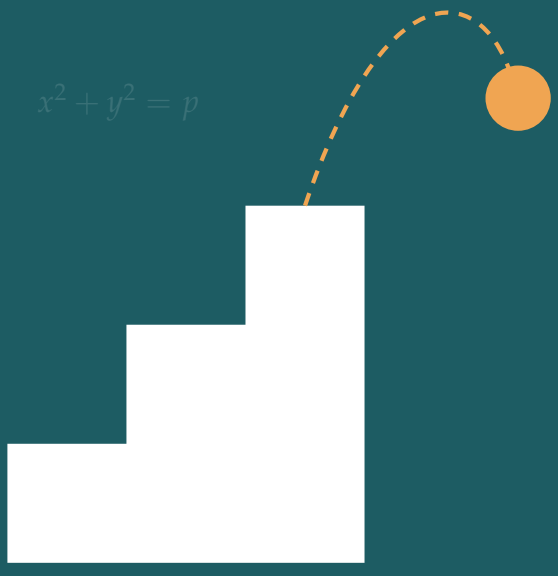
STEP 备试 · 大学数学

Beyond STEP

第 12 章

概率



KEITH McNULTY

Beyond STEP

以历届 STEP 试题为门
走进大学数学

第 12 章

概率

Keith McNulty

Beyond STEP：以历届 STEP 试题为门，走进大学数学

第 12 章：概率

本册是 *Beyond STEP* 全书 13 章中的第 12 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

12

概率

不确定的事情，可以精确地推理。直觉呢，原来不行。

本章内容

12.1	精确地取条件	STEP I 2007
12.2	基本比率谬误	STEP II 1998
12.3	“随机”是甚么意思?	STEP III 2022
12.4	两两独立，却非相互独立	STEP III 2021
12.5	最后一位乘客的座位	STEP II 2021
12.6	多一枚硬币	STEP I 2016
12.7	淘汰赛中谁会遇上谁	STEP II 样本试卷
12.8	Penney 游戏	STEP III 2001
12.9	用不公平的硬币作公平的决定	STEP I 2019
12.10	把绳头系起来	STEP III 2008
12.11	邮票、杯子与 e	STEP I 1996
12.12	何时停止寻找	STEP I 2001
12.13	三人决斗	STEP II 1995
12.14	混合策略	STEP I 1989
12.15	灭绝还是爆发	STEP II 2003
12.16	顺向与逆向的马可夫链	STEP III 1996
12.17	巴拿赫的火柴盒	STEP II 1992
12.18	无记忆性	STEP III 2021
12.19	平面上的随机点	STEP II 2012
12.20	蒲丰投针	STEP I 1993
12.21	折断一根棒	STEP III 1990
12.22	没有平均值的分布	STEP III 1995
12.23	由概率得出斯特灵公式	STEP II 2016
12.24	中心极限定理	STEP III 2008
12.25	众多之中的最大值	STEP III 2025
A	练习答案	第 149 页

机会的推理

概率论不过是化为计算的常识而已。

— 拉普拉斯（Pierre-Simon Laplace），*Théorie analytique des probabilités*（1812）

拉普拉斯这句名言只说对了一半。概率（probability）确实是精确化了的常识；但正如本章一再显示的，任由常识自行其是，就会出错。人们以为，罕见疾病的检测结果呈阳性，就表示你很可能患了病；以为若第一位登机的乘客坐错了座位，最后一位乘客就很难坐到自己的座位；以为一串掷硬币的结果，不可能胜过另一串同样长度的结果；以为多次量度的平均值总会稳定下来。这些全都是错的，而每一次只需简短的计算，便能揭示真相，更可贵的是，还能揭示当中的原因。

概率始于十七世纪有关赌博的问题，见于帕斯卡（Blaise Pascal）与费马（Pierre de Fermat）之间的书信。直到 1933 年，柯尔莫哥洛夫（Andrey Kolmogorov）订立了它的公理（axiom），概率才成为严格数学的一个分支。在这期间以及其后，它已发展成数学中应用最广的部分之一。

概率为何重要

凡是要处理不确定性的领域，都用到概率：医学（诊断检测与临床试验）、法律（证据的分量）、金融（风险与保险）、工程（可靠性与洪水）、电脑科学（随机化算法、密码学、机器学习）、物理（量子力学与统计力学）以及生物学（遗传学与流行病）。现代人工智能在很大程度上就是应用概率：语言模型所做的，是预测下一个字词出现的概率。而能够正确地就风险、机会和证据进行推理，是任何受过教育的人所能拥有的最宝贵技能之一。

你在大学会遇到甚么

大学一年级的概率课程涵盖公理、条件概率（conditional probability）与贝叶斯定理（Bayes' theorem）、独立性（independence）、离散与连续随机变量（random variable）、期望值（expectation）与方差（variance）、母函数（generating function），以及两大极限定理（limit theorem）：大数定律（law of large numbers）和中心极限定理（central limit theorem）。其后的课程会研究随机过程（random process），包括马可夫链（Markov chain）、泊松过程（Poisson process）、随机漫步（random walk）和分支过程（branching process），也会研究测度论概率（measure-theoretic probability，它为极限和连续分布奠定稳固的基础）和鞅（martingale），并把概率与统计学、分析、组合数学和物理联系起来。

贯穿本章的问题

三个大问题

1. **问题究竟在问甚么？** 随机试验（random experiment）是甚么？以甚么为条件？“随机”又是甚么意思？
2. **巧妙的切入点在哪里？** 甚么时候可以用对称性（symmetry）、指示变量（indicator）、取条件（conditioning）或精心选定的状态，来取代冗长的计算？
3. **长远而言会怎样？** 随机过程随时间如何演变？许多随机量的和、平均值和最大值又是甚么模样？

本章分为五部分。*清晰思考*（Section 12.1 至 Section 12.4）处理各种陷阱：取条件、基础比率（base rate）、“随机”的含义，以及独立的含义。*对称性与巧妙论证*（Section 12.5 至 Section 12.9）收集了一些解法既短且美的问题。*期望值、博弈与停止*（Section 12.10 至 Section 12.14）转向平均值与决策：何时停止、向谁开枪、怎样随机化。*随机过程*（Section 12.15 至 Section 12.19）追踪随机性在时间和空间中的表现。而*极限与连续概率*（Section 12.20 至 Section 12.25）则以几何概率（geometric probability）和几个重要的极限定理作结。

本章的思想

§	课题	思想
12.1	取条件	“指定的一个”与“至少一个”之别
12.2	贝叶斯定理	基础比率与自然频率（natural frequency）
12.3	贝特朗悖论（Bertrand's paradox）	“随机”需要定义
12.4	独立性	两两独立（pairwise independent）不等于相互独立（mutually independent）
12.5	飞机座位	两个座位之间的对称性
12.6	多一枚硬币	互补而对称的事件
12.7	锦标赛	数算比赛场数与配对
12.8	彭尼博弈（Penney's game）	首步分析（first-step analysis）；非传递性（non-transitivity）
12.9	公平硬币	从偏差中提取公平
12.10	随机环圈	无记忆的指示变量之和
12.11	杯子与 e	阶乘（factorial）与指数级数（exponential series）
12.12	最优停止（optimal stopping）	$\frac{1}{e}$ 法则
12.13	三人决斗（truel）	策略上的弱势
12.14	博弈论（game theory）	随机化策略；极小化极大（minimax）
12.15	分支过程	灭绝作为不动点（fixed point）
12.16	马可夫链	平稳分布（stationary distribution）；逆转
12.17	巴拿赫（Banach）的火柴盒	负二项分布（negative binomial distribution）
12.18	指数分布（exponential distribution）	无记忆性（memorylessness）与独立性
12.19	空间随机性	最近邻；瑞利（Rayleigh）分布
12.20	蒲丰投针（Buffon's needle）	几何概率与 π
12.21	随机三角形	以面积表示概率
12.22	柯西分布（Cauchy distribution）	没有平均值的分布
12.23	斯特灵公式（Stirling's formula）	由正态曲线（normal curve）得出阶乘
12.24	中心极限定理	矩母函数（moment generating function）
12.25	极值（extreme value）	冈贝尔分布（Gumbel distribution）

其中一节，即 Section 12.24，用到矩母函数，而矩母函数已于 2019 年从 STEP 考试范围中删去；该节会解释所需的内容。

怎样阅读本章

当一道概率题令人摸不着头脑时，不妨就一个细小的情况写出样本空间（sample space），然后逐一计数。当答案简单得令人起疑时，就去找出解释它的对称性。而当你的直觉与

计算结果不一致时，请相信计算，然后弄清楚直觉为何出错：真正的学习正在于此。

12.1 精确地取条件

题目

STEP I 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-s1-q13`

引子

“已知”（given that）是概率中最危险的字眼。这道题最后四个部分的说法听起来几乎一模一样：编号为 3 的圆片最先被取出；编号为 3 的圆片被取出；最先取出的是一块有编号的圆片；取出的圆片中有一块有编号。其中三个的答案都是 $\frac{1}{2}$，第四个却不是。

这个分别，正正就是著名的“男孩还是女孩”谜题背后的分别（一个家庭有两个孩子，其中至少一个是男孩；两个都是男孩的机会有多大?）。这道谜题难倒了一代又一代的学生，也难倒了不少专业数学家。当中的教训是：条件（condition）是一个精确的陈述，说明哪些结果仍然可能出现；而信息是怎样得来的，与信息本身同样重要。

动笔之前

? **想一想** 从十一块圆片中取出四块。对于“四块都有编号”这个事件（event）而言，抽取的次序有没有影响？四块圆片组成的等可能（equally likely）组合共有多少个?

每一个由四块圆片组成的组合被抽中的可能性都相同，所以我们可以不理会次序：共有 $\binom{11}{4} = 330$ 个等可能的组合。

逐步解题

阶段 1 没有条件

由四块有编号圆片组成的组合有 $\binom{5}{4} = 5$ 个，所以

$$\mathrm{P}(\text{四块都有编号}) = \frac{5}{330} = \frac{1}{66}.$$

阶段 2 3 号最先取出

想一想 编号为 3 的圆片最先被抽出之后，余下的三次抽取是怎样的?

其余三块圆片是从余下十块（四块有编号、六块空白）中随机抽出的三块，全部 $\binom{10}{3} = 120$ 个组合都是等可能的。所以

$$\text{(ii)}\quad \frac{\binom{4}{3}}{120} = \frac{4}{120} = \frac{1}{30}, \qquad \text{(iii)}\quad \frac{\binom{4}{1}\binom{6}{2}}{120} = \frac{4 \times 15}{120} = \frac{1}{2},$$

因为在 (iii) 中，其余三块之中必须恰好有一块有编号。

阶段 3 3 号被取出（在某一次）

想一想 编号为 3 的圆片是在哪一次抽出的，有没有关系？已知 3 号在四块之中，对其余三块你能说些甚么?

根据对称性（symmetry），已知这四块圆片中包含 3 号，其余三块就是从余下十块中随机抽出的三块，与第 2 阶段完全一样。所以 (iv) 的答案也是 $\frac{1}{2}$。

阶段 4 最先取出的是有编号的圆片

已知第一块圆片有编号（不论是哪个号码），其余三块同样是从余下十块圆片中随机抽出的三块，当中四块有编号、六块空白。所以 (v) 的答案也是 $\frac{1}{2}$。

阶段 5 取出的圆片中有一块有编号

想一想 现在的条件是“四块之中至少有一块有编号”。这与之前任何一个条件相同吗？330 个组合之中有多少个符合这个条件?

这个条件有所不同：它没有指明任何一块特定的圆片。只有 $\binom{6}{4} = 15$ 个组合完全没有编号圆片，所以有 315 个符合条件。其中 $\binom{5}{2}\binom{6}{2} = 10 \times 15 = 150$ 个恰好包含两块有编号的圆片。所以

$$\text{(vi)}\quad \frac{150}{315} = \frac{10}{21}. \qquad \blacksquare$$

停一停，试一试 为甚么 (vi) 与 (v) 不同？在 (v) 中，我们知道哪一块圆片有编号（第一块），而其余三块不受限制。在 (vi) 中，我们只知道有某一块圆片有编号。含有三块或四块编号圆片的组合，可以用更多方式满足“某一块圆片有编号”，所以它们不会像在 (v) 中那样比重偏低。

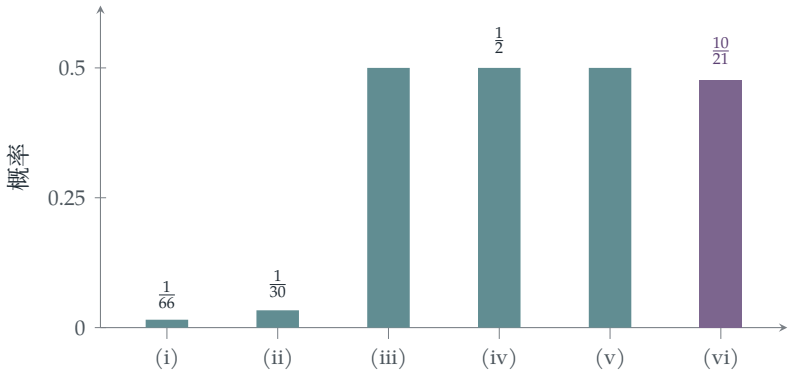


图 12.1.1 六个答案。第 (iii)、(iv) 和 (v) 部分的条件都是某一块特定的圆片有编号，答案是 $\frac{1}{2}$。第 (vi) 部分的条件只是有某一块圆片有编号，答案是 $\frac{10}{21}$。

回顾

- **条件会收窄样本空间（sample space）。** $P(A \mid B)$ 是 B 的结果之中同时属于 A 的比例。
- **对称性可以省功夫。** 3 号是最先抽出，还是在某一次抽出，并无分别，因为所有位置的地位都相同。
- **“特定的一个”不等于“至少一个”。** 听起来相似的条件，选出来的结果集合可以大不相同。

背后的数学

男孩还是女孩的谜题

马丁·加德纳（Martin Gardner）在 1959 年提出这个问题：史密斯先生有两个孩子，其中至少一个是男孩；两个都是男孩的概率是多少？列出等可能的家庭组合 BB、BG、GB、GG，剔除 GG 后剩下三个，其中一个是 BB：答案是 $\frac{1}{3}$。但如果我们知道年长的孩子是男孩，答案就是 $\frac{1}{2}$。这正是我们这道题第 (vi) 和第 (v) 部分的缩影。加德纳后来承认，他最初提出的问题含糊不清：答案取决于我们是怎样得知有一个男孩的。如果我们随机遇上其中一个孩子，而他是男孩，答案是 $\frac{1}{2}$；如果我们问“你有没有至少一个儿子？”而对

方回答有，答案就是 $\frac{1}{3}$。概率取决于产生信息的过程，而不仅是信息本身。

蒙提霍尔问题（Monty Hall）

概率中争议最多的谜题，背后也是同一个教训。在一个游戏节目中，三道门的其中一道后面有一辆汽车。你选了 1 号门。主持人知道汽车在哪里，他必定会打开另外两道门的其中一道，让你看到门后是一只山羊，并且必定会让你选择是否换门；当汽车就在你所选的门后，以致他有选择余地时，他会在另外两道门之间随机挑选。他打开了 3 号门，并让你改选 2 号门。你应该换吗？应该：换门后胜出的概率是 $\frac{2}{3}$。主持人选哪道门并非随机；这取决于汽车的位置，因而带有信息。确切的规则十分重要。如果主持人只要可以就一定打开 3 号门，那么看到 3 号门打开，1 号门和 2 号门就变得同样可能，换门后胜出的概率只有 $\frac{1}{2}$；然而，一个事先决定必定换门的参赛者，仍然有 $\frac{2}{3}$ 的机会胜出，因为换门胜出，恰好就是第一次选错的时候。Marilyn vos Savant 在 1990 年于杂志专栏中给出这个答案时，收到数以千计的来信指她错了，其中有些更是出自数学家之手。用电脑模拟一次，或者仔细列出各个等可能的情况，就可以解决争议。

贝叶斯（Bayes）与拉普拉斯（Laplace）

条件概率(conditional probability)的系统化基础,是由托马斯·贝叶斯(Thomas Bayes)和皮埃尔–西蒙·拉普拉斯（Pierre-Simon Laplace）奠定的：贝叶斯的论文在他死后于 1763 年发表，而拉普拉斯则在 1774 年以一般形式写出了现称为贝叶斯定理（Bayes' theorem）的法则。安德烈·柯尔莫哥洛夫（Andrey Kolmogorov）在 1933 年提出的公理（axioms），是现今每一个大学课程的起点，当中定义 $\mathrm{P}(A \mid B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$，而概率论的其余部分都建基于这个定义之上。下一节 Section 12.2 会展示贝叶斯定理在实际应用中的作用。

思维技巧

准确把握条件

计算条件概率之前，先确切写下哪些结果符合条件。问一问信息是怎样得来的。

在大学里，条件概率和条件期望（conditional expectation）是概率与统计中最重要的工具之一，也是最常出错的地方。把样本空间写出来，至少就一个小规模的情况写出来，是最好的防范方法。

延伸练习

1 **推广。**设有 b 块空白圆片和 n 块有编号的圆片，从中抽出 k 块，证明 (iv) 和 (v) 的答案永远相等，并求出 (vi) 在甚么情况下与它们相等。

2 **星期二出生的男孩。**一个家庭有两个孩子，其中至少一个是在星期二出生的男孩。证明两个都是男孩的概率是 $\frac{13}{27}$。（列出 14×14 个等可能的情况，或者用巧妙的方法计数。）

3 **蒙提霍尔问题。**列出各个结果，验证 $\frac{2}{3}$ 这个答案。如果主持人不知道汽车在哪里，随机打开另外两道门的其中一道，而门后刚好是一只山羊，那又如何?

4 **两张 A。**一手 13 张的桥牌中至少有一张 A。这手牌中至少有两张 A 的概率是多少?试与已知这手牌中有葵扇 A 时的概率比较。

延伸阅读

- **Jason Rosenhouse, *The Monty Hall Problem***（Oxford University Press）。讲述这道谜题、它的历史，以及它的众多变化。
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），有关条件概率的各章。
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***（CRC Press；可于网上免费阅读）。对条件概率的处理十分出色，并载有许多悖论。

12.2 基本比率谬误

题目

STEP II 1998，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/98-s2-q12`

引子

某种疾病的检测，对患者的准确度达 90%，而且甚少误报。Frankie 的检测结果呈阳性。Frankie 应该有多担心？大多数人都会答“大约 90%”，而在许多研究中，大多数医生也是这样回答。在这道题目中，正确答案是 $\frac{1}{11}$，约为 9%，足足小了十倍。

错处在于忽略了*基础比率*（base rate）：这种疾病很罕见，而健康的人为数众多，所以即使假阳性（false positive）率很低，误报的数目仍会远远多于真正的病例。这种*基础比率谬误*（base-rate fallacy）在医学、法律（在法律上称为检控官谬误（prosecutor's fallacy））、保安检查和机器学习中都举足轻重。贝叶斯定理（Bayes' theorem）正是对症良药。

动笔之前

想一想 试想像有 1000 人。当中有多少人患病？这些患者之中有多少人检测呈阳性？健康的人之中又有多少人检测呈阳性？

在 1000 人当中，有 10 人患病，其中 9 人检测呈阳性。在 990 个健康的人当中，有 $\frac{1}{11}$ 检测呈阳性，即 90 个假阳性。因此共有 99 人检测呈阳性，其中只有 9 人真正患病。

逐步解题

阶段 1 贝叶斯定理

想一想　以 D 表示“Frankie 患有此病”，以 $+$ 表示“AL 检测呈阳性”。你已知 $\mathrm{P}(+\mid D)$，想求 $\mathrm{P}(D\mid +)$。两者有甚么关系?

根据条件概率（conditional probability）的定义，$\mathrm{P}(D\mid +)=\frac{\mathrm{P}(D\cap +)}{\mathrm{P}(+)}$，而 $\mathrm{P}(+)$ 可按 Frankie 是否患病拆成两部分：

$$\begin{aligned}\mathrm{P}(D\mid +)&=\frac{\mathrm{P}(+\mid D)\mathrm{P}(D)}{\mathrm{P}(+\mid D)\mathrm{P}(D)+\mathrm{P}(+\mid \bar{D})\mathrm{P}(\bar{D})}\\&=\frac{0.9\times 0.01}{0.9\times 0.01+\frac{1}{11}\times 0.99}=\frac{0.009}{0.009+0.09}=\frac{1}{11}.\end{aligned}$$

这正是那一千人当中的 $\frac{9}{99}$。

阶段 2　第二次检测

想一想　只有 AL 检测呈阳性的人才会接受第二次检测。在这群人之中，患者占多大比例？把这个比例当作新的“基础比率”。

在 AL 检测呈阳性的人当中，有 $\frac{1}{11}$ 是患者，$\frac{10}{11}$ 不是。以这两个数作为先验概率（prior probability），再次应用贝叶斯定理，得

$$\mathrm{P}(D\mid \text{两次均呈阳性})=\frac{0.9\times\frac{1}{11}}{0.9\times\frac{1}{11}+p\times\frac{10}{11}}=\frac{0.9}{0.9+10p}.$$

令上式等于 $\frac{1}{2}$，得 $10p=0.9$，所以 $p=0.09$。

（这里假设在已知 Frankie 是否患病的条件下，两项检测的误差互相独立（independent），题目亦隐含这个假设。）

停一停，试一试　当 $p=0.09$ 时，STEP 的假阳性率与 AL 的相若，但第二次阳性结果却令概率由 $\frac{1}{11}$ 升至 $\frac{1}{2}$。为甚么质素相同的检测，第二次的影响会大得多?

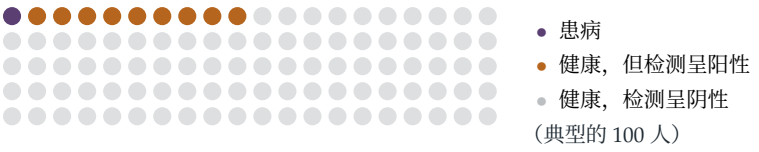


图 12.2.1　典型的一百人。其中一人患病（紫色），十次检测中有九次会呈阳性。在 99 个健康的人当中，约有 9 人会被误判为阳性（琥珀色）。在检测呈阳性的人当中，健康的人与患者的数目之比约为十比一。

回顾

- **自然频率（natural frequencies）。**想像 1000 人，答案便一目了然，也容易验算。
- **先验乘以似然（likelihood）。**后验概率（posterior probability）既取决于检测的准确度，也取决于基础比率。
- **分阶段更新。**昨天的后验就是今天的先验：第二次检测是由 $\frac{1}{11}$ 开始，而不是 $\frac{1}{100}$。

背后的数学

胜算与证据

以胜算（odds）来表达，贝叶斯定理最为清晰。D 的胜算是 $\mathrm{P}(D) : \mathrm{P}(\bar{D})$。一项证据 E 会把胜算乘以似然比（likelihood ratio）$\frac{\mathrm{P}(E|D)}{\mathrm{P}(E|\bar{D})}$：

$$\text{后验胜算} = \text{先验胜算} \times \text{似然比}.$$

这里的先验胜算是 $1:99$，而 AL 的似然比是 $0.9 \div \frac{1}{11} = 9.9$，于是后验胜算为 $9.9:99 = 1:10$，即概率为 $\frac{1}{11}$。第二项独立检测的似然比若为 10（$p = 0.09$ 时的 STEP 正是如此），便会把胜算带到 $1:1$。取对数之后，证据就可以相加：每项独立检测所贡献的，是其似然比的对数。图灵（Alan Turing）在布莱切利园（Bletchley Park）破解恩尼格玛（Enigma）密码电文时，用的正是这个方法，并以“ban”和“deciban”作为量度证据的单位。

检控官谬误

1999 年，事务律师 Sally Clark 被裁定谋杀两名年幼儿子罪名成立，定罪的部分依据是一项证据，指同一家庭发生两宗婴儿猝死的机会是“7300 万分之一”。姑且不论这个数字本身有错（它假设两宗死亡互相独立，忽略了遗传和环境因素），这个推论还把 P(证据 | 无辜) 与 P(无辜 | 证据) 混为一谈。母亲连杀两子同样极为罕见，真正应该比较的是这两种解释。皇家统计学会（Royal Statistical Society）发表公开信指出这个错误，她的定罪于 2003 年获撤销。当中的数学与本题完全相同。

Gigerenzer 与自然频率

在 1990 年代，心理学家 Gerd Gigerenzer 指出，医生若获得以条件概率表达的检测统计数字，遇到本题这类问题时大多错得离谱；但同样的资料若以自然频率表达（“在 1000 名女性当中，有 10 人患乳癌；其中 8 人检测呈阳性…”），大多数医生都能答对。他的研究改变了许多国家向病人传达医疗风险的方式。

思维技巧

用贝叶斯定理把条件概率倒转过来

$\mathrm{P}(A \mid B)$ 与 $\mathrm{P}(B \mid A)$ 是两个不同的数。要由其中一个求出另一个，你需要基础比率 $\mathrm{P}(A)$。

贝叶斯推理，即根据证据更新信念，是现代统计学和机器学习的基础，由垃圾邮件过滤器以至医学诊断都用得上。到了大学，你会遇到它的进阶形式——贝叶斯推断（Bayesian inference），届时处理的不再是单一的概率，而是先验分布和后验分布。

延伸练习

1 **雪柜里的尸体。**试做 STEP II 1987，第 16 题（`step.maths.org/questions/87-s2-q16`）：贝叶斯定理应用于两名可靠程度不同的证人，是 STEP 中较令人难忘的题目情境之一。

2 **要多罕见？**某项检测的灵敏度（sensitivity）为 0.9，假阳性率为 $\frac{1}{11}$。疾病要有多普遍，阳性结果才表示患病的机会高于一半？

3 **重复检测。**假如重复进行 AL 检测，而在已知受测者真实状况的条件下各次结果互相独立，那么两次 AL 检测均呈阳性后，患病的概率是多少？三次之后呢？

4 **筛查。**某项筛查计划为一百万人检测某种病症，其患病率（prevalence）为 10 000 分之 1，所用检测的灵敏度为 99%，假阳性率为 1%。有多少人会被告知可能患有此病症？当中又有多少人真的患病？

延伸阅读

- **Gerd Gigerenzer, *Reckoning with Risk***（Penguin）。自然频率、医学检测与法庭。
- **David Spiegelhalter, *The Art of Statistics***（Pelican）。内容包括医学筛查和 Sally Clark 案。
- **Sharon Bertsch McGrayne, *The Theory That Would Not Die***（Yale University Press）。贝叶斯定理的历史，包括图灵如何运用这个定理。

12.3 “随机”是甚么意思?

题目

STEP III 2022，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 2022 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

`www.ocr.org.uk/Images/696587-2022-past-papers.zip`

引子

“在圆上随机选一条弦（chord），它的期望长度（expected length）是多少?”这条问题听起来好像有一个答案。其实没有，除非你先说清楚“随机”是甚么意思。在圆周上随机取两点再连起来，期望长度是 $\frac{4a}{\pi} \approx 1.27a$。在圆内随机取一点，再随机取一个方向作过该点的弦，答案却是 $\frac{16a}{3\pi} \approx 1.70a$。两种做法都是选取“一条随机弦”的合理方法。

这就是贝特朗悖论（Bertrand's paradox），由约瑟夫·贝特朗（Joseph Bertrand）于 1889 年发表，用意是提出警告。只有在指明随机试验（random experiment）之后，概率（probability）才有明确的定义。这个悖论促使数学家在二十世纪为概率论建立严格的基础：在这套基础之中，每一个概率都定义在一个精确界定的结果空间之上。

动笔之前

? **想一想** 若角 AOB 为 θ，弦 AB 的长度是多少?（由 O 作垂线。）

由 O 所作的垂线平分该弦，也平分该角，所以 $AB = 2a\sin\frac{\theta}{2}$。

逐步解题

阶段 1　圆周上的两点

$$\mathrm{E}(AB) = \frac{1}{2\pi}\int_0^{2\pi} 2a\sin\frac{\theta}{2}\,\mathrm{d}\theta = \frac{a}{\pi}\Big[-2\cos\frac{\theta}{2}\Big]_0^{2\pi} = \frac{a}{\pi}\cdot 4 = \frac{4a}{\pi}. \qquad \blacksquare$$

阶段 2 圆内的一个随机点

想一想 对于 $0 \le r \le a$，$\mathrm{P}(R \le r)$ 是多少？对它求导，便得到密度。

$\mathrm{P}(R \le r)$ 是半径为 r 的圆盘面积除以整个圆盘的面积，即 $\frac{r^2}{a^2}$。所以 R 在 $[0,a]$ 上的密度（density）是 $\frac{2r}{a^2}$。（点较有可能离圆心较远，因为那里的面积较大。）

想一想 一条过 C 的弦与 OC 成角 t。这条弦与 O 相距多远？

由 O 到该直线的距离是 $R\sin t$，所以弦长为 $2\sqrt{a^2 - R^2\sin^2 t}$。

阶段 3 对 C 取平均

设 $s = \sin t$，

$$\begin{aligned} L(t) &= \int_0^a 2\sqrt{a^2 - r^2 s^2}\cdot\frac{2r}{a^2}\,\mathrm{d}r = \frac{4}{a^2}\Big[-\frac{(a^2 - r^2s^2)^{3/2}}{3s^2}\Big]_0^a \\ &= \frac{4}{a^2}\cdot\frac{a^3 - a^3(1-s^2)^{3/2}}{3s^2} = \frac{4a(1-\cos^3 t)}{3\sin^2 t}. \qquad \blacksquare \end{aligned}$$

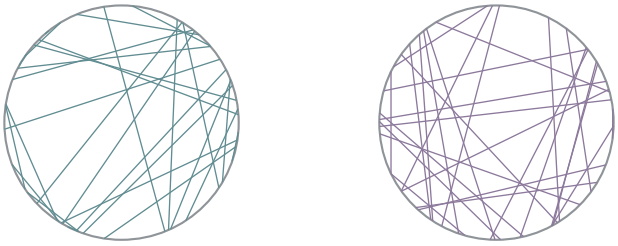
想一想 把 $1 - \cos^3 t$ 和 $\sin^2 t = 1 - \cos^2 t$ 因式分解。有甚么可以约去？

设 $c = \cos t$，则 $\frac{1-c^3}{1-c^2} = \frac{(1-c)(1+c+c^2)}{(1-c)(1+c)} = \frac{1+c+c^2}{1+c} = c + \frac{1}{1+c}$，而 $1 + \cos t = 2\cos^2\frac{t}{2}$。所以

$$L(t) = \frac{4a}{3}\Big(\cos t + \tfrac{1}{2}\sec^2\tfrac{t}{2}\Big). \qquad \blacksquare$$

阶段 4 方向也随机

$$\begin{aligned} \mathrm{E}\big(L(T)\big) &= \frac{2}{\pi}\int_0^{\pi/2}\frac{4a}{3}\Big(\cos t + \tfrac{1}{2}\sec^2\tfrac{t}{2}\Big)\,\mathrm{d}t \\ &= \frac{8a}{3\pi}\Big[\sin t + \tan\tfrac{t}{2}\Big]_0^{\pi/2} = \frac{8a}{3\pi}(1+1) = \frac{16a}{3\pi}. \qquad \blacksquare \end{aligned}$$



两个随机端点
平均长度 $\frac{4a}{\pi} \approx 1.27a$

圆盘内的随机中点
平均长度 $\frac{4a}{3} \approx 1.33a$

图 12.3.1 以两种不同方法选出的二十八条“随机弦”。左：圆周上两个独立的均匀分布点。右：以圆盘内一个均匀随机的点作为弦的中点。右边的方法在边缘附近产生的短弦较少。不同的方法给出不同的分布，也给出不同的答案。

回顾

- **指明试验。**“一条随机弦”在每一部分都化成了一个精确的步骤，而每个步骤给出不同的答案。
- **由面积求密度。**$\mathrm{P}(R \le r)$ 直接来自几何；求导便得到密度。
- **分阶段取平均。**先对 C 的位置取平均（t 固定），再对 t 取平均。

背后的数学

贝特朗的三个答案

贝特朗原本的问题是：一条随机弦比内接等边三角形的边长 $\sqrt{3}a$ 更长的概率是多少？他给出了三种方法和三个答案。

- 随机端点：固定一个端点；若另一个端点落在圆周中间的三分之一之内，弦便是长的。概率为 $\frac{1}{3}$。
- 随机取一条半径，再在其上随机取一点作为中点：若弦与 O 的距离小于 $\frac{a}{2}$，弦便是长的。概率为 $\frac{1}{2}$。
- 在圆盘内任意位置随机取中点：若中点与 O 的距离在 $\frac{a}{2}$ 之内，弦便是长的。概率为 $\frac{1}{4}$。

每一种方法都是弦的集合上一个不同的概率分布（probability distribution），当中并无矛盾：它们回答的是不同的问题。这里的教训是，“均匀随机”只有相对于某个选定的参数化（parametrisation）才有意义，而对于连续的集合，通常并没有唯一自然的参数化。

有没有正确答案?

