2|$ 当且仅当 $|A_1| + |S \setminus A_2| = n$。）证明当 n 很大时，这个概率约为 $1/\sqrt{\pi n}$。

3 **分成三份。**证明 S 的不相交子集所组成的有序对 (A, B) 的数目是 3^n，并求由不相交子集所组成的无序对 $\{A, B\}$ 的数目。

4 **小 n 的施佩纳定理。**对 $n = 3$ 和 $n = 4$，求 $\{1, \ldots, n\}$ 的子集族中，没有一个集合包含另一个集合的最大一族。

延伸阅读

- **Miklós Bóna, *A Walk Through Combinatorics***（World Scientific），其中关于子集和偏序集（partially ordered sets）的章节。
- **Béla Bollobás, *Combinatorics: Set Systems, Hypergraphs, Families of Vectors and Combinatorial Probability***（Cambridge University Press）。一本简短而优雅的书，讲述子集族，由施佩纳定理说起。
- **Peter Cameron, *Combinatorics: Topics, Techniques, Algorithms***（Cambridge University Press）。一门入门课程，特别着重用两种方法计数。

9.13 做了手脚的骰子

题目

STEP I 1999，第 12 题

继续阅读之前，请先看题目（原题为英文）。题目载于 STEP Support Programme 的试题资料库：

`step.maths.org/questions/99-s1-q12`

引子

掷两颗公平骰子时，点数总和为 7 的可能性是总和为 2 的六倍。桌上游戏的设计者有时会希望情况并非如此：能否在骰子上做手脚（加上配重，使某些面较常出现），令 2 至 12 的每个总和都有相同的机会，即 $\frac{1}{11}$？那十二个概率（probability）可以任你选择，看来回旋的余地相当充裕。

答案是不能，而证明堪称一绝：它只用到十一个总和中的三个，以及不等式（inequality）$x+\frac{1}{x}\geq 2$。这是证明某件事不可能的典范；与证明某件事办得到相比，这往往更难，也更有趣。

动笔之前

想一想 对于 $x>0$，你会怎样证明 $x+\frac{1}{x}\geq 2$？等号何时成立？

两边同乘 $x>0$：不等式变成 $x^2+1\geq 2x$，即 $(x-1)^2\geq 0$，这显然成立。等号只在 $x=1$ 时成立。（另一做法是 $x+\frac{1}{x}-2=(\sqrt{x}-\frac{1}{\sqrt{x}})^2$。）这是算术平均—几何平均不等式（AM–GM inequality）最简单的形式：x 与 $\frac{1}{x}$ 的平均值不小于它们的几何平均（geometric mean）1。

逐步解题

阶段 1 两端的总和

想一想　掷出总和 2 有多少种方法？总和 12 呢？试以 q_i 和 r_j 表示它们的概率。

总和为 2 需要两颗骰子都掷出 1，总和为 12 则需要两颗都掷出 6，所以 $\mathrm{P}(2) = q_1 r_1$ 及 $\mathrm{P}(12) = q_6 r_6$。

为了引出矛盾，假设全部十一个总和的可能性相等，每个的概率都是 $\frac{1}{11}$。那么

$$q_1 r_1 = \frac{1}{11}, \qquad q_6 r_6 = \frac{1}{11}.$$

特别地，q_1、r_1、q_6、r_6 四者全为正数。

阶段 2　中间的总和

想一想　总和 7 可以由六种方法得出。其中哪两种只涉及 1 和 6 这两个面？试利用上面的等式写出它们的概率。

总和 7 包括结果 $(1,6)$ 和 $(6,1)$，所以

$$\mathrm{P}(7) \geq q_1 r_6 + q_6 r_1.$$

现在利用第 1 步的等式引入一个比。由于 $q_1 = \frac{1}{11 r_1}$ 及 $q_6 = \frac{1}{11 r_6}$，

$$q_1 r_6 + q_6 r_1 = \frac{1}{11}\left(\frac{r_6}{r_1} + \frac{r_1}{r_6}\right) \geq \frac{1}{11} \times 2 = \frac{2}{11},$$

这里用了 $x + \frac{1}{x} \geq 2$，其中 $x = r_6 / r_1 > 0$。因此 $\mathrm{P}(7) \geq \frac{2}{11}$，与 $\mathrm{P}(7) = \frac{1}{11}$ 矛盾。各个总和不可能全部等可能。 ■

停一停，试一试　看看这个证明为何行得通。要令 2 和 12 与 7 同样可能出现，1 和 6 就必须有相当高的概率，于是单凭 $(1,6)$ 和 $(6,1)$ 这两个“混合”结果，已足以令 7 的概率过高。图 9.13.1 显示了所涉及的方格。



