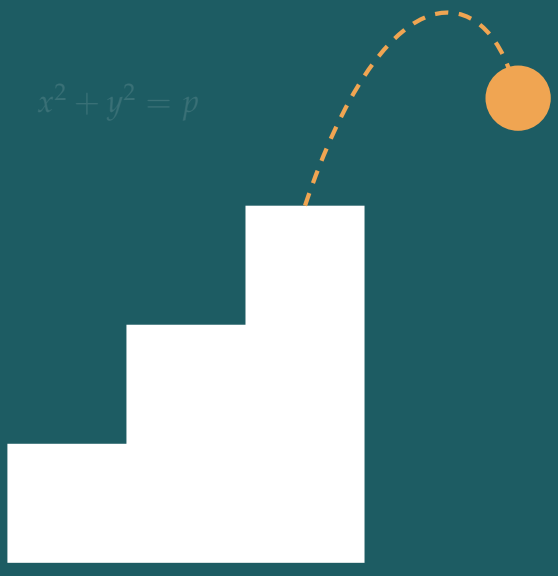
STEP 備試 · 大學數學

Beyond STEP

第 12 章

概率



KEITH McNULTY

Beyond STEP

以歷屆 STEP 試題為門
走進大學數學

第 12 章

概率

Keith McNulty

Beyond STEP：以歷屆 STEP 試題為門，走進大學數學

第 12 章：概率

本冊是 *Beyond STEP* 全書 13 章中的第 12 章。全書、各章的單行本、每一節的可列印工作紙，以及每一道 STEP 試題的詳細解答，均可在 `beyondstep.keithmcnulty.org` 免費閱讀及下載。

2026 年 11 月首次於網上發表。

本 PDF 內凡屬作者本人創作的內容，包括正文、解答、插圖、練習及答案，均按照 Creative Commons CC0 1.0 Universal 授權條款貢獻予公有領域：你可以為任何目的複製、改編及分享，毋須徵求許可。詳見 `creativecommons.org/publicdomain/zero/1.0`。

本書討論的 STEP 試題，版權屬 Cambridge University Press & Assessment 所有。本書並無轉載這些試題：每一題均附連結，通往官方來源的題目，即 STEP Support Programme 的試題資料庫或 OCR 的歷屆試卷。

如有錯漏，歡迎到 `github.com/keithmcnulty/step_book_mcnulty` 提交 issue 指正。

章

12

概率

不確定的事情，可以精確地推理。直覺呢，原來不行。

本章內容

12.1	精確地取條件	STEP I 2007
12.2	基本比率謬誤	STEP II 1998
12.3	「隨機」是甚麼意思？	STEP III 2022
12.4	兩兩獨立，卻非相互獨立	STEP III 2021
12.5	最後一位乘客的座位	STEP II 2021
12.6	多一枚硬幣	STEP I 2016
12.7	淘汰賽中誰會遇上誰	STEP II 樣本試卷
12.8	Penney 遊戲	STEP III 2001
12.9	用不公平的硬幣作公平的決定	STEP I 2019
12.10	把繩頭繫起來	STEP III 2008
12.11	郵票、杯子與 e	STEP I 1996
12.12	何時停止尋找	STEP I 2001
12.13	三人決鬥	STEP II 1995
12.14	混合策略	STEP I 1989
12.15	滅絕還是爆發	STEP II 2003
12.16	順向與逆向的馬可夫鏈	STEP III 1996
12.17	巴拿赫的火柴盒	STEP II 1992
12.18	無記憶性	STEP III 2021
12.19	平面上的隨機點	STEP II 2012
12.20	蒲豐投針	STEP I 1993
12.21	折斷一根棒	STEP III 1990
12.22	沒有平均值的分佈	STEP III 1995
12.23	由概率得出斯特靈公式	STEP II 2016
12.24	中心極限定理	STEP III 2008
12.25	眾多之中的最大值	STEP III 2025
A	練習答案	第 151 頁

機會的推理

概率論不過是化爲計算的常識而已。

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

拉普拉斯這句名言只說對了一半。概率（probability）確實是精確化了的常識；但正如本章一再顯示的，任由常識自行其是，就會出錯。人們以為，罕見疾病的檢測結果呈陽性，就表示你很可能患了病；以為若第一位登機的乘客坐錯了座位，最後一位乘客就很難坐到自己的座位；以為一串擲硬幣的結果，不可能勝過另一串同樣長度的結果；以為多次量度的平均值總會穩定下來。這些全都是錯的，而每一次只需簡短的計算，便能揭示真相，更可貴的是，還能揭示當中的原因。

概率始於十七世紀有關賭博的問題，見於帕斯卡（Blaise Pascal）與費馬（Pierre de Fermat）之間的書信。直到 1933 年，柯爾莫哥洛夫（Andrey Kolmogorov）訂立了它的公理（axiom），概率才成為嚴格數學的一個分支。在這期間以及其後，它已發展成數學中應用最廣的部分之一。

概率為何重要

凡是要處理不確定性的領域，都用到概率：醫學（診斷檢測與臨牀試驗）、法律（證據的分量）、金融（風險與保險）、工程（可靠性與洪水）、電腦科學（隨機化演算法、密碼學、機器學習）、物理（量子力學與統計力學）以及生物學（遺傳學與流行病）。現代人工智能在很大程度上就是應用概率：語言模型所做的，是預測下一個字詞出現的概率。而能夠正確地就風險、機會和證據進行推理，是任何受過教育的人所能擁有的最寶貴技能之一。

你在大學會遇到甚麼

大學一年級的概率課程涵蓋公理、條件概率（conditional probability）與貝葉斯定理（Bayes' theorem）、獨立性（independence）、離散與連續隨機變量（random variable）、期望值（expectation）與方差（variance）、母函數（generating function），以及兩大極限定理（limit theorem）：大數定律（law of large numbers）和中心極限定理（central limit theorem）。其後的課程會研究隨機過程（random process），包括馬可夫鏈（Markov chain）、泊松過程（Poisson process）、隨機漫步（random walk）和分支過程（branching process），也會研究測度論概率（measure-theoretic probability，它為極限和連續分佈

奠定穩固的基礎）和鞅（martingale），並把概率與統計學、分析、組合數學和物理聯繫起來。

貫穿本章的問題

三個大問題

1. **問題究竟在問甚麼？**隨機試驗（random experiment）是甚麼？以甚麼為條件？「隨機」又是甚麼意思？
2. **巧妙的切入點在哪裏？**甚麼時候可以用對稱性（symmetry）、指示變量（indicator）、取條件（conditioning）或精心選定的狀態，來取代冗長的計算？
3. **長遠而言會怎樣？**隨機過程隨時間如何演變？許多隨機量的和、平均值和最大值又是甚麼模樣？

本章分為五部分。*清晰思考*（Section 12.1 至 Section 12.4）處理各種陷阱：取條件、基礎比率（base rate）、「隨機」的含義，以及獨立的含義。*對稱性與巧妙論證*（Section 12.5 至 Section 12.9）收集了一些解法既短且美的問題。*期望值、博弈與停止*（Section 12.10 至 Section 12.14）轉向平均值與決策：何時停止、向誰開槍、怎樣隨機化。*隨機過程*（Section 12.15 至 Section 12.19）追蹤隨機性在時間和空間中的表現。*而極限與連續概率*（Section 12.20 至 Section 12.25）則以幾何概率（geometric probability）和幾個重要的極限定理作結。

本章的思想

§	課題	思想
12.1	取條件	「指定的一個」與「至少一個」之別
12.2	貝葉斯定理	基礎比率與自然頻率（natural frequency）
12.3	貝特朗悖論（Bertrand's paradox）	「隨機」需要定義
12.4	獨立性	兩兩獨立（pairwise independent）不等於相互獨立（mutually independent）
12.5	飛機座位	兩個座位之間的對稱性
12.6	多一枚硬幣	互補而對稱的事件
12.7	錦標賽	數算比賽場數與配對
12.8	彭尼博弈（Penney's game）	首步分析（first-step analysis）；非傳遞性（non-transitivity）
12.9	公平硬幣	從偏差中提取公平
12.10	隨機環圈	無記憶的指示變量之和
12.11	杯子與 e	階乘（factorial）與指數級數（exponential series）
12.12	最優停止（optimal stopping）	$\frac{1}{e}$ 法則
12.13	三人決鬥（truel）	策略上的弱勢
12.14	博弈論（game theory）	隨機化策略；極小化極大（minimax）
12.15	分支過程	滅絕作為不動點（fixed point）
12.16	馬可夫鏈	平穩分佈（stationary distribution）；逆轉
12.17	巴拿赫（Banach）的火柴盒	負二項分佈（negative binomial distribution）
12.18	指數分佈（exponential distribution）	無記憶性（memorylessness）與獨立性
12.19	空間隨機性	最近鄰；瑞利（Rayleigh）分佈
12.20	蒲豐投針（Buffon's needle）	幾何概率與 π
12.21	隨機三角形	以面積表示概率
12.22	柯西分佈（Cauchy distribution）	沒有平均值的分佈
12.23	斯特靈公式（Stirling's formula）	由正態曲線（normal curve）得出階乘
12.24	中心極限定理	矩母函數（moment generating function）
12.25	極值（extreme value）	岡貝爾分佈（Gumbel distribution）