1973 年，物理学家 Edwin Jaynes 主张第二种方法才是“正确”的，理由是把圆稍为移动或缩放，答案都不应改变：一条随机弦应该就像一根抛到地上所画圆圈上的饮管。只有与圆心距离均匀分布的那个分布才具有这种不变性（invariance），而抛饮管的实验也证实了 $\frac{1}{2}$。同一个想法，即正确的分布就是尊重情境中各种对称性的那一个，正是积分几何（integral geometry）的基础；积分几何是研究随机直线的数学，它也解释了 Section 12.20 中的布丰投针（Buffon's needle）。

从悖论到公理

贝特朗在 *Calcul des probabilités*（1889）中发表了他的弦悖论，部分原因是要嘲讽他眼中对“几何概率”的草率运用。庞加莱（Henri Poincaré）也讨论过它。1900 年，希尔伯特（David Hilbert）把概率的公理化（axiomatisation）列入他为新世纪提出的著名问题之中，而柯尔莫哥洛夫（Andrey Kolmogorov）在 1933 年解决了这个问题：概率是结果空间上的一个测度（measure），任何关于“随机”对象的问题，都必须先说明是哪一个空间、哪一个测度。贝特朗悖论正是说明为何必须如此的标准例子。

思维技巧

界定随机试验

计算概率之前，先准确描述随机对象是怎样产生的。不同的描述可以给出不同的答案。

在大学里，每一个概率计算都由一个样本空间（sample space）和一个概率测度（probability measure）出发。在统计学和物理学中，选择合适的随机模型（往往借助对称性）是工作的一大部分。

延伸练习

1 **贝特朗的概率。**验证上文的三个概率 $\frac{1}{3}$、$\frac{1}{2}$、$\frac{1}{4}$。

2 **另一个平均值。**证明若弦的中点在圆盘内均匀选取（右边的图），期望长度是 $\frac{4a}{3}$。（提示：这就是 $L(\frac{\pi}{2})$。为甚么?）

3 **随机距离。**若弦与 O 的距离在 $[0, a]$ 上均匀分布，证明期望长度是 $\frac{\pi a}{2}$。

4 **三种方法，三个分布。**对三种方法中的每一种，求弦长 ℓ 在 $[0, 2a]$ 上的概率密度（probability density）。

延伸阅读

- **Edwin Jaynes, “The well-posed problem”**, *Foundations of Physics* 3 (1973)。不变性论证，易读而又发人深省。
- **Gábor Székely, *Paradoxes in Probability Theory and Mathematical Statistics*** (Reidel)。贝特朗悖论及其他许多悖论。
- **Daniel Klain & Gian-Carlo Rota, *Introduction to Geometric Probability*** (Cambridge University Press)。积分几何，留待日后研读。

12.4 两两独立，却非相互独立

题目

STEP III 2021，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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

`www.ocr.org.uk/Images/696588-2021-past-papers.zip`

引子

玩家 1 和玩家 2 掷出相同的点数。这能告诉你玩家 2 和玩家 3 是否掷出相同点数吗？出人意料的是，不能：这两个事件是独立（independent）的。玩家 1 和玩家 3 的情况也一样。但如果 1 与 2 相同，而 2 又与 3 相同，那么 1 必定与 3 相同。三个事件，每两个之间都独立，合起来却不独立。

两两独立(pairwise independence)比完全独立弱,而这个差别至关重要。就方差(variance)而言，两两独立已经足够，本题正是利用了这一点。但在许多其他用途上，这并不足够。电脑科学家会刻意使用两两独立的随机数，因为产生它们的成本远低于产生完全独立的随机数，而对许多算法来说已经够用。

动笔之前

想一想 $\mathrm{P}(X_{12}=1)$ 是多少？$\mathrm{P}(X_{12}=1$ 及 $X_{23}=1)$ 又是多少？

无论玩家 1 掷出甚么，玩家 2 与之相同的概率都是 $\frac{1}{6}$。$X_{12}=1$ 和 $X_{23}=1$ 同时成立，即三人掷出相同的点数：概率为 $\frac{6}{216}=\frac{1}{36}$。

逐步解题

阶段 1 两两独立

由于 $\mathrm{P}(X_{12}=1, X_{23}=1)=\frac{1}{36}=\frac{1}{6}\cdot\frac{1}{6}=\mathrm{P}(X_{12}=1)\mathrm{P}(X_{23}=1)$，而每个变量只取两个值，所以 X_{12} 与 X_{23} 独立。（对于只取两个值的变量，联合分布表中的一格便决定了其余各格；练习 1 会把这一点说清楚。）没有共同玩家的两对，例如 X_{12} 与 X_{34}，显然是独立的。因此所有 X_{ij} 都两两独立。

阶段 2 得分的平均值与方差

得分为 $S=\sum_{i<j}X_{ij}$，是 $\binom{n}{2}$ 个变量之和，每个变量的平均值为 $\frac{1}{6}$，方差为 $\frac{1}{6}\cdot\frac{5}{6}=\frac{5}{36}$。期望值（expectation）总是可以相加的，所以 $\mathrm{E}(S)=\frac{1}{6}\binom{n}{2}$。

> **想一想** 和的方差在甚么时候等于方差之和？

和的方差等于方差之和，再加上协方差（covariance）之和的两倍，而独立的一对变量协方差为零。两两独立已经足够：

$$\mathrm{E}(S)=\frac{n(n-1)}{12},\qquad \mathrm{Var}(S)=\frac{5}{36}\binom{n}{2}=\frac{5n(n-1)}{72}.$$

阶段 3 方差公式

若各 Y_i 的平均值为零，则 $\mathrm{Var}(\sum Y_i)=\mathrm{E}((\sum Y_i)^2)$，把平方展开后，每个 Y_i^2 出现一次，每个乘积 Y_iY_j（其中 $i<j$）出现两次。取期望值便得到公式。 ■

阶段 4 奇偶游戏

> **想一想** $\mathrm{E}(Z_{12})$ 是多少？$Z_{12}Z_{23}$ 可以取哪些值？它在甚么时候不等于零？

$Z_{12}=1$ 的概率为 $\frac{3}{36}=\frac{1}{12}$（两人都掷出 2、都掷出 4 或都掷出 6），而取值 -1 的概率为 $\frac{1}{12}$；所以 $\mathrm{E}(Z_{12})=0$ 而 $\mathrm{E}(Z_{12}^2)=\frac{1}{6}$。

只有当三位玩家都掷出相同点数时，才有 $Z_{12}Z_{23}\neq 0$，而此时 $Z_{12}=Z_{23}$（同为 $+1$ 或同为 -1），所以乘积为 1。因此 $\mathrm{E}(Z_{12}Z_{23})=\frac{1}{36}$，这并不等于 $\mathrm{E}(Z_{12})\mathrm{E}(Z_{23})=0$：它们并不独立。 ■

（更直接的看法是，$\mathrm{P}(Z_{12}=1, Z_{23}=-1)=0$，而每个事件单独发生的概率都是 $\frac{1}{12}$。）

阶段 5 方差

总得分 $T=\sum Z_{ij}$ 的平均值为 0。根据第 (ii) 部分，$\mathrm{Var}(T)$ 等于 $\binom{n}{2}\cdot\frac{1}{6}$，再加上 $\mathrm{E}(Z_{ij}Z_{kl})$ 对所有“两对玩家”的组合求和后的两倍。没有共同玩家的两对，贡献为 0（它们独立，而且各自的平均值为零）。恰好有一位共同玩家的两对，每组贡献 $\frac{1}{36}$，而这样的组合共有

$n\binom{n-1}{2}$ 个（先选共同的玩家，再选另外两人）。所以

$$\begin{aligned}\operatorname{Var}(T) &= \frac{n(n-1)}{12} + 2\cdot\frac{1}{36}\cdot\frac{n(n-1)(n-2)}{2} \\ &= \frac{n(n-1)}{36}\big(3+(n-2)\big) = \frac{n(n^2-1)}{36}. \qquad \blacksquare\end{aligned}$$

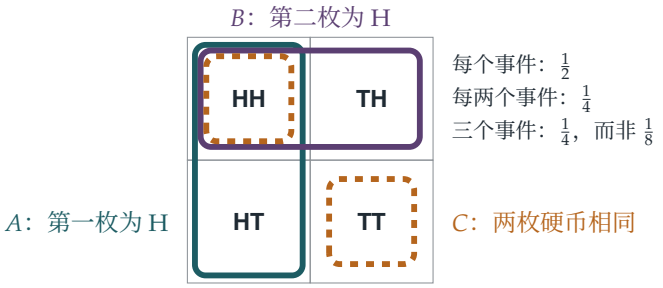


图 12.4.1 伯恩斯坦（Bernstein）的例子：两枚公平硬币。事件“第一枚硬币是正面”（蓝绿色）、“第二枚硬币是正面”（紫红色）和“两枚硬币结果相同”（琥珀色虚线）的概率各为 $\frac{1}{2}$，而每两个事件都相交于概率为 $\frac{1}{4}$ 的单一方格，所以它们两两独立。但三个事件全部相交于 HH，其概率为 $\frac{1}{4}$，而非 $\frac{1}{8}$：其中任何两个事件都决定了第三个。

回顾

- **指示变量（indicator variables）。**把得分写成指示变量之和，一切便化为对一对对玩家的考虑。
- **方差只需要看成对的变量。**和的方差只取决于两两之间的协方差，所以两两独立已足以令方差相加。
- **独立性要验证，不要假设。**两个看似相同的游戏，却有不同的相依结构。

背后的数学

多个事件的独立性

如果对每一个子族，交集的概率都等于各概率之积，则称事件 $A_1,\dots,A_n$ 独立。对三个事件来说，这意味着四个条件：三个是关于每一对的，另一个是 $\mathrm{P}(A\cap B\cap C)=\mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C)$。伯恩斯坦（Sergei Bernstein）在 1920 年代提出图 12.4.1 中的例子，说明两两之间的条件并不能推出三者的条件。（反过来，单凭三者的条件也不能推出两两之间的条件。）题目中“点数相同”的事件 X_{12}、X_{23}、X_{13} 与伯恩斯坦的硬币如出一辙。

不相关不等于独立

协方差为零，即 $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$，是比独立更弱的条件，称为不相关（uncorrelated）。STEP III 2000 Q14（见练习）说明，对于只取两个值的变量，这两个概念是一致的，但取三个值时则不然。一个简单的例子：设 X 在 $\{-1, 0, 1\}$ 上均匀分布，并设 $Y = X^2$。那么 $\mathrm{E}(XY) = \mathrm{E}(X^3) = 0 = \mathrm{E}(X)\mathrm{E}(Y)$，然而 Y 完全由 X 决定。协方差只量度线性关联。

廉价的随机性

对电脑来说，产生随机数的成本很高。1979 年，Larry Carter 和 Mark Wegman 证明，对许多算法而言，例如把数据储存在伪随机位置的*杂凑*（hashing），两两独立的随机选择与完全独立的同样有效。两两独立的数值可以用极低的成本产生：例如，取随机的 a 和 b，则 $a + bk \pmod{p}$ 对 $k = 0, 1, \ldots, p-1$ 给出 p 个数值，其中任何两个都是独立的。本题的方差计算正是它行得通的原因：切比雪夫（Chebyshev）不等式只需要方差，而方差只需要看成对的变量。

思维技巧

分解为指示变量

要求一个计数的平均值和方差，可把它写成指示变量之和，每一个可能被数到的对象对应一个指示变量。求平均值只需要考虑单个指示变量；求方差则需要考虑成对的指示变量。

这是概率论中最有用的技巧之一，早在 Section 9.8 已经出现过。在大学里，随机图、算法和统计检验的分析都以它为基础。

延伸练习

1 **两个值。**试做 STEP III 2000，第 14 题（`step.maths.org/questions/00-s3-q14`）。该题证明，对于各自只取两个值的随机变量，$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ 必然推出独立，并要求给出一个取三个值的反例。

2 **人数更多的队伍。**在第 (i) 部分中取 $n = 7$，求得分的平均值和标准差，并把例子中的得分 7 与它们比较。

3 **再谈生日。**利用指示变量，求 k 个人当中生日相同的人数对的平均值和方差（Section 9.9）。这些指示变量是否两两独立？

4 **三者成立但两两不独立。**找出三个事件，每个的概率都是 $\frac{1}{2}$，满足 $\mathrm{P}(A \cap B \cap C) = \frac{1}{8}$，但并非两两独立。

延伸阅读

- **Geoffrey Grimmett & David Stirzaker, *Probability and Random Processes***（Oxford University Press），开首几章讨论独立性。
- **Michael Mitzenmacher & Eli Upfal, *Probability and Computing***（Cambridge University Press）。两两独立及其在算法中的应用。
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***（CRC Press；可于网上免费阅读），关于指示变量与协方差。

12.5 最后一位乘客的座位

题目

STEP II 2021，第 11 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 2021 年 STEP 试卷（zip 文件）内的 STEP II 试卷：

`www.ocr.org.uk/Images/696588-2021-past-papers.zip`

引子

一百名乘客登上一班满座的飞机。第一位乘客遗失了登机证，于是随便找个座位坐下。其后每位乘客，若自己的座位仍然空着就坐自己的座位，否则就在余下的空位中随机选一个。最后一位乘客坐到自己座位的概率（probability）是多少？混乱或会蔓延至整架飞机，所以你可能预期这个概率很小；又或者大部分乘客根本不受影响，概率应该接近 1。

无论乘客人数是多少，答案都恰好是 $\frac{1}{2}$。本题以递推关系（recurrence）和归纳法（induction）证明这个结果。不过，还有一个利用对称性（symmetry）的漂亮论证，一句话就能说完；它解释了为甚么答案是 $\frac{1}{2}$，更可让你立即写出任何一位乘客的答案。

动笔之前

停一停，试一试 列出所有可能发生的情况，求出 P_2 和 P_3。

若有两个座位，T_1 会坐 S_1（这时 T_2 坐到 S_2）或 S_2，两者的概率各为 $\frac{1}{2}$：$P_2 = \frac{1}{2}$。若有三个座位：如果 T_1 坐 S_1（概率为 $\frac{1}{3}$），人人都没有问题；如果 T_1 坐 S_3，T_3 就坐不到自己的座位；如果 T_1 坐 S_2，则 T_2 要在 S_1 和 S_3 之间随机选择，而 T_3 只有在 T_2 选中 S_1 时才没有问题。所以 $P_3 = \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2}$。

逐步解题

阶段 1 递推关系

想一想 假设 T_1 坐了 S_k，其中 $2 \le k \le n-1$。乘客 $T_2, \dots, T_{k-1}$ 会怎样？当 T_k 到达时，哪些座位是空的？这个局面与一列较小的列车相比又如何？

乘客 $T_2, \dots, T_{k-1}$ 发现自己的座位空着，便坐了下来。当 T_k 到达时，空着的座位是 S_1 和 $S_{k+1}, \dots, S_n$：共有 $n-k+1$ 个。T_k 从中随机选一个，此后每位乘客若自己的座位空着就坐自己的座位，否则就随机选一个。这正是有 $n-k+1$ 个座位的原来问题，T_k 扮演第一位乘客的角色，而 S_1 则扮演 T_k“自己的”座位的角色（即一旦被选中，一切便回复正常的那个座位）。因此，T_n 坐到 S_n 的条件概率（conditional probability）是 P_{n-k+1}。

现在按 T_1 所坐的座位分情况考虑，每个座位的概率都是 $\frac{1}{n}$：坐 S_1 必定成功，坐 S_n 必定失败，而坐 S_k 则给出 P_{n-k+1}。所以

$$P_n = \frac{1}{n}\Big(1 + 0 + \sum_{k=2}^{n-1} P_{n-k+1}\Big) = \frac{1}{n}\Big(1 + \sum_{r=2}^{n-1} P_r\Big). \qquad \blacksquare$$

阶段 2 答案

想一想 $P_2 = P_3 = \frac{1}{2}$。猜出 P_n，并用强归纳法（strong induction）证明。

断言：对所有 $n \ge 2$，$P_n = \frac{1}{2}$。当 $n = 2$ 时断言成立。若 $P_r = \frac{1}{2}$ 对 $2 \le r \le n-1$ 成立，则

$$P_n = \frac{1}{n}\Big(1 + \frac{n-2}{2}\Big) = \frac{1}{n} \cdot \frac{n}{2} = \frac{1}{2}. \qquad \blacksquare$$

阶段 3 倒数第二位乘客

想一想 试找一个更巧妙的论证。到最后，留给 T_n 的是哪个座位？有可能是 S_2 吗？

最后留给 T_n 的座位不是 S_1 就是 S_n。它不可能是 S_k（其中 $2 \le k \le n-1$），因为 T_k 早就会把它坐了。而在此之前的每一次随机选择，对 S_1 和 S_n 都是对称的：每当有被占了座位的乘客要选位，S_1 和 S_n 被选中的机会均等。两者之中先被选中的那个决定了结局，所以两者同样可能成为最后剩下的座位：$P_n = \frac{1}{2}$。

同一个论证适用于任何一位乘客 T_k，$k \ge 2$。只有当某位较早的乘客占了 T_k 的座位，他们才会失去自己的座位。座位被占的连锁反应会一直持续，直至有人选中 S_1 为止；在 $S_1, S_k, S_{k+1}, \dots, S_n$ 这些座位之中，T_k 到达之前的选位者对它们一视同仁，所以最先被选

中的座位同样可能是这 $n-k+2$ 个座位中的任何一个，而 T_k 只有在最先被选中的是 S_k 时才会失去座位。所以

$$\mathrm{P}(T_k \text{ 坐在} S_k) = \frac{n-k+1}{n-k+2}.$$

当 $k = n-1$（其中 $n \geq 3$）时，这给出 $Q_n = \frac{2}{3}$；而 $Q_2 = \frac{1}{2}$，因为这时 $T_{n-1} = T_1$，他们坐在 S_1 的概率是 $\frac{1}{2}$。 ■

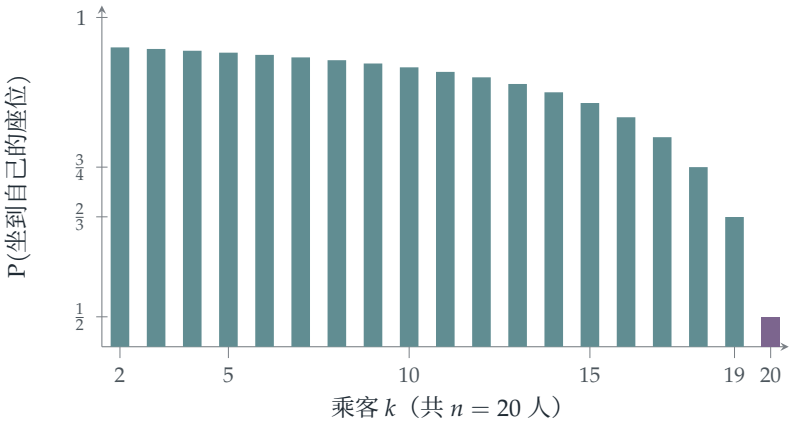


图 12.5.1 乘客 k 坐到自己座位的概率，列车共有 $n = 20$ 个座位。当 $2 \leq k \leq n$ 时，概率是 $\frac{n-k+1}{n-k+2}$：由 T_2 的 $\frac{19}{20}$ 一直下降至 T_{19} 的 $\frac{2}{3}$，以及最后一位乘客的 $\frac{1}{2}$（紫红色）。

回顾

- **化为较小的同类问题。**在 T_1 坐了 S_k 之后，余下的问题就是座位较少的同一个问题。
- **强归纳法。**递推关系牵涉所有较小的情况，所以归纳时要假设结果对所有这些情况都成立。
- **找出对称性。**座位 S_1 和 S_n 自始至终可以互换，这样不用任何代数运算就解释了 $\frac{1}{2}$ 的由来。

背后的数学

对称论证

第二个证明体现了概率论中一个威力强大的想法：若两个结果可由随机过程的某种对称性互相对应，它们的概率就相同。这里的过程在 S_1 和 S_n 其中一个被选中之前，并不区分这两个座位，所以两者成为最后剩下的座位的可能性必定相等。对称论证往往只凭一

个观察，就能取代冗长的计算。接下来的几节有很多这样的例子：Section 12.6 中多掷一次硬币，Section 12.7 中的淘汰赛抽签。

有多少位乘客被迫换位?

被迫换位的乘客（即坐不到自己座位的乘客）的数目同样有很好的规律。T_1 除非选中 S_1，否则就算是换了位；而 T_k（其中 $k \geq 2$）被迫换位的概率是 $\frac{1}{n-k+2}$。把这些概率相加，被迫换位的乘客数目的期望值（expected number）是

$$\left(1-\frac{1}{n}\right)+\left(\frac{1}{n}+\frac{1}{n-1}+\cdots+\frac{1}{2}\right)=H_n-\frac{1}{n},$$

大约是 $\ln n$：即使飞机上有 300 个座位，这场混乱平均也只影响约六位乘客。调和数（harmonic numbers）又再出现了（Section 9.11）。

思维技巧

寻找对称性

如果随机过程无法分辨两个结果，它们的可能性就相等。找出它们最初变得可以分辨的那一刻。

在大学，对称论证在概率论中随处可见，由反射原理（reflection principle）（Section 9.10）到可交换性（exchangeability）皆是；可交换性是指一列随机变量（random variables）无论以任何次序排列都有相同的分布，这个性质是不少现代统计学的基础。

延伸练习

1 **两位健忘的乘客。**如果最先的两位乘客都随机选位，最后一位乘客坐到自己座位的概率是多少?

2 **被迫换位的乘客。**列出所有可能情况，验证 $n=3$ 时被迫换位的乘客数目的期望值。

3 **迂回的路线。**利用递推关系，证明 $\sum_{r=2}^{n-1} P_r = \frac{n-2}{2}$ 与对称论证互相吻合。

4 **最后一个座位。**证明在最后一位乘客坐不到自己座位的条件下，最后一位乘客坐到 S_1 的概率是 1。为甚么?

延伸阅读

- **Peter Winkler, *Mathematical Puzzles: A Connoisseur's Collection***（A K Peters）。收录了飞机座位谜题，以及许多其他可用对称性解决的谜题。
- **Frederick Mosteller, *Fifty Challenging Problems in Probability***（Dover）。一本经典的题集。
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***（CRC Press；可于网上免费阅览）。

12.6 多一枚硬币

题目

STEP I 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-s1-q12`

引子

Bob 比 Alice 多掷一枚硬币。他掷得的正面（heads）比 Alice 多的概率（probability）是多少？他占了优势，但多掷一次不保证多一个正面，而且平手算他输。用硬算的方法做出头两个情况，两次的答案都是 $\frac{1}{2}$。这是巧合吗？

不是巧合，而且有一个只需一句话的证明：只要留意到 Bob 掷得的正面或反面（tails）必有其一比 Alice 多，但不会两者皆多。这道题目是一个典范，示范了计算如何提示出规律，而巧妙的对称性（symmetry）又如何解释这个规律。

动笔之前

想一想 Alice 掷两次：她掷得正面的次数的分布（distribution）是甚么？掷三次的 Bob 又如何？

Alice 掷得 $0, 1, 2$ 个正面的概率分别是 $\frac{1}{4}, \frac{2}{4}, \frac{1}{4}$；Bob 掷得 $0, 1, 2, 3$ 个正面的概率分别是 $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$。

逐步解题

阶段 1 硬算

第 (i) 部分，按 Alice 掷得的正面数目逐项相加：

$$\begin{aligned}\mathrm{P}(B>A) &= \tfrac{1}{4}\mathrm{P}(B\ge 1)+\tfrac{2}{4}\mathrm{P}(B\ge 2)+\tfrac{1}{4}\mathrm{P}(B=3)\\ &= \tfrac{1}{4}\cdot\tfrac{7}{8}+\tfrac{1}{2}\cdot\tfrac{4}{8}+\tfrac{1}{4}\cdot\tfrac{1}{8}=\frac{7+8+1}{32}=\frac{1}{2}.\end{aligned}$$

第 (ii) 部分是同样的计算，只是项数较多；数一数 $8\times 16=128$ 个等可能的结果，其中 64 个是 Bob 领先（图 12.6.1），所以答案同样是 $\frac{1}{2}$。

阶段 2 拆开 Bob 的投掷

想一想 把 Bob 的 $n+1$ 次投掷看成先掷 n 次，再额外掷一次。掷完头 n 次之后，他可能领先 Alice、与她平手，或者落后。每种情况下会发生甚么事?

二人各掷 n 次之后，Bob 领先的概率是 p_2，平手的概率是 p_1，落后的概率是 p_2（由 Alice 与 Bob 之间的对称性可知）。如果他领先，他会保持领先。如果平手，他额外的一掷有 $\frac{1}{2}$ 的概率令他领先。如果他落后，多一个正面最多只能追成平手。所以

$$\mathrm{P}(\text{Bob 的正面较多})=p_2+\tfrac{1}{2}p_1.$$

由于这三种情况涵盖了所有可能，$2p_2+p_1=1$，因此 $p_2+\frac{1}{2}p_1=\frac{1}{2}$。多了一枚硬币，Bob 掷得较多正面的概率对每个 n 都刚好是 $\frac{1}{2}$。 ■

停一停，试一试 以下就是那个一句话的证明。Bob 掷 $n+1$ 次，所以他的正面和反面数目加起来是 $n+1$，比 Alice 的多一。试证明 Bob 的正面比 Alice 多，当且仅当他的反面不比 Alice 多。由于正面与反面是对称的，这两个事件的概率相等，所以各为 $\frac{1}{2}$。

事实上，记 $H_A+T_A=n$ 及 $H_B+T_B=n+1$：若 $H_B>H_A$，则 $H_B\ge H_A+1$，所以 $T_B=n+1-H_B\le n-H_A=T_A$；反过来，若 $T_B\le T_A$，则 $H_B\ge H_A+1$。所以“正面较多”与“反面较多”是互补（complementary）的，而把正面与反面对调（由于硬币是公平的，这不会改变概率）就会把两者互换。

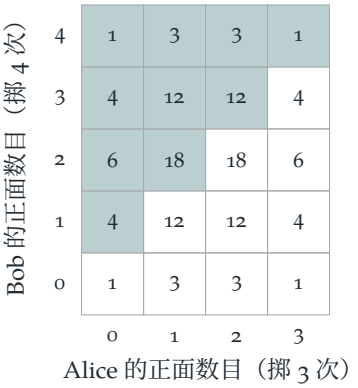


图 12.6.1 第 (ii) 部分：在 128 个等可能的结果之中，每种正面数目组合所占的个数。Bob 在有底色的格子中领先，这些格子刚好包含一半的结果。

回顾

- **计算小的情况。**两次计算都提示答案是 $\frac{1}{2}$。
- **按自然的划分来取条件。**把 Bob 额外的一掷分开，问题便化为一场势均力敌的比试再加一枚硬币。
- **找出互补的一对。**“正面较多”与“反面较多”互为补事件，并且在一个对称之下互换。

背后的数学

对称与补事件

那个漂亮的证明结合了两个事实：在一个对称之下互换的两个事件概率相等，而两个互补事件的概率之和为 1。若两者同时成立，则每个事件的概率都是 $\frac{1}{2}$。同样的推理在概率论中随处可见。在航班座位问题（Section 12.5）中，S_1 与 S_n 在一个对称之下互换，而最后剩下的恰好是其中一个。在六贯棋（Hex）中，由棋盘的对称性，加上恰有一方必胜这个事实，可知先手一方占优，不过在大棋盘上无人知道必胜策略是甚么。

平手的可能性有多大？

当二人投掷次数相同时，平手的概率 p_1 就很重要。Alice 与 Bob 平手，即 Alice 的正面数目等于 Bob 的正面数目；利用对称性，把 Bob 的正面换成他的反面，这个概率就等于 Alice 的正面数目加上 Bob 的正面数目等于 n 的概率，亦即在 $2n$ 次投掷中共有 n 个正面。

所以

$$p_1 = \binom{2n}{n}\frac{1}{4^n} \approx \frac{1}{\sqrt{\pi n}},$$

这里用了斯特灵近似（Stirling's approximation）（Section 12.23）。平手会变得罕见，但变得很慢：各掷 100 次，平手的机会仍然约有 5.6%。

思维技巧

先计算，再猜想，然后解释

当计算得出一个简单得可疑的答案时，就要找出原因。简单的答案通常有简单的解释。

在大学里，这个习惯正是理解与计算的分野。这里的第一个证明确认了答案；第二个证明则解释了答案，而且可以立即推广（试做练习 2）。

延伸练习

1 **多掷两枚硬币。**若 Bob 掷 $n+2$ 枚硬币而 Alice 掷 n 枚，证明 $\mathrm{P}(\text{Bob 的正面较多}) = \frac{1}{2} + \frac{1}{2}\mathrm{P}(H' = H_A)$，其中 H' 是 Bob 头 $n+1$ 次投掷中的正面数目。

2 **不公平的硬币。**若二人所用的硬币掷出正面的概率都是 p，多掷一次时答案是否仍然是 $\frac{1}{2}$？那个漂亮的证明中哪一步不再成立？

3 **平手。**用正文中的技巧证明 $p_1 = \binom{2n}{n}/4^n$，并求它在 $n = 1, 2, 3$ 时的值。

4 **骰子。**Alice 掷 n 粒骰子，Bob 掷 $n+1$ 粒。Bob 掷得较多六点的概率是否等于 $\frac{1}{2}$？

延伸阅读

- **Peter Winkler, *Mathematical Mind-Benders***（A K Peters）。收录大量可用对称性解决的概率谜题。
- **Gabor Székely, *Paradoxes in Probability Theory and Mathematical Statistics***（Reidel）。
- **Frederick Mosteller, *Fifty Challenging Problems in Probability***（Dover）。

12.7 淘汰赛中谁会遇上谁

题目

STEP II Specimen paper，第 16 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/spec-s2-q16`

引子

温布顿网球锦标赛（Wimbledon）有 128 名选手参加抽签。随机选出其中两人：他们在赛事中某个阶段相遇的机会有多大？你也许会尝试把每一轮相遇的概率（probability）加起来，而每一项都是“双方晋级”与“被抽在一起”这些机会的乘积。这样做是可行的，但有一个好得多的方法。数一数比赛的场数：每场比赛都由一对选手对赛，而比赛恰好有 127 场。根据对称性（symmetry），每一对选手成为其中任何一场的对赛双方的可能性都相同。因此所求的机会就是 127 除以选手配对的总数。

这道题是“数对了东西，概率问题便迎刃而解”的绝佳例子。答案 $\frac{1}{2^{n-1}}$ 只需一行便可得出。

动笔之前

想一想 有 2^n 名选手的淘汰赛（knockout tournament）共要进行多少场比赛？（每场比赛淘汰一名选手。）

每场比赛恰好淘汰一名选手，而除了一人之外所有选手都被淘汰，所以共有 $2^n - 1$ 场比赛。

逐步解题

阶段 1 第一轮

把选出的两名选手称为 X 和 Y。在第一轮，X 的对手是其余 2^n-1 名选手中任何一人的可能性都相同，所以

$$\mathrm{P}(\text{相遇于轮次1}) = \frac{1}{2^n-1}.$$

阶段 2 关键的对称性

> ? **想一想** 每一轮的抽签都是随机的，而每场比赛都相当于掷一次公平的硬币。这个过程对待某一对选手的方式，会否与对待另一对有所不同？

不会。由于抽签是随机的，而且每场比赛的胜负机会都是 50–50，把选手重新标号并不会改变任何概率。所以任何一场指定的比赛（例如决赛）由 $\binom{2^n}{2}$ 对选手中任何一对对赛的可能性都相同。

阶段 3 决赛

决赛是一场特定的比赛，所以它由 X 和 Y 对赛的概率是

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

（验算：两人都必须打进决赛，这需要他们分处不同的半区，并且各自胜出 $n-1$ 场比赛。）

阶段 4 在任何阶段

> ? **想一想** X 和 Y 最多只能相遇一次。这一点怎样让你可以把各场比赛的概率相加？

设 I_m 在第 m 场比赛由 X 和 Y 对赛时取值 1。在 2^n-1 场比赛中，每一场都有 $\mathrm{P}(I_m=1)=\binom{2^n}{2}^{-1}$，而这些事件是互斥（mutually exclusive）的（他们不可能相遇两次），所以

$$\mathrm{P}(\text{相遇}) = \frac{2^n-1}{\binom{2^n}{2}} = \frac{2(2^n-1)}{2^n(2^n-1)} = \frac{1}{2^{n-1}}. \qquad \blacksquare$$

以温布顿为例，$n=7$，两名指定选手相遇的概率是 $\frac{1}{64}$。作为验算，在 2^n-1 场比赛中，第一轮占了 2^{n-1} 场，由此得出 $\frac{2^{n-1}}{\binom{2^n}{2}} = \frac{1}{2^n-1}$，与第 1 阶段的结果一致。

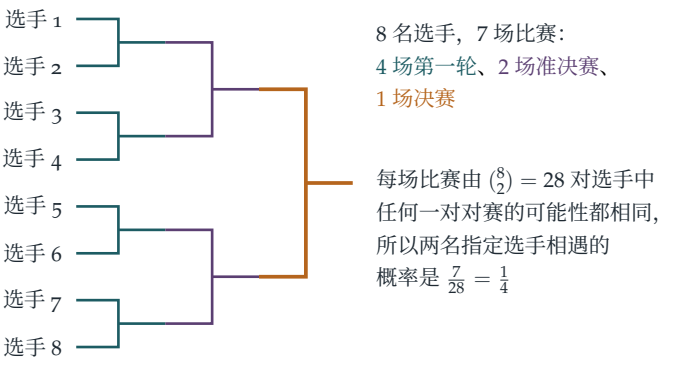


图 12.7.1 $2^3 = 8$ 名选手的淘汰赛（图中画成固定的赛程表，但题目中每一轮都会重新随机抽签）。比赛共有 7 场。根据对称性，每一场由 28 对选手中任何一对对赛的可能性都相同，所以选出的两名选手相遇的概率是 $\frac{7}{28} = \frac{1}{4} = \frac{1}{2^{n-1}}$。

回顾

- **数比赛的场数。**比赛的场数 $2^n - 1$ 是关键的事实。
- **各对选手之间的对称性。**每一对选手成为任何一场指定比赛的对赛双方的可能性都相同。
- **对各场比赛的线性性质。**由于两名选手最多相遇一次，相遇的概率就是相遇次数的期望值，也就是对各场比赛求和。

背后的数学

期望次数

第 4 阶段的论证其实是一个期望值（expectation）的计算：X 和 Y 相遇的次数是若干个指示变量（indicator）之和，每场比赛对应一个；而它的期望值（由于它只取 0 或 1，所以等于相遇的概率）就是各个指示变量的期望值之和。这就是期望值的线性性质（linearity of expectation），在 Section 9.11 和 Section 12.4 中都曾用到。它最大的好处是完全不必理会各轮之间错综复杂的相依关系。

设有种子的赛事

现实中的赛事设有种子（seeded）：最强的选手会被编排在适当的位置，使他们要到后段轮次才可能相遇。而且现实中的比赛也不是 50–50 的。这两点都破坏了对称性，于是计算两名指定选手相遇的机会便成了一项货真价实的运算。为体育运动建立模型的统计学

家正是运用这类计算来预测赛事，当中每场比赛的概率都是根据选手的评分估计出来的。对称的情况则是一条基线，用来衡量各种现实因素的影响。

思维技巧

数对的东西

只要选对了对象，概率往往可以用（有利对象的数目）÷（等可能对象的数目）求得。在这里，对象就是比赛和选手配对。

选择数甚么，正是组合概率（combinatorial probability）的艺术。在大学里，同一个想法会以算两次（double counting）的形式出现（Section 7.9），也会出现在概率方法（probabilistic method）之中：在那里，某种结构的期望数目可以用来证明它存在。

延伸练习

1 **逐轮计算。**直接证明 X 和 Y 在第 k 轮相遇的概率是 $\frac{2^{n-k}}{\binom{2^n}{2}}$，并验证这些概率加起来等于 $\frac{1}{2^{n-1}}$。

2 **固定的赛程表。**假如赛程表改为预先固定（如图所示），但所有选手都被随机编入其中，答案会否改变？

3 **一名强手。**假设选手 X 每场比赛都胜出。现在 X 与 Y 相遇的概率是多少？

4 **联赛。**在一个有 $2n$ 支球队的联赛中，每两队都对赛一次，共有多少场比赛？相应的“相遇概率”是甚么？

延伸阅读

- **Frederick Mosteller, *Fifty Challenging Problems in Probability***（Dover），书中收录了有关淘汰赛的问题。
- **David Spiegelhalter, *The Art of Statistics***（Pelican），讲述建立模型与预测。
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***（CRC Press；可于网上免费阅读），讲述指示变量与对称性。

12.8 Penney 游戏

题目

STEP III 2001，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/01-s3-q13`

引子

你和朋友各自选定一个由三次掷硬币结果组成的序列，然后不断掷硬币，直至其中一个序列出现为止。在任何一个时刻，每个三次投掷的序列出现的机会都相同，那么无论选哪一个，理应同样好吧？完全不是。无论朋友选哪个序列，你都可以选出一个胜过它的序列，胜算（odds）有时更高达 7 比 1。

这就是 *Penney 游戏*（Penney's game），由 Walter Penney 于 1969 年发明。本题要证明 A 胜 B，B 胜 C，C 胜 D，而 D 又胜 A：这是一个循环，就像包剪揼一样。“胜过”这种关系并不具传递性（transitive）。长远而言，后选的一方总能胜出。

动笔之前

想一想 假设至今的投掷结果以 TT 结尾，而且仍未有人胜出。THH 有没有可能比 TTH 先出现？

从 TT 出发，之后每掷出一个 T，最后两次投掷仍然是 TT，而第一个 H 一出现就凑成 TTH。要 THH 出现，必须先有 TH，而 TH 只能由一个 T 后接一个 H 而来，但这个 H 已经凑成了 TTH。所以 TT 一旦出现，A 必胜。

逐步解题

阶段 1 状态

想一想 两个序列都不可能在两次投掷内完成。掷了两次之后,从每个状态(state)出发，下一次投掷会带来甚么结果?