图 9.13.1 两颗骰子的 36 个结果，各以其总和标示。证明只用到四个角落的方格。两个琥珀色方格的概率之积，等于两个深蓝绿色方格的概率之积，即 $q_1r_1q_6r_6 = q_1r_6 \cdot q_6r_1$；而根据算术平均—几何平均不等式，蓝绿色那一对之和至少是琥珀色那一对的几何平均的两倍。

阶段 3 同一颗骰子掷两次

想一想 如果把一颗骰子掷两次，两次点数相同的概率是多少？试以 q_j 表示。对公平骰子而言，这个概率是多少？

两次掷出相同点数的概率是 $q_1^2 + q_2^2 + \cdots + q_6^2$。对公平骰子而言，这是 $6 \times \frac{1}{36} = \frac{1}{6}$，已经远大于 $\frac{1}{36}$。

想一想 在骰子上做手脚，能否令 $\sum q_j^2$ 小于 $\frac{1}{6}$？你知道 $\sum q_j = 1$。

写成 $q_j = \frac{1}{6} + e_j$，其中各偏差 e_j 之和为零。那么

$$\sum q_j^2 = \sum\left(\frac{1}{36} + \frac{e_j}{3} + e_j^2\right) = \frac{1}{6} + \frac{1}{3}\sum e_j + \sum e_j^2 = \frac{1}{6} + \sum e_j^2 \geq \frac{1}{6}.$$

所以掷出重复点数的概率必定不小于 $\frac{1}{6}$，而且只有公平骰子才令等号成立。它当然不可能小于 $\frac{1}{36}$。 ■

回顾

- **反证法（proof by contradiction）。**我们假设所有总和等可能，然后推出不可能成立的结论。
- **只用所需的条件。**十一条等式之中，三条已经足够。选对等式正是关键的一步。
- **以不等式证明不可能。**第 (i) 部分的算术平均—几何平均不等式，以及第 (ii) 部分的“平方和非负”，都一举排除了一整类的可能性。

背后的数学

用多项式来证明

以下是一个完全不同的证明，用的是 Section 9.6 的生成函数（generating function）。设 $Q(x) = q_1x + \cdots + q_6x^6$ 及 $R(x) = r_1x + \cdots + r_6x^6$。假如所有总和的概率都是 $\frac{1}{11}$，则

$$Q(x)R(x) = \frac{x^2 + x^3 + \cdots + x^{12}}{11} = \frac{x^2}{11}\big(1 + x + \cdots + x^{10}\big).$$

现在 $Q(x)/x$ 是一个 5 次实系数多项式（polynomial），而每个奇数次的实多项式都有一个实根（它的图像由 $-\infty$ 走到 $+\infty$，或者反过来）。所以左边有一个非零的实根。但 $1 + x + \cdots + x^{10} = \frac{x^{11}-1}{x-1}$，它的根是 1 以外的 11 次单位根（roots of unity），当中没有一个是实数。矛盾。同样的论证表明，只要 $n - 1$ 是奇数，即 n 是偶数，无论怎样在两颗 n 面骰子上做手脚，都不能令各个总和等可能。（当 n 是奇数时，多项式的论证需要多花些工夫，但上面的算术平均—几何平均证明对每个 $n \geq 2$ 都适用。）

碰撞概率

第 (ii) 部分中的量 $\sum q_j^2$，即两次独立抽取结果相同的机会，称为一个分布的*碰撞概率*（collision probability）。第 (ii) 部分表明，均匀分布（uniform distribution）的碰撞概率最小。它量度一个分布有多集中：在生态学中，$\sum q_j^2$ 是辛普森多样性指数（Simpson's index of diversity），其中 q_j 是物种 j 在某个生境中所占的比例；在经济学中，它是量度市场集中度的赫芬达尔指数（Herfindahl index）；在密码学中，它量度一个密码分布有多容易被猜中。它也解释了 Section 9.9 结尾的那句话：各个生日的概率不相等，只会令生日相同的情况更有可能出现。

不可能性的证明

数学中一些最著名的结果，说的正是某件事办不到：没有分数的平方等于 2；没有尺规作图法能三等分任意角；没有根式公式能解所有五次方程；没有算法能判定每一个数学命题。每一个这样的结果都改变了所属的学科，因为要证明*没有*任何方法行得通，就必须了解所有方法构成的整个空间。灌铅骰子问题是一个缩影：要排除每一种做手脚的方式，我们需要一个每种方式都必须遵守的不变性质（算术平均—几何平均不等式）。

思维技巧

以反证法和不等式证明不可能

要证明没有任何参数的选取行得通，先假设有一组行得通，找出互相抵触的那几个条件，再用不等式证明它们不可能同时成立。

在大学里，反证法配合不等式是证明某些对象不存在的标准方法，见于分析、组合数学和最优化；在这些领域中，算术平均—几何平均不等式和柯西—施瓦兹不等式（Cauchy–Schwarz inequality）这类不等式为可达到的结果设下了界限。