其中一節，即 Section 12.24，用到矩母函數，而矩母函數已於 2019 年從 STEP 考試範圍中刪去；該節會解釋所需的內容。

怎樣閱讀本章

當一道概率題令人摸不着頭腦時，不妨就一個細小的情況寫出樣本空間（sample space），然後逐一點算。當答案簡單得令人起疑時，就去找出解釋它的對稱性。而當你的直覺與

計算結果不一致時，請相信計算，然後弄清楚直覺為何出錯：真正的學習正在於此。

12.1 精確地取條件

題目

STEP I 2007，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/07-s1-q13`

引子

「已知」（given that）是概率中最危險的字眼。這道題最後四個部分的說法聽起來幾乎一模一樣：編號為 3 的圓片最先被取出；編號為 3 的圓片被取出；最先取出的是一塊有編號的圓片；取出的圓片中有一塊有編號。其中三個的答案都是 $\frac{1}{2}$，第四個卻不是。

這個分別，正正就是著名的「男孩還是女孩」謎題背後的分別（一個家庭有兩個孩子，其中至少一個是男孩；兩個都是男孩的機會有多大？）。這道謎題難倒了一代又一代的學生，也難倒了不少專業數學家。當中的教訓是：條件（condition）是一個精確的陳述，說明哪些結果仍然可能出現；而資訊是怎樣得來的，與資訊本身同樣重要。

動筆之前

? **想一想** 從十一塊圓片中取出四塊。對於「四塊都有編號」這個事件（event）而言，抽取的次序有沒有影響？四塊圓片組成的等可能（equally likely）組合共有多少個？

每一個由四塊圓片組成的組合被抽中的可能性都相同，所以我們可以不理會次序：共有 $\binom{11}{4} = 330$ 個等可能的組合。

逐步解題

階段 1　沒有條件

由四塊有編號圓片組成的組合有 $\binom{5}{4} = 5$ 個，所以

$$\mathrm{P}(\text{四塊都有編號}) = \frac{5}{330} = \frac{1}{66}.$$

階段 2 3 號最先取出

想一想 編號為 3 的圓片最先被抽出之後，餘下的三次抽取是怎樣的？

其餘三塊圓片是從餘下十塊（四塊有編號、六塊空白）中隨機抽出的三塊，全部 $\binom{10}{3} = 120$ 個組合都是等可能的。所以

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

因為在 (iii) 中，其餘三塊之中必須恰好有一塊有編號。

階段 3 3 號被取出（在某一次）

想一想 編號為 3 的圓片是在哪一次抽出的，有沒有關係？已知 3 號在四塊之中，對其餘三塊你能說些甚麼？

根據對稱性（symmetry），已知這四塊圓片中包含 3 號，其餘三塊就是從餘下十塊中隨機抽出的三塊，與第 2 階段完全一樣。所以 (iv) 的答案也是 $\frac{1}{2}$。

階段 4 最先取出的是有編號的圓片

已知第一塊圓片有編號（不論是哪個號碼），其餘三塊同樣是從餘下十塊圓片中隨機抽出的三塊，當中四塊有編號、六塊空白。所以 (v) 的答案也是 $\frac{1}{2}$。

階段 5 取出的圓片中有一塊有編號

想一想 現在的條件是「四塊之中至少有一塊有編號」。這與之前任何一個條件相同嗎？330 個組合之中有多少個符合這個條件？

這個條件有所不同：它沒有指明任何一塊特定的圓片。只有 $\binom{6}{4} = 15$ 個組合完全沒有編號圓片，所以有 315 個符合條件。其中 $\binom{5}{2}\binom{6}{2} = 10 \times 15 = 150$ 個恰好包含兩塊有編號的圓片。所以

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

停一停，試一試 為甚麼 (vi) 與 (v) 不同？在 (v) 中，我們知道哪一塊圓片有編號（第一塊），而其餘三塊不受限制。在 (vi) 中，我們只知道有某一塊圓片有編號。含有三塊或四塊編號圓片的組合，可以用更多方式滿足「某一塊圓片有編號」，所以它們不會像在 (v) 中那樣比重偏低。

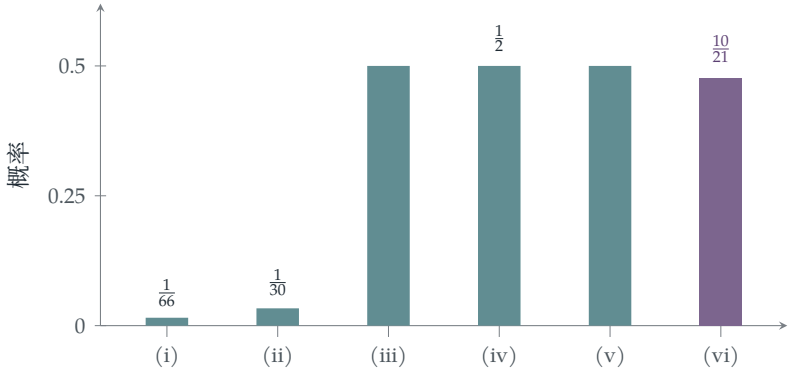


圖 12.1.1 六個答案。第 (iii)、(iv) 和 (v) 部分的條件都是某一塊特定的圓片有編號，答案是 $\frac{1}{2}$。第 (vi) 部分的條件只是有某一塊圓片有編號，答案是 $\frac{10}{21}$。

回顧

- **條件會收窄樣本空間（sample space）。** $\mathrm{P}(A \mid B)$ 是 B 的結果之中同時屬於 A 的比例。
- **對稱性可以省功夫。** 3 號是最先抽出，還是在某一次抽出，並無分別，因為所有位置的地位都相同。
- **「特定的一個」不等於「至少一個」。** 聽起來相似的條件，選出來的結果集合可以大不相同。

背後的數學

男孩還是女孩的謎題

馬丁·加德納（Martin Gardner）在 1959 年提出這個問題：史密斯先生有兩個孩子，其中至少一個是男孩；兩個都是男孩的概率是多少？列出等可能的家庭組合 BB、BG、GB、GG，剔除 GG 後剩下三個，其中一個是 BB：答案是 $\frac{1}{3}$。但如果我們知道年長的孩子是男孩，答案就是 $\frac{1}{2}$。這正是我們這道題第 (vi) 和第 (v) 部分的縮影。加德納後來承認，他最初提出的問題含糊不清：答案取決於我們是怎樣得知有一個男孩的。如果我們隨機遇上其中一個孩子，而他是男孩，答案是 $\frac{1}{2}$；如果我們問「你有沒有至少一個兒子？」而

對方回答有，答案就是 $\frac{1}{3}$。概率取決於產生資訊的過程，而不僅是資訊本身。

蒙提霍爾問題（Monty Hall）

概率中爭議最多的謎題，背後也是同一個教訓。在一個遊戲節目中，三道門的其中一道後面有一輛汽車。你選了 1 號門。主持人知道汽車在哪裏，他必定會打開另外兩道門的其中一道，讓你看到門後是一隻山羊，並且必定會讓你選擇是否換門；當汽車就在你所選的門後，以致他有選擇餘地時，他會在另外兩道門之間隨機挑選。他打開了 3 號門，並讓你改選 2 號門。你應該換嗎？應該：換門後勝出的概率是 $\frac{2}{3}$。主持人選哪道門並非隨機；這取決於汽車的位置，因而帶有資訊。確切的規則十分重要。如果主持人只要可以就一定打開 3 號門，那麼看到 3 號門打開，1 號門和 2 號門就變得同樣可能，換門後勝出的概率只有 $\frac{1}{2}$；然而，一個事先決定必定換門的參賽者，仍然有 $\frac{2}{3}$ 的機會勝出，因為換門勝出，恰好就是第一次選錯的時候。Marilyn vos Savant 在 1990 年於雜誌專欄中給出這個答案時，收到數以千計的來信指她錯了，其中有些更是出自數學家之手。用電腦模擬一次，或者仔細列出各個等可能的情況，就可以解決爭議。

貝葉斯（Bayes）與拉普拉斯（Laplace）

條件概率（conditional probability）的系統化基礎，是由托馬斯·貝葉斯（Thomas Bayes）和皮埃爾–西蒙·拉普拉斯（Pierre-Simon Laplace）奠定的：貝葉斯的論文在他死後於 1763 年發表，而拉普拉斯則在 1774 年以一般形式寫出了現稱為貝葉斯定理（Bayes' theorem）的法則。安德烈·柯爾莫哥洛夫（Andrey Kolmogorov）在 1933 年提出的公理（axioms），是現今每一個大學課程的起點，當中定義 $\mathrm{P}(A \mid B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$，而概率論的其餘部分都建基於這個定義之上。下一節 Section 12.2 會展示貝葉斯定理在實際應用中的作用。

思維技巧

準確把握條件

計算條件概率之前，先確切寫下哪些結果符合條件。問一問資訊是怎樣得來的。

在大學裏，條件概率和條件期望（conditional expectation）是概率與統計中最重要的工具之一，也是最常出錯的地方。把樣本空間寫出來，至少就一個小規模的情況寫出來，是最好的防範方法。