由于游戏只取决于最后两次投掷，我们可以追踪状态（即最后两次投掷的结果）：

- 从 TT 出发：A 胜（上面已证），所以 $p_{\mathrm{TT}} = 1$。
- 从 HT 出发：下一次投掷会得出 TH 或 TT，所以 $p_{\mathrm{HT}} = \frac{1}{2}p_{\mathrm{TH}} + \frac{1}{2}p_{\mathrm{TT}}$。
- 从 TH 出发：掷出 H 便凑成 THH（B 胜）；掷出 T 则得出 HT。所以 $p_{\mathrm{TH}} = \frac{1}{2}\cdot 0 + \frac{1}{2}p_{\mathrm{HT}}$。
- 从 HH 出发：掷出 H 仍是 HH；掷出 T 则得出 HT。所以 $p_{\mathrm{HH}} = \frac{1}{2}p_{\mathrm{HH}} + \frac{1}{2}p_{\mathrm{HT}}$，由此得 $p_{\mathrm{HH}} = p_{\mathrm{HT}}$。

阶段 2 求解

把 $p_{\mathrm{TH}} = \frac{1}{2}p_{\mathrm{HT}}$ 代入 HT 的方程：$p_{\mathrm{HT}} = \frac{1}{4}p_{\mathrm{HT}} + \frac{1}{2}$，所以 $p_{\mathrm{HT}} = \frac{2}{3}$、$p_{\mathrm{HH}} = \frac{2}{3}$ 及 $p_{\mathrm{TH}} = \frac{1}{3}$。四个起始状态的概率（probability）各为 $\frac{1}{4}$，所以

$$\mathrm{P}(A \text{ 胜}) = \tfrac{1}{4}\left(\tfrac{2}{3} + \tfrac{2}{3} + \tfrac{1}{3} + 1\right) = \tfrac{2}{3}. \qquad \blacksquare$$

阶段 3 B 对 C

想一想 C 想要 HHT，B 想要 THH。C 在甚么情况下胜出?（留意第一个 HH 之前紧接着的必定是甚么。）

如果头两次投掷是 HH，那么其后第一个 T 便凑成 HHT，THH 来不及出现：C 胜。否则，在 HH 首次出现之前的某处必有一个 T，因此第一个 HH 的前一次投掷必定是 T：THH 就在那一刻出现，早于 HHT（HHT 还需要在 HH 之后再掷出一个 T）。所以只有当游戏以 HH 开始时 C 才会胜出：

$$\mathrm{P}(B \text{ 胜}) = \tfrac{3}{4}. \qquad \blacksquare$$

阶段 4 循环

用同样的方法（或“背后的数学”一节的捷径）可以证明 C（HHT）胜 D（HTT）的概率是 $\frac{2}{3}$，而 D（HTT）胜 A（TTH）的概率是 $\frac{3}{4}$；后者正是第 3 阶段把正面与反面对调后的镜像。所以

$$A \text{ 胜} B, \qquad B \text{ 胜} C, \qquad C \text{ 胜} D, \qquad D \text{ 胜} A. \qquad \blacksquare$$

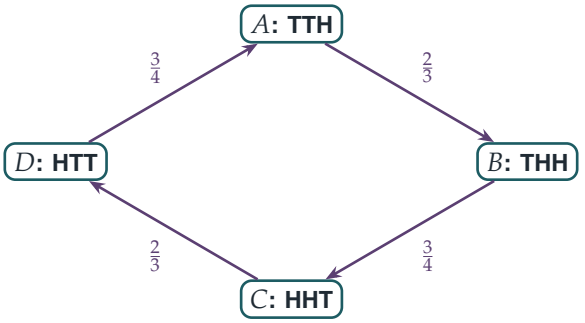


图 12.8.1 Penney 循环。由 X 指向 Y 的箭头表示 X 的序列胜过 Y 的序列，箭头旁为相应的概率。每个序列都被另一个序列击败，所以后选的一方总能挑出胜算较高的序列。

回顾

- **选对状态。**只有最后两次投掷才有影响，所以四个状态、四条方程便概括了整个游戏。
- **首步分析（first-step analysis）。**每条方程都是以下一次投掷为条件：从某状态出发的概率，就是下一次投掷所到达的各个状态的概率的平均值。
- **找出决定性的事件。**在第 3 阶段，“第一个 HH 之前是 T”这一点无须列方程便决定了胜负。

背后的数学

为甚么有些序列较好

在任何一个位置，所有三次投掷的序列出现的机会都相同，而每个序列的期望等待时间（expected waiting time）也不难计算。但各序列的等待时间并不相同：TTH 平均要等 8 次投掷，TTT 却要等 14 次，因为与自身重叠的序列（例如 TTT，只差一步的 TTH 会令已累积的两次投掷前功尽废）出现得较慢。而在两个序列的竞赛中，关键在于每个序列与对方如何重叠。THH 胜 HHT，是因为 HHT 每次出现（除了在最开头之外），前面都有一个 T，这表示 THH 刚刚已经出现。后选一方的窍门总是一样：取对手序列的头两个符号，再在前面加上与对手中间符号相反的那个符号。

康威公式（Conway's formula）

约翰·康威发现了一条计算胜算的绝妙公式。对于长度为 n 的两个序列 X 和 Y，定义相关数（correlation）XY 为一个 n 位的二进制数：由左至右，若 X 中由第 i 个符号开始的部

分与 Y 开头相同长度的部分相同，则第 i 位为 1。那么 Y 比 X 先出现的胜算是

$$(XX - XY) : (YY - YX).$$

取 $X = \text{HHT}$ 及 $Y = \text{THH}$: $XX = 100_2 = 4$（只有整个 HHT 与自身吻合）; $XY = 001_2 = 1$（只有 HHT 末尾的 T 与 THH 的开头吻合）; $YY = 100_2 = 4$; 而 $YX = 011_2 = 3$（THH 末尾的 HH 和 H 都与 HHT 的开头吻合）。所以 THH 先出现的胜算是 $(4-1) : (4-3) = 3 : 1$，即第 3 阶段的 $\frac{3}{4}$。康威公式适用于任何长度、任何字母表的序列，其证明用到鞅（martingale）的理论：想像有一间赌场，赌客就序列的出现下注，再利用公平游戏的期望利润为零这一事实。

Penney 与加德纳（Gardner）

Walter Penney 于 1969 年在 *Journal of Recreational Mathematics* 的一篇短文中介绍了这个游戏。马丁·加德纳于 1974 年在他的 *Scientific American* 专栏中介绍了它，使它广为人知，文中还包括康威公式，这条公式是康威发现的，但他一直没有发表。这类非传递关系（埃弗龙骰子（Efron's dice）是另一个例子：骰子 A 胜 B，B 胜 C，C 胜 D，而 D 胜 A）一再令人惊讶，在投票理论中也很重要，因为多数人的偏好可以形成循环：这就是孔多塞悖论（Condorcet's paradox）。

思维技巧

首步分析

若一个过程在各状态之间转移，要求出与它有关的概率或期望值（*expectation*），可以第一步为条件。这样每个状态都给出一条线性方程。

这是处理马可夫链（Markov chain）的基本技巧，你会在 Section 12.16 以及大学概率课程中一再遇到。同样的方法可以求出先到达某一状态而非另一状态的概率、被吸收前的期望时间，以及赌博问题中的破产概率。

延伸练习

1 **四人同场。**试做 STEP II 2015，第 12 题（`step.maths.org/questions/15-s2-q12`）：四名玩家同时参与的 Penney 游戏。

2 **等待时间。**证明直至 HH 首次出现的期望投掷次数是 6，而直至 HT 首次出现的则是 4。（使用首步分析。）

3 **最佳对策。**对于八个长度为三的序列，逐一找出以最大概率胜过它的序列。哪个序列最容易被击败?

4 **康威公式。**用康威公式验证 HHT 胜 HTT 的概率是 $\frac{2}{3}$。

延伸阅读

- **Martin Gardner, *Time Travel and Other Mathematical Bewilderments***（W. H. Freeman）。书中包括 Penney 游戏和康威公式。
- **Ronald Graham, Donald Knuth & Oren Patashnik, *Concrete Mathematics***（Addison-Wesley），第 8.4 节。以生成函数（generating function）处理 Penney 游戏。
- **David Williams, *Probability with Martingales***（Cambridge University Press）。以鞅给出的证明，留待日后再读。

12.9 用不公平的硬币作公平的决定

题目

STEP I 2019，第 11 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 I 试卷：

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

引子

你要作出一个公平（fair）的决定，但手上唯一的硬币是弯的，而且你不知道它偏差（bias）有多大。这样还能公平地作出选择吗？冯纽曼（John von Neumann）在 1951 年给出的答案是：掷两次。若结果是 HT，就选酒店；若是 TH，就选茶室；若是 HH 或 TT，就再掷两次。由于 HT 和 TH 的概率（probability）都是 pq，所以不论 p 是多少，这个决定都完全公平。

本题的程序刚好相反：它在出现 HH 和 TT 时作决定，这样做很快，却不公平（选酒店的概率是 $\frac{p^2}{p^2+q^2}$）。本题要证明的是，不论硬币如何，这个程序都不会拖得很久。比较这两个程序，可以看到产生随机性时的一个基本取舍（trade-off）：速度与公平之间的取舍。

动笔之前

想一想 在一轮之中，作出决定的概率是多少？没有决定的概率又是多少？$2pq$ 最大可以取甚么值？

若两枚硬币结果相同，便作出决定，概率是 p^2+q^2；否则他们继续下去，概率是 $2pq$。由于 $pq = p(1-p) \le \frac{1}{4}$（当 $p = \frac{1}{2}$ 时等号成立），所以 $2pq \le \frac{1}{2}$。

逐步解题

阶段 1 在第 n 轮作出决定

各轮之间互相独立（independent）。要在第 n 轮作出决定，须先有 $n-1$ 轮结果不同，然后一轮结果相同：

$$\mathrm{P}(\text{作出决定的轮次为}n) = (2pq)^{n-1}(p^2+q^2). \qquad \blacksquare$$

阶段 2 在第 n 轮或之前作出决定

? **想一想** 在首 n 轮中没有作出决定的概率是多少?

n 轮之内没有决定的概率是 $(2pq)^n \le \left(\frac{1}{2}\right)^n$，所以不论 p 是多少，在第 n 轮或之前作出决定的概率至少是 $1-\frac{1}{2^n}$。 ■

阶段 3 三个人

? **想一想** 第一轮出现两个正面和一个反面之后，第三个人把反面翻成正面。第二轮要发生甚么事才会有决定？如果第一轮是两个反面和一个正面呢?

第一轮作出决定的概率是 p^3+q^3。否则，有两枚硬币结果相同，另一枚不同。若多数是正面（概率为 $3p^2q$），那枚不同的硬币会被翻成正面，而第二轮要作出决定，就需要另外两人都掷出正面：概率是 p^2（此时已不可能全是反面）。同理，若多数是反面（概率为 $3pq^2$），要作出决定就需要两个反面：q^2。所以

$$\mathrm{P}(\text{作出决定的轮次不迟于}2) = p^3+q^3+3p^2q\cdot p^2+3pq^2\cdot q^2.$$

? **想一想** 把所有东西用 $s=pq$ 表示。（利用 $p+q=1$。）

我们有 $p^3+q^3 = (p+q)^3-3pq(p+q) = 1-3s$，以及 $3p^4q+3pq^4 = 3s(p^3+q^3) = 3s(1-3s)$。所以

$$\mathrm{P} = 1-3s+3s-9s^2 = 1-9s^2 \ge 1-\frac{9}{16} = \frac{7}{16},$$

因为 $s \le \frac{1}{4}$。 ■

回顾

- **补事件（complement）。**“在第 n 轮或之前至少发生一次”最容易的算法是 1−（“n 轮之内从未发生”）。
- **为最坏情况定界。**未知的 p 是靠不等式（inequality）$pq \le \frac{1}{4}$ 来处理的。
- **对称函数（symmetric functions）。**对 p 和 q 对称的式子，用 $p+q=1$ 和 pq 来表示时便会简化。

背后的数学

冯纽曼的妙计

题目中的程序很快，但有偏差。冯纽曼的程序则是公平的：掷两次，出现 HT 或 TH 就作决定，出现 HH 或 TT 就重复。每次尝试成功的概率是 $2pq$，所以平均需要 $\frac{1}{2pq}$ 次尝试；当硬币偏差很大时，这会很慢（图 12.9.1）。题目的程序每轮作出决定的概率是 $p^2+q^2 \geq \frac{1}{2}$，但当 $p > \frac{1}{2}$ 时，它的结果偏向酒店。

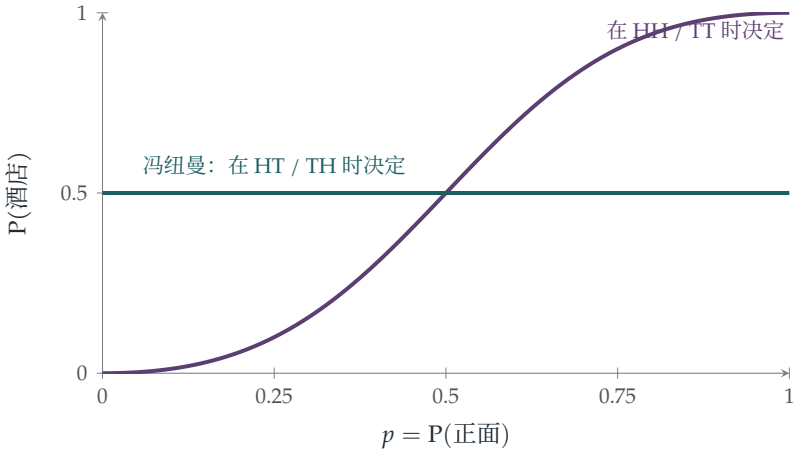


图 12.9.1 选中酒店的概率，对应硬币的偏差。在两枚硬币结果相同时作决定（紫红色）很快，但偏向较可能出现的一面。冯纽曼的程序（蓝绿色）对每一个 p 都完全公平，代价是当硬币偏差很大时需时较长。

提取随机性

冯纽曼的妙计是第一个*随机性提取器*（randomness extractor）：一种把不完美的随机性变成完美随机性的程序。它浪费很多（每两次投掷最多只得到一个公平的位元，而且往往一个也得不到）；更好的方法会把被弃掉的 HH 和 TT 重新利用，从而提取硬币所含的几乎全部随机性，而这随机性的多少是用硬币的*熵*（entropy）来量度的，即每次投掷 $-p\log_2 p - q\log_2 q$ 个位元。提取器用于电脑的硬件随机数产生器：这些产生器收集不可预测但有偏差的物理杂讯，并必须把它变成无偏差的位元，供密码学（cryptography）使用。

冯纽曼与随机数

冯纽曼在 1951 年的论文“Various techniques used in connection with random digits”中描述了这个妙计。当时他和同事正在为设计核武器而进行一些最早期的电脑模拟（即

“蒙地卡罗方法”（Monte Carlo method））。在同一篇论文中，他就用算术方法产生随机数说了一句名言：“任何人若考虑用算术方法来产生随机数字，当然就是身处罪恶之中。”

思维技巧

利用对称性消去未知量

当某个参数未知时，找出那些不论参数取何值、概率都相等的事件，并以它们作为决定的依据。

HT 和 TH 有相同的概率 pq，所以比较这两者就把 p 完全消去了。在大学里，同样的想法在统计学中以枢轴量（pivotal quantities）的形式出现，也出现于实验设计之中：在那里，随机化（randomisation）和对称性会把未知的效应抵消掉。

延伸练习

1 **轮数的期望值。**求第 (i) 部分的程序所需轮数的期望值（expected number），以及冯纽曼的程序所需轮数的期望值。对哪些 p 而言，冯纽曼的程序较慢?

2 **公平的三选一。**利用一枚有偏差的硬币，设计一个程序，从三个选项中选出一个，使每个选项的概率都恰好是 $\frac{1}{3}$。

3 **用公平硬币造出有偏差的骰子。**利用一枚公平的硬币，设计一个程序，模拟一枚掷出正面的概率恰好是 $\frac{1}{3}$ 的硬币。所需投掷次数的期望值是多少?

4 **再看三个人。**在第 (ii) 部分中，他们去酒店的概率是多少？请以 p 的函数表示。当 $p = \frac{1}{2}$ 时，这个程序公平吗?

延伸阅读

- **John von Neumann，“Various techniques used in connection with random digits”**，*National Bureau of Standards Applied Mathematics Series* 12 (1951)。篇幅短小而著名。
- **Donald Knuth，*The Art of Computer Programming*, Vol. 2**（Addison-Wesley），第 3 章，论随机数。
- **Michael Mitzenmacher & Eli Upfal，*Probability and Computing***（Cambridge University

Press），论提取随机性与熵。

12.10 把绳头系起来

题目

STEP III 2008，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/08-s3-q13`

引子

把一百条绳子放进袋中，只让绳头露在外面。随机选两个绳头绑在一起，如此重复，直至再没有松开的绳头为止。最后你会得到一堆纠缠的绳圈。究竟有多少个？一百条绳子原则上可以结成一百个小环，也可以结成一个巨大的圈。出人意表的是，绳环数目的期望值（expected number）只有约 3.3；即使有四万条绳子，也不过约 6。

只要看出一个关键之处，问题便几乎不费吹灰之力：每打一个结之后，不论刚才发生了甚么事，箱中都只是少了一条绳子。这个过程会“忘记”自己的历史，于是绳环的数目就变成一连串独立（independent）掷硬币结果之和，只是每次的概率（probability）各有不同。

动笔之前

想一想 把两个绳头绑在一起之后，还剩下多少条绳子（结成环的不再算作一“条”）？这是否取决于有没有结成环？

如果两个绳头属于同一条绳子，这条绳子便成为一个环，余下 $n-1$ 条。如果两个绳头属于不同的绳子，这两条绳子便接成较长的一条，同样余下 $n-1$ 条。无论是哪一种情况，我们都剩下 $n-1$ 条绳子，每条各有两个绳头。

逐步解题

阶段 1 第一个结

绳头共有 $2n$ 个。先选定一个；与它相绑的绳头等可能地是其余 $2n-1$ 个绳头中的任何一个，而当中恰好有一个属于同一条绳子。因此第一个结结成环的概率是 $\frac{1}{2n-1}$，而第一步所产生的绳环数目的期望值就是 $\frac{1}{2n-1}$。

阶段 2 所有的结

> ? **想一想** 根据上面的观察，打第二个结之前的情况是怎样的？第三个呢？设 I_k 为指示变量，表示在剩下 k 条绳子时所打的结是否结成环。各个 $\mathrm{E}(I_k)$ 是多少？

每次打结之前，箱中有 k 条绳子，其中 $k = n, n-1, \ldots, 1$，而这个结结成环的概率是 $\frac{1}{2k-1}$。（最后一个结在只剩一条绳子时打，必定结成环。）绳环的总数是 $R = I_n + I_{n-1} + \cdots + I_1$，所以

$$\mathrm{E}(R) = \sum_{k=1}^{n} \frac{1}{2k-1} = 1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n-1}. \qquad \blacksquare$$

阶段 3 方差

> ? **想一想** 各个指示变量（indicator）I_k 是否独立？（某一步结成环的概率，是否取决于之前发生过的事？）

在每一步，不论之前发生过甚么事，结成环的概率都是 $\frac{1}{2k-1}$，因为当时的格局永远是“k 条绳子”。所以各个 I_k 互相独立，方差（variance）可以相加：

$$\mathrm{Var}(R) = \sum_{k=1}^{n} \frac{1}{2k-1}\left(1 - \frac{1}{2k-1}\right) = \sum_{k=1}^{n} \frac{2k-2}{(2k-1)^2}. \qquad \blacksquare$$

阶段 4 四万条绳子

> ? **想一想** 试用调和数（harmonic numbers）表示直至 $2n-1$ 为止所有奇数的倒数之和。

把直至 $2n$ 为止所有正整数的倒数之和，减去其中偶数的倒数之和，便得

$$\sum_{k=1}^{n} \frac{1}{2k-1} = H_{2n} - \tfrac{1}{2}H_n \approx \ln(2n) - \tfrac{1}{2}\ln n = \tfrac{1}{2}\ln(4n).$$

当 $n = 40\,000$ 时，$4n = 160\,000 = 20^4$，所以 $\mathrm{E}(R) \approx \frac{1}{2} \cdot 4 \ln 20 \approx 6$。$\blacksquare$（准确值是 6.28；两者之差来自欧拉常数（Euler's constant），因为 $H_n \approx \ln n + 0.577$。）

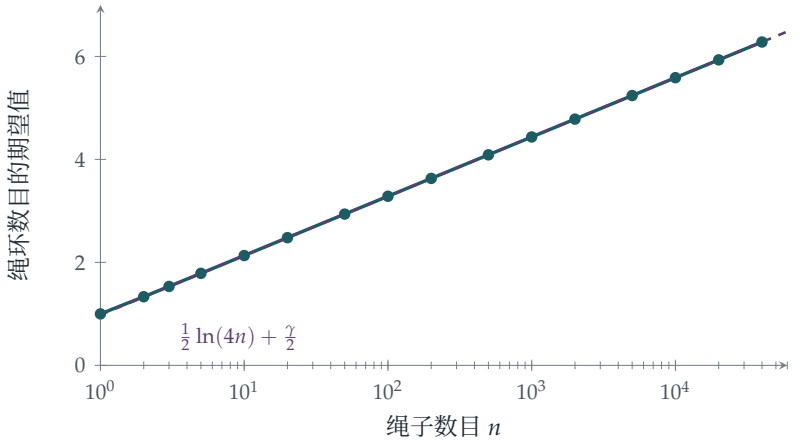


图 12.10.1 绳环数目的期望值与绳子数目的关系，横轴采用对数尺度。准确值（蓝绿色）落在直线 $\frac{1}{2}\ln(4n)+\frac{\gamma}{2}$（紫红色虚线）上，其中 $\gamma \approx 0.577$ 是欧拉常数。把绳子数目乘以 10，绳环只会多出约 1.15 个。

回顾

- **找出过程忘记了甚么。**每打一个结之后，不论过去的经过如何，箱中都是“k 条绳子”。
- **又是指示变量。**绳环的数目是指示变量之和，每个结对应一个。
- **有独立性便有方差。**由于每一步的概率都不取决于过去，各个指示变量互相独立。

背后的数学

对数增长

绳环数目的期望值以 $\frac{1}{2}\ln n$ 的速度增长，而方差与平均值相当接近（两者都是一些很小的概率之和）。根据中心极限定理（central limit theorem）（Section 12.24），当 n 很大时，绳环数目近似服从正态分布（normal distribution），平均值和方差都约为 $\frac{1}{2}\ln(4n)$。若有 40 000 条绳子，你通常会得到 2 至 11 个环。

随机置换的循环

同一个对数也出现在 n 件物件的随机置换（random permutation）的循环（cycle）数目之中：平均有 $H_n \approx \ln n$ 个（绳子问题是它的近亲，对应的是 $\frac{1}{2}\ln n$）。随机置换的循环很少：通常有一个巨大的循环包含了大部分元素，而最大循环所占元素的比例，平均约为 0.62，即 Golomb–Dickman 常数。随机地把绳子绑在一起，情况也是一样：大部分绳子最终都落在一两个长圈之中。这些事实属于随机组合结构（random combinatorial structures）

的研究范围，而这门学问又与数论（number theory）有关（一个随机整数的质因数，其表现就像一个随机置换的循环）。

随机纽结与聚合物

绑绳子问题是某些物理系统的简化模型。溶液中的长链聚合物分子会首尾相接，形成环和链，而成环的统计规律也遵从类似的法则。在纽结理论（knot theory）中，数学家问的是一个更难的问题：一个随机的闭合圈有多大机会打了结？对于很长的随机圈，答案趋向 1，这个结果由 De Witt Sumners 和 Stuart Whittington 于 1988 年证明。

思维技巧

找出无记忆的结构

如果每一步之后，系统所处的状态只取决于已经走了多少步，那么各步的表现便互相独立，期望值和方差都可以直接相加。

这正是优惠券收集问题（coupon collector）（Section 9.11）背后的想法，当中每个阶段都是一次全新的几何等待；它也在马可夫链（Markov chains）的理论中再次出现，在那里只有当前的状态才起作用。

延伸练习

1 **小的情况。**对于 $n = 2$，列出所有可能情况，并验证绳环数目的期望值是 $\frac{4}{3}$。

2 **一个大圈。**证明整个过程最终得到单一个包含所有绳子的圈的概率是 $\prod_{k=2}^{n}\left(1 - \frac{1}{2k-1}\right)$，并证明当 n 很大时，这个概率约为 $\sqrt{\frac{\pi}{4n}}$。（提示：找出这个乘积与 $\binom{2n}{n}$ 的关系。）

3 **置换的循环。**证明 n 件物件的随机置换的循环数目的期望值是 H_n，方法是逐个元素地构造这个置换（即“中国餐馆”构造法）。

4 **方差。**证明 $\mathrm{Var}(R) = \mathrm{E}(R) - \sum_{k=1}^{n} \frac{1}{(2k-1)^2}$，并证明被减去的那个和趋向 $\frac{\pi^2}{8}$。

延伸阅读

- **Peter Winkler, *Mathematical Puzzles: A Connoisseur's Collection***（A K Peters），书中收录了这道绑绳子谜题。

- **Richard Arratia, Andrew Barbour & Simon Tavaré, *Logarithmic Combinatorial Structures***（European Mathematical Society）。讲述循环、质数与随机结构，可留待日后阅读。
- **Colin Adams, *The Knot Book***（American Mathematical Society）。纽结理论入门，内容包括随机纽结。

12.11 邮票、杯子与 e

题目

STEP I 1996，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/96-s1-q13`

引子

有这样一种古怪的卖邮票方法：买家不断猜邮票藏在哪一个杯子下面，每一轮多加一个杯子；到他第一次猜错时，桌上有多少个杯子，他就要付多少英镑。平均而言，他要付 £e，约为 £2.72。

e 这个数在概率（probability）中出现之频繁，简直不可思议：没有人拿回自己衬衫的机会（Section 9.8）、秘书问题（secretary problem，Section 12.12）、泊松分布（Poisson distribution）。本题也许是它众多现身之中最简单的一次，而且原来它与另一道著名谜题有着完全相同的分布（distribution）：要把多少个介乎 0 与 1 之间的随机数相加，总和才会超过 1？平均而言，是 e 个。

动笔之前

想一想 若 $N = 4$，必须发生甚么事？概率是多少？

Jim 必须在有 2 个杯子时猜中（概率为 $\frac{1}{2}$），在有 3 个杯子时也猜中（$\frac{1}{3}$），然后在有 4 个杯子时猜错（$\frac{3}{4}$）。所以 $\mathrm{P}(N = 4) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{3}{4} = \frac{3}{4!} = \frac{1}{8}$。

逐步解题

阶段 1 **分布**

若 $N = k$（其中 $k \geq 2$），Jim 在有 $2, 3, \ldots, k-1$ 个杯子时都猜中，而在有 k 个杯子时猜错：

$$\mathrm{P}(N = k) = \frac{1}{2} \cdot \frac{1}{3} \cdots \frac{1}{k-1} \cdot \frac{k-1}{k} = \frac{k-1}{k!}.$$

（验算：$\frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \cdots = \sum_{k\geq 2} \left(\frac{1}{(k-1)!} - \frac{1}{k!} \right) = 1$，这是一个裂项和（telescoping sum）。）

阶段 2 期望值

? **想一想** 化简 $k \cdot \frac{k-1}{k!}$。你会得到甚么级数（series）？

$$\mathrm{E}(N) = \sum_{k\geq 2} \frac{k(k-1)}{k!} = \sum_{k\geq 2} \frac{1}{(k-2)!} = \frac{1}{0!} + \frac{1}{1!} + \cdots = e. \qquad \blacksquare$$

阶段 3 方差

? **想一想** 求 $\mathrm{E}(N^2)$：可写成 $k^2 = k(k-1) + k$，或直接化简 $\frac{k^2(k-1)}{k!}$。

$$\mathrm{E}(N^2) = \sum_{k\geq 2} \frac{k^2(k-1)}{k!} = \sum_{k\geq 2} \frac{k}{(k-2)!} = \sum_{j\geq 0} \frac{j+2}{j!} = \sum_{j\geq 1} \frac{1}{(j-1)!} + 2\sum_{j\geq 0} \frac{1}{j!} = e + 2e = 3e.$$

所以 $\mathrm{Var}(N) = 3e - e^2 = e(3-e) \approx 0.766$。 $\blacksquare$

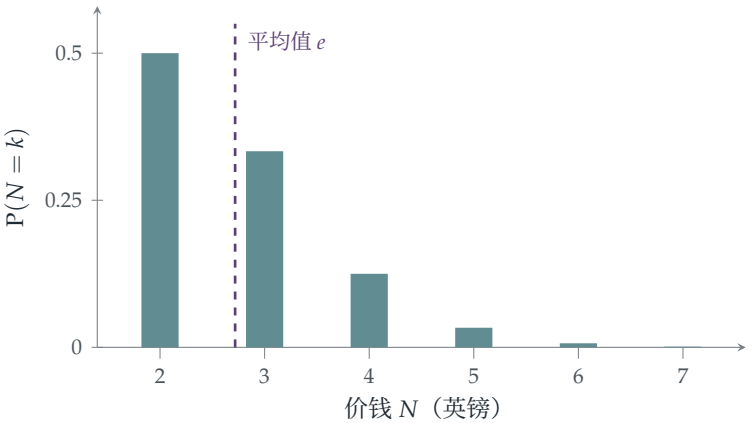


图 12.11.1 价钱的分布，$\mathrm{P}(N = k) = \frac{k-1}{k!}$。有一半机会 Jim 只需付 £2；那条又长又细的尾巴把平均值拉高至 £$e \approx$ £2.72。