延伸练习

1 **柯西—施瓦兹不等式。**利用柯西—施瓦兹不等式 $\left(\sum a_j b_j\right)^2 \le \sum a_j^2 \sum b_j^2$，取 $a_j = q_j$ 及 $b_j = 1$，为第 (ii) 部分给出一行的证明。

2 **两颗不同的骰子。**如果把两颗不同的灌铅骰子各掷一次，它们点数相同的概率能否小于 $\frac{1}{6}$？能否等于 0？

3 **面上数字不同时的均匀总和。**证明如果两颗骰子各面的数字另作安排，它们可以给出等可能的总和：试取一颗骰子的数字为 $0, 1, 2, 3, 4, 5$，另一颗为 $0, 6, 12, 18, 24, 30$。会出现哪些总和？为甚么这与题目并不矛盾？

4 **三颗骰子。**三颗各面为 1 至 6 的六面灌铅骰子，能否令 3 至 18 的所有总和等可能？（提示：利用总和 3、8、13 和 18，以及六个数的算术平均—几何平均不等式。为甚么多项式证明这次行不通？）

延伸阅读

- **J. Michael Steele, *The Cauchy–Schwarz Master Class***（Cambridge University Press）。一本出色的不等式入门书，由算术平均—几何平均不等式讲起。
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press）。讲述生成函数，以及它们在上述这类证明中的应用。
- **Ian Stewart, *Galois Theory***（CRC Press），开首数章。最著名的不可能性证明：五次方程。

我们学到了甚么

本章开首说过，计数是最古老的数学，而要数得巧妙，至今仍是最难的事情之一。我们提出了三个问题：怎样不逐一列举而数出答案？甚么时候某件事必然发生？计数与概率（probability）怎样互相帮助？以下是本章十三道题目所揭示的答案。

1. 怎样不逐一列举而数出答案?

本章逐步建立了一套工具。

- **递归关系（recurrences）。**在平面上多加一条直线（Section 9.1），或在围栏上多加一根柱子（Section 9.3），数目都只会以简单的方式改变，由此得出可以求解的递归关系。只要多记录一点资料，即最后一根柱子的颜色，这些递归关系便能自成一体。
- **生成函数（generating functions）。**把数列（sequence）包装成幂级数（power series），递归关系便化为代数运算（Section 9.4），既解释了为何 $1/89 = 0.01123\ldots$，也把数算组合变成多项式相乘：Section 9.5 的硬币、Section 9.6 的骰子，以及 Section 9.7 的等幂和，其中 $x = 1$ 处的重根（repeated root）包办了一切。
- **双射（bijections）。**反射原理（reflection principle）（Section 9.10）把不合格的队列与不受限制的路径一一配对，从而数出前者的数目；而取随机集合的补集（complement），则把子集变成互不相交的集合对（Section 9.12）。
- **容斥原理（inclusion–exclusion）。**先多算、后修正，便求得错排（derangements）的数目（Section 9.8）。
- **重复计数（double counting）。**以欧拉公式（Euler's formula）数算区域（Section 9.1），以及按大小数算子集对（Section 9.12），都是循不同途径得出相同答案，并顺道证明了一些恒等式。

2. 甚么时候某件事必然发生?

有好几节显示，单凭规模已足以迫出结构。根据鸽巢原理（pigeonhole principle），六个人当中必有三人互相认识，或有三人互不相识（Section 9.2），但五个人则不足够。二十三人当中有人同一天生日的机会已过半（Section 9.9），因为配对的数目随人数的平方增长。一个大集合的随机子集几乎必定互相重叠（Section 9.12）。还有一些事情，无论怎样尝试都不可能发生：不论怎样为两颗骰子灌铅，都无法令每个点数总和出现的机会均等

(Section 9.13)；而除了 Sicherman 的编号方式外，再没有其他骰子编号能重现普通骰子的点数总和分布（Section 9.6）。

3. 计数与概率怎样互相帮助?