延伸練習

1 **推廣。**設有 b 塊空白圓片和 n 塊有編號的圓片，從中抽出 k 塊，證明 (iv) 和 (v) 的答案永遠相等，並求出 (vi) 在甚麼情況下與它們相等。

2 **星期二出生的男孩。**一個家庭有兩個孩子，其中至少一個是在星期二出生的男孩。證明兩個都是男孩的概率是 $\frac{13}{27}$。(列出 14×14 個等可能的情況，或者用巧妙的方法計數。)

3 **蒙提霍爾問題。**列出各個結果，驗證 $\frac{2}{3}$ 這個答案。如果主持人不知道汽車在哪裏，隨機打開另外兩道門的其中一道，而門後剛好是一隻山羊，那又如何？

4 **兩張 A。**一手 13 張的橋牌中至少有一張 A。這手牌中至少有兩張 A 的概率是多少？試與已知這手牌中有葵扇 A 時的概率比較。

延伸閱讀

- **Jason Rosenhouse, *The Monty Hall Problem***（Oxford University Press）。講述這道謎題、它的歷史，以及它的眾多變化。
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），有關條件概率的各章。
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***（CRC Press；可於網上免費閱讀）。對條件概率的處理十分出色，並載有許多悖論。

12.2 基本比率謬誤

題目

STEP II 1998，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/98-s2-q12`

引子

某種疾病的檢測，對患者的準確度達 90%，而且甚少誤報。Frankie 的檢測結果呈陽性。Frankie 應該有多擔心？大多數人都會答「大約 90%」，而在許多研究中，大多數醫生也是這樣回答。在這道題目中，正確答案是 $\frac{1}{11}$，約為 9%，足足小了十倍。

錯處在於忽略了*基礎比率*（base rate）：這種疾病很罕見，而健康的人為數眾多，所以即使假陽性（false positive）率很低，誤報的數目仍會遠遠多於真正的病例。這種*基礎比率謬誤*（base-rate fallacy）在醫學、法律（在法律上稱為檢控官謬誤（prosecutor's fallacy））、保安檢查和機器學習中都舉足輕重。貝葉斯定理（Bayes' theorem）正是對症良藥。

動筆之前

想一想 試想像有 1000 人。當中有多少人患病？這些患者之中有多少人檢測呈陽性？健康的人之中又有多少人檢測呈陽性？

在 1000 人當中，有 10 人患病，其中 9 人檢測呈陽性。在 990 個健康的人當中，有 $\frac{1}{11}$ 檢測呈陽性，即 90 個假陽性。因此共有 99 人檢測呈陽性，其中只有 9 人真正患病。

逐步解題

階段 1 貝葉斯定理

想一想　以 D 表示「Frankie 患有此病」，以 $+$ 表示「AL 檢測呈陽性」。你已知 $\mathrm{P}(+ \mid D)$，想求 $\mathrm{P}(D \mid +)$。兩者有甚麼關係？

根據條件概率（conditional probability）的定義，$\mathrm{P}(D \mid +) = \frac{\mathrm{P}(D \cap +)}{\mathrm{P}(+)}$，而 $\mathrm{P}(+)$ 可按 Frankie 是否患病拆成兩部分：

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

這正是那一千人當中的 $\frac{9}{99}$。

階段 2　第二次檢測

想一想　只有 AL 檢測呈陽性的人才會接受第二次檢測。在這羣人之中，患者佔多大比例？把這個比例當作新的「基礎比率」。

在 AL 檢測呈陽性的人當中，有 $\frac{1}{11}$ 是患者，$\frac{10}{11}$ 不是。以這兩個數作為先驗概率（prior probability），再次應用貝葉斯定理，得

$$\mathrm{P}(D \mid \text{兩次均呈陽性}) = \frac{0.9 \times \frac{1}{11}}{0.9 \times \frac{1}{11} + p \times \frac{10}{11}} = \frac{0.9}{0.9 + 10p}.$$

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

（這裏假設在已知 Frankie 是否患病的條件下，兩項檢測的誤差互相獨立（independent），題目亦隱含這個假設。）

停一停，試一試　當 $p = 0.09$ 時，STEP 的假陽性率與 AL 的相若，但第二次陽性結果卻令概率由 $\frac{1}{11}$ 升至 $\frac{1}{2}$。為甚麼質素相同的檢測，第二次的影響會大得多？

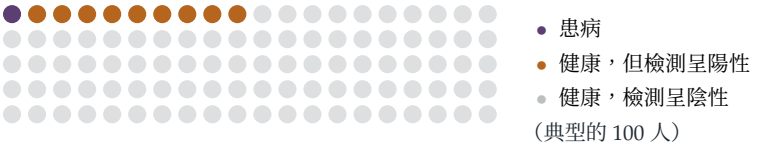


圖 12.2.1　典型的一百人。其中一人患病（紫色），十次檢測中有九次會呈陽性。在 99 個健康的人當中，約有 9 人會被誤判為陽性（琥珀色）。在檢測呈陽性的人當中，健康的人與患者的數目之比約為十比一。

回顧

- **自然頻率（natural frequencies）。**想像 1000 人，答案便一目了然，也容易驗算。
- **先驗乘以似然（likelihood）。**後驗概率（posterior probability）既取決於檢測的準確度，也取決於基礎比率。
- **分階段更新。**昨天的後驗就是今天的先驗：第二次檢測是由 $\frac{1}{11}$ 開始，而不是 $\frac{1}{100}$。

背後的數學

勝算與證據

以勝算（odds）來表達，貝葉斯定理最為清晰。D 的勝算是 $\mathrm{P}(D) : \mathrm{P}(\bar{D})$。一項證據 E 會把勝算乘以*似然比*（likelihood ratio）$\frac{\mathrm{P}(E|D)}{\mathrm{P}(E|\bar{D})}$：

$$\text{後驗勝算} = \text{先驗勝算} \times \text{似然比}.$$

這裏的先驗勝算是 $1:99$，而 AL 的似然比是 $0.9 \div \frac{1}{11} = 9.9$，於是後驗勝算為 $9.9:99 = 1:10$，即概率為 $\frac{1}{11}$。第二項獨立檢測的似然比若為 10（$p = 0.09$ 時的 STEP 正是如此），便會把勝算帶到 $1:1$。取對數之後，證據就可以*相加*：每項獨立檢測所貢獻的，是其似然比的對數。圖靈（Alan Turing）在布萊切利園（Bletchley Park）破解恩尼格瑪（Enigma）密碼電文時，用的正是這個方法，並以「ban」和「deciban」作為量度證據的單位。

檢控官謬誤

1999 年，事務律師 Sally Clark 被裁定謀殺兩名年幼兒子罪名成立，定罪的部分依據是一項證據，指同一家庭發生兩宗嬰兒猝死的機會是「7300 萬分之一」。姑且不論這個數字本身有錯（它假設兩宗死亡互相獨立，忽略了遺傳和環境因素），這個推論還把 P(證據 | 無辜) 與 P(無辜 | 證據) 混為一談。母親連殺兩子同樣極為罕見，真正應該比較的是這兩種解釋。皇家統計學會（Royal Statistical Society）發表公開信指出這個錯誤，她的定罪於 2003 年獲撤銷。當中的數學與本題完全相同。

Gigerenzer 與自然頻率

在 1990 年代，心理學家 Gerd Gigerenzer 指出，醫生若獲得以條件概率表達的檢測統計數字，遇到本題這類問題時大多錯得離譜；但同樣的資料若以自然頻率表達（「在 1000 名女性當中，有 10 人患乳癌；其中 8 人檢測呈陽性…」），大多數醫生都能答對。他的研究改變了許多國家向病人傳達醫療風險的方式。

思維技巧

用貝葉斯定理把條件概率倒轉過來

$\mathrm{P}(A \mid B)$ 與 $\mathrm{P}(B \mid A)$ 是兩個不同的數。要由其中一個求出另一個，你需要基礎比率 $\mathrm{P}(A)$。

貝葉斯推理，即根據證據更新信念，是現代統計學和機器學習的基礎，由垃圾郵件過濾器以至醫學診斷都用得上。到了大學，你會遇到它的進階形式——貝葉斯推斷（Bayesian inference），屆時處理的不再是單一的概率，而是先驗分佈和後驗分佈。

延伸練習

1 **雪櫃裏的屍體。**試做 STEP II 1987，第 16 題（`step.maths.org/questions/87-s2-q16`）：貝葉斯定理應用於兩名可靠程度不同的證人，是 STEP 中較令人難忘的題目情境之一。

2 **要多罕見？**某項檢測的靈敏度（sensitivity）為 0.9，假陽性率為 $\frac{1}{11}$。疾病要有多普遍，陽性結果才表示患病的機會高於一半？

3 **重複檢測。**假如重複進行 AL 檢測，而在已知受測者真實狀況的條件下各次結果互相獨立，那麼兩次 AL 檢測均呈陽性後，患病的概率是多少？三次之後呢？

4 **篩查。**某項篩查計劃為一百萬人檢測某種病症，其患病率（prevalence）為 10 000 分之 1，所用檢測的靈敏度為 99%，假陽性率為 1%。有多少人會被告知可能患有此病症？當中又有多少人真的患病？