回顾

▶ **概率的乘积。**一连串独立（independent）的成功之后接着一次失败，所得的乘积可化

简为 $\frac{k-1}{k!}$。

- **约去阶乘。**乘以 k 或 k^2 再约简，那些和式就变成了指数级数（exponential series）。
- **以平移求 $\mathrm{E}(N^2)$。**以 $j=k-2$ 重新编排下标，和式便分拆成两份 e。

背后的数学

换了装扮的同一个分布

从 $[0,1]$ 中均匀（uniformly）地随机选取数 $U_1, U_2, \ldots$，并设 M 为令它们的总和超过 1 所需的个数。首 k 个数之和小于 1 的概率，就是区域 $\{u_1+\cdots+u_k<1, u_i\geq 0\}$ 的体积，即 $\frac{1}{k!}$（面积为 $\frac{1}{2}$ 的三角形、体积为 $\frac{1}{6}$ 的四面体，如此类推）。所以

$$\mathrm{P}(M>k)=\frac{1}{k!}, \qquad \mathrm{P}(M=k)=\frac{1}{(k-1)!}-\frac{1}{k!}=\frac{k-1}{k!},$$

正正就是 N 的分布。在两个问题中都有 $\mathrm{P}(N>k)=\frac{1}{k!}$：就杯子而言，这是连续猜中 $k-1$ 次的概率。利用 $\mathrm{E}(N)=\sum_{k\geq 0}\mathrm{P}(N>k)$（Section 9.9 中的一条公式），一行便得出 $\mathrm{E}(N)=\sum\frac{1}{k!}=e$。

为何 e 不断出现

每当概率中含有 $\frac{1}{k!}$，不论它来自 k 件物件按某个特定次序排列、k 个概率依次为 $\frac{1}{2}, \frac{1}{3}, \ldots$ 的相继事件，还是单纯形（simplex）的体积，对 k 求和都会产生指数级数。错排（derangements，Section 9.8）给出 $\sum\frac{(-1)^k}{k!}=e^{-1}$；这里我们则有 $\sum\frac{1}{k!}=e$。至于形如 $\left(1-\frac{1}{n}\right)^n\to e^{-1}$ 的极限（limit），它们源自大量独立而不大可能发生的事件，正好解释了秘书问题和泊松分布中的 e。

蒙地卡罗 e

“总和超过 1”这个问题提供了一个以实验估计 e 的方法：产生随机数，数一数需要多少个才超过 1，重复多次，然后取平均值。它是蒲丰投针（Buffon's needle，Section 12.20）的近亲，后者用来估计 π。这类蒙地卡罗（Monte Carlo）方法以赌场命名，利用随机抽样来估计难以直接计算的量；如今它们广泛应用于科学的各个范畴，由粒子物理学到金融，主要用于计算高维积分。

思维技巧

认出级数

当期望值是一个含有阶乘的和式时，把各项化简，直至出现一个标准级数（指数级数、几何级数、二项级数）。

在大学，生成函数（generating function）把这种做法系统化：概率生成函数（probability generating function）$\mathrm{E}(t^N)$ 把整个分布包装起来，而它在 $t = 1$ 处的导数则给出平均值和方差。

延伸练习

1 **生成函数。**证明 $\mathrm{E}(t^N) = 1 + (t-1)e^t$，并用它求 $\mathrm{E}(N)$ 和 $\mathrm{E}(N(N-1))$。

2 **尾部公式。**对于取非负整数值的随机变量（random variable），证明 $\mathrm{E}(N) = \sum_{k\geq 0} \mathrm{P}(N > k)$，并用它一行求出 $\mathrm{E}(N)$。

3 **由一个杯子开始。**假如游戏由一个杯子开始（Jim 不可能猜错），分布会有甚么改变？价钱的期望值是多少？

4 **单纯形。**证明 $\{u_1 + u_2 + u_3 < 1, u_i \geq 0\}$ 的体积为 $\frac{1}{6}$，并用归纳法（induction）证明 k 维的情况为 $\frac{1}{k!}$。

延伸阅读

- **Eli Maor, *e: The Story of a Number***（Princeton University Press）。e 的历史，包括它在概率中的各次现身。
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），关于生成函数。
- **Paul Nahin, *Digital Dice***（Princeton University Press）。以蒙地卡罗模拟解决的概率谜题。

12.12 何时停止寻找

题目

STEP I 2001，第 14 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/01-s1-q14`

引子

你正在逐一面试应征者，而且必须即场决定：取录眼前这一位，然后停止；还是永远拒绝这一位，再见下一位。你想要的是最好的那一位。如果太早取录，你很可能错过更好的人；如果等得太久，最好的那一位很可能已经走了。你应该怎样做？

答案是概率（probability）论中最著名的结果之一：先拒绝最初 37% 的应征者，然后取录第一位比这些人都好的应征者。这个做法大约有 37% 的机会成功，而 37% 正是 $\frac{1}{e}$。本题计算其中的基本概率，而与它配对的 STEP II 2006 题目（见练习）则分析这个策略（strategy）最简单的版本：只拒绝第一位应征者。

动笔之前

想一想 在随机次序之下，全体之中最好的应征者落在最先面试的 n 位之中的概率是多少？

最好的应征者落在 N 个位置中任何一个的可能性均等，所以概率是 $\frac{n}{N}$。

逐步解题

阶段 1 最先 n 位中的最佳者就是全体最佳者

最先 n 位中的最佳者是全体最佳者，当且仅当全体最佳者在最先 n 位之中：概率为 $\frac{n}{N}$。■

阶段 2 最佳或次佳

想一想 最先 n 位中的最佳者是全体次佳者，当且仅当次佳者在最先 n 位之中而最佳者不在其中。这个概率是多少?

次佳者在最先 n 位之中的概率是 $\frac{n}{N}$；在此条件下，最佳者在其余 $N-1$ 个位置的其中一个，当中有 $N-n$ 个位置排在最先 n 位之后。所以

$$\mathrm{P}(\text{最佳或次佳}) = \frac{n}{N} + \frac{n}{N}\cdot\frac{N-n}{N-1}. \qquad \blacksquare$$

当 $N=4$，$n=2$ 时：$\frac{1}{2}+\frac{1}{2}\cdot\frac{2}{3}=\frac{5}{6}$。把最好的两位应征者在四个位置中的排法逐一列出（全部 12 种有序排法的可能性均等），除了两位顶尖应征者同时落在第 3 和第 4 位的 2 种排法之外，最先两位中的最佳者都是顶尖两位之一：$\frac{10}{12}=\frac{5}{6}$。

阶段 3 最优策略

想一想 策略：先拒绝最初 r 位应征者，然后取录第一位比他们全部都好的应征者。如果最佳应征者在位置 $k>r$，这个策略在甚么情况下会选中此人?

设最佳应征者在位置 k，这个策略选中此人，当且仅当位置 $r+1,\dots,k-1$ 上没有任何应征者胜过最初 r 位全部的人，即是说，最先 $k-1$ 位应征者中的最佳者在最初 r 位之中。根据第 1 阶段（把 N 换成 $k-1$），这件事的概率是 $\frac{r}{k-1}$。最佳应征者落在每个位置 k 的概率都是 $\frac{1}{N}$，所以

$$\mathrm{P}(\text{成功}) = \frac{1}{N}\sum_{k=r+1}^{N}\frac{r}{k-1} = \frac{r}{N}\left(\frac{1}{r}+\frac{1}{r+1}+\cdots+\frac{1}{N-1}\right).$$

想一想 当 N 很大时，记 $x=\frac{r}{N}$。这个和大约等于 $\ln\frac{N}{r}$。x 取甚么值时 $x\ln\frac{1}{x}$ 达到最大?

成功的概率大约是 $x\ln\frac{1}{x}$。求导可得，$\ln\frac{1}{x}-1=0$ 于 $x=\frac{1}{e}$ 时成立，而最大值是 $\frac{1}{e}\approx 0.368$（图 12.12.1）。

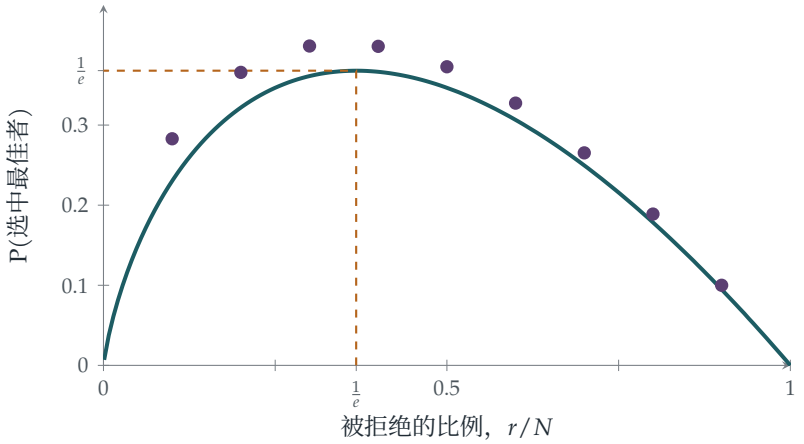


图 12.12.1 若先拒绝占比例 r/N 的最初一批应征者，选中最佳应征者的概率：极限曲线 $x \ln \frac{1}{x}$（蓝绿色），以及 $N = 10$ 时的精确值（紫红色）。最佳策略是拒绝大约 $\frac{1}{e} \approx 37\%$ 的应征者，成功的概率大约是 $\frac{1}{e}$，不论 N 有多大。

回顾

- **位置的对称性。**在随机次序之下，每位应征者落在任何位置的可能性均等，由此立即得出 $\frac{n}{N}$。
- **以最佳者的位置为条件。**计算策略的成功概率时，我们以最佳应征者所在的位置为条件分情况处理。
- **求和变成积分。**当 N 很大时，调和和（harmonic sum）变成了对数，再经一次求最大值便得出 $\frac{1}{e}$。

背后的数学

最优停止

秘书问题（secretary problem）是*最优停止*（optimal stopping）的奠基例子：根据随时间陆续到来的信息，决定何时采取行动，以使期望回报最大。可以证明，在*所有*策略之中（而不只是“先拒绝 r 位，然后取录下一位至今最佳者”这一类），这种阈值规则（threshold rule）是最优的，论证的方法是由最后一位应征者开始向前逆推。同样的数学也决定了何时卖出股票、何时行使金融期权，以及应该花多久去找泊车位。一个值得注意的特点是，在极限之下答案与 N 无关：不论应征者有一百位还是一百万位，你成功的概率都是 $\frac{1}{e}$。

各种变化

如果你只要选中顶尖两位应征者的其中一位便满意，最佳策略大约有 57% 的机会成功。如果你想做到的不是使选中最佳者的机会最大，而是使所选应征者的期望*名次*最优，最优策略便复杂得多，而期望名次趋向大约 3.87（Chow、Moriguti、Robbins 和 Samuels，1964）。如果你有一定的概率可以回头召回已被拒绝的应征者，又或者你知道应征者分数的分布（即“完全信息”问题），答案又会不同。练习中的雪糕题是最简单的情形：只拒绝一个，然后取下一个比它好的。

是谁解决的?

这个问题在 1950 年代流传于数学家之间，马丁·加德纳（Martin Gardner）于 1960 年以“googol 游戏”之名把它普及开来（在纸条上写下数字；你的对手逐张翻开，并且必须在最大的那个数出现时停下）。Thomas Ferguson 的文章“Who solved the secretary problem?”（1989）追溯了它错综复杂的历史；有一种说法指，天文学家开普勒（Johannes Kepler）在 1611 年挑选第二任妻子时，会见了十一位人选，最后向第五位求婚。

思维技巧

把策略参数化，然后求最优

当最佳策略未知时，考虑一族带有参数的自然策略，把成功概率写成该参数的函数，然后求最优。

在大学里，这种方法用于决策论、博弈论和统计学。要证明这一族之中的最佳策略就是全体之中的最佳策略，是更困难也更有趣的一步，通常要用*逆向归纳法*（backward induction）来完成。

延伸练习

1 **雪糕。**试做 STEP II 2006，第 13 题（`step.maths.org/questions/06-s2-q13`）：这是秘书问题的小型版本，规则很简单。拒绝第一个，然后取第一个胜过它的。

2 **较小的 N。**当 $N = 4$ 时，求 r 的最佳值和成功的概率。用列举法验算。

3 **最佳或次佳。**利用第 (ii) 部分，论证“最先 n 位中的最佳者是顶尖两位之一”的可能性远高于“最先 n 位中的最佳者就是最佳者”。这对于选中顶尖两位之一的策略有甚么启示?

4 **精确的最大值。**当 $N = 10$ 时，从图中验证 $r = 3$ 是最优的。当 $N = 100$ 时，证明 $r = 37$ 是最优的（准确至四位小数，$r = 36$、37、38 的成功概率分别是 0.3710、0.3710、0.3708）。

延伸阅读

- **Thomas Ferguson，“Who solved the secretary problem?”**，*Statistical Science* 4（1989）。历史与各种变化；可在网上免费阅读。
- **Brian Christian & Tom Griffiths，*Algorithms to Live By***（William Collins）。把最优停止及其他算法应用于日常决策。
- **Martin Gardner，*New Mathematical Diversions***（Mathematical Association of America）。googol 游戏。

12.13 三人决斗

题目

STEP II 1995，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/95-s2-q12`

引子

三个人用餐包决斗。Cuthbert 百发百中，Bertie 每五次掷中三次，而 Algernon 准头最差，每五次只掷中两次。Algernon 先掷。他应该瞄准谁？答案出人意表：谁也不瞄准。他应该故意把餐包掷偏。

原因在于策略。只要三人都仍在场，Bertie 和 Cuthbert 这两个危险的对手就会互相瞄准。假如 Algernon 淘汰了其中一人，余下的那个便会转而对付他。他故意掷失，就能让那两人先行分出胜负，然后由他向幸存者先掷一次。事实上，只要采用最佳方案，准头最差的人反而最有可能胜出。这场“三人决斗”（truel）是博弈论（game theory）中常被引用的例子，说明弱势有时反而是优势。

动笔之前

想一想 在一场二人决斗中，Percy 先掷。他要么一掷即中，要么两人都掷失，于是局面完全重演。试写出一条关于 Percy 胜出概率（probability）的方程。

设 P 为 Percy 胜出的概率。他一掷即中的概率是 p；否则（概率为 $1-p$）轮到 Rodney 掷，而 Rodney 掷失的概率是 $1-r$，此时局面回到起点。所以 $P = p + (1-p)(1-r)P$。

逐步解题

阶段 1 二人决斗

解方程，得 $P(1-(1-p)(1-r))=p$，而 $1-(1-p)(1-r)=p+r-pr$，所以

$$P=\frac{p}{p+r-pr}. \qquad \blacksquare$$

若 Algernon $(p=\frac{2}{5})$ 对 Bertie $(r=\frac{3}{5})$，由 Algernon 先掷，这个概率是 $\frac{2/5}{2/5+3/5-6/25}=\frac{10}{19}$；若由 Bertie 先掷，Algernon 胜出的概率是 $\left(1-\frac{3}{5}\right)\cdot\frac{10}{19}=\frac{4}{19}$。

阶段 2 Algernon 的三个选择

想一想 在每种情况下，还剩下谁？下一个轮到谁掷？

(i) *Algernon 掷中 Cuthbert*。接着由 Bertie 掷，目标是 Algernon。Algernon 要在 Bertie 先掷的二人决斗中胜出：概率为 $\frac{4}{19}$。

(ii) *Algernon 掷中 Bertie*。接着由 Cuthbert 掷，目标是 Algernon，而他从不失手。Algernon 的胜出机会是 0。

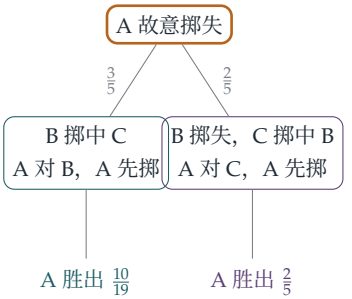
(iii) *Algernon 掷失*。Bertie 向 Cuthbert 掷。他有 $\frac{3}{5}$ 的概率掷中，剩下 Algernon 和 Bertie，并由 Algernon 先掷：Algernon 胜出的概率是 $\frac{10}{19}$。Bertie 有 $\frac{2}{5}$ 的概率掷失，然后 Cuthbert 掷中 Bertie，Algernon 便要面对 Cuthbert，而且只有一次出手机会：他必须掷中才能胜出，概率为 $\frac{2}{5}$。合计为

$$\frac{3}{5}\cdot\frac{10}{19}+\frac{2}{5}\cdot\frac{2}{5}=\frac{6}{19}+\frac{4}{25}=\frac{150+76}{475}=\frac{226}{475}\approx 0.476.$$

阶段 3 建议

想一想 如果 Algernon 瞄准 Cuthbert，他可能掷中（概率为 $\frac{2}{5}$），也可能掷失。他整体的胜出机会是多少？试与故意掷失比较。

瞄准 Cuthbert 的胜出概率是 $\frac{2}{5}\cdot\frac{4}{19}+\frac{3}{5}\cdot\frac{226}{475}=\frac{200+678}{2375}=\frac{878}{2375}\approx 0.370$。瞄准 Bertie 则更差。故意掷失的胜出概率是 $\frac{226}{475}\approx 0.476$，是三者中最高的。所以 Algernon 应该把餐包掷偏，这样他胜出的概率是 $\frac{226}{475}$。 $\blacksquare$



采用这个方案：
A 胜出 $\frac{226}{475} \approx 0.48$
B 胜出 $\frac{135}{475} \approx 0.28$
C 胜出 $\frac{114}{475} \approx 0.24$

图 12.13.1 Algernon 的最佳方案。他故意掷失，让两个较强的对手先行交锋，并确保自己能向幸存者先掷一次。结果准头最差的人最有可能胜出，而百发百中的人反而最不可能胜出。

回顾

- **重复出现的局面可以列出方程。**两人都掷失后，决斗“重新开始”，由此得出一条关于胜出概率的线性方程（linear equation）。
- **考虑对手的反应。**Algernon 的每个选择，都会改变 Bertie 和 Cuthbert 接下来的行动。
- **比较所有选择，包括甚么都不做。**最好的一着，正是看来最差的那一着。

背后的数学

策略与博弈论

三人决斗展示了博弈论的核心思想：最佳行动取决于其他参与者会如何回应。掷中最危险的对手 Cuthbert 看似理所当然，但这样做会改变局势，反而对 Algernon 不利。在本题中，Bertie 和 Cuthbert 的策略（strategy）是固定的（他们互相瞄准）；在完整的博弈论分析中，每个参与者都会自行选择策略，并且知道其他人也在这样做。这类分析始于冯·诺伊曼（John von Neumann）与摩根斯坦（Oskar Morgenstern）在 1944 年的工作，以及纳殊（John Nash）在 1950 年的工作，如今已成为经济学、演化生物学和政治学的基础。

强大反成弱点

三人决斗中谁会胜出？若 Algernon 采用最佳方案，Algernon 胜出的概率是 $\frac{226}{475}$，Bertie 是 $\frac{135}{475}$，Cuthbert 是 $\frac{114}{475}$。准头最好的人最不可能幸存，因为凡是可以选择目标的人都会瞄准他。同样的规律也见于有多名候选人的选举、企业之间的竞争，以及自然界：最强的竞争者可能成为首个被攻击的目标。Martin Shubik 在 1950 年代把三人决斗引入博弈论，

并借此探问：最适者是否必然生存？

思维技巧

多想一步

选择行动之前，先弄清楚它会怎样改变其他各人面对的局面，以及他们会如何回应。

到了大学，这会发展成博弈论中的逆向归纳法（backward induction）：先分析博弈的终局，再往回推，从而决定最佳的第一步。

延伸练习

1 **Bertie 与 Cuthbert。**验证：若 Algernon 故意掷失，Bertie 胜出的概率是 $\frac{135}{475}$，Cuthbert 胜出的概率是 $\frac{114}{475}$。

2 **准头较好的投手。**假如 Algernon 掷中的概率是 a 而不是 $\frac{2}{5}$，当 a 取哪些值时，故意掷失仍然是他最佳的第一步？

3 **Bertie 应该瞄准 Cuthbert 吗？**已知 Algernon 故意掷失，Bertie 瞄准 Cuthbert 而不瞄准 Algernon，是否对他自己有利？（假设 Cuthbert 的行为不变。）

4 **投掷次数有限的决斗。**在一场每人只可掷两次的二人决斗中，求 Percy 胜出的概率，并与不限次数的决斗比较。

延伸阅读

- **D. Marc Kilgour & Steven Brams，“The truel”**，*Mathematics Magazine* 70 (1997)。一篇浅白易读的三人决斗综述。
- **Avinash Dixit & Barry Nalebuff，*Thinking Strategically***（Norton）。写给一般读者的博弈论读物，当中也谈到三人决斗。
- **Ken Binmore，*Game Theory: A Very Short Introduction***（Oxford University Press）。

12.14 混合策略

题目

STEP I 1989，第 16 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-q16`

引子

玩“包剪揾”时，任何固定的出法都会被对手利用：每次都出揾（石头），对手很快便学会出包（布）。唯一稳妥的策略就是随机出手。本题是一个更精细的版本：各种结果的得失并不相等，怎样随机化才对也并不明显。结果发现，只要以概率 $\frac{1}{4}$、$\frac{1}{2}$、$\frac{1}{4}$ 分别猜 1、2、3，不论对手怎样做，玩家都可以保证长远而言不会输钱。

这让我们初次窥见冯纽曼（John von Neumann）在 1928 年提出的**极小极大定理**（minimax theorem），它是博弈论（game theory）的奠基结果：在任何一方所得即另一方所失的二人博弈中，双方各有一个随机化策略（randomised strategy），能保证在对手发挥得完美无瑕时，仍取得可能范围内的最佳结果。

动笔之前

想一想 写出一个表，列出每一对猜测之下 B 赢得的金额（B 的猜测列于左侧，A 的猜测列于顶端）。

按规则逐一推算（例如，若 B 说 1 而 A 说 3，两数相差 2，说出较小数的 B 要付 £2）：

B \ A	1	2	3
1	0	+1	−2
2	−1	0	+1
3	+2	−1	0

这个表是**反对称**（antisymmetric）的（把两位玩家对调，正负号便会改变），所以这个博

弈是公平的，意思是双方都没有与生俱来的优势。

逐步解题

阶段 1　对付固定的猜测

若 A 每次都说 1，B 应该说 3，每局赢 £2。若 A 每次都说 2，B 说 1，赢 £1。若 A 每次都说 3，B 说 2，赢 £1。任何固定的猜测都会被对手利用。

阶段 2　对付随机的猜测

想一想　若 A 以均匀的概率随机猜测，当 B 说 j 时，B 的期望收益（expected gain）是多少？（求表中第 j 行的平均值。）

各行的平均值分别是：$j=1$ 时为 $\frac{1}{3}(0+1-2)=-\frac{1}{3}$，$j=2$ 时为 0，$j=3$ 时为 $\frac{1}{3}(2-1+0)=\frac{1}{3}$。所以 B 每局赢得金额的期望值是

$$-\tfrac{1}{3}b_1+0\cdot b_2+\tfrac{1}{3}b_3,$$

当 $b_3=1$ 时取得最大值：每次都说 3，平均每局赢 $\frac{1}{3}$ 镑。均匀的随机化并不能保护 A。

阶段 3　一般情况

想一想　现在设 A 以概率 a_j 说 j。把 B 赢得金额的期望值按 A 的猜测归项：a_1 的系数（即第 1 列以 b_j 为权的加权平均）是甚么？

沿着表中各列由上而下读，

$$\text{B 赢得金额的期望值}=(-b_2+2b_3)a_1+(b_1-b_3)a_2+(-2b_1+b_2)a_3,$$

所以 $X=-b_2+2b_3$，$Y=b_1-b_3$，$Z=-2b_1+b_2$。

想一想　你能令三者同时为零吗？（共有三条方程，但它们并非独立；为甚么？）

它们并非独立：$X+2Y+Z=0$ 恒成立（请自行验证）。所以由 $Y=0$ 得 $b_1=b_3$，由 $X=0$ 得 $b_2=2b_3$，两者合起来便迫使 $Z=0$ 亦成立。再配合 $b_1+b_2+b_3=1$：

$$b_1=\tfrac{1}{4},\qquad b_2=\tfrac{1}{2},\qquad b_3=\tfrac{1}{4},$$

于是 $X=Y=Z=0$。这样，不论 a_j 取甚么值，B 赢得金额的期望值都是 0，B 长远而言不会有任何损失。 ■

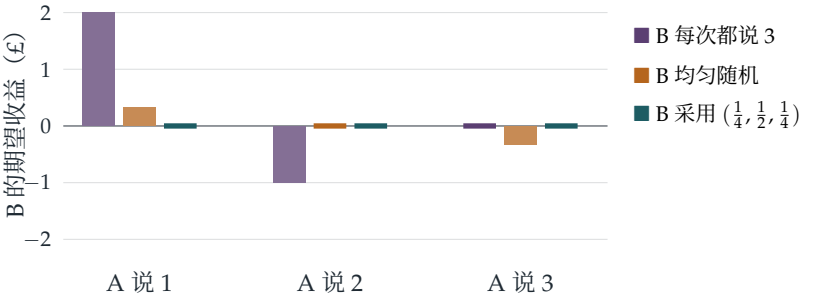


图 12.14.1 B 采用三种不同策略时，面对 A 每一种选择的期望收益。每次都说 3，遇上 A 说 1 会赢，遇上 A 说 2 却会输；均匀随机猜测，遇上 A 说 3 会输。策略 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ 则不论 A 怎样做，收益都恰好是 0（图中以幼条表示），所以永远不会被对手利用。

回顾

- **写出得失表。**所有推理都围绕一个 3×3 矩阵的行和列。
- **纯策略（pure strategy）会被对手利用。**面对任何固定的猜测，甚至面对均匀随机的猜测，对手都有一个有利可图的应对方法。
- **拉平。**稳妥的策略令对手选甚么都没有分别：A 的每一种选择都带来相同的期望结果。

背后的数学

极小极大定理

在一个二人零和（zero-sum）博弈中，设得失矩阵（payoff matrix）为 M，B 选取概率 $\mathbf{b}$，A 选取概率 $\mathbf{a}$，则 B 赢得金额的期望值是 $\mathbf{b}^{\mathsf{T}} M \mathbf{a}$。冯纽曼的极小极大定理指出

$$\max_{\mathbf{b}} \min_{\mathbf{a}} \mathbf{b}^{\mathsf{T}} M \mathbf{a} = \min_{\mathbf{a}} \max_{\mathbf{b}} \mathbf{b}^{\mathsf{T}} M \mathbf{a}.$$

用文字说：B 预先公开自己的随机化策略而仍能保证取得的金额，等于 A 预先公开自己的策略而能把 B 压制到的金额。这个共同的数值称为博弈的值（value）。在本题中，由反对称性可知博弈的值是 0，而策略 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ 对双方都是最优的。这个定理可用线性代数（linear algebra）和凸性（convexity）证明（它等价于线性规划（linear programming）的对偶定理），而它推广到多人博弈的版本，即纳殊均衡（Nash equilibrium），为纳殊（John Nash）赢得诺贝尔经济学奖。

为甚么要随机化?

这里的随机性无关运气，而是关乎信息。如果 B 的选择可以被预测，A 便能加以利用；B 以恰当的概率随机选择，既令自己的选择无从预测，又确保任何预测都帮不了对手。现实中的例子包括足球的十二码罚球——对职业球员的研究发现，主射球员和守门员在左右两边之间随机选择，比例与博弈论的最优值十分接近——以及网球的发球。

冯纽曼 1928 年的论文

冯纽曼在 1928 年证明了极小极大定理，论文题为“On the theory of games of strategy”（论策略博弈的理论），研究动机来自扑克，因为扑克中的“偷鸡”（bluffing，即虚张声势）正是一种随机化。1944 年，他与经济学家 Oskar Morgenstern 出版了 *Theory of Games and Economic Behavior*（博弈论与经济行为），博弈论自此成为一门学科。冯纽曼后来忆述，他一直视极小极大定理为关键结果，没有它，就没有值得发表的博弈理论。

思维技巧

以随机化消除对手的优势

当任何固定的选择都会被对手利用时，就随机选择，而所用的概率要令对手无论选哪一项都没有分别。

在大学里，博弈论、线性规划和算法理论都会用到这个想法。例如，随机算法（randomised algorithm）会作出任何对手都无法预测的随机选择，借此防范最坏情况的输入。

延伸练习

1. **A 的最佳应对。**若 B 采用 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$，证明 A 的每一种策略所得的期望值都是 0。若 B 改用 $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$，求 A 的最佳应对。

2. **包剪揼。**写出“包剪揼”的得失矩阵，并求最优策略。为甚么这里的最优策略是均匀的，而本题的却不是?