在本章的后半部分, 概率是靠计数求得的: 错排用容斥原理, 生日问题用连乘积, 选票问题 (ballot problem) 用反射。但反过来, 概率也帮了计数的忙。期望值的线性性质 (linearity of expectation) 给出了集换券收集问题 (coupon collector) 的 nH_n (Section 9.11), 也给出了洗牌后配对数目的期望值, 完全毋须计数。把随机子集看成掷 n 次硬币（Section 9.12），3^n 和 $(m+1)^n$ 这类数目便成了一行就能完成的计算。而艾狄胥（Erdős）的概率方法（probabilistic method）（Section 9.2）则通过证明随机着色有正概率符合要求，证明了理想的着色确实存在。有两个常数凭空出现：e 出现在错排概率的极限中，而欧拉常数 γ 则出现在集齐一整套所需的代价中。

4. 计数的工具箱

方法	用于何处
多加一个时有何改变	直线（9.1）；围栏柱子（9.3）
鸽巢原理	拉姆齐（Ramsey）（9.2）
耦合递归关系	围栏柱子（9.3）
生成函数	9.4、9.5、9.6、9.7、9.13
从极端情况入手分类讨论	奇异骰子（9.6）；灌铅骰子（9.13）
容斥原理	错排（9.8）
补集与指数近似	生日（9.9）
反射原理	选票问题（9.10）
期望值的线性性质	集换券（9.11）；配对（9.8）
每次处理一个元素	随机子集（9.12）
以算术—几何平均不等式（AM–GM）导出矛盾	灌铅骰子（9.13）

下一步

- 错排的泊松极限（Poisson limit）会在 Section 12.11 再次出现，而随机变量（random variables）的生成函数则推动 Section 12.15 对家族树的研究。
- 耦合递归关系在 Section 12.16 会变成马可夫链（Markov chains）。
- 集换券收集问题中的冈贝尔分布（Gumbel distribution），在 Section 12.25 用来描述众多随机量中的最大者。

如果你想在入读大学前再进一步

- **Miklós Bóna, *A Walk Through Combinatorics***：一本平易近人的入门教材，几乎涵盖本章的每一个概念。
- **Herbert Wilf, *generatingfunctionology***：生成函数的经典著作，可于网上免费阅读。
- **Ronald Graham, Donald Knuth & Oren Patashnik, *Concrete Mathematics***：求和、递归关系与生成函数，笔调风趣，习题丰富。
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***：大学一年级概率论中与计数有关的部分。

怀特黑德（Whitehead）曾称组合数学（combinatorics）为“拓扑学的贫民窟”。本章尝试说明，为何今天再没有人会这样说。每一个巧妙的计数背后，都有一个影响远超计数本身的想法：一个生成函数、一个双射、一个递归关系、一个不等式。这些想法把计数与代数、分析和概率连系起来；学会寻找这些想法，而不是再多找一个窍门，正是令计数成为数学的关键。

延伸练习答案

以下只是简短的答案和提示，并非完整解答：这里只给出结果和关键步骤，方便你核对自己的做法。请先认真做过每道练习，然后才翻看。

9.1 直线分割平面

1. 一个圆把平面分成 2 个区域。第 n 个圆与其余各圆相交于 $2(n-1)$ 点，这些点把它分成 $2(n-1)$ 段弧，每段弧把一个区域一分为二。所以 $R_n = R_{n-1} + 2(n-1)$，由此得 $n^2 - n + 2$。

2. 每条直线所增加的区域数目，是 1 加上该直线上的交点数目，所以总数是 $1 + n +$ (交点数目)。那 k 条平行线少了 $\binom{k}{2}$ 个交点：$1 + n + \binom{n}{2} - \binom{k}{2}$。

3. 第 n 个平面被其余平面截出 $n-1$ 条处于一般位置（general position）的直线，这些直线把它分成 $\frac{1}{2}(n^2 - n + 2)$ 块，每一块把空间中的一个区域一分为二。把这个数加到 $\frac{1}{6}\big((n-1)^3 + 5(n-1) + 6\big)$ 上，便得 $\frac{1}{6}(n^3 + 5n + 6)$。

4. 当 $n = 6$ 时有 31 个区域，而不是 32 个。在圆内，这个图有 $V = n + \binom{n}{4}$ 个顶点（vertex）（每个交点由四个点决定）和 $E = n + \binom{n}{2} + 2\binom{n}{4}$ 条边（edge）（圆上的各段弧，加上各条弦，而每个交点再多添两段）。欧拉（Euler）公式 $V - E + F = 2$ 把圆外的区域也算在内，所以区域的数目是 $F - 1 = 1 + \binom{n}{2} + \binom{n}{4}$。

9.2 小型版的拉姆齐定理

1. 双色角（bichromatic angle）最多有 $6 \times 6 = 36$ 个。每个非单色的三角形恰好含有两个双色角，所以这类三角形最多有 18 个。由于 $\binom{6}{3} = 20$，最少有 2 个三角形是单色的。

2. 每个点有 16 条边，所以其中最少有 6 条同色，设为红色，并设它们连到 $P_1, \ldots, P_6$ 各点。若 P_i 之间有任何一条边是红色，便有一个红色三角形。若没有，则各 P_i 构成一个以两种颜色著色的 K_6，其中必有单色三角形。