延伸閱讀

- **Gerd Gigerenzer, *Reckoning with Risk***（Penguin）。自然頻率、醫學檢測與法庭。
- **David Spiegelhalter, *The Art of Statistics***（Pelican）。內容包括醫學篩查和 Sally Clark 案。
- **Sharon Bertsch McGrayne, *The Theory That Would Not Die***（Yale University Press）。貝葉斯定理的歷史，包括圖靈如何運用這個定理。

12.3 「隨機」是甚麼意思？

題目

STEP III 2022，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2022 年 STEP 試卷（zip 檔案）內的 STEP III 試卷：

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

引子

「在圓上隨機選一條弦（chord），它的期望長度（expected length）是多少？」這條問題聽起來好像有一個答案。其實沒有，除非你先說清楚「隨機」是甚麼意思。在圓周上隨機取兩點再連起來，期望長度是 $\frac{4a}{\pi} \approx 1.27a$。在圓內隨機取一點，再隨機取一個方向作過該點的弦，答案卻是 $\frac{16a}{3\pi} \approx 1.70a$。兩種做法都是選取「一條隨機弦」的合理方法。

這就是貝特朗悖論（Bertrand's paradox），由約瑟夫·貝特朗（Joseph Bertrand）於 1889 年發表，用意是提出警告。只有在指明隨機試驗（random experiment）之後，概率（probability）才有明確的定義。這個悖論促使數學家在二十世紀為概率論建立嚴格的基礎：在這套基礎之中，每一個概率都定義在一個精確界定的結果空間之上。

動筆之前

? **想一想** 若角 AOB 為 θ，弦 AB 的長度是多少？（由 O 作垂線。）

由 O 所作的垂線平分該弦，也平分該角，所以 $AB = 2a\sin\frac{\theta}{2}$。

逐步解題

階段 1　圓周上的兩點

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

階段 2　圓內的一個隨機點

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

$\mathrm{P}(R \le r)$ 是半徑為 r 的圓盤面積除以整個圓盤的面積，即 $\frac{r^2}{a^2}$。所以 R 在 $[0,a]$ 上的密度（density）是 $\frac{2r}{a^2}$。（點較有可能離圓心較遠，因為那裏的面積較大。）

想一想　一條過 C 的弦與 OC 成角 t。這條弦與 O 相距多遠？

由 O 到該直線的距離是 $R\sin t$，所以弦長為 $2\sqrt{a^2 - R^2\sin^2 t}$。

階段 3　對 C 取平均

設 $s = \sin t$，

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

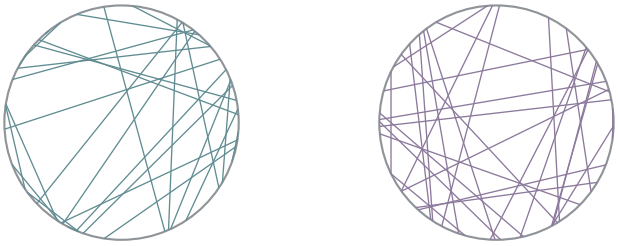
想一想　把 $1 - \cos^3 t$ 和 $\sin^2 t = 1 - \cos^2 t$ 因式分解。有甚麼可以約去？

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

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

階段 4　方向也隨機

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



兩個隨機端點
平均長度 $\frac{4a}{\pi} \approx 1.27a$

圓盤內的隨機中點
平均長度 $\frac{4a}{3} \approx 1.33a$

圖 12.3.1 以兩種不同方法選出的二十八條「隨機弦」。左：圓周上兩個獨立的均勻分佈點。右：以圓盤內一個均勻隨機的點作為弦的中點。右邊的方法在邊緣附近產生的短弦較少。不同的方法給出不同的分佈，也給出不同的答案。

回顧

- **指明試驗。**「一條隨機弦」在每一部分都化成了一個精確的步驟，而每個步驟給出不同的答案。
- **由面積求密度。**$\mathrm{P}(R \leq r)$ 直接來自幾何；求導便得到密度。
- **分階段取平均。**先對 C 的位置取平均（t 固定），再對 t 取平均。

背後的數學

貝特朗的三個答案

貝特朗原本的問題是：一條隨機弦比內接等邊三角形的邊長 $\sqrt{3}a$ 更長的概率是多少？他給出了三種方法和三個答案。

- 隨機端點：固定一個端點；若另一個端點落在圓周中間的三分之一之內，弦便是長的。概率為 $\frac{1}{3}$。
- 隨機取一條半徑，再在其上隨機取一點作為中點：若弦與 O 的距離小於 $\frac{a}{2}$，弦便是長的。概率為 $\frac{1}{2}$。
- 在圓盤內任意位置隨機取中點：若中點與 O 的距離在 $\frac{a}{2}$ 之內，弦便是長的。概率為 $\frac{1}{4}$。

每一種方法都是弦的集合上一個不同的概率分佈（probability distribution），當中並無矛盾：它們回答的是不同的問題。這裏的教訓是，「均勻隨機」只有相對於某個選定的參數化（parametrisation）才有意義，而對於連續的集合，通常並沒有唯一自然的參數化。

有沒有正確答案？

1973 年，物理學家 Edwin Jaynes 主張第二種方法才是「正確」的，理由是把圓稍為移動或縮放，答案都不應改變：一條隨機弦應該就像一根拋到地上所畫圓圈上的飲管。只有與圓心距離均勻分佈的那個分佈才具有這種不變性（invariance），而拋飲管的實驗也證實了 $\frac{1}{2}$。同一個想法，即正確的分佈就是尊重情境中各種對稱性的那一個，正是積分幾何（integral geometry）的基礎；積分幾何是研究隨機直線的數學，它也解釋了 Section 12.20 中的布豐投針（Buffon's needle）。

從悖論到公理

貝特朗在 *Calcul des probabilités*（1889）中發表了他的弦悖論，部分原因是要嘲諷他眼中對「幾何概率」的草率運用。龐加萊（Henri Poincaré）也討論過它。1900 年，希爾伯特（David Hilbert）把概率的公理化（axiomatisation）列入他為新世紀提出的著名問題之中，而柯爾莫哥洛夫（Andrey Kolmogorov）在 1933 年解決了這個問題：概率是結果空間上的一個測度（measure），任何關於「隨機」對象的問題，都必須先說明是哪一個空間、哪一個測度。貝特朗悖論正是說明為何必須如此的標準例子。

思維技巧

界定隨機試驗

計算概率之前，先準確描述隨機對象是怎樣產生的。不同的描述可以給出不同的答案。

在大學裏，每一個概率計算都由一個樣本空間（sample space）和一個概率測度（probability measure）出發。在統計學和物理學中，選擇合適的隨機模型（往往藉助對稱性）是工作的一大部分。

延伸練習

1 **貝特朗的概率。**驗證上文的三個概率 $\frac{1}{3}$、$\frac{1}{2}$、$\frac{1}{4}$。

2 **另一個平均值。**證明若弦的中點在圓盤內均勻選取（右邊的圖），期望長度是 $\frac{4a}{3}$。（提示：這就是 $L\left(\frac{\pi}{2}\right)$。為甚麼？）

3 **隨機距離。**若弦與 O 的距離在 $[0, a]$ 上均勻分佈，證明期望長度是 $\frac{\pi a}{2}$。

4 **三種方法，三個分佈。**對三種方法中的每一種，求弦長 ℓ 在 $[0, 2a]$ 上的概率密度（probability density）。

延伸閱讀

- **Edwin Jaynes, "The well-posed problem"**, *Foundations of Physics* 3 (1973)。不變性論證，易讀而又發人深省。
- **Gábor Székely, *Paradoxes in Probability Theory and Mathematical Statistics*** (Reidel)。貝特朗悖論及其他許多悖論。
- **Daniel Klain & Gian-Carlo Rota, *Introduction to Geometric Probability*** (Cambridge University Press)。積分幾何，留待日後研讀。

12.4 兩兩獨立，卻非相互獨立

題目

STEP III 2021，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2021 年 STEP 試卷（zip 檔案）內的 STEP III 試卷：

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

引子

玩家 1 和玩家 2 擲出相同的點數。這能告訴你玩家 2 和玩家 3 是否擲出相同點數嗎？出人意料的是，不能：這兩個事件是獨立（independent）的。玩家 1 和玩家 3 的情況也一樣。但如果 1 與 2 相同，而 2 又與 3 相同，那麼 1 必定與 3 相同。三個事件，每兩個之間都獨立，合起來卻不獨立。

兩兩獨立（pairwise independence）比完全獨立弱，而這個差別至關重要。就方差（variance）而言，兩兩獨立已經足夠，本題正是利用了這一點。但在許多其他用途上，這並不足夠。電腦科學家會刻意使用兩兩獨立的隨機數，因為產生它們的成本遠低於產生完全獨立的隨機數，而對許多演算法來說已經夠用。