3. **唯一性。**证明 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ 是 B 能保证不输的唯一策略。（提示：它必须令 X、Y、Z 全部非负；利用 $X + 2Y + Z = 0$。）

4. **改动后的博弈。**若相差 2 时的罚款由 £2 改为 £3，求最优策略。

延伸阅读

- **Ken Binmore, *Playing for Real***（Oxford University Press）。一本详尽而平易近人的博弈论入门书。
- **Avinash Dixit & Barry Nalebuff, *The Art of Strategy***（Norton）。以体育和商业的例子讲解博弈论。
- **William Poundstone, *Prisoner's Dilemma***（Anchor）。冯纽曼、博弈论与冷战。

12.15 灭绝还是爆发

题目

STEP II 2003，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/03-s2-q12`

引子

1873 年，高尔顿（Francis Galton）在《*Educational Times*》上提出一个问题：为甚么贵族的姓氏那么容易失传？每个男人所生的儿子数目是随机的，而每个儿子都会把姓氏传下去。即使每个男人平均有多于一个儿子，这个姓氏就能永远流传下去吗？这个问题开创了分支过程（branching process）的理论，如今这套理论用来模拟核连锁反应、传染病的扩散、细胞群体的增长，以及突变能否存活下去。

本题建立的是最简单的模型：每个粒子分裂成 0、1 或 2 个粒子。粒子总数的期望值（expectation）按几何级数增长或缩减。但期望值并不是故事的全部。即使期望的粒子总数急剧膨胀，整个族系完全灭绝的概率仍然是正数，而一个漂亮的不动点（fixed point）论证就能把这个概率求出来。

动笔之前

? **想一想** 如果某一阶段有两个粒子，下一阶段粒子数目的分布是甚么？

两个粒子各自独立地产生后代，后代数目服从 $\mathrm{B}(2, p)$，所以总数服从 $\mathrm{B}(4, p)$：共有四次独立的“机会”，每次成功的概率为 p。

逐步解题

阶段 1　先是两个，然后又是两个

$\mathrm{P}(X_1 = 2) = p^2$，然后 $\mathrm{P}(X_2 = 2 \mid X_1 = 2) = \binom{4}{2} p^2 q^2 = 6p^2q^2$。所以

$$\mathrm{P}(X_1 = 2, X_2 = 2) = 6p^4q^2. \qquad \blacksquare$$

阶段 2 逆向推理

> ? **想一想** $X_2 = 2$ 可以有多少种方式发生？（X_1 可以是甚么？）

要使 $X_2 = 2$，可以是 $X_1 = 1$（概率为 $2pq$），其后产生 2 个后代（p^2）；也可以是 $X_1 = 2$，其后由 $\mathrm{B}(4, p)$ 得出 2 个。（$X_1 = 0$ 则甚么也不剩。）所以 $\mathrm{P}(X_2 = 2) = 2p^3q + 6p^4q^2$，而

$$\mathrm{P}(X_1 = 2 \mid X_2 = 2) = \frac{6p^4q^2}{2p^3q + 6p^4q^2} = \frac{3pq}{1 + 3pq}.$$

令此式等于 $\frac{9}{25}$：$75pq = 9 + 27pq$，所以 $pq = \frac{3}{16}$，即 $p(1-p) = \frac{3}{16}$，由此得 $p = \frac{1}{4}$ 或 $p = \frac{3}{4}$。 $\blacksquare$

阶段 3 粒子总数的期望值

> ? **想一想** 如果在第 $n-1$ 阶段有 k 个粒子，第 n 阶段粒子数目的期望值是多少？然后对 k 取平均。

每个粒子平均有 $\mathrm{E}(X) = 2p$ 个后代，所以 k 个粒子平均有 $2pk$ 个后代：$\mathrm{E}(X_n \mid X_{n-1} = k) = 2pk$。对 k 取平均，得 $\mathrm{E}(X_n) = 2p\,\mathrm{E}(X_{n-1})$。由于 $\mathrm{E}(X_0) = 1$，

$$\mathrm{E}(X_n) = (2p)^n.$$

当 $p = \frac{1}{4}$ 时，$(2p)^n = 2^{-n} \to 0$；当 $p = \frac{3}{4}$ 时，$(2p)^n = \left(\frac{3}{2}\right)^n \to \infty$。 $\blacksquare$

回顾

- **独立二项变量之和。**k 个粒子各自给出 $\mathrm{B}(2, p)$，合起来就是 $\mathrm{B}(2k, p)$。
- **时间上的贝叶斯定理（Bayes' theorem）。**$\mathrm{P}(X_1 = 2 \mid X_2 = 2)$ 是由较后的观察结果逆向推断较早的状态。
- **条件期望值（conditional expectation）。**以上一代的情况为条件，期望值就化成一个等比数列（geometric sequence）。

背后的数学

灭绝的概率

当 $p = \frac{3}{4}$ 时，粒子总数的期望值无止境地增长。这是否表示群体一定能存活下去？不一定：如果第一个粒子没有后代（概率为 $q^2 = \frac{1}{16}$），这个族系便立即终结。设 e 为最终灭绝的概率。第一个粒子有 k 个后代，而每个后代的族系各自独立地以概率 e 灭绝，所以

$$e = q^2 + 2pq\,e + p^2e^2 = (q + pe)^2 = \mathrm{G}(e),$$

其中 $\mathrm{G}(s) = (q + ps)^2$ 是后代数目的*概率生成函数*（probability generating function）。灭绝概率是 G 的一个不动点：事实上是最小的非负不动点。当 $p = \frac{3}{4}$ 时，由 $(1 + 3e)^2 = 16e$ 得 $9e^2 - 10e + 1 = 0$，所以 $e = \frac{1}{9}$ 或 1，而灭绝概率是 $\frac{1}{9}$。当 $p = \frac{1}{4}$ 时，在 $[0, 1]$ 内唯一的不动点是 $e = 1$：灭绝是必然的（图 12.15.1）。

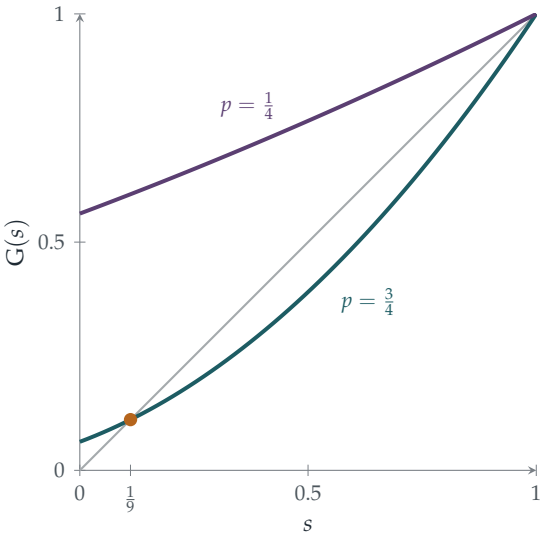


图 12.15.1 后代数目的生成函数 $\mathrm{G}(s) = (q + ps)^2$ 与直线 $y = s$。灭绝概率就是两者相交的最小一点。当 $p = \frac{1}{4}$（紫红色，平均值 $\frac{1}{2}$）时，该点是 $s = 1$：灭绝是必然的。当 $p = \frac{3}{4}$（蓝绿色，平均值 $\frac{3}{2}$）时，该点是 $s = \frac{1}{9}$（琥珀色）：群体以概率 $\frac{8}{9}$ 永远存活下去。

临界门槛

一般而言，如果后代数目的平均值小于 1，分支过程必然灭绝；平均值等于 1 时也是如此，唯一的例外是每个个体都恰好有一个后代的平凡情况（这时族系只是一个接一个地永远延续下去）；如果平均值大于 1，则分支过程以正概率存活下去。后代数目的平均值就是流行病学中的*繁殖数*（reproduction number）R：$R > 1$ 的疫症有可能大规模爆发，但即使如此，早期的爆发仍可能纯粹因为偶然而无疾而终，其概率正好由上面的不动点方程给出。在核物理学中，如果每次裂变平均恰好引发另一次裂变，连锁反应就处于*临界*

（critical）状态，而操作反应堆的学问，就在于让它一直保持在这个状态。

高尔顿、Watson 与 Bienaymé

高尔顿的问题在 1874 年由 Henry William Watson 牧师作答，但答案有部分是错的：他找到了不动点方程，却漏掉了第二个根，因而断定每个姓氏最终都会失传。其实正确的分析早在 1845 年已由法国统计学家 Irénée-Jules Bienaymé 提出，只是他的研究一直被人遗忘，直到 1970 年代才重见天日。这个模型现在称为*高尔顿—Watson 过程*（Galton–Watson process，有时称为 Bienaymé–Galton–Watson 过程），是每一门随机过程（random processes）课程中的核心例子。

思维技巧

以第一代为条件

一个会自我更新的过程（每个后代都开始一个独立的、与整个过程相同的复本），必定满足一条以第一步为条件而得出的方程。

方程 $e = \mathrm{G}(e)$ 是一个不动点问题，与 Section 2.2 和 Section 8.7 中遇到的属同一类。在大学里，生成函数和不动点是研究分支过程、随机树和随机网络的主要工具。

延伸练习

1 **一般情况下的灭绝。**设后代数目的分布为 $\mathrm{B}(2, p)$，证明当 $p \le \frac{1}{2}$ 时灭绝概率为 1，而当 $p > \frac{1}{2}$ 时灭绝概率为 $\left(\frac{q}{p}\right)^2$。

2 **方差（variance）。**证明 $\mathrm{Var}(X_n)$ 满足 $\mathrm{Var}(X_n) = (2p)^2\mathrm{Var}(X_{n-1}) + 2pq\,\mathrm{E}(X_{n-1})$，并求 $p = \frac{1}{2}$ 时的方差。

3 **在第 n 代或之前灭绝。**设 $e_n = \mathrm{P}(X_n = 0)$。证明 $e_n = \mathrm{G}(e_{n-1})$，并就 $p = \frac{3}{4}$ 计算 e_1、e_2、e_3。它们是否趋近 $\frac{1}{9}$？

4 **姓氏。**如果每个男人有 0、1、2 或 3 个儿子的概率各为 $\frac{1}{4}$，他的姓氏最终失传的概率是多少？

延伸阅读

- **Geoffrey Grimmett & David Stirzaker, *Probability and Random Processes***（Oxford University Press），论分支过程的一节。
- **Adam Kucharski, *The Rules of Contagion***（Profile）。讲述繁殖数与疫症爆发，适合一般读者。
- **David Kendall, "Branching processes since 1873"**，*Journal of the London Mathematical Society* 41 (1966)。讲述这段历史，包括 Bienaymé 的贡献。

12.16 顺向与逆向的马可夫链

题目

STEP III 1996，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/96-s3-q12`

引子

Ecks 教授的定理有一个规律：她下一个证明甚么，只取决于她上一个证明了甚么。一个过程的未来若只取决于现在而与过去无关，就称为**马可夫链**（*Markov chain*）。马可夫链无处不在：天气模型、蛇梯棋这种图版游戏、DNA 模型、Google 最初的 PageRank 算法，以至统计学家用来模拟复杂概率分布的方法，全都用得上它。

本题要求出各类定理在长远而言所占的比例，即**平稳分布**（*stationary distribution*）：长远来说，她的定理每六个之中有一个是新而正确的，而有一半是错的。然后题目把这条链在时间上倒过来运行，并问：已知今天的定理，昨天的定理是哪一类？

动笔之前

? **想一想** “每一行（row）之和为 1”这个条件是甚么意思？各列（column）之和是否也是 1？

第 i 行列出的是：已知现时的类型为 i，下一个定理属于各类型的概率；三者必有其一发生，所以每一行之和为 1。各列则不一定：第二列之和恰好是 1，但第一列之和是 0.6，第三列之和是 1.4。

逐步解题

阶段 1 **一步**

根据全概率定律（law of total probability），以现时定理的类型 i 为条件，

$$b_j = \sum_i \mathrm{P}(\text{现时}i)\,\mathrm{P}(\text{下一个}j \mid \text{现时}i) = a_1p_{1j} + a_2p_{2j} + a_3p_{3j}. \qquad \blacksquare$$

用矩阵的语言来说，行向量（row vector）$\mathbf{b} = \mathbf{a}P$。

阶段 2 平稳分布

想一想 我们要求 $\mathbf{a} = \mathbf{a}P$，且 $a_1 + a_2 + a_3 = 1$。把各方程写出来。哪些最简单？

各方程为

$$\begin{aligned} a_1 &= 0.3a_1 + 0.2a_2 + 0.1a_3, \\ a_2 &= 0.3a_1 + 0.4a_2 + 0.3a_3, \\ a_3 &= 0.4a_1 + 0.4a_2 + 0.6a_3. \end{aligned}$$

（它们并非独立：把三式相加，两边都得 $a_1 + a_2 + a_3$。）由第二式得 $0.6a_2 = 0.3(a_1 + a_3) = 0.3(1 - a_2)$，所以 $a_2 = \frac{1}{3}$。由第一式得 $7a_1 = 2a_2 + a_3 = \frac{2}{3} + a_3$，而 $a_1 + a_3 = \frac{2}{3}$，所以 $8a_1 = \frac{4}{3}$，$a_1 = \frac{1}{6}$，$a_3 = \frac{1}{2}$：

$$(a_1, a_2, a_3) = \left(\tfrac{1}{6}, \tfrac{1}{3}, \tfrac{1}{2}\right). \qquad \blacksquare$$

阶段 3 把链倒过来运行

想一想 $q_{ij} = \mathrm{P}(\text{上一个}i \mid \text{现时}j)$。运用贝叶斯定理（Bayes' theorem），上一个定理和现时的定理都采用平稳分布。

根据贝叶斯定理，

$$q_{ij} = \frac{\mathrm{P}(\text{上一个}i)\,\mathrm{P}(\text{现时}j \mid \text{上一个}i)}{\mathrm{P}(\text{现时}j)} = \frac{a_ip_{ij}}{a_j}.$$

例如 $q_{11} = \frac{\frac{1}{6}\times 0.3}{\frac{1}{6}} = 0.3$ 及 $q_{21} = \frac{\frac{1}{3}\times 0.2}{\frac{1}{6}} = 0.4$。全部结果如下（以 i 标示行，j 标示列）：

$$(q_{ij}) = \begin{pmatrix} 0.3 & 0.15 & \frac{2}{15} \\ 0.4 & 0.4 & \frac{4}{15} \\ 0.3 & 0.45 & 0.6 \end{pmatrix}. \qquad \blacksquare$$

现在是每一列之和为 1，因为第 j 列列出的是上一个定理各种可能的类型。

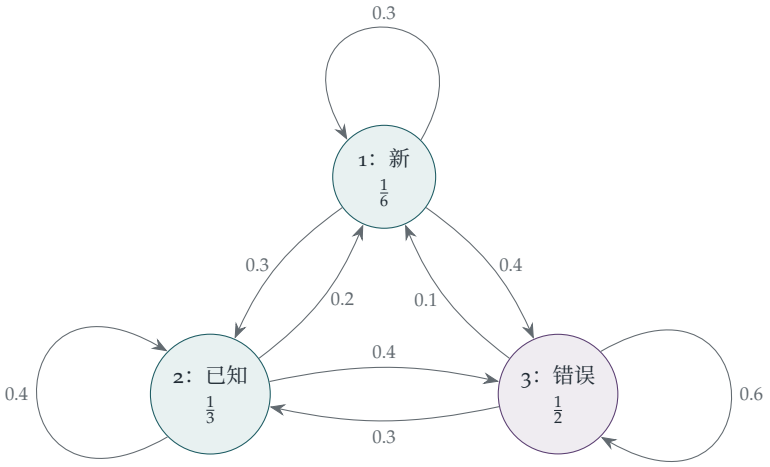


图 12.16.1 把 Ecks 教授的定理看成一条马可夫链。箭头显示下一个定理属于各类型的概率。圆圈内的分数是平稳分布：长远来说，每六个定理有一个是新而正确的，每三个有一个是已知的，而有一半是错的。

回顾

- **以矩阵表达全概率定律。**链每走一步，就是把一个行向量乘以矩阵 P。
- **平稳即特征向量（eigenvector）。**$\mathbf{a} = \mathbf{a}P$ 说明 $\mathbf{a}$ 是 P 的左特征向量，对应的特征值（eigenvalue）为 1。
- **用贝叶斯定理倒推。**逆向概率构成另一个随机矩阵（stochastic matrix），由正向的随机矩阵和平稳分布构造而成。

背后的数学

收敛至平衡

让 Ecks 教授从任何一个定理开始；再证明 n 个之后，类型的分布是 $\mathbf{a}_0 P^n$。对于像这样的一条链，即每种类型都可以紧接在任何一种类型之后，不论起点为何，P^n 都会收敛至一个每一行均等于平稳分布 $(\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$ 的矩阵。收敛的速度由 P 的其他特征值（这里是 0.2 和 0.1）决定，它们以几何方式缩小：只需几个定理之后，她的过去就被遗忘了。这个关于马可夫链的*遍历定理*（*ergodic theorem*）是大学二年级概率课程的核心结果，马可夫链蒙地卡罗（Markov chain Monte Carlo）方法之所以行得通，靠的正是它：要从一个复杂的分布中抽样，就设计一条以该分布为平稳分布的马可夫链，然后让它运行一段长时间。

可逆性

逆向链 (q_{ij}) 与正向链 (p_{ij}) 并不相同：从她的定理的一份长纪录中，你可以分辨出时间的方向。若一条链的逆向版本与正向版本一致，即对所有 i、j 都有 $a_i p_{ij} = a_j p_{ji}$，就称为*可逆*（*reversible*）；网络上的随机游走（random walk）以及处于热平衡的物理系统的模型都属此类。可逆性是一个威力强大的性质，因为它令平稳分布很容易求出。

马可夫与普希金

马可夫（Andrey Markov）在 1906 年引入他的链，部分原因是要证明大数定律（law of large numbers）并不需要独立性；当时他正与数学家 Pavel Nekrasov 争论，后者声称独立性是必需的。1913 年，马可夫把他的想法应用于普希金（Pushkin）的诗体小说《叶甫盖尼·奥涅金》（*Eugene Onegin*）的首 20 000 个字母，把每个字母分为元音或辅音，并证明每个字母的类型取决于前一个字母的类型，而这种关系可用一条两个状态的链很好地描述。这是史上第一次对文本作统计分析，也是所有预测输入系统的始祖。

思维技巧

概率作向量，转移作矩阵

当一个系统以固定的概率在各状态之间转移时，把分布写成向量，把转移写成矩阵。这样，长远行为就变成一个关于特征向量的问题。

在大学里，这样做会把第 7 章的线性代数带进概率论。佩龙－弗罗宾尼斯定理（Perron–Frobenius theorem）保证平稳分布存在，而其他特征值则量度到达平稳分布有多快。

延伸练习

1 **兔子窝。**试做 STEP III 2003，第 13 题（`step.maths.org/questions/03-s3-q13`）：在一个正方形的四个顶点上的随机游走，每个顶点各有一个出口，以首步方程求解。

2 **两步。**若 Ecks 教授上一个定理是新而正确的，求再下一个定理属于三种类型的概率。

3 **特征值。**证明 P 的特征值为 1、0.2 和 0.1。

4 **可逆吗？**检查 Ecks 教授的链是否满足 $a_i p_{ij} = a_j p_{ji}$。答案告诉你甚么？

延伸阅读

- **James Norris, *Markov Chains***（Cambridge University Press）。标准的大学本科教科书。
- **Olle Häggström, *Finite Markov Chains and Algorithmic Applications***（Cambridge University Press）。篇幅短小，讲马可夫链蒙地卡罗尤其出色。
- **Brian Hayes, “First links in the Markov chain”**，*American Scientist* 101 (2013)。重述马可夫对普希金作品的分析。

12.17 巴拿赫的火柴盒

题目

STEP II 1992，第 16 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-s2-q16`

引子

相传数学家巴拿赫（Stefan Banach）烟瘾很大，两边口袋各放一盒火柴。每次要用火柴，他就随机选一个口袋。有一天，他发现其中一盒空了。另一盒还剩多少根火柴？这就是*巴拿赫火柴盒问题*（Banach's matchbox problem），由斯坦豪斯（Hugo Steinhaus）为向巴拿赫致意而设计；本题则是的士司机的版本，把火柴换成了糖果。

微妙之处在于“发现”一词。司机取走一包糖的最后一颗时，并不会察觉那包已经空了，要到下一次伸手去拿那包糖时才知道。所以这个事件是：她伸手去拿太妃糖，而全部 n 颗太妃糖早已吃完。计算这件事在第 N 次伸手时发生的方法数目，是二项分布（binomial distribution）的一道巧妙练习；而巴拿赫问题的答案有一个出人意表的特点：另一盒最可能剩下的火柴数目是 0（与 1 并列）。

动笔之前

想一想 若司机在第 N 次伸手时发现太妃糖已经吃完，第 N 次伸手必须是甚么？之前的 $N-1$ 次伸手又必须包含甚么？

第 N 次伸手必须是去拿太妃糖，而之前的 $N-1$ 次伸手必须刚好取走了 n 颗太妃糖（即全部）和 $N-1-n$ 颗薄荷糖。此外，她在此之前不可以已经发现薄荷糖空了。

逐步解题

阶段 1 发现太妃糖吃完了

想一想 头 $N-1$ 次伸手可以按甚么次序出现？$N-1-n$ 要符合甚么条件，才能确保她从未发现那包薄荷糖是空的？

那 n 次拿太妃糖的伸手，可以是头 $N-1$ 次伸手中的任何 n 次。只要她最多取了 m 颗薄荷糖，即 $N-1-n \le m$，她就从未在薄荷糖吃完之后再伸手去拿薄荷糖。因此，对于 $n+1 \le N \le n+m+1$，

$$\mathrm{P}(T_N) = \binom{N-1}{n} p^{n+1} q^{N-1-n},$$

其中 T_N 是事件“她在第 N 次伸手时发现太妃糖吃完了”（因子 p^{n+1} 已包括第 N 次伸手本身）。

阶段 2 发现薄荷糖吃完了

把两者的角色对调，用同样的论证可得，对于 $m+1 \le N \le n+m+1$，

$$\mathrm{P}(M_N) = \binom{N-1}{m} q^{m+1} p^{N-1-m}.$$

阶段 3 条件概率

想一想 已知在第 N 次伸手时有所发现。那次拿的是太妃糖的概率是多少？

事件 T_N 与 M_N 互斥（exclusive），所以

$$\mathrm{P}(\text{太妃糖} \mid \text{有所发现的伸手次数为} N) = \frac{\binom{N-1}{n} p^{n+1} q^{N-1-n}}{\binom{N-1}{n} p^{n+1} q^{N-1-n} + \binom{N-1}{m} q^{m+1} p^{N-1-m}},$$

其中二项式系数（binomial coefficient）在上述范围以外一律视为 0。（若 $N \le m$，则只有太妃糖可能已经吃完，概率为 1；同样，若 $N \le n$ 但 $N > m$，概率为 0。） ■

停一停，试一试 当 $n = m$ 且 $p = q = \frac{1}{2}$ 时，验算一下这条公式。根据对称性，对每一个可能的 N，答案都应该是 $\frac{1}{2}$。是否如此？

回顾

- **准确理解事件。**“发现”是指伸手去拿一包已空的糖，而不是取走最后一颗糖。

- **固定最后一步，再数其余。**第 N 次伸手已经固定；头 $N-1$ 次可以按任何次序出现，于是得出一个二项式系数。
- **检查边界。**限制条件 $N-1-n \le m$ 和 $N-1-m \le n$ 排除了不可能出现的过程。

背后的数学

巴拿赫的火柴盒

在巴拿赫的版本中，每盒起初有 n 根火柴，选中每个口袋的机会均等。当他首次发现其中一盒空了，设 K 为另一盒的火柴数目。若 $K=k$，他必定已从空的那盒取了 n 根，从另一盒取了 $n-k$ 根，次序不拘，然后再一次伸手去拿空的那盒；而两盒之中任何一盒都可以是空的那盒：

$$\mathrm{P}(K=k) = 2\binom{2n-k}{n}\left(\frac{1}{2}\right)^{2n-k+1} = \binom{2n-k}{n}\left(\frac{1}{2}\right)^{2n-k}.$$

这些概率随 k 增大而递减，唯一例外是 $\mathrm{P}(K=0)=\mathrm{P}(K=1)$（图 12.17.1）。剩余火柴数目的期望值（expected value）约为 $2\sqrt{n/\pi}-1$：若每盒有 50 根火柴，约为 7。这个 $\sqrt{n}$ 是经过 $2n$ 次随机选择之后，两盒之间差距的典型大小，与生日问题（Section 9.9）和随机漫步（random walk）中的平方根定律如出一辙。

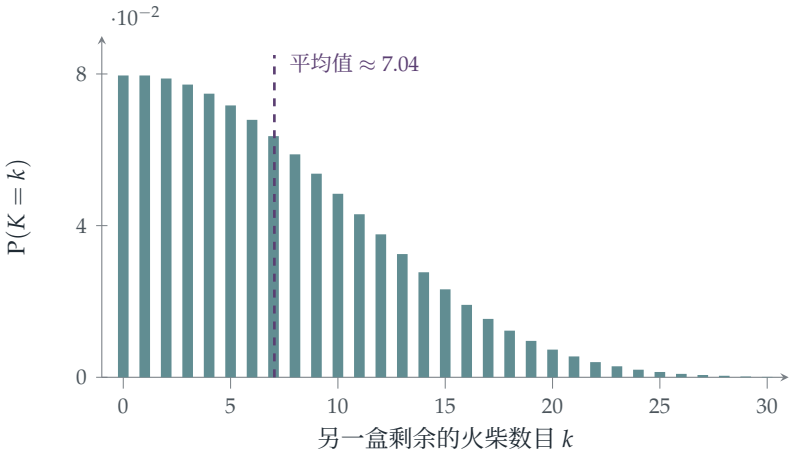


图 12.17.1 每盒有 50 根火柴的巴拿赫火柴盒问题：首次发现其中一盒空了的时候，另一盒剩余火柴数目的分布。最可能的数值是 0 和 1，两者概率相同，但平均值约为 7。

负二项分布

在一连串独立试验（independent trials）中，每次试验成功的概率为 p，则取得第 r 次成功所需的试验次数服从*负二项分布*（negative binomial distribution）：$\mathrm{P}(N=k)=$

$\binom{k-1}{r-1}p^r q^{k-r}$。的士司机的问题，是两个负二项过程（太妃糖和薄荷糖）之间的竞赛，每个过程所需的“成功”次数都比该包糖的数目多一次。负二项分布见于保险（索偿宗数）、生态学（动物聚集分布的数目）和遗传学，往往用来代替泊松分布（Poisson distribution），因为它更有弹性。

巴拿赫与苏格兰咖啡馆

巴拿赫是泛函分析（functional analysis）的奠基人之一，他的数学研究有不少是在 Lwów（今称 Lviv）的苏格兰咖啡馆（Scottish Café）里完成的。他和同事把问题写在一本由侍应保管的笔记簿上，这就是《苏格兰书》（Scottish Book），解题奖品从一杯啤酒到一只活鹅不等。斯坦豪斯是巴拿赫的恩师，他为了一篇向巴拿赫的吸烟习惯致意的演说，设计了这道火柴盒问题。最先在书刊中分析这个问题的，是费勒（William Feller）1950 年出版的著名概率论教科书。

思维技巧

固定最终事件，数出通往它的路径

要求一个过程在某一时刻以某种方式结束的概率，可先固定最后一步发生的事，再数出通往这一步的所有过程。

在大学里，这是负二项分布、随机漫步的首达时间（first-passage time），以及 Section 9.10 中反射原理（reflection principle）的基础。

延伸练习

1 **小规模情况。**设 $n = m = 1$ 及 $p = \frac{1}{2}$，列出所有可能的序列，并求 N 的分布。

2 **巴拿赫公式。**对 $n = 1$ 和 $n = 2$，验证 $\sum_{k=0}^{n} \binom{2n-k}{n} 2^{-(2n-k)} = 1$。

3 **并列最高。**证明在巴拿赫问题中，对每一个 n 都有 $\mathrm{P}(K = 0) = \mathrm{P}(K = 1)$。

4 **哪一包先吃完？**就的士司机的问题，当 $n = m = 2$ 及 $p = \frac{2}{3}$ 时，求太妃糖是首先被发现空了的那一包（不论何时发现）的概率。

延伸阅读

► **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1**（Wiley），

第 VI 章，正是这本书把巴拿赫问题介绍给全世界。

- **R. Daniel Mauldin (ed.), *The Scottish Book***（Birkhäuser）。收录苏格兰咖啡馆的问题，并附评注。
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），内容涉及负二项分布。

12.18 无记忆性

题目

STEP III 2021，第 11 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 2021 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

`www.ocr.org.uk/Images/696588-2021-past-papers.zip`

引子

一颗放射性原子在实验室里待了一百万年，一直没有衰变。它是否“快要”衰变了？不是：它在下一秒内衰变的机会，与一颗刚刚产生的原子完全相同。放射性原子不会老化。直至衰变为止的等候时间服从*指数分布*（*exponential distribution*），而这个分布的决定性特征，就是它没有记忆。

这道题以一种令人印象深刻的形式展示了这种无记忆性（memorylessness）。把一个指数等候时间拆成整数个单位和小数部分。原来这两部分是*独立*（*independent*）的：知道你等了多少个完整的小时，对于分钟数毫无提示。这道题还会在练习中解释，为甚么你等巴士的时间，可以显得比两班巴士之间的间隔还要长。

动笔之前

想一想 对于 $x \geq 0$，$\mathrm{P}(X > x)$ 是甚么？$\mathrm{P}(X > s+t \mid X > s)$ 又是甚么？

$\mathrm{P}(X > x) = \int_x^\infty \lambda e^{-\lambda u}\,\mathrm{d}u = e^{-\lambda x}$。于是 $\mathrm{P}(X > s+t \mid X > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathrm{P}(X > t)$：已经等了 s 之后，余下的等候时间与从头开始等候的分布相同。这就是*无记忆性质*（*memoryless property*）。

逐步解题

阶段 1 整数部分

$Y = n$ 当且仅当 $n \le X < n+1$：

$$\mathrm{P}(Y = n) = e^{-\lambda n} - e^{-\lambda(n+1)} = (1 - e^{-\lambda})e^{-n\lambda}. \qquad \blacksquare$$

所以 Y 服从几何分布（geometric distribution）：每一个完整单位都以概率 $e^{-\lambda}$ 被“挨过”。

阶段 2 小数部分

想一想 $Z < z$ 表示 X 落在其中一个区间 $[n, n+z)$ 内。把这些区间的概率加起来。

$$\mathrm{P}(Z < z) = \sum_{n \ge 0} \left(e^{-\lambda n} - e^{-\lambda(n+z)}\right) = (1 - e^{-\lambda z}) \sum_{n \ge 0} e^{-\lambda n} = \frac{1 - e^{-\lambda z}}{1 - e^{-\lambda}}. \qquad \blacksquare$$

阶段 3 Z 的平均值

想一想 对于取值于 $[0,1]$ 的随机变量（random variable），$\mathrm{E}(Z) = \int_0^1 \mathrm{P}(Z > z)\,\mathrm{d}z$。运用这条公式。

设 $c = 1 - e^{-\lambda}$，则 $\mathrm{P}(Z > z) = 1 - \frac{1-e^{-\lambda z}}{c}$，所以

$$\mathrm{E}(Z) = 1 - \frac{1}{c}\left(1 - \frac{1 - e^{-\lambda}}{\lambda}\right) = 1 - \frac{1}{c} + \frac{1}{\lambda} = \frac{1}{\lambda} - \frac{e^{-\lambda}}{1 - e^{-\lambda}} = \frac{1}{\lambda} - \frac{1}{e^{\lambda} - 1}. \qquad \blacksquare$$

（当 λ 很小，即等候时间很长时，这个值趋向 $\frac{1}{2}$：小数部分几乎是均匀分布的。当 λ 很大时，它接近 $\frac{1}{\lambda}$，因为 X 几乎总是小于 1。）

阶段 4 独立性

$$\mathrm{P}(Y = n, z_1 < Z < z_2) = \mathrm{P}(n + z_1 < X < n + z_2) = e^{-\lambda n}\left(e^{-\lambda z_1} - e^{-\lambda z_2}\right).$$

想一想 利用第 (i) 及第 (ii) 部分，与 $\mathrm{P}(Y = n)\,\mathrm{P}(z_1 < Z < z_2)$ 作比较。

由 (i) 及 (ii)，$\mathrm{P}(Y = n)\mathrm{P}(z_1 < Z < z_2) = (1 - e^{-\lambda})e^{-n\lambda} \cdot \frac{e^{-\lambda z_1} - e^{-\lambda z_2}}{1 - e^{-\lambda}}$，两者相同。所以 Y 与 Z 独立。 $\blacksquare$

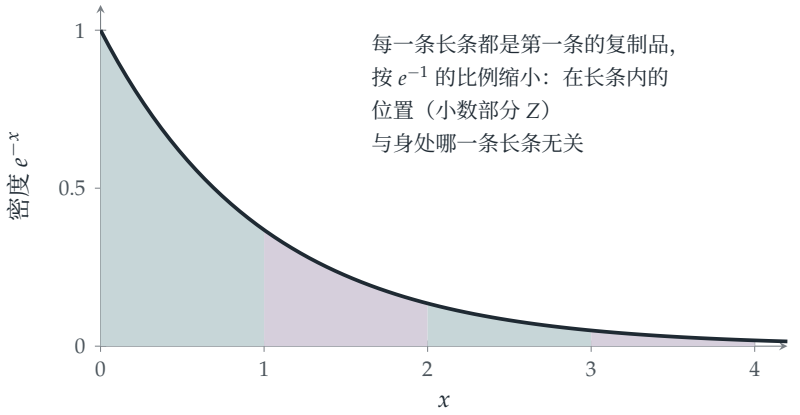


图 12.18.1 $\lambda = 1$ 的指数密度，分割成单位宽度的长条。由于 $e^{-(n+z)} = e^{-n}e^{-z}$，每一条长条的形状都相同。$X$ 落在哪一条长条（整数部分 Y），与它落在长条内的甚么位置（小数部分 Z），两者是独立的。

回顾

- **逐段求和。** Z 的分布和联合概率（joint probability），都是把各区间 $[n, n+1)$ 上的概率相加而得的。
- **尾部公式。** $\mathrm{E}(Z) = \int \mathrm{P}(Z > z)\,\mathrm{d}z$ 让我们不必求出密度。
- **能够分解即是独立。** 联合概率分拆成一个 n 的函数乘以一个 z_1、z_2 的函数。

背后的数学

指数分布为何特别

Y 与 Z 的独立性直接来自 $e^{-\lambda(n+z)} = e^{-\lambda n}e^{-\lambda z}$，而这正是无记忆性质的另一副面貌。事实上，指数分布是唯一没有记忆的连续分布：若对所有 $s, t \geq 0$ 都有 $\mathrm{P}(X > s+t) = \mathrm{P}(X > s)\mathrm{P}(X > t)$，则 $g(t) = \ln \mathrm{P}(X > t)$ 满足 $g(s+t) = g(s) + g(t)$，而（对于单调（monotonic）的 g）这就迫使 $g(t) = -\lambda t$。正因如此，指数分布可用来描述那些“随机”发生、没有任何累积过程的事件的等候时间：放射性衰变、顾客的到达、不会损耗的电子元件的故障。

最小值与等候时间悖论

若 X_1 与 X_2 是独立的指数随机变量，率分别为 λ_1 和 λ_2，则 $\mathrm{P}(\min > x) = e^{-\lambda_1 x}e^{-\lambda_2 x}$，所以最小值服从率为 $\lambda_1 + \lambda_2$ 的指数分布。等候两个独立随机事件中最先发生的一个，就

如同以合并后的率等候单一事件（STEP III 2002 Q13，见练习）。无记忆性也解释了*等候时间悖论*（*waiting-time paradox*）：若巴士随机到站，平均间隔为 10 分钟，则一位在随机时刻到达的乘客平均要等 10 分钟，而不是 5 分钟，因为余下的等候时间是无记忆的；而乘客到达时所处的那个间隔，平均长达 20 分钟，因为较长的间隔更容易被碰上。

卢瑟福（Rutherford）与泊松过程（Poisson process）

1910 年，欧内斯特·卢瑟福与汉斯·盖革（Hans Geiger）以 7.5 秒为一个时段，数算一个放射源所放出的阿尔法粒子（alpha particles），发现这些计数服从泊松分布（Poisson distribution）；数学家 Harry Bateman 证明了，只要各次放射以恒定的率独立发生，结果就理应如此。这样一来，相邻两次放射之间的时间便服从指数分布。*泊松过程*，即间隔服从指数分布的一连串事件，如今已成为描述时间上随机性的基本模型，应用于电话通话、保险索偿以及神经元的放电；下一节 Section 12.19 会把它推广到空间。

思维技巧

以因式分解检验独立性

若两个随机变量的联合概率可分解为其中一个变量的函数乘以另一个变量的函数，则这两个随机变量独立。

在大学里，联合密度（joint density）能否分解是检验独立性的标准方法，你亦会借此在多变量问题中辨认出独立的分量，例如 Section 12.22 中服从正态分布（normal distribution）的点的半径与角度。

延伸练习

1 **两条巴士路线。**试做 STEP III 2002，第 13 题（`step.maths.org/questions/02-s3-q13`）。这道题利用两个指数随机变量的最小值，求出等候下一班巴士的时间。

2 **离散的无记忆性。**证明几何分布是无记忆的：$\mathrm{P}(Y \geq m+n \mid Y \geq m) = \mathrm{P}(Y \geq n)$。

3 **等候时间悖论。**巴士的到站是一个率为 λ 的泊松过程。证明包含某固定时刻 t（t 很大）的那个间隔，其期望长度为 $\frac{2}{\lambda}$，即平均间隔的两倍。

4 Z **的密度。**求 Z 的概率密度（probability density），并就 $\lambda = 1$ 和 $\lambda = 5$ 绘出其草图。

延伸阅读

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），论及指数分布与泊松过程。
- **Sheldon Ross, *Introduction to Probability Models***（Academic Press）。指数分布、泊松过程与排队。
- **David Spiegelhalter, *The Art of Uncertainty***（Pelican）。日常生活中的随机性与等候。

12.19 平面上的随机点

题目

STEP II 2012，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/12-s2-q13`

引子

假如超级市场、树木、恒星或流动电话发射站“随机”散布在某个区域内，你要走多远才到达最近的一个？答案只取决于密度 k（每单位面积的平均数目），而且有一个独特的形状，就是瑞利分布（*Rayleigh distribution*）。平均而言，最近的一个位于距离 $\frac{1}{2\sqrt{k}}$ 处：若每平方公里有一间超级市场，这距离就是 500 米。

这个模型称为空间泊松过程（*spatial Poisson process*），即时间上随机到达（Section 12.18）的二维版本。计算过程简短而优美，但它真正的价值在于充当基准：生态学家把实际的最近邻距离与这条公式比较，以判断树木或鸟巢究竟是聚集成群、随机散布，还是刻意彼此隔开。

动笔之前

想一想 一个平均值为 μ 的泊松随机变量（Poisson random variable），取值 0 的概率是多少？

$\mathrm{P}(0) = e^{-\mu}$。半径为 y 的圆面积是 πy^2，所以圆内超级市场数目的平均值为 $k\pi y^2$。

逐步解题

阶段 1 附近没有超级市场

在距离 y 以内没有超级市场的概率是 $e^{-\pi k y^2}$。

阶段 2 最近距离的分布

? **想一想** $Y \geq y$ 表示在距离 y 以内没有超级市场。那么 $\mathrm{P}(Y < y)$ 是甚么?