3. 点 v 有 9 条边，所以要么最少有 4 条是金色，要么最少有 6 条是银色。若有 4 个金色邻点：它们之间只要有一条金色边，便与 v 构成金色三角形，否则它们构成一个银色的 K_4。若有 6 个银色邻点：它们之中必有一个单色三角形；若是金色，证明完成，若是银色，它便与 v 构成一个银色的 K_4。

4. 给每一项标上 (i,d)：分别是以该项结尾的最长递增子序列（subsequence）和最长递减子序列的长度。对任意两项，较后的一项不是 i 较大，就是 d 较大，所以各个标记互不相同。假如没有长度为 3 的子序列，那么 5 项就只有 4 个可能的标记。两者皆无的例子：$2,1,4,3$。

9.3 为篱笆桩着色

1. 与之前一样，$r_{n+1}=s_n$（第 n 根柱不是红色，所以第 $n+1$ 根柱可以是红色），而 $r_n+s_n=(k-1)^{n-1}$，因为第一根之后的每根柱都有 $k-1$ 种选择。所以 $r_{n+1}+r_n=(k-1)^{n-1}$，而 $r_n=\frac{(k-1)^{n-1}+(k-1)(-1)^{n-1}}{k}$。至于围成一圈的情况：第 1 根柱的颜色有 k 种选择，然后是一排 $n+1$ 根柱，最后一根与第一根同色，由此得 $kr_{n+1}=(k-1)^n+(-1)^n(k-1)$。

2. $2,3,5,8,\ldots$：设共有 a_n 种方法,以白色结尾的一排可接在任何一排 $n-1$ 根柱之后,而以红色结尾的一排则接在以白色结尾的一排 $n-2$ 根柱之后,所以 $a_n=a_{n-1}+a_{n-2}$。答案是 F_{n+2}。

3. 中心有 4 种选择，然后是 n-圈的 3-著色：$4\big(2^n+2(-1)^n\big)$。

4. 当 $n=4$ 时：有 18 种著色。转四分之一圈不会保持任何一种著色不变，转半圈则保持其中 6 种（即相对的柱同色的那些）不变，所以有 $\frac14(18+0+6+0)=6$ 条项链。当 $n=5$ 时：有 30 种著色，没有任何旋转能保持其中一种不变，所以是 6。

9.4 把数列看成函数

1. 余下各项之和最多是 $F_7\sum_{r\ge7}2^{r-7}10^{-r}=\frac{13\times10^{-7}}{0.8}<2\times10^{-6}$。由于 $\sum F_r10^{-r}=\frac{10}{89}$，而首六项给出 0.112358，所以 $\frac{10}{89}$ 介乎 0.112358 与 0.112360 之间，于是 $\frac{1}{89}=0.011235\ldots$。若改以两位数为一组，则 $\sum F_r100^{-r-1}=\frac{1}{9899}$。

2. 把递推关系（recurrence）求和，得 $f(x)(1-7x+10x^2)=2-7x$，而 $\frac{2-7x}{(1-2x)(1-5x)}$ 可如题目所述分拆。至于 b_n：由 $10p+5q=40$ 和 $40p+10q=100$ 得 $p=1$，$q=6$；$t^2=t+6$ 的根是 3 和 -2，而 $b_n=4\cdot3^n+(-2)^n$。

3. $\sum nx^n=\frac{x}{(1-x)^2}$；在 $x=\frac{1}{100}$ 时它等于 $\frac{100}{9801}$，所以 $\frac{1}{9801}=\sum n\cdot100^{-n-1}$。这个规律在 98 这一组出错：100 那一组把 1 进位到 99 那一组，后者又再进位到 98 那一组，于是展开式是 $\ldots96\,97\,99\,00\,01\ldots$。

4. 乘以 $\frac{1}{1-x}=1+x+x^2+\cdots$ 之后，x^n 的系数变成 $u_0+\cdots+u_n$。平方数的生成函数（generating function）是 $\frac{x(1+x)}{(1-x)^3}$，所以其部分和的生成函数是 $\frac{x+x^2}{(1-x)^4}$，而 x^n 的系数是 $\binom{n+2}{3}+\binom{n+1}{3}=\frac{n(n+1)(2n+1)}{6}$。

9.5 凑零钱

1. x^{24} 的系数就是 $6a+6b+3c=24$ 的解的数目：对由 0 到 4 的每个 $s=a+b$，各有 $s+1$ 种选择，合共 15 种。一般而言，x^n 的系数是 0，除非 $n=3m$，这时它等于 $\frac{1}{2}(k+1)(k+2)$，其中 $k=\lfloor m/2\rfloor$。加上额外的因式后，系数变成部分和：x^{24}、x^{25} 和 x^{66} 的系数分别是 55、55 和 650。它们所数的，是用 1p 和 3p 硬币以及两种不同的 6p 硬币付款的方法数目。