動筆之前

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

無論玩家 1 擲出甚麼，玩家 2 與之相同的概率都是 $\frac{1}{6}$。$X_{12}=1$ 和 $X_{23}=1$ 同時成立，即三人擲出相同的點數：概率為 $\frac{6}{216}=\frac{1}{36}$。

逐步解題

階段 1 **兩兩獨立**

由於 $\mathrm{P}(X_{12}=1, X_{23}=1)=\frac{1}{36}=\frac{1}{6}\cdot\frac{1}{6}=\mathrm{P}(X_{12}=1)\mathrm{P}(X_{23}=1)$，而每個變量只取兩個值，所以 X_{12} 與 X_{23} 獨立。（對於只取兩個值的變量，聯合分佈表中的一格便決定了其餘各格；練習 1 會把這一點說清楚。）沒有共同玩家的兩對，例如 X_{12} 與 X_{34}，顯然是獨立的。因此所有 X_{ij} 都兩兩獨立。

階段 2 得分的平均值與方差

得分為 $S=\sum_{i<j}X_{ij}$，是 $\binom{n}{2}$ 個變量之和，每個變量的平均值為 $\frac{1}{6}$，方差為 $\frac{1}{6}\cdot\frac{5}{6}=\frac{5}{36}$。期望值（expectation）總是可以相加的，所以 $\mathrm{E}(S)=\frac{1}{6}\binom{n}{2}$。

> **想一想** 和的方差在甚麼時候等於方差之和？

和的方差等於方差之和，再加上協方差（covariance）之和的兩倍，而獨立的一對變量協方差為零。兩兩獨立已經足夠：

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

階段 3 方差公式

若各 Y_i 的平均值為零，則 $\mathrm{Var}(\sum Y_i)=\mathrm{E}((\sum Y_i)^2)$，把平方展開後，每個 Y_i^2 出現一次，每個乘積 Y_iY_j（其中 $i<j$）出現兩次。取期望值便得到公式。 ■

階段 4 奇偶遊戲

> **想一想** $\mathrm{E}(Z_{12})$ 是多少？$Z_{12}Z_{23}$ 可以取哪些值？它在甚麼時候不等於零？

$Z_{12}=1$ 的概率為 $\frac{3}{36}=\frac{1}{12}$（兩人都擲出 2、都擲出 4 或都擲出 6），而取值 -1 的概率為 $\frac{1}{12}$；所以 $\mathrm{E}(Z_{12})=0$ 而 $\mathrm{E}(Z_{12}^2)=\frac{1}{6}$。

只有當三位玩家都擲出相同點數時，才有 $Z_{12}Z_{23}\neq 0$，而此時 $Z_{12}=Z_{23}$（同為 $+1$ 或同為 -1），所以乘積為 1。因此 $\mathrm{E}(Z_{12}Z_{23})=\frac{1}{36}$，這並不等於 $\mathrm{E}(Z_{12})\mathrm{E}(Z_{23})=0$：它們並不獨立。 ■

（更直接的看法是，$\mathrm{P}(Z_{12}=1, Z_{23}=-1)=0$，而每個事件單獨發生的概率都是 $\frac{1}{12}$。）

階段 5 方差

總得分 $T=\sum Z_{ij}$ 的平均值為 0。根據第 (ii) 部分，$\mathrm{Var}(T)$ 等於 $\binom{n}{2}\cdot\frac{1}{6}$，再加上 $\mathrm{E}(Z_{ij}Z_{kl})$ 對所有「兩對玩家」的組合求和後的兩倍。沒有共同玩家的兩對，貢獻為 0（它們獨立，而且各自的平均值為零）。恰好有一位共同玩家的兩對，每組貢獻 $\frac{1}{36}$，而這樣的組合共

有 $n\binom{n-1}{2}$ 個（先選共同的玩家，再選另外兩人）。所以

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

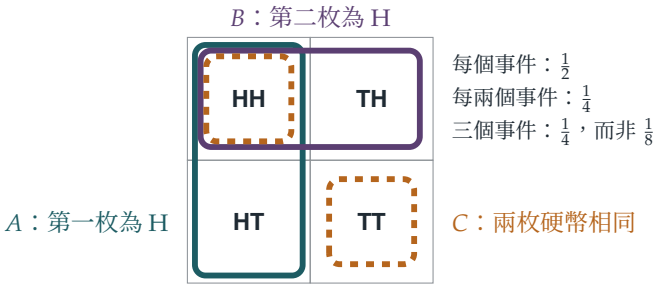


圖 12.4.1 伯恩斯坦（Bernstein）的例子：兩枚公平硬幣。事件「第一枚硬幣是正面」（藍綠色）、「第二枚硬幣是正面」（紫紅色）和「兩枚硬幣結果相同」（琥珀色虛線）的概率各為 $\frac{1}{2}$，而每兩個事件都相交於概率為 $\frac{1}{4}$ 的單一方格，所以它們兩兩獨立。但三個事件全部相交於 HH，其概率為 $\frac{1}{4}$，而非 $\frac{1}{8}$：其中任何兩個事件都決定了第三個。

回顧

- **指示變量（indicator variables）。**把得分寫成指示變量之和，一切便化為對一對對玩家的考慮。
- **方差只需要看成對的變量。**和的方差只取決於兩兩之間的協方差，所以兩兩獨立已足以令方差相加。
- **獨立性要驗證，不要假設。**兩個看似相同的遊戲，卻有不同的相依結構。

背後的數學

多個事件的獨立性

如果對每一個子族，交集的概率都等於各概率之積，則稱事件 $A_1,\dots,A_n$ *獨立*。對三個事件來說，這意味着四個條件：三個是關於每一對的，另一個是 $\mathrm{P}(A\cap B\cap C) = \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C)$。伯恩斯坦（Sergei Bernstein）在 1920 年代提出圖 12.4.1 中的例子，說明兩兩之間的條件並不能推出三者的條件。（反過來，單憑三者的條件也不能推出兩兩之間的條件。）題目中「點數相同」的事件 X_{12}、X_{23}、X_{13} 與伯恩斯坦的硬幣如出一轍。

不相關不等於獨立

協方差為零，即 $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$，是比獨立更弱的條件，稱為不相關（uncorrelated）。STEP III 2000 Q14（見練習）說明，對於只取兩個值的變量，這兩個概念是一致的，但取三個值時則不然。一個簡單的例子：設 X 在 $\{-1, 0, 1\}$ 上均勻分佈，並設 $Y = X^2$。那麼 $\mathrm{E}(XY) = \mathrm{E}(X^3) = 0 = \mathrm{E}(X)\mathrm{E}(Y)$，然而 Y 完全由 X 決定。協方差只量度線性關聯。

廉價的隨機性

對電腦來說，產生隨機數的成本很高。1979 年，Larry Carter 和 Mark Wegman 證明，對許多演算法而言，例如把數據儲存在偽隨機位置的雜湊（hashing），兩兩獨立的隨機選擇與完全獨立的同樣有效。兩兩獨立的數值可以用極低的成本產生：例如，取隨機的 a 和 b，則 $a + bk \pmod{p}$ 對 $k = 0, 1, \ldots, p-1$ 給出 p 個數值，其中任何兩個都是獨立的。本題的方差計算正是它行得通的原因：切比雪夫（Chebyshev）不等式只需要方差，而方差只需要看成對的變量。

思維技巧

分解為指示變量

要求一個計數的平均值和方差，可把它寫成指示變量之和，每一個可能被數到的對象對應一個指示變量。求平均值只需要考慮單個指示變量；求方差則需要考慮成對的指示變量。

這是概率論中最有用的技巧之一，早在 Section 9.8 已經出現過。在大學裏，隨機圖、演算法和統計檢驗的分析都以它為基礎。

延伸練習

1 **兩個值。**試做 STEP III 2000，第 14 題（`step.maths.org/questions/00-s3-q14`）。該題證明，對於各自只取兩個值的隨機變量，$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ 必然推出獨立，並要求給出一個取三個值的反例。

2 **人數更多的隊伍。**在第 (i) 部分中取 $n = 7$，求得分的平均值和標準差，並把例子中的得分 7 與它們比較。

3 **再談生日。**利用指示變量，求 k 個人當中生日相同的人數對的平均值和方差（Section 9.9）。這些指示變量是否兩兩獨立？

4 **三者成立但兩兩不獨立。**找出三個事件，每個的概率都是 $\frac{1}{2}$，滿足 $\mathrm{P}(A \cap B \cap C) = \frac{1}{8}$，但並非兩兩獨立。

延伸閱讀

- **Geoffrey Grimmett & David Stirzaker, *Probability and Random Processes***（Oxford University Press），開首幾章討論獨立性。
- **Michael Mitzenmacher & Eli Upfal, *Probability and Computing***（Cambridge University Press）。兩兩獨立及其在演算法中的應用。
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***（CRC Press；可於網上免費閱讀），關於指示變量與協方差。

12.5 最後一位乘客的座位

題目

STEP II 2021，第 11 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2021 年 STEP 試卷（zip 檔案）內的 STEP II 試卷：