$\mathrm{P}(Y < y) = 1 - e^{-\pi k y^2}$，求导后便得到密度函数（density）

$$\mathrm{f}(y) = 2\pi k y e^{-\pi k y^2}, \qquad y \geq 0. \qquad \blacksquare$$

阶段 3 平均值

? **想一想** 可以用分部积分法（integration by parts）计算 $y\,\mathrm{f}(y)$ 的积分，也可以利用 $\mathrm{E}(Y) = \int_0^\infty \mathrm{P}(Y > y)\,\mathrm{d}y$。然后作代换，以便使用题目给出的高斯积分（Gaussian integral）。

$$\mathrm{E}(Y) = \int_0^\infty e^{-\pi k y^2}\,\mathrm{d}y.$$

代入 $x = y\sqrt{2\pi k}$，使得 $\pi k y^2 = \frac{1}{2}x^2$，便得 $\mathrm{E}(Y) = \frac{1}{\sqrt{2\pi k}}\sqrt{\frac{\pi}{2}} = \frac{1}{2\sqrt{k}}$。 $\blacksquare$

阶段 4 方差

$$\mathrm{E}(Y^2) = \int_0^\infty y^2 \cdot 2\pi k y e^{-\pi k y^2}\,\mathrm{d}y.$$

代入 $u = \pi k y^2$（因此 $\mathrm{d}u = 2\pi k y\,\mathrm{d}y$），上式变成 $\int_0^\infty \frac{u}{\pi k} e^{-u}\,\mathrm{d}u = \frac{1}{\pi k}$。因此

$$\mathrm{Var}(Y) = \frac{1}{\pi k} - \frac{1}{4k} = \frac{4-\pi}{4\pi k}. \qquad \blacksquare$$

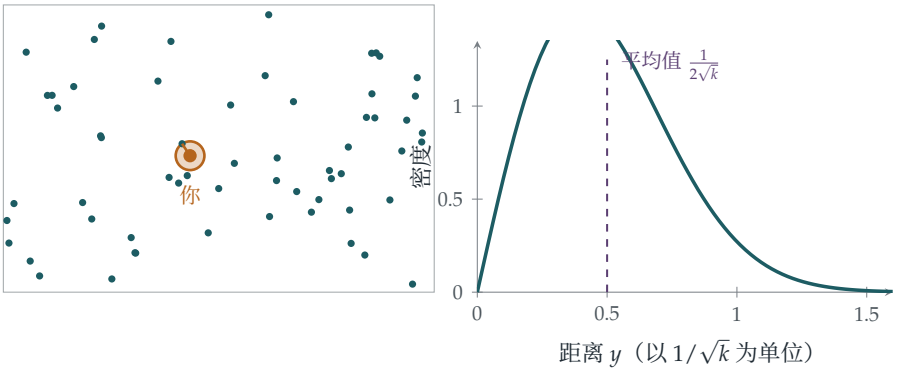


图 12.19.1 左：六十个点随机散布在面积为 24 的区域内（所以 $k = 2.5$），以及由选定的一点到其中最近一点的距离（琥珀色）。右：该距离的瑞利密度函数 $2\pi k y e^{-\pi k y^2}$，距离以 $1/\sqrt{k}$ 为单位。其平均值为 $\frac{1}{2\sqrt{k}}$。

回顾

- **把距离问题化为计数问题。**“最近的一个比 y 更远”就是“在 y 以内一个也没有”。
- **先求累积分布，再求密度。**$\mathrm{P}(Y < y)$ 可以直接写出；求导后便得到密度函数。
- **以代换化为已知积分。**每个矩（moment）都化成了高斯积分或指数积分。

背后的数学

平面上的泊松过程

强度（intensity）为 k 的空间泊松过程，是一种随机散布的点集，满足：互不相交的区域内的点数彼此独立，而面积为 A 的区域内的点数服从平均值为 kA 的泊松分布。把极大量的点独立而均匀地投放到一片极大的区域上，并保持适当的平均密度，得到的就是这种过程：这正是“稀有事件定律”（law of rare events），它同样解释了错排（derangement）的现象（Section 9.8）。它是完全空间随机性（complete spatial randomness）的自然模型，实际的分布形态都以它为比较标准。*Clark–Evans 检验*（*Clark–Evans test*）由两位生态学家于 1954 年提出，把观察所得的平均最近邻距离与 $\frac{1}{2\sqrt{k}}$ 比较：小得多表示有聚集现象（种子落在母树附近），大得多则表示间隔规则（有领域行为的鸟类，或植物之间争夺水分）。

三维的情况

在三维空间中作同样的计算，可得到最近的恒星、分子或星系的距离满足 $\mathrm{P}(Y > y) = e^{-\frac{4}{3}\pi k y^3}$。钱德拉塞卡（Subrahmanyan Chandrasekhar）在 1943 年用它来研究一颗恒星受随机分布的邻近恒星所施的引力，物理学家则用它来计算气体分子的平均自由程（mean free path）。

瑞利

密度函数 $\frac{y}{\sigma^2} e^{-y^2/(2\sigma^2)}$ 以瑞利勋爵（Lord Rayleigh）命名，他在 1880 年研究声学理论时遇上它：它描述了许多相位随机的波叠加后的振幅。若一点的两个坐标是独立的正态随机变量（normal random variable），它也给出该点与原点距离的分布，这是理解它为何在这里出现的另一个角度（练习 3）。今天，它用来模拟流动电话信号的衰落，因为信号是沿着许多相位随机的路径到达接收器的。

思维技巧

先求分布函数

要求最小值、最大值或最近距离的分布时，$\mathrm{P}(Y > y)$ 往往很容易直接写出。求导便得到密度函数。

这个方法会在 Section 12.25 再次出现，用来处理许多随机寿命中最长的一个。在大学里，这是处理任何“顺序统计量”（order statistic）的标准第一步。

延伸练习

1 **中位数。**求到最近超级市场的距离的中位数（median），并与平均值比较。

2 **第二近的。**求到第二近的超级市场的距离的分布。（提示：$\mathrm{P}(Y_2 > y)$ 是在 y 以内最多只有一间超级市场的概率。）

3 **正态坐标。**若 U 和 V 是独立的 $\mathrm{N}(0, \sigma^2)$ 随机变量，证明 $\sqrt{U^2 + V^2}$ 的密度函数为 $\frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)}$。$\sigma$ 取何值时会得到超级市场的分布？

4 **Clark–Evans。**在一项调查中，一片面积 1 km^2 的树林内有 400 棵树，而一棵树到其最近邻的平均距离为 35 m。这与完全空间随机性相符吗？还是显示有聚集或间隔的现象？

延伸阅读

- **Peter Diggle, *Statistical Analysis of Spatial Point Patterns***（Arnold）。这方面的标准参考书，载有 Clark–Evans 检验及其后续发展的方法。
- **John Kingman, *Poisson Processes***（Oxford University Press）。一本优美的短篇著作，讨论任意维数的泊松过程。
- **Geoffrey Grimmett & David Stirzaker, *Probability and Random Processes***（Oxford University Press），论泊松过程的一章。

12.20 蒲丰投针

题目

STEP I 1993，第 16 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-q16`

引子

把一根针抛到铺有平行地板条的地板上。针横跨板缝的机会竟然与 π 有关，尽管眼前根本没有任何圆形。抛很多次并数一数横跨板缝的次数，就可以用实验来估算 π。这就是*布丰投针问题*（*Buffon's needle*），由布丰伯爵乔治-路易·勒克莱尔（Georges-Louis Leclerc, Comte de Buffon）于 1733 年提出，并于 1777 年由他本人解决：它是*几何概率*（*geometric probability*）的第一个问题，也是所有蒙地卡罗方法（Monte Carlo method）的始祖。

这个 π 来自角度：针的方向在半圈之内均匀分布（uniform），把针横跨平行线方向的“影子”长度对所有方向取平均，便得出因子 $\frac{2}{\pi}$。本题则以 $\cos^{-1}$ 的积分来完成这个计算。

动笔之前

想一想 针的中心与最近一条直线的距离为 x，而针与这些直线的垂直线所成的角为 y。针朝着该直线能伸展多远？

针的每一半长度为 a，它在垂直于直线方向上的分量是 $a\cos y$。因此，针与最近的直线相交，当且仅当 $a\cos y \geq x$。（由于 $a < b$，针不可能触及其他直线。）

逐步解题

阶段 1 积分

设 $t=\cos y$，当 t 由 0 变到 1 时，$y=\cos^{-1}t$ 由 $\frac{\pi}{2}$ 变到 0，而 $\mathrm{d}t=-\sin y\,\mathrm{d}y$。所以

$$\int_0^1 \cos^{-1}t\,\mathrm{d}t=\int_0^{\pi/2} y\sin y\,\mathrm{d}y=[-y\cos y]_0^{\pi/2}+\int_0^{\pi/2}\cos y\,\mathrm{d}y=0+1=1. \qquad \blacksquare$$

（或者：这个积分是 $\cos^{-1}$ 的图像与两条坐标轴之间的面积，它等于 $y=\cos x$ 由 0 到 $\frac{\pi}{2}$ 的曲线下面积，两者只是沿 $y=x$ 反射而已。）

阶段 2 给定位置

想一想 对于固定的 x，哪些角度 y 会使针与直线相交？这些角度的概率是多少？

若 $x\geq a$，针无法触及直线。若 $x<a$，相交的条件是 $\cos y\geq\frac{x}{a}$，即 $y\leq\cos^{-1}\frac{x}{a}$。由于 Y 在 $[0,\frac{\pi}{2}]$ 上均匀分布，

$$\mathrm{P}(\text{相交}\mid X=x)=\frac{2}{\pi}\cos^{-1}\frac{x}{a}\qquad(0\leq x<a). \qquad \blacksquare$$

阶段 3 对位置取平均

X 在 $[0,b]$ 上均匀分布，其密度（density）为 $\frac{1}{b}$，所以

$$\mathrm{P}(\text{相交})=\frac{1}{b}\int_0^a\frac{2}{\pi}\cos^{-1}\frac{x}{a}\,\mathrm{d}x=\frac{2a}{\pi b}\int_0^1\cos^{-1}t\,\mathrm{d}t=\frac{2a}{\pi b}, \qquad \blacksquare$$

其中用了代换 $x=at$。

停一停，试一试 若针的长度恰好等于半间距（$a=\frac{b}{2}$，即针的长度 $2a$ 等于 b），相交的概率是 $\frac{1}{\pi}$。要把 π 估算至小数点后两位，你需要抛多少次？



图 12.20.1 三十根长度为 $2a=1.4$ 的针，抛在相距 $2b=2$ 的直线上。与直线相交的针以紫红色表示。这里有 17 根相交，而期望值是 13.4 根：小样本只能给出粗略的估算，而布丰投针正是出名缓慢的 π 计算方法。

回顾

- **两个独立的坐标。**针的位置和角度互相独立（independent）且均匀分布，所以所求概率是一个二重积分（double integral），可以逐个变量计算。
- **先取条件，再求平均。**先固定 x，对角度取平均；然后再对 x 取平均。
- **代换显露出常数。**以 $x = at$ 作缩放，答案便化约为单一个通用的积分。

背后的数学

布丰的面条

以下的方法几乎不用计算就能看出答案。把针屈曲成任意形状，成为一根“面条”，并设 C 为它与直线相交的次数。根据期望值的线性（linearity of expectation），一条曲线相交次数的期望值，等于它各小段的期望值之和，所以与长度成正比：$\mathrm{E}(C) = \kappa L$，其中 κ 是某个常数。要求出 κ，可取面条为一个直径等于直线间距 $2b$ 的圆：它总是恰好相交两次，而其长度为 $2\pi b$。所以 $\kappa \cdot 2\pi b = 2$，得出 $\kappa = \frac{1}{\pi b}$。对于长度为 $2a < 2b$ 的直针，它最多相交一次，故 $\mathrm{P}(\text{相交}) = \mathrm{E}(C) = \frac{2a}{\pi b}$。这个论证出自 Joseph-Émile Barbier（1860 年），是概率推理中的瑰宝：那个 π 来自用来校准常数的圆。

估算 π，以及错误示范

若抛出 n 根针，其中 c 根相交，则 $\frac{c}{n} \approx \frac{2a}{\pi b}$，而 $\pi \approx \frac{2an}{bc}$。误差只会按 $\frac{1}{\sqrt{n}}$ 的速度缩小，所以要把 π 算至小数点后三位，大约需要抛一百万次。1901 年，意大利数学家 Mario Lazzarini 声称抛了 3408 次便得出 $\pi \approx \frac{355}{113} = 3.1415929\ldots$（准确至小数点后六位）。统计学家指出，这个结果好得令人难以置信：抛 3408 次，再配合一根精心挑选的针，恰好 1808 次相交便会得出 $\frac{355}{113}$，而只要在适当的时刻停止实验，就能制造出这个结果。这是关于选择性报告的经典警世故事。

布丰

布丰是一位博物学家，曾任巴黎皇家植物园园长，并著有一套 36 卷的自然史。他于 1733 年向法国科学院（Académie des Sciences）提出投针问题（以及一个使用方形地砖的版本，即“franc-carreau”游戏），并于 1777 年发表解答。他的目的是要说明，概率中不单可以运用算术，也可以运用几何。积分几何（integral geometry）——研究随机直线及其与曲线和曲面的交点的理论——正是由此发展而来，如今应用于体视学（stereology），即从二维切片估算三维结构（例如组织中的细胞）的大小和形状的科学。

思维技巧

把概率看成区域上的积分

当结果取决于数个独立的连续随机变量时，一个事件的概率就是联合密度（*joint density*）在该事件发生的区域上的积分。

在大学里，这成为连续概率的基础：联合密度、边际分布（marginal distribution）和变量代换（change of variable），这些全都会在 Section 12.21 和 Section 12.22 再次出现。

延伸练习

1 **长针。**若 $a > b$，针可能与多于一条直线相交。证明相交次数的期望值仍然是 $\frac{2a}{\pi b}$。

2 **面条。**把一根长度为 $2a$ 的针屈成直角（两臂各长 a）。相交次数的期望值是多少？

3 **方形地砖。**把一枚半径为 r 的硬币抛到铺有边长为 $s > 2r$ 的方形地砖的地板上。求硬币完全落在一块地砖之内的概率。

4 **Lazzarini 的针。**设 $a = \frac{5}{6}b$（使得 $\frac{2a}{\pi b} = \frac{5}{3\pi}$），证明抛 3408 次而有 1808 次相交，会给出估算值 $\pi \approx \frac{355}{113}$。

延伸阅读

- **Daniel Klain & Gian-Carlo Rota, *Introduction to Geometric Probability***（Cambridge University Press），此书由布丰投针问题讲起。
- **Lee Badger, "Lazzarini's lucky approximation of π"**，*Mathematics Magazine* 67 (1994)。一个可疑实验背后的统计学。
- **Paul Nahin, *Digital Dice***（Princeton University Press）。以蒙地卡罗方法处理概率问题。

12.21 折断一根棒

题目

STEP III 1990，第 16 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-s3-q16`

引子

在一根木棒上随机选两点，把它折断。三段木棒能否组成三角形？著名的答案是：概率（probability）为 $\frac{1}{4}$。如果能组成三角形，它是锐角（acute）三角形还是钝角（obtuse）三角形？通常是钝角三角形：所组成的三角形中，只有大约三分之一是锐角三角形。这类关于随机图形的问题，自十九世纪起便令概率学家乐此不疲，也是练习处理连续概率（continuous probability）的绝佳场地：区域内的一个随机点、一个界定子区域的条件，以及一个计算其面积的积分。

本题采用两种不同的折棒方法，并在每种情况下问出现钝角的可能性有多大。答案 $3-2\sqrt{2}$ 和 $3-4\ln 2$ 都是出人意表的数，却源自简单的几何。

动笔之前

想一想 三个长度能组成三角形，当且仅当每一个都小于其余两个之和。如果三个长度之和为 L，这对最长的一个有何含意？边 c 所对的角何时是钝角？

$c < a+b$ 连同 $a+b+c=L$，即 $c<\frac{L}{2}$；所以三段能组成三角形，正正是当每一段都短于总长度的一半。根据余弦公式（cosine rule），c 所对的角是钝角，正正是当 $c^2 > a^2+b^2$。

逐步解题

阶段 1 先折一次，再对半分

三段的长度是 X、$\frac{1-X}{2}$、$\frac{1-X}{2}$。两段相等的木棒必定短于 $\frac{1}{2}$，所以能组成三角形当且仅当 $X<\frac{1}{2}$。这个三角形是等腰（isosceles）三角形；两个相等的角都是锐角，所以只有当 X 所对的角是钝角时，它才是钝角三角形：

$$X^2>2\Big(\frac{1-X}{2}\Big)^2=\frac{(1-X)^2}{2},\qquad \text{即}\qquad \sqrt{2}X>1-X,\quad X>\frac{1}{1+\sqrt{2}}=\sqrt{2}-1.$$

已知 $X<\frac{1}{2}$ 时，X 在 $[0,\frac{1}{2}]$ 上均匀分布（uniform），所以

$$\mathrm{P}(\text{钝角}\mid\text{三角形})=\frac{\frac{1}{2}-(\sqrt{2}-1)}{\frac{1}{2}}=3-2\sqrt{2}\approx 0.172.\qquad\blacksquare$$

阶段 2 第二种构造方法

想一想 设 (x,y) 位于以 $(1,0)$、$(1,1)$、$(0,1)$ 为顶点的三角形内，对边长 x、y 和 $2-x-y$ 检查三条三角不等式（triangle inequality）。

该区域是 $\{x\le 1, y\le 1, x+y\ge 1\}$。三条不等式分别是：$x+y\ge 2-x-y$，即 $x+y\ge 1$；$x+(2-x-y)\ge y$，即 $y\le 1$；以及 $y+(2-x-y)\ge x$，即 $x\le 1$。它们在区域内全部成立（在区域的边上等号成立，此时三角形是退化（degenerate）的）。 ■

想一想 B 处的角所对的边是 $CA=y$。把条件 $y^2>x^2+(2-x-y)^2$ 写成关于 y 的不等式。

展开得 $y^2>x^2+4+x^2+y^2-4x-4y+2xy$，所以 $4y-2xy>2x^2-4x+4$，即 $2y(2-x)>2(x^2-2x+2)$。由于 $2-x>0$，

$$\angle ABC\text{ 为钝角}\iff y>\frac{x^2-2x+2}{2-x}.\qquad\blacksquare$$

阶段 3 面积

想一想 以 $2-x$ 除 x^2-2x+2。然后求区域中位于曲线上方那部分的面积，再除以区域的面积 $\frac{1}{2}$。

由于 $x^2-2x+2=-x(2-x)+2$，曲线是 $y=-x+\frac{2}{2-x}$。它由 $(0,1)$ 伸延至 $(1,1)$，而当 $0<x<1$ 时，它位于直线 $y=1-x$ 之上、$y=1$ 之下（图 12.21.1）。钝角区域位于曲线与 $y=1$ 之间，面积为

$$\int_0^1\Big(1+x-\frac{2}{2-x}\Big)\mathrm{d}x=1+\frac{1}{2}+\big[2\ln(2-x)\big]_0^1=\frac{3}{2}-2\ln 2.$$

除以 $\frac{1}{2}$：$\mathrm{P}(\angle ABC\text{ 为钝角})=3-4\ln 2\approx 0.227$。 ■

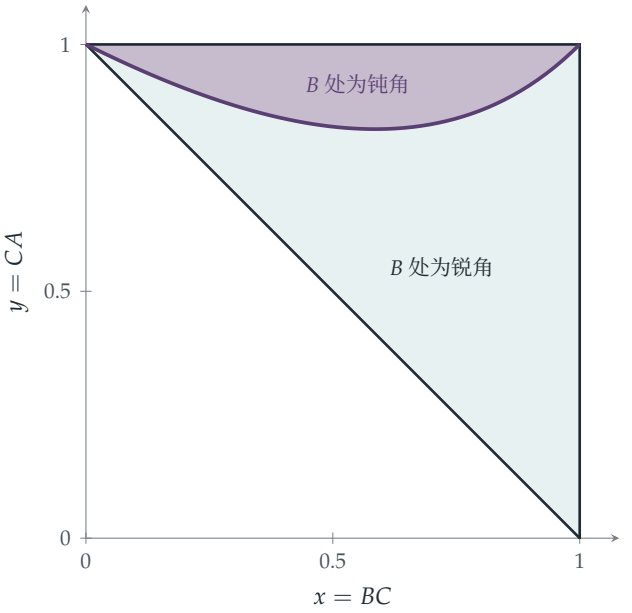


图 12.21.1 第 (ii) 部分中 (x, y) 所有可能取值构成的三角形。在曲线 $y = \frac{x^2-2x+2}{2-x}$ 上方（紫色区域），B 处的角是钝角；这个区域占三角形的比例为 $3 - 4\ln 2 \approx 0.23$。

回顾

- **把几何转化为不等式。**“能组成三角形”和“是钝角三角形”都化成了各长度之间的不等式。
- **连续变量的条件概率（conditional probability）。**已知 $X < \frac{1}{2}$ 时，X 在较小的区间上均匀分布。
- **概率即面积之比。**对于区域内均匀分布的点，概率等于面积除以总面积。

背后的数学

经典的折棒问题

在两个独立而均匀分布的点 U 和 V 处把木棒折断。数对 (U, V) 在单位正方形上均匀分布，而三段能组成三角形，正正是当每一段都短于 $\frac{1}{2}$：这个条件切去四个小三角形，总面积为 $\frac{3}{4}$，余下的概率是 $\frac{1}{4}$。已知能组成三角形时，它是锐角三角形的概率是 $\frac{1}{4}$ 当中的 $3\ln 2 - 2 \approx 0.079$，即大约 32%。这三段木棒是均匀随机分割（uniform random partition）的一个例子，它们的联合分布（joint distribution）是狄利克雷分布（Dirichlet distribution），

这种分布出现在贝叶斯统计学（Bayesian statistics）中，也出现在描述资源如何随机分配的模型中。

怎样折断木棒，大有关系

请留意，题目中的两种方法，对听起来相似的问题给出了不同的答案 $3-2\sqrt{2}$ 和 $3-4\ln 2$，而两者又都与经典方法的答案不同。正如贝特朗（Bertrand）的弦（Section 12.3）一样，在指明具体程序之前，“随机”一词毫无意义。

随机三角形

关于随机三角形的问题，是维多利亚时代流行的趣题。路易斯·卡罗（Lewis Carroll）的 *Pillow Problems*（1893 年）问及“随机”选取的三角形是钝角三角形的概率，而他的答案，正如本节的答案一样，完全取决于他所选择的随机性模型。关于随机三角形的问题一直带来惊喜：例如，在正方形内均匀随机选取三点，它们组成钝角三角形的概率约为 0.725，这个数值到二十世纪才有人准确算出。

思维技巧

样本空间（sample space）的几何

把随机结果表示为区域内的一点，把事件表示为子区域，并以面积之比（或密度（*density*）的积分）计算概率。

在大学里，这幅图像是多元概率（multivariate probability）的基础，也是变量代换公式（change-of-variables formula）的基础；这条公式说明坐标改变时，密度如何随之变换。

延伸练习

1 **经典问题。**证明：若在两个独立而均匀分布的点处把木棒折断，则各段能组成三角形的概率是 $\frac{1}{4}$。

2 **折断较长的一段。**在一个均匀分布的点处把木棒折断，然后在较长的一段上取一个均匀分布的点，再把它折断。证明三段能组成三角形的概率是 $2\ln 2-1\approx 0.386$。

3 **锐角还是钝角。**在第 (ii) 部分中，求三角形 ABC 在 C 处（对边为 $AB=2-x-y$）有钝角的概率。由此求它有任何一个角是钝角的概率。

4 **最长一段的期望长度。**对于经典的折棒问题，证明最长一段的期望（expected）长度是 $\frac{11}{18}$。

延伸阅读

- **Martin Gardner, *The Colossal Book of Mathematics***（Norton），书中收录了折棒问题。
- **Frederick Mosteller, *Fifty Challenging Problems in Probability***（Dover）。
- **Lewis Carroll, *Pillow Problems***（Dover，与 *A Tangled Tale* 合订）。维多利亚时代的趣题，包括随机三角形。

12.22 没有平均值的分布

题目

STEP III 1995，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/95-s3-q12`

引子

把许多次量度的结果取平均，随机误差便会互相抵消：平均值会稳定下来，趋向真实的数值。这条*大数定律*（*law of large numbers*）是整个统计学的基石。但它暗藏一个假设：这些量度必须有有限的平均值（mean）。本题构造了一个再自然不过的随机变量（random variable），即两个独立（independent）标准正态（standard normal）变量之比，而它根本没有平均值。它的平均永远不会稳定下来；事实上，一百万个这样的变量的平均，其分布与单独一个的分布完全相同。

这就是*柯西分布*（*Cauchy distribution*）。它的密度（density）看来像一个人畜无害的钟形，但它的尾部实在太厚，偶然出现的极大数值足以左右任何平均。本题用一个漂亮的几何论证把它找出来：两个坐标之比就是一条直线的斜率，而径向对称令这条直线的角度呈均匀分布。

动笔之前

想一想 若 (U, V) 是一个随机点，其密度只取决于它与原点的距离，那么由原点到该点的直线与坐标轴所成的角 Θ，你能说些甚么？

把平面旋转并不会改变密度，所以每个方向的可能性都相同：角 Θ 在 $[0, 2\pi)$ 上呈均匀（uniform）分布。而 $\frac{U}{V}$ 只取决于方向，与距离无关。

逐步解题

阶段 1 联合密度（joint density）

由于 X 与 Y 独立，它们的联合密度就是各自密度的乘积：

$$\mathrm{f}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}.$$

它只通过 x^2+y^2（即与原点距离的平方）取决于 x 和 y：在以原点为圆心的圆上，密度是常数。这就是“径向对称”（radially symmetrical）的意思。

阶段 2 运用对称性

> ? **想一想** $\frac{U}{V} < k$ 可以在两种情况下发生，视乎 V 的正负号而定。这两种情况与旋转 180° 的变换 $(u,v) \mapsto (-u,-v)$ 有甚么关系？

$\frac{U}{V} < k$ 的意思是：要么 $V > 0$ 且 $U < kV$，要么 $V < 0$ 且 $U > kV$。旋转 $(u,v) \mapsto (-u,-v)$ 把第一个区域映射到第二个区域上，并且保持密度不变。所以两者的概率相同，于是

$$\mathrm{P}\Big(\frac{U}{V} < k\Big) = 2\mathrm{P}(U < kV, V > 0). \qquad \blacksquare$$

阶段 3 一个角

> ? **想一想** 在以 u 为横轴、v 为纵轴的平面上，画出区域 $\{V > 0, U < kV\}$。它是一个顶点在原点的楔形。它的角度是多少？

区域 $V > 0$ 是上半平面。通过原点的直线 $u = kv$ 与正 v 轴所成的角为 $\arctan k$（当 $k > 0$ 时偏向正 u 的一方）。上半平面中位于这条直线左方的部分，是一个角度为 $\frac{\pi}{2} + \arctan k$ 的楔形。根据径向对称，楔形的概率等于它的角度除以 2π，所以

$$\mathrm{P}\Big(\frac{U}{V} < k\Big) = 2 \cdot \frac{\frac{\pi}{2} + \arctan k}{2\pi} = \frac{1}{2} + \frac{1}{\pi} \arctan k.$$

对 k 求导，便得到密度

$$\mathrm{g}(k) = \frac{1}{\pi(1+k^2)}. \qquad \blacksquare$$

回顾

- **独立的正态变量具有径向对称性。**两个正态密度的乘积只取决于距离。
- **比值即方向。**$\frac{U}{V}$ 只取决于该点的角度，而这个角度呈均匀分布。
- **楔形的概率。**在径向对称之下，楔形的概率与它的角度成正比。

背后的数学

没有平均值

柯西分布的平均值本应是 $\int_{-\infty}^{\infty} \frac{k}{\pi(1+k^2)}\,\mathrm{d}k$，但 $\int_0^K \frac{k}{1+k^2}\,\mathrm{d}k = \frac{1}{2}\ln(1+K^2) \to \infty$：这个积分并不收敛（converge），所以平均值不存在。（对称性或会令人以为答案是 0，但正负两部分都是无穷大，而 $\infty - \infty$ 并没有值。）方差（variance）同样是无穷大。

更糟的是，n 个独立柯西变量的平均仍然是柯西变量，而且分布完全相同。（其中一个看出这点的方法，是利用特征函数（characteristic function）$\mathrm{E}(e^{itX}) = e^{-|t|}$。）所以取平均完全无补于事：一百万次量度的平均，并不比单独一次可靠。图 12.22.1 显示了实际情况：正态样本的累计平均会稳定下来，而柯西样本的累计平均却永远东歪西倒，不时被偶然出现的极大数值撞得偏离原位。

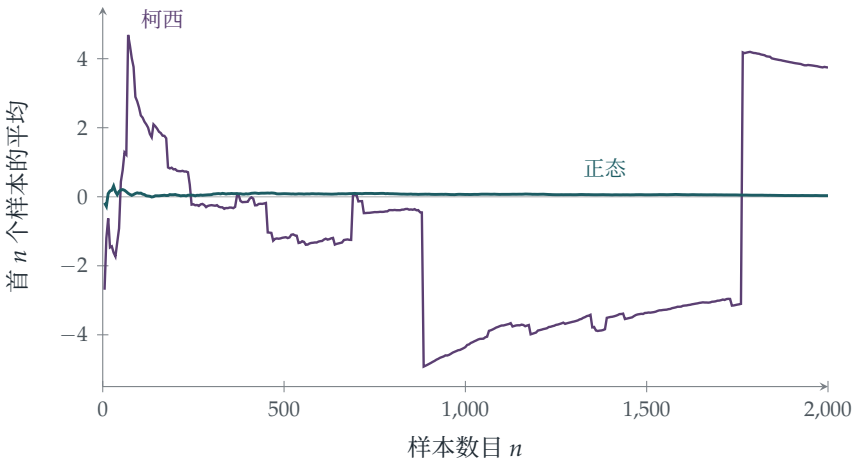


图 12.22.1 2000 个独立样本的累计平均。对于标准正态样本（蓝绿色），平均值一如大数定律所预测，稳定下来并趋向 0。对于柯西样本（紫红色），它永远不会稳定：每隔一段时间，单单一个极大的数值就会把它抛到另一个位置。

柯西变量从何而来

拿一支电筒，朝一个均匀随机的方向照向一面墙：光点在墙上的位置便服从柯西分布。在物理学中，同一个密度称为*劳仑兹分布*（*Lorentzian*），用来描述谱线和共振峰的形状（比较 Section 11.1）。在金融学中，像柯西分布这类厚尾（heavy-tailed）分布，会用来为偶尔出现的股灾建立模型，而正态模型往往严重低估这类事件。这对统计学是一个严肃的教训：建基于平均值和方差的方法，遇上厚尾数据便会失效，这时就需要稳健的方法，例如使用中位数（median）（柯西分布的中位数完全没有问题）。

泊松、柯西与一场论战

这个分布在 1824 年由泊松（Siméon Denis Poisson）研究过，他注意到它违反大数定律。柯西（Augustin-Louis Cauchy）则在 1853 年与 Irénée-Jules Bienaymé 就最小二乘法（method of least squares）公开论战时遇上它：柯西用它来说明，当误差属于厚尾时，最小二乘法可以是一个很差的方法。这个分布最终冠上了柯西的名字，正是 Stephen Stigler 所称的“史蒂格勒命名定律”（Stigler's law of eponymy）的一例：没有任何科学发现是以其最初发现者命名的。

思维技巧

检查期望值是否存在

使用平均值或方差之前，先确定定义它的积分是收敛的。对称性并不能令一个发散的积分变成零。

在大学里，可积性条件是概率论中每一条定理的一部分：大数定律需要有限的平均值，中心极限定理（central limit theorem）（Section 12.24）需要有限的方差。当这些条件不成立时，柯西分布就是标准的反例。

延伸练习

1 **电筒。**一支电筒与一面长而直的墙相距 1，所指的方向与垂直线成角 Θ，其中 Θ 在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 上呈均匀分布。证明光点的位置 $\tan\Theta$ 服从柯西分布。