2. 每个部分 k 要么用一次，要么完全不用。这个乘积可以逐项相消：$\prod\frac{1-x^{2k}}{1-x^k}$ 把分母中指数为偶数的因式全部约去，剩下 $\prod\frac{1}{1-x^{\text{奇数}}}$，也就是分拆（partition）成奇数部分的生成函数。当 $n=6$ 时：6、$5+1$、$4+2$、$3+2+1$，以及 $5+1$、$3+3$、$3+1+1+1$、$1+1+1+1+1+1$：两类各有四个。

3. 当 $n=0,1,2,\ldots$ 时，数值依次是 $1,1,2,2,3,4,5,6,7,8,10,\ldots$，而 $\frac{(n+4)^2}{20}$ 与它们的差距总不超过 $\frac{1}{2}$（已验算至 $n=60$；证明则要用到 $\frac{1}{(1-x)(1-x^2)(1-x^5)}$ 的部分分式（partial fractions））。

4. 最后一枚硬币是 1p 或 2p，所以 $c_n=c_{n-1}+c_{n-2}$：这就是斐波那契（Fibonacci）数，$c_n=F_{n+1}$。

9.6 奇怪的骰子

1. 每颗骰子都是 x 乘以 $1+x$、$1+x$、$1+x^2$、$1+x^2$ 这些因式中若干个的乘积，而且在 $x=1$ 时的值是 4，所以它要取其中两个因式。除了标准的一对之外，唯一的选择是 $x(1+x)^2$ 和 $x(1+x^2)^2$：各面是 $1,2,2,3$ 和 $1,3,3,5$。

2. 乘积是 $2x^3+x^4+x^5+4x^6+3x^7+8x^8+6x^9+5x^{10}+6x^{11}$，与表中数字吻合。

3. 可以：取 Sicherman 骰子对 $1,2,2,3,3,4$ 和 $1,3,4,5,6,8$，再加一颗标准骰子。

4. $(x+x^2)(x+\cdots+x^8)=x^2(1+x)^2(1+x^2)(1+x^4)$。硬币需要一个值为 2 的因式：除了标准的选择 $x(1+x)$ 之外，它可以是 $x(1+x^2)$（两面是 $1,3$，配上八面体（octahedron）骰子 $1,2,2,3,5,6,6,7$），也可以是 $x(1+x^4)$（两面是 $1,5$，配上 $1,2,2,3,3,4,4,5$）。

9.7 相等的幂和

1. $\sum(c+b_k)^3-\sum(c+a_k)^3=3c^2\left(\sum b_k-\sum a_k\right)+3c\left(\sum b_k^2-\sum a_k^2\right)+\sum b_k^3-\sum a_k^3$，而首两个括号都等于零。由 $\{1,4,6,7\}$、$\{2,3,5,8\}$ 出发，取 $c=8$：得 $\{1,4,6,7,10,11,13,16\}$ 和 $\{2,3,5,8,9,12,14,15\}$。

2. A 和 $B\cup\{0\}$ 都有 6 个元素，和为 33，平方和为 253。于是 $\sum_{a\in A}(n+a)^2=\sum_{b\in B\cup\{0\}}(n+b)^2$，这样便把 $\{n^2,\ldots,(n+11)^2\}$ 分成两组。

3. 用二项式定理（binomial theorem）展开 $(x+c)^r$：式中只涉及不高于 r 次的幂，而这些幂的和都相等。题目中的两个集合，在第二个加入 0 之后，合起来就是 $\{0,\dots,15\}$，再作平移便得任意 16 个连续整数。

4. 按 $ABABABAB$ 的次序，A 约得 1.676，B 约得 1.042。按 $ABBABAAB$ 的次序，A 约得 1.560，B 约得 1.158：仍不相等，但已接近得多。

9.8 没有人拿到自己的恤衫

1. $\mathrm{E}(C_1)=\frac{1}{m}$，$\mathrm{Var}(C_1)=\frac{1}{m}-\frac{1}{m^2}$，而 $\mathrm{Cov}(C_1,C_2)=\frac{1}{m(m-1)}-\frac{1}{m^2}=\frac{1}{m^2(m-1)}$。于是 $\mathrm{E}(N)=1$，而 $\mathrm{Var}(N)=m\left(\frac{1}{m}-\frac{1}{m^2}\right)+m(m-1)\frac{1}{m^2(m-1)}=1$。当 $m=4$ 时：$\mathrm{P}(N=0)=\frac{9}{24}=0.375$，相比之下 $\mathrm{e}^{-1}\approx 0.368$。