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

引子

一百名乘客登上一班滿座的飛機。第一位乘客遺失了登機證，於是隨便找個座位坐下。其後每位乘客，若自己的座位仍然空着就坐自己的座位，否則就在餘下的空位中隨機選一個。最後一位乘客坐到自己座位的概率（probability）是多少？混亂或會蔓延至整架飛機，所以你可能預期這個概率很小；又或者大部分乘客根本不受影響，概率應該接近 1。

無論乘客人數是多少，答案都恰好是 $\frac{1}{2}$。本題以遞推關係（recurrence）和歸納法（induction）證明這個結果。不過，還有一個利用對稱性（symmetry）的漂亮論證，一句話就能說完；它解釋了為甚麼答案是 $\frac{1}{2}$，更可讓你立即寫出任何一位乘客的答案。

動筆之前

停一停，試一試 列出所有可能發生的情況，求出 P_2 和 P_3。

若有兩個座位，T_1 會坐 S_1（這時 T_2 坐到 S_2）或 S_2，兩者的概率各為 $\frac{1}{2}$：$P_2 = \frac{1}{2}$。若有三個座位：如果 T_1 坐 S_1（概率為 $\frac{1}{3}$），人人都沒有問題；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，則 T_2 要在 S_1 和 S_3 之間隨機選擇，而 T_3 只有在 T_2 選中 S_1 時才沒有問題。所以 $P_3 = \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2}$。

逐步解題

階段 1　遞推關係

想一想　假設 T_1 坐了 S_k，其中 $2 \le k \le n-1$。乘客 $T_2, \dots, T_{k-1}$ 會怎樣？當 T_k 到達時，哪些座位是空的？這個局面與一列較小的列車相比又如何？

乘客 $T_2, \dots, T_{k-1}$ 發現自己的座位空着，便坐了下來。當 T_k 到達時，空着的座位是 S_1 和 $S_{k+1}, \dots, S_n$：共有 $n-k+1$ 個。T_k 從中隨機選一個，此後每位乘客若自己的座位空着就坐自己的座位，否則就隨機選一個。這正是有 $n-k+1$ 個座位的原來問題，T_k 扮演第一位乘客的角色，而 S_1 則扮演 T_k「自己的」座位的角色（即一旦被選中，一切便回復正常的那個座位）。因此，T_n 坐到 S_n 的條件概率（conditional probability）是 P_{n-k+1}。

現在按 T_1 所坐的座位分情況考慮，每個座位的概率都是 $\frac{1}{n}$：坐 S_1 必定成功，坐 S_n 必定失敗，而坐 S_k 則給出 P_{n-k+1}。所以

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

階段 2　答案

想一想　$P_2 = P_3 = \frac{1}{2}$。猜出 P_n，並用強歸納法（strong induction）證明。

斷言：對所有 $n \ge 2$，$P_n = \frac{1}{2}$。當 $n = 2$ 時斷言成立。若 $P_r = \frac{1}{2}$ 對 $2 \le r \le n-1$ 成立，則

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

階段 3　倒數第二位乘客

想一想　試找一個更巧妙的論證。到最後，留給 T_n 的是哪個座位？有可能是 S_2 嗎？

最後留給 T_n 的座位不是 S_1 就是 S_n。它不可能是 S_k（其中 $2 \le k \le n-1$），因為 T_k 早就會把它坐了。而在此之前的每一次隨機選擇，對 S_1 和 S_n 都是對稱的：每當有被佔了座位的乘客要選位，S_1 和 S_n 被選中的機會均等。兩者之中先被選中的那個決定了結局，所以兩者同樣可能成為最後剩下的座位：$P_n = \frac{1}{2}$。

同一個論證適用於任何一位乘客 T_k，$k \ge 2$。只有當某位較早的乘客佔了 T_k 的座位，他們才會失去自己的座位。座位被佔的連鎖反應會一直持續，直至有人選中 S_1 為止；在 $S_1, S_k, S_{k+1}, \dots, S_n$ 這些座位之中，T_k 到達之前的選位者對它們一視同仁，所以最先被

選中的座位同樣可能是這 $n-k+2$ 個座位中的任何一個，而 T_k 只有在最先被選中的是 S_k 時才會失去座位。所以

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

當 $k = n-1$（其中 $n \geq 3$）時，這給出 $Q_n = \frac{2}{3}$；而 $Q_2 = \frac{1}{2}$，因為這時 $T_{n-1} = T_1$，他們坐在 S_1 的概率是 $\frac{1}{2}$。 ■

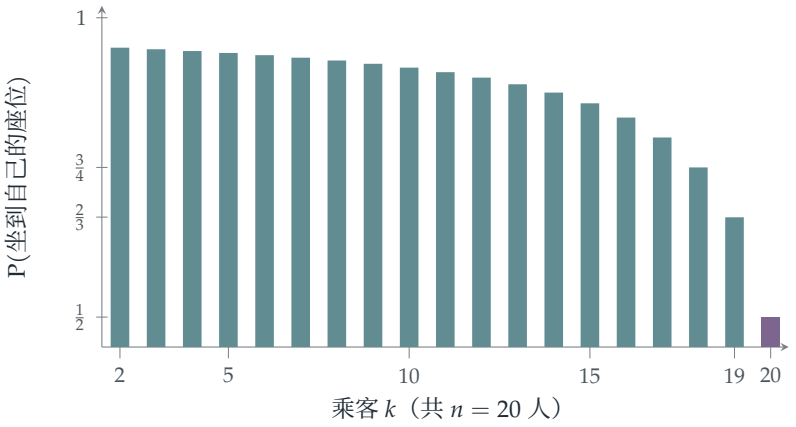


圖 12.5.1 乘客 k 坐到自己座位的概率，列車共有 $n = 20$ 個座位。當 $2 \leq k \leq n$ 時，概率是 $\frac{n-k+1}{n-k+2}$：由 T_2 的 $\frac{19}{20}$ 一直下降至 T_{19} 的 $\frac{2}{3}$，以及最後一位乘客的 $\frac{1}{2}$（紫紅色）。

回顧

- **化為較小的同類問題。**在 T_1 坐了 S_k 之後，餘下的問題就是座位較少的同一個問題。
- **強歸納法。**遞推關係牽涉所有較小的情況，所以歸納時要假設結果對所有這些情況都成立。
- **找出對稱性。**座位 S_1 和 S_n 自始至終可以互換，這樣不用任何代數運算就解釋了 $\frac{1}{2}$ 的由來。

背後的數學

對稱論證

第二個證明體現了概率論中一個威力強大的想法：若兩個結果可由隨機過程的某種對稱性互相對應，它們的概率就相同。這裏的過程在 S_1 和 S_n 其中一個被選中之前，並不區分這兩個座位，所以兩者成為最後剩下的座位的可能性必定相等。對稱論證往往只憑一

個觀察，就能取代冗長的計算。接下來的幾節有很多這樣的例子：Section 12.6 中多擲一次硬幣，Section 12.7 中的淘汰賽抽籤。

有多少位乘客被迫換位？

被迫換位的乘客（即坐不到自己座位的乘客）的數目同樣有很好的規律。T_1 除非選中 S_1，否則就算是換了位；而 T_k（其中 $k \geq 2$）被迫換位的概率是 $\frac{1}{n-k+2}$。把這些概率相加，被迫換位的乘客數目的期望值（expected number）是

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

大約是 $\ln n$：即使飛機上有 300 個座位，這場混亂平均也只影響約六位乘客。調和數（harmonic numbers）又再出現了（Section 9.11）。

思維技巧

尋找對稱性

如果隨機過程無法分辨兩個結果，它們的可能性就相等。找出它們最初變得可以分辨的那一刻。

在大學，對稱論證在概率論中隨處可見，由反射原理（reflection principle）（Section 9.10）到可交換性（exchangeability）皆是；可交換性是指一列隨機變量（random variables）無論以任何次序排列都有相同的分佈，這個性質是不少現代統計學的基礎。

延伸練習

1. **兩位健忘的乘客。**如果最先的兩位乘客都隨機選位，最後一位乘客坐到自己座位的概率是多少？

2. **被迫換位的乘客。**列出所有可能情況，驗證 $n=3$ 時被迫換位的乘客數目的期望值。

3. **迂迴的路線。**利用遞推關係，證明 $\sum_{r=2}^{n-1} P_r = \frac{n-2}{2}$ 與對稱論證互相吻合。

4. **最後一個座位。**證明在最後一位乘客坐不到自己座位的條件下，最後一位乘客坐到 S_1 的概率是 1。為甚麼？

延伸閱讀

- **Peter Winkler, *Mathematical Puzzles: A Connoisseur's Collection***（A K Peters）。收錄了飛機座位謎題，以及許多其他可用對稱性解決的謎題。
- **Frederick Mosteller, *Fifty Challenging Problems in Probability***（Dover）。一本經典的題集。
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***（CRC Press；可於網上免費閱覽）。

12.6 多一枚硬幣

題目

STEP I 2016，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/16-s1-q12`

引子

Bob 比 Alice 多擲一枚硬幣。他擲得的正面（heads）比 Alice 多的概率（probability）是多少？他佔了優勢，但多擲一次不保證多一個正面，而且平手算他輸。用硬算的方法做出頭兩個情況，兩次的答案都是 $\frac{1}{2}$。這是巧合嗎？