2 **倒数。**证明若 C 服从柯西分布，则 $\frac{1}{C}$ 亦然。（提示：$\frac{U}{V}$ 与 $\frac{V}{U}$。）

3 **中位数行得通。**证明柯西分布的中位数是 0，而它的四分位数（quartiles）是 ± 1。

4 **半径与角度。**设 X、Y 为独立的标准正态变量，通过转换成极坐标（polar coordinates），证明 $R^2 = X^2 + Y^2$ 服从平均值为 2 的指数分布（exponential distribution），并且与角度独立。（这是用来产生正态随机数的 Box–Muller 方法的基础。）

延伸阅读

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），论随机变量的变换。

- **Nassim Nicholas Taleb, *The Black Swan***（Penguin）。讲述厚尾现象及其重要性，适合一般读者。
- **Stephen Stigler, *The History of Statistics***（Harvard University Press）。书中记述了柯西与 Bienaymé 的论战。

12.23 由概率得出斯特灵公式

题目

STEP II 2016，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/16-s2-q13`

引子

100! 有多大? 要逐项乘出来得花点工夫：它有 158 位数字。斯特灵公式（*Stirling's formula*）$n! \approx \sqrt{2\pi n}\,(n/e)^n$ 只用一行，就能给出误差在 0.1% 以内的答案。它是数学中最有用的近似式之一，在概率（probability）、统计物理和算法分析中经常用到。

它有很多种证明（proof），可以用积分（integral）（Section 4.4），也可以用欧拉–麦克劳林公式（Euler–Maclaurin formula）。本题给出的是最出人意表的一种：用概率来证明。平均值（mean）为 n 的泊松分布（Poisson distribution）近似于正态分布（normal distribution），在峰顶比较两者，$n!$ 的值就被确定下来。斯特灵公式中的 $\sqrt{2\pi}$，正正就是正态密度（normal density）中的 $\sqrt{2\pi}$。

动笔之前

想一想 设离散随机变量（discrete random variable）X 的平均值为 μ、方差（variance）为 σ^2，并且近似于正态分布，那么 $\mathrm{P}(X = \mu)$ 大约是多少？（试想一个宽度为 1、以 μ 为中心的区间的概率。）

按正态近似（normal approximation），$\mathrm{P}(X = \mu) \approx \mathrm{P}(\mu - \frac{1}{2} < X < \mu + \frac{1}{2})$，这大约等于宽度 1 乘以峰顶处的密度：$\frac{1}{\sigma\sqrt{2\pi}}$。

逐步解题

阶段 1 十六枚硬币

$X \sim \mathrm{B}(16, \tfrac{1}{2})$ 的平均值为 8，方差为 $16 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 4$，所以 $\sigma = 2$。由正态近似，

$$\mathrm{P}(X = 8) \approx \frac{1}{2\sqrt{2\pi}} \approx 0.1995.$$

准确值是 $\binom{16}{8}/2^{16} = \tfrac{12870}{65536} \approx 0.1964$，两者相差不到 2%。

阶段 2 中央二项式系数

> **想一想** $\mathrm{B}(2n, \tfrac{1}{2})$ 的平均值和标准差（standard deviation）是多少？分别写出 $\mathrm{P}(X = n)$ 的准确值和近似值。

平均值是 n，方差是 $\tfrac{n}{2}$，所以 $\sigma = \sqrt{n/2}$。准确值是 $\mathrm{P}(X = n) = \binom{2n}{n}\tfrac{1}{2^{2n}}$；近似值是 $\tfrac{1}{\sqrt{n/2}\sqrt{2\pi}} = \tfrac{1}{\sqrt{\pi n}}$。令两者相等，

$$\frac{(2n)!}{(n!)^2\, 2^{2n}} \approx \frac{1}{\sqrt{\pi n}}, \qquad (2n)! \approx \frac{2^{2n}(n!)^2}{\sqrt{n\pi}}. \qquad \blacksquare$$

阶段 3 斯特灵公式

> **想一想** 参数为 n 的泊松分布，其平均值和方差都等于 n。分别写出 $\mathrm{P}(X = n)$ 的准确值和近似值。

准确值是 $\mathrm{P}(X = n) = e^{-n}\tfrac{n^n}{n!}$。当 n 很大时，泊松分布近似于正态分布（它是 n 个独立的 Poisson(1) 变量之和），其中 $\sigma = \sqrt{n}$，所以 $\mathrm{P}(X = n) \approx \tfrac{1}{\sqrt{2\pi n}}$。令两者相等，

$$n! \approx \sqrt{2\pi n}\, e^{-n} n^n. \qquad \blacksquare$$

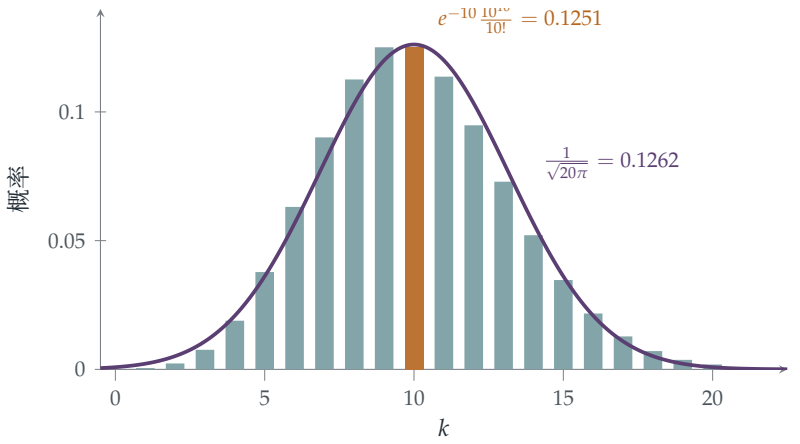


图 12.23.1 平均值为 10 的泊松分布（长条），以及平均值和方差与之相同的正态密度（紫色曲线）。在峰顶，泊松概率 $e^{-10}10^{10}/10!$（琥珀色）与正态密度 $\frac{1}{\sqrt{2\pi\cdot 10}}$ 十分接近。把这个近似等式移项整理，就是斯特灵公式。

回顾

- **一点上的概率 ≈ 密度乘以宽度。**对于格点分布（lattice distribution），某一点上的正态密度近似于该点的概率。
- **准确公式遇上近似式。**二项分布（binomial distribution）和泊松分布的概率是准确的；令它们等于各自的正态近似，便得到阶乘（factorial）的估计。
- **按公式的形状选择分布。**$\binom{2n}{n}$ 令人想到 $\mathrm{B}(2n, \frac{1}{2})$；$\frac{n^n}{n!}$ 令人想到 $\mathrm{Poisson}(n)$。

背后的数学

这个近似有多准确？

这个论证只是启发式（heuristic）的：它在*单独一点上*使用正态近似，而一般的中心极限定理（central limit theorem）并不足以支持这样做。它有一个严格的版本，称为*局部中心极限定理*（*local central limit theorem*），所做的正是这件事，并使这个论证成为真正的证明。斯特灵公式其实出奇地准确：比值 $\frac{n!}{\sqrt{2\pi n}(n/e)^n}$ 大约是 $1 + \frac{1}{12n}$，所以即使 $n = 1$，它给出的也是 0.92 而不是 1；而当 $n = 10$ 时，它只偏小了 0.8%。完整的级数（series）$1 + \frac{1}{12n} + \frac{1}{288n^2} - \cdots$ 是*渐近级数*（*asymptotic series*）的一个例子：对于大的 n，它给出的近似值越来越好；但对于固定的 n，它最终会发散（diverge）。

阶乘与正态分布为何相连

斯特灵公式和正态密度中的 $\sqrt{2\pi}$ 是同一个常数，因为 $\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}$。棣莫弗（Abraham de Moivre）在 1733 年正是循这条路找到二项分布的正态近似的：他需要在 n 很大时近似 $\binom{n}{k}$；他找到了 $n! \approx Cn^{n+1/2}e^{-n}$ 这个形式，却没有找到那个常数。斯特灵（James Stirling）证明了 $C = \sqrt{2\pi}$。正态分布是从阶乘的渐近性质中诞生的。

棣莫弗与斯特灵

棣莫弗是一位逃亡到伦敦的法国新教徒，靠当私人教师和为赌徒出谋献策维生。他在 1718 年出版了 *The Doctrine of Chances*（《机会论》），并在 1733 年的一本小册子中找到二项分布的正态近似，这是正态曲线首次面世。苏格兰数学家斯特灵在他的 *Methodus Differentialis*（1730）中发表了该常数的值，棣莫弗亦大方地把功劳归于他。阶乘的近似至今仍是活跃的课题：组合数学（combinatorics）和密码学（cryptography）都需要对极大的 n 计算 $n!$。

思维技巧

用概率去证明与概率无关的事情

把一个准确的概率与一个近似值作比较，可以得知公式中那些数的信息。

这种“概率方法（probabilistic method）”是双向的。这里概率给出了阶乘的估计；艾狄胥（Erdős）则用它来证明某些图（graph）的存在（Section 9.2）。到了大学，你会见到概率被用来证明分析、数论和组合数学中的定理。

延伸练习

1 **检验公式。**用斯特灵公式估计 10! 和 20!，并与真确值 3 628 800 和 2 432 902 008 176 640 000 比较。

2 **由 (iii) 到 (ii)。**证明第 (iii) 部分的斯特灵公式可推出第 (ii) 部分的近似式。

3 **100! 的位数。**利用 $\log_{10} n! \approx n\log_{10}\frac{n}{e} + \frac{1}{2}\log_{10}(2\pi n)$ 证明 100! 有 158 位数字。

4 **掷硬币打成平手。**利用第 (ii) 部分，证明掷一枚公平硬币 1000 次而恰好得到 500 次正面的概率约为 0.025。

延伸阅读

- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1**（Wiley），第 II 章，当中证明了斯特灵公式。
- **Ronald Graham, Donald Knuth & Oren Patashnik, *Concrete Mathematics***（Addison-Wesley），第 9 章，讲述渐近分析。
- **Stephen Stigler, *The History of Statistics***（Harvard University Press），讲述棣莫弗与正态曲线的诞生。

12.24 中心极限定理

题目

STEP III 2008，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/08-s3-q12`

引子

把许多微小而独立的随机效应加起来，不论每一个效应本身服从甚么分布，总和都会服从钟形的正态曲线。这就是*中心极限定理*（central limit theorem），它也许是概率论中最重要的定理。它解释了为何身高、量度误差和考试分数都大致服从正态分布（normal distribution），也解释了为何统计学家即使对背后的分布一无所知，仍能说出一个平均值有多可靠，只要该分布的方差（variance）有限便可（试与 Section 12.22 比较）。

本题在一个特殊情况下证明这个定理。一只青蛙随机跳跃，每次跳跃服从双边指数分布（two-sided exponential distribution），其密度函数（density）有一个尖峰，完全不像钟形。跳了许多次之后，牠的总位移却服从正态分布。所用的工具是*矩母函数*（moment generating function），它把独立变量之和化为乘积，再取对数，又把乘积化为和。

关于课程大纲的说明

矩母函数已于 2019 年从 STEP 考试大纲中删除。随机变量（random variable）X 的矩母函数是 $\mathrm{M}_X(\theta) = \mathrm{E}(e^{\theta X})$。它在 0 处的各阶导数给出各阶矩（moments）：$\mathrm{M}'(0) = \mathrm{E}(X)$，$\mathrm{M}''(0) = \mathrm{E}(X^2)$。若 X 与 Y 独立（independent），则 $\mathrm{M}_{X+Y} = \mathrm{M}_X\mathrm{M}_Y$，而且 $\mathrm{M}_{cX}(\theta) = \mathrm{M}_X(c\theta)$。若矩母函数收敛，则（在适当的意义下）分布也收敛。

动笔之前

想一想 密度函数 $\frac{1}{2}e^{-|x|}$ 是对称的。它的图像是甚么样子？它的平均值（mean）是多少？

它是一条帐篷形的曲线，在 0 处有一个高度为 $\frac{1}{2}$ 的尖峰，并向两边按指数衰减（图 12.24.1）。由对称性可知平均值为 0。

逐步解题

阶段 1 矩母函数

当 $|\theta| < 1$ 时，把积分在 0 处分开，

$$\mathrm{M}_X(\theta) = \frac{1}{2}\int_{-\infty}^{0} e^{(1+\theta)x}\,\mathrm{d}x + \frac{1}{2}\int_{0}^{\infty} e^{-(1-\theta)x}\,\mathrm{d}x = \frac{1}{2}\left(\frac{1}{1+\theta} + \frac{1}{1-\theta}\right) = \frac{1}{1-\theta^2}. \qquad \blacksquare$$

展开得 $\mathrm{M}_X(\theta) = 1 + \theta^2 + \theta^4 + \cdots$，所以 $\mathrm{E}(X) = 0$，$\mathrm{E}(X^2) = 2$（即 $\frac{\theta^2}{2!}$ 的系数）。方差为 2。

阶段 2 青蛙的位移

想一想 $Y = X_1 + \cdots + X_n$ 的矩母函数是甚么？$Y/\sqrt{2n}$ 的又是甚么？

由独立性得 $\mathrm{M}_Y(\theta) = (1-\theta^2)^{-n}$，再作缩放，

$$\mathrm{M}_{Y/\sqrt{2n}}(\theta) = \mathrm{M}_Y\left(\frac{\theta}{\sqrt{2n}}\right) = \left(1 - \frac{\theta^2}{2n}\right)^{-n}.$$

（除以 Y 的标准差（standard deviation）$\sqrt{2n}$ 作缩放，令方差变为 1。）

阶段 3 极限

想一想 取对数，并利用当 u 很小时 $\ln(1-u) \approx -u$。

$\ln \mathrm{M}_{Y/\sqrt{2n}}(\theta) = -n\ln\left(1 - \frac{\theta^2}{2n}\right) = -n\left(-\frac{\theta^2}{2n} - \frac{\theta^4}{8n^2} - \cdots\right) \to \frac{\theta^2}{2}$，所以

$$\mathrm{M}_{Y/\sqrt{2n}}(\theta) \to e^{\theta^2/2}, \qquad \blacksquare$$

这正是标准正态分布的矩母函数。

阶段 4 青蛙的机会

想一想 当 n 很大时，Y 近似服从 $\mathrm{N}(0, 2n)$。何时有 $\mathrm{P}(|Y| \geq 25) = 0.05$?

对于标准正态变量 Z，$\mathrm{P}(|Z| \geq 1.96) = 0.05$。因此我们需要 $\frac{25}{\sqrt{2n}} = 1.96$，由此得 $2n = \left(\frac{25}{1.96}\right)^2 \approx 162.7$，即 $n \approx 81.3$。由于落在远处的机会随 n 增大而增加，要有至少 5% 的机会，最少的跳跃次数约为 82。 ■

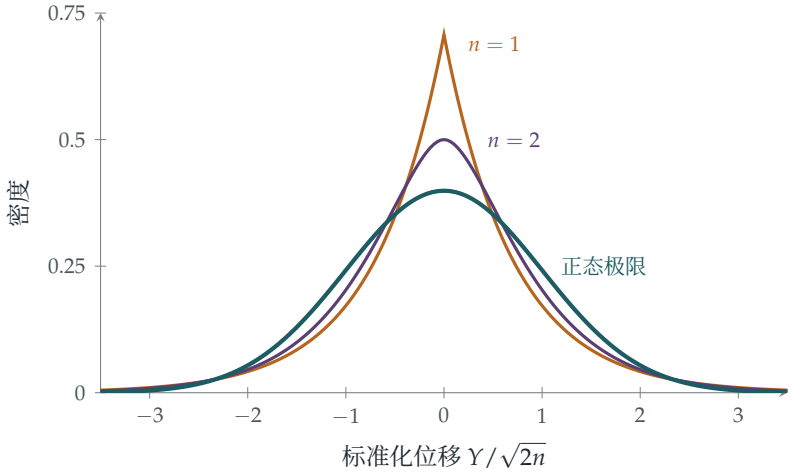


图 12.24.1 青蛙跳了 $n = 1$ 次（琥珀色：一个拉普拉斯（Laplace）尖峰）和 $n = 2$ 次（紫红色）之后，其标准化位移的密度函数，与标准正态密度函数（蓝绿色）的比较。只跳了两次，尖峰已经变得圆滑；中心极限定理指出，当 $n \to \infty$ 时，这些曲线收敛至正态曲线。

回顾

- **变换把和化为乘积。**独立变量之和的矩母函数，等于各变量矩母函数的乘积。
- **对数把乘积化为和。**$\ln\left(1 - \frac{\theta^2}{2n}\right)^{-n}$ 变成了一个和，其极限很容易求得。
- **标准化（standardise）。**除以标准差 $\sqrt{2n}$ 令方差保持为 1，所以极限既不是 0，也不是无穷大。

背后的数学

一般的定理

同样的论证适用于任何平均值为 0、方差 σ^2 有限的分布（还须有矩母函数，不过这个条件可以避开）。这时 $\mathrm{M}(\theta) = 1 + \frac{1}{2}\sigma^2\theta^2 + o(\theta^2)$，而且

$$\left(\mathrm{M}\left(\frac{\theta}{\sigma\sqrt{n}}\right)\right)^n = \left(1 + \frac{\theta^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{\theta^2/2}.$$

取极限之后只有方差留下来：原来分布的形状已被遗忘。这就是正态分布具有普遍性的原因。在大学里，这个定理是用特征函数（characteristic function）$\mathrm{E}(e^{i\theta X})$ 来证明的，特征函数必定存在，而分布的收敛则由莱维连续性定理（Lévy's continuity theorem）赋予严格的意义。

有多快?

中心极限定理说明极限情况下会发生甚么，却没有说明收敛得有多快。*Berry–Esseen* 定理（Berry–Esseen theorem）给出误差的上界：一个常数乘以 $\frac{\mathrm{E}|X|^3}{\sigma^3\sqrt{n}}$。对于像青蛙跳跃那样的对称分布，跳了几十次之后近似已相当准确。对于偏态（skewed）分布，或者在极端的尾部，即使 n 很大，近似也可能很差：正态近似会低估极罕见事件发生的机会，这个事实对保险和金融都有影响。

从棣美弗到 Lindeberg

棣美弗（De Moivre）在 1733 年发现了二项分布（binomial distribution）的正态近似（Section 12.23），拉普拉斯在 1810 年把它推广至更一般的变量之和，并首先提出量度误差是许多微小效应之和这个想法。严格的证明要到很久之后才出现：李亚普诺夫（Aleksandr Lyapunov）在 1901 年给出证明，Jarl Lindeberg 则在 1922 年给出证明，并找到了基本上是令定理成立的最弱条件。“中心极限定理”这个名称是波利亚（George Pólya）在 1920 年起的，意思是这个定理在概率论中处于中心地位。

思维技巧

变换、化简、再变换回来

一个困难的运算（把独立随机变量相加，即对它们的密度函数作卷积（*convolution*）），经过变换之后变成一个容易的运算（乘法）。在那里完成工作，然后诠释所得的结果。

同样的策略也是生成函数（generating functions）（Section 9.4）、分析学和工程学中的傅里叶（Fourier）变换和拉普拉斯变换，以及把乘法化为加法的对数背后

的原理。

延伸练习

1 **两次跳跃的精确结果。**证明 $Y = X_1 + X_2$ 的密度函数为 $\frac{1}{4}(1 + |y|)e^{-|y|}$，并验证其方差为 4。

2 **从二项到正态。**设 $X \sim \mathrm{B}(1, \frac{1}{2})$，并把它平移使平均值为 0（即取值 $\pm\frac{1}{2}$），求出 $\mathrm{M}_X(\theta) = \cosh\frac{\theta}{2}$，并证明 n 个这样的变量的标准化和的矩母函数趋向 $e^{\theta^2/2}$。

3 **青蛙的精确结果。**利用电脑（或以 $n = 2$ 时的精确密度函数作检验），看看当 $n = 82$ 时正态近似有多准确。

4 **为何柯西不行？**柯西分布（Cauchy distribution）（Section 12.22）没有矩母函数。论证中哪一步失效？n 个柯西变量的平均值又会怎样？

延伸阅读

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），关于生成函数和极限定理的一章。
- **Hans Fischer, *A History of the Central Limit Theorem***（Springer）。从拉普拉斯到 Lindeberg。
- **David Spiegelhalter, *The Art of Statistics***（Pelican），讲述为何正态分布无处不在。

12.25 众多之中的最大值

题目

STEP III 2025，第 11 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 2025 年 STEP 试卷（zip 文件）内的 STEP III 试卷：

`www.ocr.org.uk/Images/750486-2025-past-papers.zip`

引子

下一个世纪最高的洪水会有多高？一千个灯泡之中，寿命最长的那个能用多久？在 n 名参赛者之中，跑得最快的马拉松选手会有多快？这类关于极值（extremes）的问题，即许多随机数值之中最大或最小的一个，在工程、保险和气候科学中都极为重要，而极值的表现与平均值截然不同。

本题要找出 n 个指数分布（exponential distribution）寿命中最大者的分布。它的典型大小按 $\ln n$ 增长，虽然缓慢，却没有上限。然而它的离散程度却完全不会增长：一千个寿命中的最大值，其可预测的程度与十个寿命中的最大值大致相同。在极限情况下，它的分布形状就是耿贝尔分布（Gumbel distribution），也就是极值世界中与正态曲线相对应的角色。

动笔之前

想一想 $X_1, \ldots, X_n$ 中的最大者不超过 y，恰好等同于发生了甚么事？

$Y \le y$ 恰好在每个 $X_i \le y$ 时成立。利用独立性（independence），这个概率便化为一个乘积。

逐步解题

阶段 1 最大值的分布

每个 X_i 在 $y \ge 0$ 时都有 $\mathrm{F}(y) = 1 - e^{-\lambda y}$，由独立性可得

$$\mathrm{G}(y) = \mathrm{P}(X_1 \le y, \dots, X_n \le y) = (1 - e^{-\lambda y})^n. \qquad \blacksquare$$

阶段 2　分位数

> ? **想一想**　由 $\mathrm{G}(L) = \alpha$ 解出 L。然后解 $\mathrm{G}(U) = 1 - \alpha$。

由 $\left(1 - e^{-\lambda L}\right)^n = \alpha$ 得 $e^{-\lambda L} = 1 - \alpha^{1/n}$，所以

$$L(\alpha) = -\frac{1}{\lambda}\ln\!\left(1 - \alpha^{1/n}\right), \qquad U(\alpha) = -\frac{1}{\lambda}\ln\!\left(1 - (1-\alpha)^{1/n}\right). \qquad \blacksquare$$

阶段 3　n 很大的情况

> ? **想一想**　把 $\alpha^{1/n} = e^{(\ln \alpha)/n}$ 写出来。当 n 很大时，指数很小。那么 $1 - \alpha^{1/n}$ 近似于甚么？

$\alpha^{1/n} = e^{(\ln\alpha)/n} \approx 1 + \frac{\ln \alpha}{n}$，所以 $1 - \alpha^{1/n} \approx \frac{-\ln\alpha}{n} = \frac{\ln(1/\alpha)}{n}$。因此

$$\lambda L(\alpha) \approx -\ln\frac{\ln(1/\alpha)}{n} = \ln n - \ln\!\left(\ln\frac{1}{\alpha}\right). \qquad \blacksquare$$

同理，$\lambda U(\alpha) \approx \ln n - \ln\!\left(\ln \frac{1}{1-\alpha}\right)$。

阶段 4　增长与离散程度

中位数（median）是 $L(\frac{1}{2})$，而 $\lambda L(\frac{1}{2}) \approx \ln n - \ln\ln 2 \to \infty$。但是

$$\lambda\big(U(\alpha) - L(\alpha)\big) \approx \ln\!\left(\ln\frac{1}{\alpha}\right) - \ln\!\left(\ln\frac{1}{1-\alpha}\right),$$

当中的 $\ln n$ 项已互相抵消：区间的阔度与 n 无关。 $\blacksquare$

阶段 5　一个 90% 区间

> ? **想一想**　取 $\alpha = 0.05$。$\ln\frac{1}{\alpha} = \ln 20$ 和 $\ln\frac{1}{0.95}$ 大约是多少？

当 $\alpha = 0.05$ 时，$\ln 20 \approx 3$，所以 $\ln\ln\frac{1}{\alpha} \approx \ln 3$。另外 $\ln\frac{1}{0.95} = -\ln(1-0.05) \approx 0.05 = \frac{1}{20}$，所以 $\ln\ln\frac{1}{1-\alpha} \approx \ln\frac{1}{20} \approx -3$。因此阔度约为 $\frac{1}{\lambda}(\ln 3 + 3) \approx \frac{4.1}{\lambda}$，而 Y 以概率 0.9 落在 $[L(0.05), U(0.05)]$ 之内，这个区间的阔度约为 $4\lambda^{-1}$。 $\blacksquare$

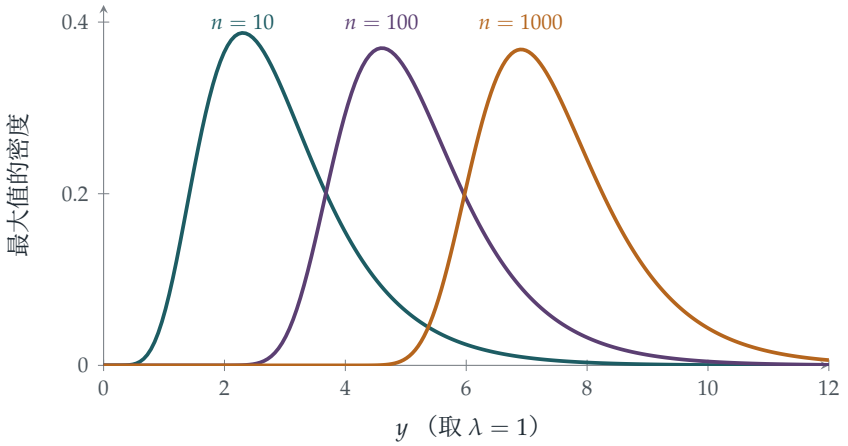


图 12.25.1 n 个指数分布寿命（平均值为 1）中最大者的密度，分别对应 $n = 10$、100 和 1000。n 每增至十倍，曲线便向右平移 $\ln 10 \approx 2.3$，但其形状和阔度基本上保持不变：这就是平移了 $\ln n$ 的耿贝尔分布。

回顾

- **所有值都不超过 y 时，最大值亦然。**独立性使最大值的分布函数成为一个幂。
- **以反函数求分位数。**解 $G(y) = \alpha$ 便得出区间两端的精确公式。
- **先近似，再比较。**L 和 U 的近似式都含有 $\ln n$，在计算阔度时互相抵消。

背后的数学

耿贝尔分布

由于 $P(\lambda Y - \ln n \le x) = \left(1 - \frac{e^{-x}}{n}\right)^n \to \exp(-e^{-x})$，平移后的最大值 $\lambda Y - \ln n$ 在极限下服从耿贝尔分布，其分布函数（distribution function）为 $\exp(-e^{-x})$。它的平均值是欧拉常数（Euler's constant）$\gamma \approx 0.577$。你其实早已遇过它：集齐全部 n 款优惠券所需的时间（Section 9.11）就是 n 个（接近独立的）等待时间中的最大值，这正是它约为 $n \ln n + \gamma n$ 的原因，而其波动属耿贝尔型，数量级为 n。

极值理论

费雪（Fisher）、Tippett 和格涅坚科（Gnedenko）有一个非凡的定理（1928–1943）：许多独立同分布（independent identically distributed）的变量，其最大值经适当平移和缩放后，极限形状只可能有三种：耿贝尔型（适用于轻尾分布，例如指数分布或正态分布）、弗雷歇（Fréchet）型（适用于重尾分布，例如柯西（Cauchy）分布）和韦布尔（Weibull）型

（适用于有硬性上限的变量）。这是中心极限定理（central limit theorem）（Section 12.24）在极值方面的对应结果，所属的学问称为极值理论（extreme value theory）。工程师用它来设计防洪设施和海上平台，以抵御“百年一遇的巨浪”；保险公司用它为巨灾保障定价；气候科学家则用它来估计热浪正在怎样变化。

耿贝尔与洪水

Emil Julius Gumbel（耿贝尔）是德国统计学家，直言不讳地反对纳粹，1932 年被迫离开大学教席，最终逃往美国。他在 1930 和 1940 年代把极值理论应用于河流泛滥的研究。他的著作 *Statistics of Extremes*（1958）使这门学问成为实用的工具。荷兰的三角洲工程（Delta Works）是在 1953 年北海洪灾之后为保护荷兰而兴建的，设计时就运用了极值统计学，以抵御预计约 10 000 年一遇的风暴潮。

思维技巧

最大值与最小值的分布

对于独立变量的最大值，$\mathrm{P}(\max \le y) = \prod \mathrm{P}(X_i \le y)$；对于最小值，$\mathrm{P}(\min > y) = \prod \mathrm{P}(X_i > y)$。

这一个想法就能处理最近的超级市场（Section 12.19）、第一班巴士（Section 12.18）和寿命最长的灯泡。在大学里，它是次序统计量（order statistics）和极值理论的起点。

延伸练习

1 **最大值的期望值。**证明 $\mathrm{E}(Y) = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right)$。（提示：根据无记忆性（memorylessness），n 个寿命中的第一个结束之后，余下的 $n-1$ 个会重新开始。）

2 **最小值。**证明 $X_1, \ldots, X_n$ 中的最小者服从参数为 $n\lambda$ 的指数分布。

3 **中位数的精确值。**求 Y 的精确中位数，并在 $n = 10$ 时与近似值 $\frac{1}{\lambda}(\ln n - \ln\ln 2)$ 比较。

4 **均匀分布的最大值。**若 $X_1, \ldots, X_n$ 在 $[0,1]$ 上均匀分布（uniform），求它们的最大值 M 的分布，并证明当 n 很大时，$n(1-M)$ 近似服从指数分布。这属于三种极值形状中的哪一种?

延伸阅读

- **Stuart Coles, *An Introduction to Statistical Modeling of Extreme Values***（Springer）。标准的入门书，附有洪水和海平面的真实数据。
- **Emil Julius Gumbel, *Statistics of Extremes***（Dover）。经典之作。
- **David Spiegelhalter, *The Art of Uncertainty***（Pelican），谈罕见事件以及应该怎样思考它们。

我们学到了甚么

本章开首我们说过：不确定性是可以精确地推理的，但单凭直觉却不行。我们提出了三个问题：题目问的究竟是甚么？巧妙的切入点在哪里？长远而言会发生甚么事？以下是本章二十五道题目所揭示的答案。

1. 题目问的究竟是甚么？

首四节谈的都是如何把问题弄清楚。“已知取出的是一块有编号的圆片”与“已知取出的是编号为 3 的圆片”，两者给出不同的答案（Section 12.1）。一种罕见疾病的检测即使准确度达 90%，检测呈阳性的病人真正患病的可能性也只有 9%，因为基本比率（base rate）举足轻重（Section 12.2）。“随机的弦”在我们说明它是如何选取之前，根本没有期望长度可言（Section 12.3），而折断的木棒（Section 12.21）说明的也是同一道理。此外，“独立”（independent）原来有好几种意思：事件可以两两独立而并非全体独立（Section 12.4），而不相关（uncorrelated）的变量也不一定独立。在每一个例子中，解决之道都是确切地写下有哪些可能的结果，以及它们是如何产生的。

2. 巧妙的切入点在哪里？

许多题目都既有冗长的计算，也有简短的洞见。

- **对称性（symmetry）。**过程本身无法分辨的两个座位（Section 12.5）；“正面较多”与“反面较多”是互补的镜像（Section 12.6）；每一对选手在任何一场比赛中相遇的可能性均等（Section 12.7）；对任何硬币而言，HT 与 TH 出现的可能性相同（Section 12.9）；径向对称的点，其方向服从均匀分布（Section 12.22）。
- **指示变量（indicator）与线性（linearity）。**把配对（Section 12.4）、绳结（Section 12.10）和比赛场数（Section 12.7）写成指示变量之和来计数，毋须求出完整的分布，便可得到平均值和方差。
- **状态与首步分析（first-step analysis）。**Penney 游戏（Section 12.8）、兔子洞穴与 Ecks 教授（Section 12.16），以及姓氏的消亡（Section 12.15），全都是以下一步为条件来求解的。
- **策略。**三人决斗（truel）中枪法最差的人应该故意射失（Section 12.13）；随机化的策略无法被对手利用（Section 12.14）；先拒绝最初 $\frac{1}{e}$ 的应征者是最优的做法（Section 12.12）。

3. 长远而言会发生甚么事?