2. 第 1 张牌落在某个位置 i（有 $n-1$ 种选择）。若第 i 张牌落在位置 1，余下的就是 $n-2$ 张牌的错排（derangement）；若不是，则把位置 1 当作第 i 张牌的禁位，得到 $n-1$ 张牌的错排。验算：$D_3=2(1+0)=2$，$D_4=3(2+1)=9$。

3. 先选出 k 张留在原位的牌，再把其余的牌错排：$\binom{n}{k}\frac{D_{n-k}}{n!}=\frac{1}{k!}\cdot\frac{D_{n-k}}{(n-k)!}=\frac{d_{n-k}}{k!}$，它趋向 $\frac{\mathrm{e}^{-1}}{k!}$。

4. $30-(15+10+6)+(5+3+2)-1=8$，它等于 $30\cdot\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{4}{5}$：容斥原理（inclusion–exclusion）的和式可以因式分解（这就是欧拉的 $\varphi(30)$）。

9.9 生日问题

1. $\mathrm{P}(X=r)=\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{r-2}{n}\right)\frac{r-1}{n}$，而这些概率之和为 1。至于期望值（expectation），$\mathrm{P}(Y=r)$ 在 $\mathrm{P}(Y\ge 1),\dots,\mathrm{P}(Y\ge r)$ 中各被计算一次。记 $Q_k=\mathrm{P}(X>k)=\prod_{j=1}^{k-1}\left(1-\frac{j}{n}\right)$。于是 $\mathrm{E}(X)=\sum(k+1)\frac{k}{n}Q_k$，同时 $\mathrm{E}(X)=1+\sum Q_k$；利用 $\sum\frac{k}{n}Q_k=1$ 并将两式相减，得 $\sum Q_k\left(1-\frac{k^2}{n}\right)=0$。

2. 由 $\binom{k}{3}\approx 365^2\ln 2$ 得 $k\approx 82$；这个近似忽略了各个三人组之间的相依关系，真正的答案是 88。

3. 14 人。（每一对人“接近”的概率是 $\frac{3}{365}$，而由 $\binom{k}{2}\frac{3}{365}\approx\ln 2$ 得 $k\approx 14$；精确计算的结果是：14 人时概率约为 0.537，13 人时约为 0.483。）

4. 两人同在第 i 天生日的概率是 p_i^2；把它们加起来。根据柯西–施瓦兹不等式（Cauchy–Schwarz），$1=\left(\sum p_i\right)^2\le 365\sum p_i^2$，等号仅在所有 p_i 都相等时成立。

9.10 总能找赎

1. 第一个人必须持有一枚 £1 硬币。若按其后几个人的情况取条件，便要分成许多情形（视乎每位 £2 顾客之后你手上有哪些硬币），每种情形最后都归结到第 (i) 和第

(ii) 部分的答案；把它们合起来，得 $\frac{m-2}{m+1}$。用反射原理（reflection principle）则一行便可完成。

2. 第一票必须投给 A。在以 A 开始的点票过程之中，坏的那些（即曾经出现平手的）可借着反射第一次平手之前的部分，与以 B 开始的点票过程一一对应。所以好的点票过程的数目是 $\binom{a+b-1}{a-1}-\binom{a+b-1}{a}$，除以 $\binom{a+b}{a}$ 便得 $\frac{a-b}{a+b}$。

3. ((()))、(()())、(())()、()(())、()()()。一个双射（bijection）：第一对括号把序列分成括号之内的部分和括号之后的部分；相应地，以某条固定的边为底的三角形把多边形分成两个较小的多边形。对两部分分别递归处理。

4. 每位 £5 顾客需要找回四枚 £1 硬币，所以路径每步上升 1 或下降 4。路径可以越过界线而不碰到它，于是反射再不能把坏的路径与一个简单的集合配对起来。

9.11 集齐一套

1. 各段等待时间的期望值之和是 $\frac{n}{n}+\frac{n}{n-1}+\cdots+\frac{n}{n/2+1}=n(H_n-H_{n/2})\approx n(\ln n-\ln\frac{n}{2})=n\ln 2$。

2. 每张卡缺少的概率是 $(1-\frac{1}{n})^k$；对 n 张卡求和。要达到 95%，需 $(1-\frac{1}{n})^k\approx \mathrm{e}^{-k/n}=0.05$，所以 $k\approx n\ln 20\approx 3n$。

3. 买了 t 包之后，某一张指定的卡出现少于两次的概率约为 $\mathrm{e}^{-t/n}(1+\frac{t}{n})$。当这类卡的期望数目 $n\mathrm{e}^{-t/n}(1+\frac{t}{n})$ 降至约 1 时，两个孩子便大致都集齐了；由此得 $\frac{t}{n}\approx \ln n+\ln\ln n$，远小于 $2\ln n$。