不是巧合，而且有一個只需一句話的證明：只要留意到 Bob 擲得的正面或反面（tails）必有其一比 Alice 多，但不會兩者皆多。這道題目是一個典範，示範了計算如何提示出規律，而巧妙的對稱性（symmetry）又如何解釋這個規律。

動筆之前

? **想一想** Alice 擲兩次：她擲得正面的次數的分佈（distribution）是甚麼？擲三次的 Bob 又如何？

Alice 擲得 $0, 1, 2$ 個正面的概率分別是 $\frac{1}{4}, \frac{2}{4}, \frac{1}{4}$；Bob 擲得 $0, 1, 2, 3$ 個正面的概率分別是 $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$。

逐步解題

階段 1 硬算

第 (i) 部分，按 Alice 擲得的正面數目逐項相加：

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

第 (ii) 部分是同樣的計算，只是項數較多；數一數 $8 \times 16 = 128$ 個等可能的結果，其中 64 個是 Bob 領先（圖 12.6.1），所以答案同樣是 $\frac{1}{2}$。

階段 2　拆開 Bob 的投擲

想一想　把 Bob 的 $n+1$ 次投擲看成先擲 n 次，再額外擲一次。擲完頭 n 次之後，他可能領先 Alice、與她平手，或者落後。每種情況下會發生甚麼事？

二人各擲 n 次之後，Bob 領先的概率是 p_2，平手的概率是 p_1，落後的概率是 p_2（由 Alice 與 Bob 之間的對稱性可知）。如果他領先，他會保持領先。如果平手，他額外的一擲有 $\frac{1}{2}$ 的概率令他領先。如果他落後，多一個正面最多只能追成平手。所以

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

由於這三種情況涵蓋了所有可能，$2p_2 + p_1 = 1$，因此 $p_2 + \frac{1}{2}p_1 = \frac{1}{2}$。多了一枚硬幣，Bob 擲得較多正面的概率對每個 n 都剛好是 $\frac{1}{2}$。 ■

停一停，試一試　以下就是那個一句話的證明。Bob 擲 $n+1$ 次，所以他的正面和反面數目加起來是 $n+1$，比 Alice 的多一。試證明 Bob 的正面比 Alice 多，當且僅當他的反面不比 Alice 多。由於正面與反面是對稱的，這兩個事件的概率相等，所以各為 $\frac{1}{2}$。

事實上，記 $H_A + T_A = n$ 及 $H_B + T_B = n+1$：若 $H_B > H_A$，則 $H_B \geq H_A + 1$，所以 $T_B = n + 1 - H_B \leq n - H_A = T_A$；反過來，若 $T_B \leq T_A$，則 $H_B \geq H_A + 1$。所以「正面較多」與「反面較多」是互補（complementary）的，而把正面與反面對調（由於硬幣是公平的，這不會改變概率）就會把兩者互換。

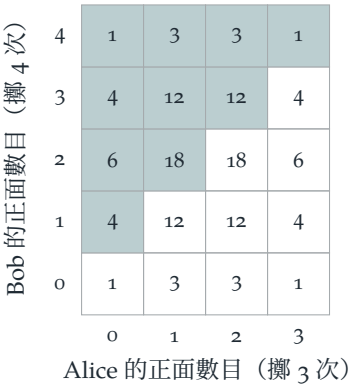


圖 12.6.1 第 (ii) 部分：在 128 個等可能的結果之中，每種正面數目組合所佔的個數。Bob 在有底色的格子中領先，這些格子剛好包含一半的結果。

回顧

- **計算小的情況。**兩次計算都提示答案是 $\frac{1}{2}$。
- **按自然的劃分來取條件。**把 Bob 額外的一擲分開，問題便化為一場勢均力敵的比試再加一枚硬幣。
- **找出互補的一對。**「正面較多」與「反面較多」互為補事件，並且在一個對稱之下互換。

背後的數學

對稱與補事件

那個漂亮的證明結合了兩個事實：在一個對稱之下互換的兩個事件概率相等，而兩個互補事件的概率之和為 1。若兩者同時成立，則每個事件的概率都是 $\frac{1}{2}$。同樣的推理在概率論中隨處可見。在航班座位問題 (Section 12.5) 中，S_1 與 S_n 在一個對稱之下互換，而最後剩下的恰好是其中一個。在六貫棋 (Hex) 中，由棋盤的對稱性，加上恰有一方必勝這個事實，可知先手一方佔優，不過在大棋盤上無人知道必勝策略是甚麼。

平手的可能性有多大？

當二人投擲次數相同時，平手的概率 p_1 就很重要。Alice 與 Bob 平手，即 Alice 的正面數目等於 Bob 的正面數目；利用對稱性，把 Bob 的正面換成他的反面，這個概率就等於 Alice 的正面數目加上 Bob 的正面數目等於 n 的概率，亦即在 $2n$ 次投擲中共有 n 個正

面。所以

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

這裏用了斯特靈近似（Stirling's approximation）（Section 12.23）。平手會變得罕見，但變得很慢：各擲 100 次，平手的機會仍然約有 5.6%。

思維技巧

先計算，再猜想，然後解釋

當計算得出一個簡單得可疑的答案時，就要找出原因。簡單的答案通常有簡單的解釋。

在大學裏，這個習慣正是理解與計算的分野。這裏的第一個證明確認了答案；第二個證明則解釋了答案，而且可以立即推廣（試做練習 2）。

延伸練習

1. **多擲兩枚硬幣。**若 Bob 擲 $n+2$ 枚硬幣而 Alice 擲 n 枚，證明 $\mathrm{P}(\text{Bob 的正面較多}) = \frac{1}{2} + \frac{1}{2}\mathrm{P}(H' = H_A)$，其中 H' 是 Bob 頭 $n+1$ 次投擲中的正面數目。
2. **不公平的硬幣。**若二人所用的硬幣擲出正面的概率都是 p，多擲一次時答案是否仍然是 $\frac{1}{2}$？那個漂亮的證明中哪一步不再成立？
3. **平手。**用正文中的技巧證明 $p_1 = \binom{2n}{n}/4^n$，並求它在 $n = 1, 2, 3$ 時的值。
4. **骰子。**Alice 擲 n 粒骰子，Bob 擲 $n+1$ 粒。Bob 擲得較多六點的概率是否等於 $\frac{1}{2}$？

延伸閱讀

- **Peter Winkler, *Mathematical Mind-Benders***（A K Peters）。收錄大量可用對稱性解決的概率謎題。
- **Gabor Székely, *Paradoxes in Probability Theory and Mathematical Statistics***（Reidel）。
- **Frederick Mosteller, *Fifty Challenging Problems in Probability***（Dover）。

12.7 淘汰賽中誰會遇上誰

題目

STEP II Specimen paper，第 16 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/spec-s2-q16`

引子

溫布頓網球錦標賽（Wimbledon）有 128 名選手參加抽籤。隨機選出其中兩人：他們在賽事中某個階段相遇的機會有多大？你也許會嘗試把每一輪相遇的概率（probability）加起來，而每一項都是「雙方晉級」與「被抽在一起」這些機會的乘積。這樣做是可行的，但有一個好得多的方法。數一數比賽的場數：每場比賽都由一對選手對賽，而比賽恰好有 127 場。根據對稱性（symmetry），每一對選手成為其中任何一場的對賽雙方的可能性都相同。因此所求的機會就是 127 除以選手配對的總數。

這道題是「數對了東西，概率問題便迎刃而解」的絕佳例子。答案 $\frac{1}{2^{n-1}}$ 只需一行便可得出。

動筆之前

想一想 有 2^n 名選手的淘汰賽（knockout tournament）共要進行多少場比賽？（每場比賽淘汰一名選手。）

每場比賽恰好淘汰一名選手，而除了一人之外所有選手都被淘汰，所以共有 $2^n - 1$ 場比賽。

逐步解題

階段 1　第一輪

把選出的兩名選手稱為 X 和 Y。在第一輪，X 的對手是其餘 $2^n - 1$ 名選手中任何一人的可能性都相同，所以

$$\mathrm{P}(\text{相遇於輪次1}) = \frac{1}{2^n - 1}.$$

階段 2 關鍵的對稱性

想一想 每一輪的抽籤都是隨機的，而每場比賽都相當於擲一次公平的硬幣。這個過程對待某一對選手的方式，會否與對待另一對有所不同？

不會。由於抽籤是隨機的，而且每場比賽的勝負機會都是 50–50，把選手重新標號並不會改變任何概率。所以任何一場指定的比賽（例如決賽）由 $\binom{2^n}{2}$ 對選手中任何一對對賽的可能性都相同。

階段 3 決賽

決賽是一場特定的比賽，所以它由 X 和 Y 對賽的概率是

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

（驗算：兩人都必須打進決賽，這需要他們分處不同的半區，並且各自勝出 $n - 1$ 場比賽。）

階段 4 在任何階段

想一想 X 和 Y 最多只能相遇一次。這一點怎樣讓你可以把各場比賽的概率相加？

設 I_m 在第 m 場比賽由 X 和 Y 對賽時取值 1。在 $2^n - 1$ 場比賽中，每一場都有 $\mathrm{P}(I_m = 1) = \binom{2^n}{2}^{-1}$，而這些事件是互斥（mutually exclusive）的（他們不可能相遇兩次），所以