本章最后的部分追踪随机性如何随时间演变,以及在大量试验中的表现。分支过程(branching process）要么消亡，要么爆发式增长，而即使是爆发式增长的过程，也可能因偶然而消亡（Section 12.15）。马可夫链（Markov chain）会趋于一个平稳分布（stationary distribution），并忘却自己的过去（Section 12.16）。服从指数分布的等候时间则完全没有记忆（Section 12.18）。平均值通常会稳定下来，但柯西（Cauchy）分布却不然，因为它没有平均值（Section 12.22）。根据中心极限定理（central limit theorem），大量独立效应之和会趋于正态分布（Section 12.24），而这种正态性之中更隐藏着 $n!$ 的斯特灵（Stirling）公式（Section 12.23）。最大值的表现又截然不同：它们按 $\ln n$ 的速度增长，离散程度却完全不会增长（Section 12.25）。而常数 e、π 和 γ 则不断出现，在杯子与邮票之中，也在针与绳圈之中。

4. 概率的工具箱

方法	你在哪里用过
写下样本空间（sample space）	条件概率（12.1）；贝特朗（Bertrand）（12.3）
贝叶斯定理（Bayes' theorem）与赔率（odds）	基本比率（12.2）；马可夫逆推（12.16）
对称性	12.5、12.6、12.7、12.22
指示变量与线性	12.4、12.7、12.10
首步分析	Penney（12.8）；兔子洞穴（12.16）；分支过程（12.15）
补事件（complement）与界	公平的决定（12.9）
以级数求期望值	邮票与 e（12.11）
参数化并求最优	秘书问题（12.12）
博弈论（game theory）式推理	三人决斗（12.13）；混合策略（12.14）
计数通往终点的路径	火柴盒（12.17）
先求分布函数（distribution function）	12.18、12.19、12.25
把概率看成面积或积分	蒲丰（Buffon）（12.20）；折断的木棒（12.21）
正态近似	斯特灵（12.23）；中心极限定理（12.24）

下一步

- 第 13 章由概率转向统计：从数据估计未知的量，由 Section 13.1 开始。
- Section 9.8、Section 9.9 和 Section 9.11 的计数问题，是本章内容在组合数学方面的一面。
- 马可夫链用到第 7 章的线性代数，而分支过程的不动点则用到 Section 8.7 的动力系统理论。

如果你想在入读大学之前再进一步

- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***：一流的入门课程，可在网上免费阅读，并附有讲课影片。
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***：简洁而严谨，属大学一年级程度。
- **Frederick Mosteller, *Fifty Challenging Problems in Probability***：经典谜题，附完整解答。
- **David Spiegelhalter, *The Art of Statistics* 及 *The Art of Uncertainty***：现实世界中的概率与风险。

拉普拉斯（Laplace）说过，概率就是化为计算的常识。本章说明了，之所以往往需要计算，正正是因为常识会失灵；而最好的计算不止于得出数字，它们还能解释个中道理。一种对称性、一个指示变量、一个选得恰当的状态，或一条极限定理，都能把一道谜题化为一个洞见，而大学的概率课程正是建基于这些洞见之上。

延伸练习答案

以下只是简短的答案和提示，并非完整解答：它们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，然后才看答案。

12.1 精确地取条件

1. 在 (iv) 和 (v) 中，其余 $k-1$ 块圆片都是从余下的 $b+n-1$ 块中随机抽出的，其中 $n-1$ 块有编号，所以两者都等于 $\frac{(n-1)\binom{b}{k-2}}{\binom{b+n-1}{k-1}}$。(vi) 的答案是 $\frac{\binom{n}{2}\binom{b}{k-2}}{\binom{b+n}{k}-\binom{b}{k}}$。设 $N=b+n$，两者只在 $(2N-nk)\binom{N-1}{k-1}=2k\binom{b}{k}$ 时才相等。这只会在个别特殊情况下发生，例如 $(b,n,k)=(2,4,3)$、$(3,3,4)$ 或 $(6,11,3)$，又或者是 $n=2$ 而所有圆片都被抽出这种平凡情况。对于题目所给的数值，等式并不成立。

2. 两个孩子的（性别，星期几）组合共有 $14^2=196$ 种，各种组合可能性相等，其中 $196-13^2=27$ 种包含至少一个星期二出生的男孩。在这些组合之中，$7^2-6^2=13$ 种是两个男孩。所以概率是 $\frac{13}{27}$。

3. 你选了 1 号门。若汽车在 1 号门后（概率 $\frac{1}{3}$），换门会输；若汽车在 2 号或 3 号门后，主持人必须打开另一扇后面是山羊的门，换门便会赢：概率为 $\frac{2}{3}$。若主持人随机打开一扇门，而门后碰巧是山羊，那么根据你所见，“汽车在你选的门后”和“汽车在另一扇关着的门后”这两个结果的可能性相等：概率为 $\frac{1}{2}$。

4. $\mathrm{P}(\geq 2 \mid \geq 1)=\frac{1-\mathrm{P}(0)-\mathrm{P}(1)}{1-\mathrm{P}(0)}\approx 0.370$，但 $\mathrm{P}(\geq 2 \mid \text{葵扇 A})=1-\binom{48}{12}/\binom{51}{12}\approx 0.561$。知道是哪一张 A，会令手上有第二张 A 的可能性增加。

12.2 基本比率谬误

1. 若尸体在那里，两者都说“有尸体”的概率是 pq；若不在，则是 $(1-p)(1-q)$。先验胜算（prior odds）为 1 : 1，所以后验概率（posterior probability）是 $\frac{pq}{pq+(1-p)(1-q)}=\frac{pq}{1-p-q+2pq}$。至于第二部分，在三种可能性相等的情况中，有两种情况下“A 在那里”为真，所以 $\mathrm{P}(\text{两者都在} \mid \text{两者都说}A)=\frac{pq}{2pq+(1-p)(1-q)}$。

2. 我们需要 $0.9\pi > \frac{1}{11}(1-\pi)$，即 $\pi > \frac{10}{109}\approx 9.2\%$。

3. 每一次 AL 测试呈阳性，都会把胜算乘以 $\frac{0.9}{1/11}=9.9$。由先验胜算 $\frac{1}{99}$ 开始：一次阳性之后，胜算为 0.1（概率 $\frac{1}{11}$）；两次之后，胜算为 0.99（概率约 0.497）；三次之后，胜算为 9.8（概率约 0.907）。

4. 在 100 名患者之中，99 人测试呈阳性；在 999 900 名没有患病的人之中，约有 9999 人呈阳性。所以约有 10 100 人被告知可能患病，而当中只有 99 人（约 1%）真正患病。

12.3 “随机”是甚么意思？

1. 弦（chord）比内接等边三角形的边长 $\sqrt{3}a$ 长的条件分别是：它的远端落在与近端相对的那三分之一圆周上（$\frac{1}{3}$）；它与 O 的距离小于 $\frac{a}{2}$（$\frac{1}{2}$）；它的中点落在半径为 $\frac{a}{2}$ 的圆内（$\frac{1}{4}$）。

2. 以 C 为中点的弦垂直于 OC，所以它就是通过 C 而角度为 $t = \frac{\pi}{2}$ 的弦，而 $L(\frac{\pi}{2}) = \frac{4a}{3}\left(0 + \frac{1}{2}\sec^2\frac{\pi}{4}\right) = \frac{4a}{3}$。

3. 弦长为 $2\sqrt{a^2 - d^2}$，平均值为 $\frac{2}{a}\int_0^a \sqrt{a^2 - d^2}\,\mathrm{d}d = \frac{2}{a}\cdot\frac{\pi a^2}{4} = \frac{\pi a}{2}$。

4. 随机端点：$\frac{2}{\pi\sqrt{4a^2 - \ell^2}}$。随机距离：$\frac{\ell}{2a\sqrt{4a^2 - \ell^2}}$。随机中点：$\frac{\ell}{2a^2}$。

12.4 两两独立，却非相互独立

1. 展开后得 $\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = (a - pq)(x_1 - x_2)(y_1 - y_2)$，所以 $a = pq$；其余三格随之确定，于是两者独立。例子：在 $(0,0)$ 放概率 $\frac{1}{3}$，在 $(\pm1, \pm1)$ 各点放概率 $\frac{1}{6}$。两个边际分布（marginal distribution）都是均匀的，而且 $\mathrm{E}(AB) = 0$，但 $\mathrm{P}(A = 0, B = 1) = 0 \neq \frac{1}{9}$。

2. 共有 21 对，每对相同的概率是 $\frac{1}{6}$，而各指示变量（indicator）两两独立（pairwise independent）。所以平均值是 3.5，方差（variance）是 $21 \cdot \frac{1}{6} \cdot \frac{5}{6} \approx 2.92$，标准差约为 1.71。得分 7 大约比平均值高出两个标准差。

3. 平均值是 $\binom{k}{2}\frac{1}{365}$。各指示变量两两独立，与骰子的情况完全一样，所以方差是 $\binom{k}{2}\frac{1}{365}\cdot\frac{364}{365}$。（它们并非相互独立（mutually independent）：若 1, 2 相同，2, 3 也相同，则 1, 3 亦相同。）

4. 取八个可能性相等的结果，$A = B = \{1,2,3,4\}$，$C = \{1,5,6,7\}$。那么 $\mathrm{P}(A \cap B \cap C) = \frac{1}{8}$，但 $\mathrm{P}(A \cap B) = \frac{1}{2} \neq \frac{1}{4}$。

12.5 最后一位乘客的座位

1. 对所有 $n \geq 3$，答案都是 $\frac{1}{3}$。最后一个座位现在是 S_1、S_2 和 S_n 其中之一，而每一次随机选择对这三个座位都是对称的。

2. 若 T_1 坐了 S_1，没有人要让位。若 T_1 坐了 S_3，则 T_1 和 T_3 都坐错了位。若 T_1 坐了 S_2，便由 T_2 选择：选 S_1 有 2 人坐错位，选 S_3 则有 3 人。期望值（expectation）是 $\frac{1}{3}\left(0 + 2 + \frac{5}{2}\right) = \frac{3}{2}$，与公式给出的 $\frac{2}{3} + \frac{1}{3} + \frac{1}{2}$ 一致。

3. 当 $r \geq 2$ 时 $P_r = \frac{1}{2}$，于是 $\frac{1}{n}\left(1 + \frac{n-2}{2}\right) = \frac{1}{2}$，所以递推关系（recurrence）与 $P_n = \frac{1}{2}$ 相容。

4. 最后一个空位必定是 S_1 或 S_n（任何其他 S_k 最迟在 T_k 登机时已被占去）。所以若它不是 S_n，就是 S_1。

12.6 多一枚硬币

1. 设 H' 为 Bob 首 $n+1$ 次掷币的结果，X 为他最后一次掷币的结果。若 $H' > H_A$，或者 $H' = H_A$ 而 X 是正面，Bob 便胜出。所以概率是 $\mathrm{P}(H' > H_A) + \frac{1}{2}\mathrm{P}(H' = H_A)$，而根据“多一枚硬币”的结果，$\mathrm{P}(H' > H_A) = \frac{1}{2}$。

2. 不成立。当 $n = 0$ 时，Bob 胜出的概率是 p。那个巧妙的证明把正面和反面对调，而这只有对公平硬币才是一种对称。

3. $H_A = H_B$ 恰好在 $H_A + (\text{Bob 的反面数目}) = n$ 时成立，即一个 $\mathrm{B}(2n, \frac{1}{2})$ 变量等于 n。数值：$\frac{1}{2}$、$\frac{3}{8}$、$\frac{5}{16}$。

4. 不成立：掷出六点就像一枚 $p = \frac{1}{6}$ 的硬币。当 $n = 0$ 时概率是 $\frac{1}{6}$，而一般而言概率低于 $\frac{1}{2}$。

12.7 淘汰赛中谁会遇上谁

1. 第 k 轮有 2^{n-k} 场比赛，而根据对称性，每场比赛由 $\binom{2^n}{2}$ 对球员中任何一对对赛的可能性都相等。把整个淘汰赛的 $2^n - 1$ 场比赛加起来，得 $\frac{2^n-1}{\binom{2^n}{2}} = \frac{2}{2^n}$。

2. 不会：在固定的对阵表中随机编排位置，对称性依然保持，同样的计数方法仍然适用。

3. X 在第 k 轮的对手来自一组 2^{k-1} 名球员，每人从该组胜出的可能性相等。所以 Y 是该对手的概率为 $\frac{2^{k-1}}{2^n-1} \cdot \frac{1}{2^{k-1}} = \frac{1}{2^n-1}$，而 n 轮的总和是 $\frac{n}{2^n-1}$。

4. 共有 $\binom{2n}{2} = n(2n-1)$ 场比赛；每一对都会相遇，所以概率是 1。

12.8 Penney 游戏

1. HHT 只有在头两次掷出 HH 时才会胜出；否则在凑成 HH 之前已出现一个 T，THH 必定先出现。所以 HHT 胜出的概率是 $\frac{1}{4}$。四人一起玩时：HH 之后，HHT 必胜；TT 之后，TTH 必胜；由 HT 和 TH 开始，HTT 和 THH 的胜出机会分别是 $\frac{2}{3}$ 和 $\frac{1}{3}$（或倒过来），于是每名玩家的胜出概率是 $\frac{1}{4} \cdot 1$ 或 $\frac{1}{4}\left(\frac{2}{3} + \frac{1}{3}\right)$，即 $\frac{1}{4}$。TTH 击败 THH 的概率是 $\frac{2}{3}$（第 (i) 部分）。

2. 对于 HH：$e_0 = 1 + \frac{1}{2}e_{\mathrm{H}} + \frac{1}{2}e_0$ 及 $e_{\mathrm{H}} = 1 + \frac{1}{2}e_0$，得 $e_0 = 6$。对于 HT：一旦有了一个 H，便等待一个 T（平均掷两次），而第一个 H 平均要掷两次：共 4 次。

3. HHH 和 TTT 分别被 THH 和 HTT 击败，概率为 $\frac{7}{8}$；HHT 和 TTH 分别被 THH 和 HTT 击败，概率为 $\frac{3}{4}$；其余四个（HTH、HTT、THH、THT）分别被 HHT、HHT、TTH、TTH 击败，概率为 $\frac{2}{3}$。HHH 和 TTT 最容易被击败。

4. 设 $A = \mathrm{HTT}$，$B = \mathrm{HHT}$，各相关数（correlation）为 $AA = 4$、$AB = 0$、$BB = 4$、$BA = 2$。B 胜出的胜算是 $(AA - AB) : (BB - BA) = 4 : 2$，即概率为 $\frac{2}{3}$。

12.9 用不公平的硬币作公平的决定

1. 所需轮数服从几何分布（geometric distribution）：第 (i) 部分的平均值是 $\frac{1}{p^2+q^2}$，而冯·诺伊曼（von Neumann）的方法（HT 或 TH）的平均值是 $\frac{1}{2pq}$。由于 $p^2 + q^2 - 2pq = (p-q)^2$，对所有 $p \neq \frac{1}{2}$，冯·诺伊曼的方法都较慢。不过，公平的却是它。

2. 用冯·诺伊曼的技巧制造公平的位元（bit）。每次取两个：00、01 和 10 分别代表三个选项，11 则表示重新开始。

3. 把掷币结果逐次与二进制展开式 $\frac{1}{3} = 0.010101\ldots$ 比较（把 H 读作 1，T 读作 0）。在第一次出现不同的那一掷，若掷币结果小于该位数字（T 对着 1），便宣布“正面”，否则宣布“反面”。这样得出正面的概率恰好是 $\frac{1}{3}$，平均用 2 次掷币。

4. 在未能决定的一轮之后，其中一面有两个，另一面有一个，而整个过程是这两个状态上的马可夫链（Markov chain）。解之得 $\mathrm{P}(\text{酒店}) = \frac{p^3}{p^3+q^3}$，当 $p = \frac{1}{2}$ 时等于 $\frac{1}{2}$：这时是公平的，其他情况则不然。

12.10 把绳头系起来

1. 第一次打结形成一个环的概率是 $\frac{1}{3}$（其余三个绳端之中，有一个属于同一条绳），而最后一次打结则必定形成一个环。环的数目：2 个的概率是 $\frac{1}{3}$，1 个的概率是 $\frac{2}{3}$；平均值是 $\frac{4}{3}$。

2. 余下 k 条绳时，一次打结不形成环的概率是 $1 - \frac{1}{2k-1}$。乘积 $\prod_{k=2}^{n} \frac{2k-2}{2k-1}$ 等于 $\frac{4^n}{2n\binom{2n}{n}}$，约为 $\frac{\sqrt{\pi n}}{2n} = \sqrt{\frac{\pi}{4n}}$。

3. 依次插入元素 $1, 2, \ldots, n$：元素 k 要么开始一个新的循环（cycle）（概率 $\frac{1}{k}$），要么放在之前 $k-1$ 个元素其中一个之后。把各指示变量加起来便得 H_n。

4. 各次打结的成环指示变量互相独立，概率为 $\frac{1}{2k-1}$，所以方差是 $\sum \frac{1}{2k-1}\left(1 - \frac{1}{2k-1}\right)$。另外 $\sum_{k\geq 1} \frac{1}{(2k-1)^2} = \frac{\pi^2}{6} - \frac{1}{4} \cdot \frac{\pi^2}{6} = \frac{\pi^2}{8}$。

12.11 邮票、杯子与 e

1. $\sum_{k\geq 2} \frac{k-1}{k!} t^k = te^t - (e^t - 1) = 1 + (t-1)e^t$。求导后得 $\mathrm{E}(N) = G'(1) = \mathrm{e}$ 及 $\mathrm{E}(N(N-1)) = G''(1) = 2\mathrm{e}$，所以 $\mathrm{Var}(N) = 3\mathrm{e} - \mathrm{e}^2$。

2. $\mathrm{P}(N=j)$ 在 $k=0,\ldots,j-1$ 的 $\mathrm{P}(N>k)$ 中出现，即共 j 次。这里 $\mathrm{P}(N>k)=\frac{1}{k!}$，所以 $\mathrm{E}(N)=\sum\frac{1}{k!}=\mathrm{e}$。

3. 完全没有改变：只有一个杯时，Jim 必定“猜”中。期望价格仍然是 e。

4. $\int_0^1\frac{(1-u)^2}{2}\,\mathrm{d}u=\frac{1}{6}$。一般而言，在 $u_k=u$ 处的截面是一个按 $1-u$ 缩放的 $(k-1)$ 维单纯形（simplex），所以 $V_k=V_{k-1}\int_0^1(1-u)^{k-1}\,\mathrm{d}u=\frac{V_{k-1}}{k}$。

12.12 何时停止寻找

1. 若最大的在第 $k\geq 2$ 位，你恰好在第一个雪糕是首 $k-1$ 个之中最大时才会得到它，概率为 $\frac{1}{k-1}$。对 k 取平均得 $\frac{1}{n}\sum_{k=2}^{n}\frac{1}{k-1}$，当 $n=4$ 时是 $\frac{11}{24}$。列出全部 24 种次序，可知得到第二、第三和第四大的概率分别是 $\frac{7}{24}$、$\frac{1}{6}$ 和 $\frac{1}{12}$。

2. 拒绝 $r=0,1,2,3$ 人分别给出 $\frac{1}{4}$、$\frac{11}{24}$、$\frac{5}{12}$、$\frac{1}{4}$。所以应拒绝一名候选人，成功概率为 $\frac{11}{24}$。

3. 首 n 人之中最好的一个位列头两名的概率是 $\frac{n}{N}+\frac{n}{N}\cdot\frac{N-n}{N-1}$，当 n 很小时接近 $\frac{n}{N}$ 的两倍。若头两名任何一个都可以接受，你便可以少看一些候选人才作决定。（以此为目标的最佳策略，成功概率约为 0.57。）

4. 设 $\mathrm{P}(r)=\frac{r}{N}\sum_{k=r}^{N-1}\frac{1}{k}$：当 $N=10$ 时，$r=2,3,4$ 给出 0.366、0.399、0.398；当 $N=100$ 时，$r=36,37,38$ 给出 0.3710、0.3710、0.3708，其中 $r=37$ 仅仅领先。

12.13 三人决斗

1. Algernon 掷失后，Bertie 掷向 Cuthbert。若他掷中 $(\frac{3}{5})$，便是 Algernon 对 Bertie，由 Algernon 先掷，Bertie 胜出的概率是 $1-\frac{2/5}{2/5+3/5-6/25}=\frac{9}{19}$。若他掷失，Cuthbert 会掷中 Bertie，然后除非 Algernon 先掷中 $(\frac{3}{5})$，否则 Cuthbert 胜出。所以 Bertie 胜出的概率是 $\frac{3}{5}\cdot\frac{9}{19}=\frac{135}{475}$，Cuthbert 则是 $\frac{2}{5}\cdot\frac{3}{5}=\frac{114}{475}$。

2. 对所有 $a<1$ 都成立。掷失给出 $\frac{3}{5}\cdot\frac{a}{a+\frac{3}{5}-\frac{3}{5}a}+\frac{2}{5}a$；掷中 Cuthbert 给出 $1-\frac{3/5}{3/5+a-\frac{3}{5}a}$（这时 Bertie 先向他掷），这个值总是较小；掷中 Bertie 则给出 0。

3. 不应该。若 Bertie 掷向 Algernon 而掷中，Cuthbert 随即掷中 Bertie；若他掷失，Cuthbert 照样掷中 Bertie。无论如何 Bertie 都一无所得，而掷向 Cuthbert 则有 $\frac{27}{95}$。

4. $p+(1-p)(1-r)p$，它小于 $\frac{p}{p+r-pr}$，即整个等比级数（geometric series）之和。

12.14 混合策略

1. 取 $b=(\frac{1}{4},\frac{1}{2},\frac{1}{4})$，则 $X=Y=Z=0$，所以无论对方怎样做，B 都赢得 0。对着 $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$，A 猜 1、2 和 3 的期望赢款分别是 $-\frac{1}{3}$、0 和 $\frac{1}{3}$，所以 A 应该每次都猜 3。

2. 收益矩阵（payoff matrix）的对角线上是 0，其余位置是 ± 1，把三个策略循环置换后矩阵不变，所以最优策略是 $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$。在原题中，相差 1 和相差 2 的罚则不同，所以各策略不能互换。

3. 若 X、Y、$Z \geq 0$ 且 $X + 2Y + Z = 0$，则三者都是 0。由此必有 $b_1 = b_3$ 及 $b_2 = 2b_3$。

4. 现在 $X = -b_2 + 3b_3$、$Y = b_1 - b_3$ 及 $Z = -3b_1 + b_2$，而 $X + 3Y + Z = 0$。令三者都等于零，得 $(\frac{1}{5}, \frac{3}{5}, \frac{1}{5})$。

12.15 灭绝还是爆发

1. 灭绝概率（extinction probability）是 $s = (q + ps)^2$ 在 $[0, 1]$ 内的最小根。两个根是 1 和 $\left(\frac{q}{p}\right)^2$，而第二个根恰好在 $p > \frac{1}{2}$ 时小于 1。

2. 给定 X_{n-1}, X_n 的平均值是 $2pX_{n-1}$，方差是 $2pqX_{n-1}$，所以 $\mathrm{Var}(X_n) = \mathrm{E}(2pqX_{n-1}) + (2p)^2\mathrm{Var}(X_{n-1})$。当 $p = \frac{1}{2}$ 时：$\mathrm{Var}(X_n) = \mathrm{Var}(X_{n-1}) + \frac{1}{2}$，所以 $\mathrm{Var}(X_n) = \frac{n}{2}$。

3. 在第 n 代或之前灭绝，表示第一代每一颗粒子的后代都在 $n - 1$ 代之内消亡。取 $p = \frac{3}{4}$：$e_1 = 0.0625$，$e_2 \approx 0.0881$，$e_3 \approx 0.0999$，递增并趋向 $\frac{1}{9} \approx 0.111$。

4. $s = \frac{1}{4}(1 + s + s^2 + s^3)$ 可因式分解为 $(s - 1)(s^2 + 2s - 1) = 0$，所以灭绝概率是 $\sqrt{2} - 1 \approx 0.414$。

12.16 顺向与逆向的马可夫链

1. 根据对称性 $p_B = p_D$；另外 $p_A = \frac{1}{3} + \frac{2}{3}p_B$、$p_B = \frac{1}{3}p_A + \frac{1}{3}p_C$ 及 $p_C = \frac{2}{3}p_B$。所以 $p_A = \frac{7}{15}$。由 A 出发，不经过 C 而在 A 离开：$x_A = \frac{1}{3} + \frac{2}{3}x_B$ 及 $x_B = \frac{1}{3}x_A$，所以 $x_A = \frac{3}{7}$。条件概率是 $\frac{3/7}{7/15} = \frac{45}{49}$。

2. 把第一行乘以 P：$(0.19, 0.33, 0.48)$。

3. 迹（trace）是 1.3，行列式（determinant）是 0.02，而 1 是一个特征值（eigenvalue）（各行之和为 1）。另外两个特征值之和为 0.3，积为 0.02：它们是 0.2 和 0.1。

4. $a_1 p_{12} = \frac{1}{20}$ 但 $a_2 p_{21} = \frac{1}{15}$，所以这条链不是可逆的（reversible）：倒过来运行时，它的转移概率并不相同（第 (iii) 部分的 q_{ij} 与 p_{ij} 不同）。

12.17 巴拿赫的火柴盒

1. 若第二次伸手要拿的与第一次是同一种，她当下便发现那包已空（$N = 2$）；否则她拿走另一颗糖果，并在第三次伸手时发现有一包已空。所以 $N = 2$ 或 3，概率各为 $\frac{1}{2}$。

2. $n = 1$：$\frac{1}{2} + \frac{1}{2} = 1$。$n = 2$：$\frac{6}{16} + \frac{6}{16} + \frac{4}{16} = 1$。

3. $\binom{2n-1}{n}2^{-(2n-1)} = \frac{1}{2}\binom{2n}{n}\cdot 2\cdot 4^{-n} = \binom{2n}{n}4^{-n}$。

4. 就余下的拖肥糖和薄荷糖数目作首步递推（first-step recursion），得 $\frac{64}{81} \approx 0.79$。

12.18 无记忆性

1. $\mathrm{P}(\min > t) = \mathrm{e}^{-\lambda_1 t}\mathrm{e}^{-\lambda_2 t}$。等候下一班巴士的时间服从指数分布（exponential distribution），率为每分钟 $\frac{1}{15}+\frac{1}{30}=\frac{1}{10}$，所以其中位数（median）是 $10\ln 2 \approx 6.9$ 分钟。

2. $\mathrm{P}(Y \geq k) = (1-p)^k$，所以比值是 $(1-p)^n$。

3. 包含 t 的那段间隔，由回溯至上一班巴士的时间和前望至下一班巴士的时间组成，两者都服从平均值为 $\frac{1}{\lambda}$ 的指数分布（当 t 很大时），所以间隔的期望长度是 $\frac{2}{\lambda}$。较长的间隔较有可能包含某个指定时刻。

4. 在 $[0,1]$ 上为 $\frac{\lambda \mathrm{e}^{-\lambda z}}{1-\mathrm{e}^{-\lambda}}$。当 $\lambda = 1$ 时，它由约 1.58 缓缓下降至 0.58；当 $\lambda = 5$ 时，它集中在 0 附近，由约 5.03 下降至 0.03。

12.19 平面上的随机点

1. 由 $1-\mathrm{e}^{-\pi k y^2} = \frac{1}{2}$ 得中位数为 $\sqrt{\frac{\ln 2}{\pi k}} \approx \frac{0.470}{\sqrt{k}}$，比平均值 $\frac{1}{2\sqrt{k}}$ 略小。

2. $\mathrm{P}(Y_2 > y) = \mathrm{e}^{-\pi k y^2}(1+\pi k y^2)$，所以概率密度（density）是 $2\pi^2 k^2 y^3 \mathrm{e}^{-\pi k y^2}$。

3. 在极坐标下，R 的密度是 $\frac{1}{2\pi\sigma^2}\mathrm{e}^{-r^2/(2\sigma^2)}\cdot 2\pi r$。当 $\sigma^2 = \frac{1}{2\pi k}$ 时，它与超级市场的分布吻合。

4. 取 $k = 4\times 10^{-4}$ 每 m^2，完全随机的假设预测平均值为 $\frac{1}{2\sqrt{k}} = 25\,\mathrm{m}$。观察所得的 35 m 较大，这显示树木是互相疏开的（竞争），而不是聚集成丛。

12.20 蒲丰投针

1. 把针切成很多小段。相交次数的期望值具有可加性，而每一小段的贡献与其长度成正比，所以无论 a 是多少，总和都是 $\frac{2a}{\pi b}$。

2. 根据同样的可加性（蒲丰投面条，Buffon's noodle），仍然是 $\frac{2a}{\pi b}$。

3. 圆心与每条边的距离都必须至少为 r：$\frac{(s-2r)^2}{s^2}$。

4. 由 $\frac{1808}{3408} \approx \frac{5}{3\pi}$ 得 $\pi \approx \frac{5\times 3408}{3\times 1808} = \frac{17\,040}{5424} = \frac{355}{113}$：完美得令人起疑。

12.21 折断一根棒

1. 要组成三角形，每一段都必须短于 $\frac{1}{2}$。在由折断点组成的单位正方形中，这个条件划出一个面积为 $\frac{1}{4}$ 的区域。

2. 假设第一个折断点在 $x < \frac{1}{2}$（另一情况是对称的），而较长的一段 $1-x$ 再被均匀地折断。新的两段都必须短于 $\frac{1}{2}$，其概率为 $\frac{x}{1-x}$。所以答案是 $2\int_0^{1/2} \frac{x}{1-x}\,\mathrm{d}x = 2\ln 2 - 1$。

3. 当 $(2-x-y)^2 > x^2+y^2$，即 $y < \frac{2-2x}{2-x}$ 时，C 处的角是钝角。这条曲线与 $x+y=1$ 之间的面积是 $\frac{3}{2} - 2\ln 2$，所以概率是 $3-4\ln 2$，与 B 处的相同。这并非巧合：三条边各自最多为 1，总和为 2，并且是均匀分布的，所以它们可以互换。因此 $\mathrm{P}(\text{钝角三角形}) = 9 - 12\ln 2 \approx 0.68$。

4. 根据容斥原理（inclusion–exclusion），最长的一段 M 满足 $\mathrm{P}(M > m) = 3(1-m)^2 - 3(1-2m)_+^2 + (1-3m)_+^2$，其中 $x_+ = \max(x, 0)$。在 $[0,1]$ 上积分得 $\mathrm{E}(M) = 1 - \frac{1}{2} + \frac{1}{9} = \frac{11}{18}$。

12.22 没有平均值的分布

1. $\mathrm{P}(\tan\Theta \le k) = \frac{1}{\pi}\left(\tan^{-1}k + \frac{\pi}{2}\right)$，求导后得 $\frac{1}{\pi(1+k^2)}$。

2. 把 U 和 V 对调，径向对称的密度保持不变，所以 $\frac{V}{U}$ 与 $\frac{U}{V}$ 有相同的分布。

3. 密度是对称的，所以中位数是 0；而 $\mathrm{P}(C \le 1) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}1 = \frac{3}{4}$。

4. $\frac{1}{2\pi}\mathrm{e}^{-r^2/2}\,r\,\mathrm{d}r\,\mathrm{d}\theta$，而令 $s = r^2$ 后这就是 $\frac{1}{2\pi}\,\mathrm{d}\theta \cdot \frac{1}{2}\mathrm{e}^{-s/2}\,\mathrm{d}s$：角度是均匀分布的，$R^2$ 服从平均值为 2 的指数分布，而且两者独立。

12.23 由概率得出斯特灵公式

1. $\sqrt{20\pi}(10/\mathrm{e})^{10} \approx 3\,598\,696$（约低 0.8%）及 $\sqrt{40\pi}(20/\mathrm{e})^{20} \approx 2.4228 \times 10^{18}$（约低 0.4%）。

2. $\frac{(2n)!}{(n!)^2} \approx \frac{\sqrt{4\pi n}(2n/\mathrm{e})^{2n}}{2\pi n(n/\mathrm{e})^{2n}} = \frac{4^n}{\sqrt{\pi n}}$。

3. $100\log_{10}\frac{100}{\mathrm{e}} + \frac{1}{2}\log_{10}(200\pi) \approx 157.97$，所以 100! 有 158 位数字。

4. $\binom{1000}{500}2^{-1000} \approx \frac{1}{\sqrt{500\pi}} \approx 0.0252$。

12.24 中心极限定理

1. 作卷积（convolution），得 $\frac{1}{4}\int \mathrm{e}^{-|x|-|y-x|}\,\mathrm{d}x = \frac{1}{4}(1+|y|)\mathrm{e}^{-|y|}$。方差是 $2+2=4$。

2. $\mathrm{E}(\mathrm{e}^{\theta X}) = \frac{1}{2}(\mathrm{e}^{\theta/2} + \mathrm{e}^{-\theta/2}) = \cosh\frac{\theta}{2}$。标准化后的和 $\frac{2}{\sqrt{n}}\sum X_i$ 的母函数（generating function）是 $\cosh^n\frac{\theta}{\sqrt{n}}$，其对数为 $n\ln\left(1 + \frac{\theta^2}{2n} + \cdots\right) \to \frac{\theta^2}{2}$。

3. 把 Y 写成两个独立的 $\Gamma(82,1)$ 变量之差，再作数值积分，得 $\mathrm{P}(|Y| \geq 25) \approx 0.051$：正态近似（normal approximation）非常准确。

4. 它没有矩母函数（moment generating function），甚至连平均值也没有。事实上，n 个独立柯西（Cauchy）变量的平均值本身也服从柯西分布：取平均完全不会令它变得集中。

12.25 众多之中的最大值

1. n 个寿命之中，最先完结的一个平均需时 $\frac{1}{n\lambda}$；之后余下 $n-1$ 个，一切重新开始，如此类推。所以 $\mathrm{E}(Y) = \frac{1}{\lambda}\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + 1\right)$。

2. $\mathrm{P}(\min > t) = (\mathrm{e}^{-\lambda t})^n = \mathrm{e}^{-n\lambda t}$。

3. 由 $(1-\mathrm{e}^{-\lambda y})^n = \frac{1}{2}$ 得 $\lambda y = -\ln(1-2^{-1/n})$。当 $n=10$ 时这是 2.704，而近似值为 $\ln 10 - \ln\ln 2 \approx 2.669$。

4. $\mathrm{P}(M \leq m) = m^n$，所以 $\mathrm{P}\big(n(1-M) > s\big) = \left(1-\frac{s}{n}\right)^n \to \mathrm{e}^{-s}$。这属于韦布尔（Weibull）型：各变量有上界，而最大值会挤向这个上界。

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 12 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。