4. $\frac{1}{n+1}<\ln\frac{n+1}{n}$（$\frac{1}{x}$ 在 $[n,n+1]$ 上的曲线下面积），所以 $H_n-\ln n$ 递减。由凸性（convexity）可知梯形（trapezium）高估了面积：$\ln\frac{n+1}{n}<\frac{1}{2}\left(\frac{1}{n}+\frac{1}{n+1}\right)$，所以 $H_n-\ln n-\frac{1}{2n}$ 递增。取 $n=1$，得 $\gamma>\frac{1}{2}$；取 $n=4$，得 $\gamma<H_4-\ln 4\approx 0.697$。

9.12 随机子集

1. 每个元素必须属于最少一个集合：$\left(\frac{3}{4}\right)^n$，一般情况则是 $\left(1-2^{-m}\right)^n$。

2. $|S\setminus A_2|$ 与 $|A_2|$ 有相同的分布（distribution），并且与 A_1 独立（independent），所以所求概率就是 $|A_1|+|S\setminus A_2|=n$ 的概率。$(A_1,S\setminus A_2)$ 这一对集合对应于一个有 $2n$ 个元素的集合的随机子集，其大小为 n 的概率是 $\binom{2n}{n}/4^n$。由斯特灵（Stirling）公式（或沃利斯（Wallis）乘积，见 Section 4.4）得 $\approx\frac{1}{\sqrt{\pi n}}$。

3. 每个元素放进 A、放进 B，或两者都不放：3^n。满足 $A=B$ 的有序对只有 $(\varnothing,\varnothing)$，所以共有 $\frac{3^n-1}{2}+1=\frac{3^n+1}{2}$ 个无序对。

4. 当 $n=3$ 时，是 3 个集合（全部单元素集，或全部二元素集）。当 $n=4$ 时，是 6 个集合（全部二元素集）。一般情况是 $\binom{n}{\lfloor n/2\rfloor}$（Sperner 定理，Sperner's theorem）。

9.13 做了手脚的骰子

1. $1 = \left(\sum q_j\right)^2 \le 6\sum q_j^2$，所以 $\mathrm{P}(\text{相同}) = \sum q_j^2 \ge \frac{1}{6} > \frac{1}{36}$。

2. 两问的答案都是“可以”：所求概率是 $\sum q_j r_j$，而比方说，若一颗骰子永远掷出 1，另一颗永远掷出 2，它便等于 0。

3. 点数总和是 $0, 1, \ldots, 35$，每个恰好出现一次：是均匀分布。题目要求两颗骰子都标上 1 至 6，而这里的骰子并非如此。

4. 不可以。记 $\mathrm{P}(3) = q_1 r_1 s_1$ 和 $\mathrm{P}(18) = q_6 r_6 s_6$，两者都是 $\frac{1}{16}$。$\mathrm{P}(8) + \mathrm{P}(13)$ 包含 $q_1 r_1 s_6, \ldots, q_1 r_6 s_6$ 这六项（两个 1 和一个 6，或两个 6 和一个 1），它们的乘积是 $(\mathrm{P}(3)\mathrm{P}(18))^3$。由算术–几何平均不等式（AM–GM）得 $\mathrm{P}(8) + \mathrm{P}(13) \ge 6 \cdot \frac{1}{16} > \frac{2}{16}$。原来那个多项式证明照搬过来并不成立，因为 $1 + x + \cdots + x^{15}$ 确实有一个实根 -1。（不过它可以补救：三个实系数五次多项式需要三个实根，而 -1 只是单根。）

Beyond STEP

STEP 是大多数中学生一生中应考过最艰深的数学考试，也是最出色的考试之一：题目简短、陌生而深刻，当中最精彩的，其实是乔装了的大学数学小品。

本书精选最有趣的 STEP 试题，由 1987 年首次开考之前的样本试卷（specimen papers），到 2025 年的试卷，并以每一道题目作为一扇门。每道题目都会仔细解答，沿途提出优秀学生会问的问题，然后带你穿过这扇门：这个想法从何而来，到了大学会演变成甚么，以及接下来该读些甚么。

本书为正在准备 STEP、有天分又有志投身数学的学生而写。他们想要的不止是模拟试卷：他们想开始像数学系本科生那样思考，并一探前方有甚么正等着自己。

- 共 13 章，涵盖纯数学（pure mathematics）、力学（mechanics）、概率（probability）与统计（statistics）
- 136 道 STEP 历届试题，每题均附引导式解答
- 每道题目背后的大学数学
- 附答案的延伸练习，以及延伸阅读

全书共 13 章，本册为第 9 章。 全书各章及所有详细解答均可免费阅读、下载及分享，网址为

beyondstep.keithmcnulty.org

文字内容已贡献予公有领域（CC0 1.0）。STEP 试题 © Cambridge University Press & Assessment。