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

以溫布頓為例，$n = 7$，兩名指定選手相遇的概率是 $\frac{1}{64}$。作為驗算，在 $2^n - 1$ 場比賽中，第一輪佔了 2^{n-1} 場，由此得出 $\frac{2^{n-1}}{\binom{2^n}{2}} = \frac{1}{2^n - 1}$，與第 1 階段的結果一致。

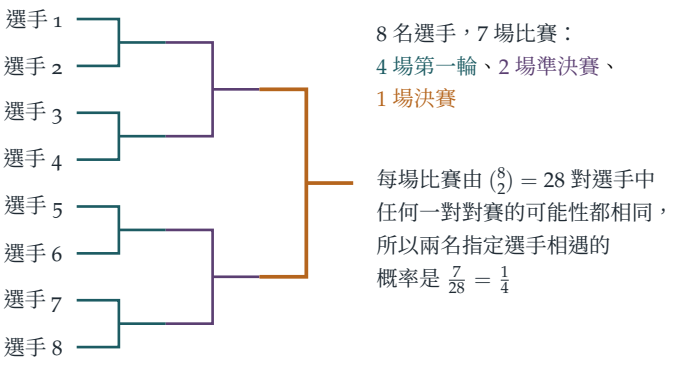


圖 12.7.1 $2^3 = 8$ 名選手的淘汰賽（圖中畫成固定的賽程表，但題目中每一輪都會重新隨機抽籤）。比賽共有 7 場。根據對稱性，每一場由 28 對選手中任何一對對賽的可能性都相同，所以選出的兩名選手相遇的概率是 $\frac{7}{28} = \frac{1}{4} = \frac{1}{2^{n-1}}$。

回顧

- **數比賽的場數。**比賽的場數 $2^n - 1$ 是關鍵的事實。
- **各對選手之間的對稱性。**每一對選手成為任何一場指定比賽的對賽雙方的可能性都相同。
- **對各場比賽的線性性質。**由於兩名選手最多相遇一次，相遇的概率就是相遇次數的期望值，也就是對各場比賽求和。

背後的數學

期望次數

第 4 階段的論證其實是一個期望值（expectation）的計算：X 和 Y 相遇的次數是若干個指示變量（indicator）之和，每場比賽對應一個；而它的期望值（由於它只取 0 或 1，所以等於相遇的概率）就是各個指示變量的期望值之和。這就是期望值的線性性質（linearity of expectation），在 Section 9.11 和 Section 12.4 中都曾用到。它最大的好處是完全不必理會各輪之間錯綜複雜的相依關係。

設有種子的賽事

現實中的賽事設有種子（seeded）：最強的選手會被編排在適當的位置，使他們要到後段輪次才可能相遇。而且現實中的比賽也不是 50–50 的。這兩點都破壞了對稱性，於是計算兩名指定選手相遇的機會便成了一項貨真價實的運算。為體育運動建立模型的統計

學家正是運用這類計算來預測賽事，當中每場比賽的概率都是根據選手的評分估計出來的。對稱的情況則是一條基線，用來衡量各種現實因素的影響。

思維技巧

數對的東西

只要選對了對象，概率往往可以用（有利對象的數目）÷（等可能對象的數目）求得。在這裏，對象就是比賽和選手配對。

選擇數甚麼，正是組合概率（combinatorial probability）的藝術。在大學裏，同一個想法會以算兩次（double counting）的形式出現（Section 7.9），也會出現在概率方法（probabilistic method）之中：在那裏，某種結構的期望數目可以用來證明它存在。

延伸練習

1 **逐輪計算。**直接證明 X 和 Y 在第 k 輪相遇的概率是 $\frac{2^{n-k}}{\binom{2^n}{2}}$，並驗證這些概率加起來等於 $\frac{1}{2^{n-1}}$。

2 **固定的賽程表。**假如賽程表改為預先固定（如圖所示），但所有選手都被隨機編入其中，答案會否改變？

3 **一名強手。**假設選手 X 每場比賽都勝出。現在 X 與 Y 相遇的概率是多少？

4 **聯賽。**在一個有 $2n$ 支球隊的聯賽中，每兩隊都對賽一次，共有多少場比賽？相應的「相遇概率」是甚麼？

延伸閱讀

- **Frederick Mosteller, *Fifty Challenging Problems in Probability***（Dover），書中收錄了有關淘汰賽的問題。
- **David Spiegelhalter, *The Art of Statistics***（Pelican），講述建立模型與預測。
- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***（CRC Press；可於網上免費閱讀），講述指示變量與對稱性。

12.8 Penney 遊戲

題目

STEP III 2001，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/01-s3-q13`

引子

你和朋友各自選定一個由三次擲硬幣結果組成的序列，然後不斷擲硬幣，直至其中一個序列出現為止。在任何一個時刻，每個三次投擲的序列出現的機會都相同，那麼無論選哪一個，理應同樣好吧？完全不是。無論朋友選哪個序列，你都可以選出一個勝過它的序列，勝算（odds）有時更高達 7 比 1。

這就是 *Penney 遊戲*（Penney's game），由 Walter Penney 於 1969 年發明。本題要證明 A 勝 B，B 勝 C，C 勝 D，而 D 又勝 A：這是一個循環，就像包剪揼一樣。「勝過」這種關係並不具傳遞性（transitive）。長遠而言，後選的一方總能勝出。

動筆之前

想一想 假設至今的投擲結果以 TT 結尾，而且仍未有人勝出。THH 有沒有可能比 TTH 先出現？

從 TT 出發，之後每擲出一個 T，最後兩次投擲仍然是 TT，而第一個 H 一出現就湊成 TTH。要 THH 出現，必須先有 TH，而 TH 只能由一個 T 後接一個 H 而來，但這個 H 已經湊成了 TTH。所以 TT 一旦出現，A 必勝。

逐步解題

階段 1 **狀態**

想一想 兩個序列都不可能在兩次投擲內完成。擲了兩次之後，從每個狀態（state）出發，下一次投擲會帶來甚麼結果？

由於遊戲只取決於最後兩次投擲，我們可以追蹤狀態（即最後兩次投擲的結果）：

- 從 TT 出發：A 勝（上面已證），所以 $p_{\mathrm{TT}} = 1$。
- 從 HT 出發：下一次投擲會得出 TH 或 TT，所以 $p_{\mathrm{HT}} = \frac{1}{2}p_{\mathrm{TH}} + \frac{1}{2}p_{\mathrm{TT}}$。
- 從 TH 出發：擲出 H 便湊成 THH（B 勝）；擲出 T 則得出 HT。所以 $p_{\mathrm{TH}} = \frac{1}{2}\cdot 0 + \frac{1}{2}p_{\mathrm{HT}}$。
- 從 HH 出發：擲出 H 仍是 HH；擲出 T 則得出 HT。所以 $p_{\mathrm{HH}} = \frac{1}{2}p_{\mathrm{HH}} + \frac{1}{2}p_{\mathrm{HT}}$，由此得 $p_{\mathrm{HH}} = p_{\mathrm{HT}}$。

階段 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 $p_{\mathrm{TH}} = \frac{1}{3}$。四個起始狀態的概率（probability）各為 $\frac{1}{4}$，所以

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

階段 3 B 對 C

想一想 C 想要 HHT，B 想要 THH。C 在甚麼情況下勝出？（留意第一個 HH 之前緊接着的必定是甚麼。）

如果頭兩次投擲是 HH，那麼其後第一個 T 便湊成 HHT，THH 來不及出現：C 勝。否則，在 HH 首次出現之前的某處必有一個 T，因此第一個 HH 的前一次投擲必定是 T：THH 就在那一刻出現，早於 HHT（HHT 還需要在 HH 之後再擲出一個 T）。所以只有當遊戲以 HH 開始時 C 才會勝出：

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

階段 4 循環

用同樣的方法（或「背後的數學」一節的捷徑）可以證明 C（HHT）勝 D（HTT）的概率是 $\frac{2}{3}$，而 D（HTT）勝 A（TTH）的概率是 $\frac{3}{4}$；後者正是第 3 階段把正面與反面對調後的鏡像。所以

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

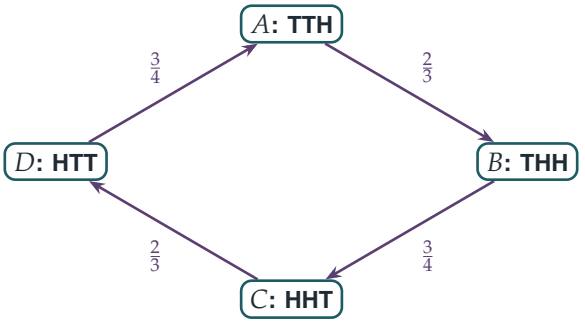


圖 12.8.1 Penney 循環。由 X 指向 Y 的箭頭表示 X 的序列勝過 Y 的序列，箭頭旁為相應的概率。每個序列都被另一個序列擊敗，所以後選的一方總能挑出勝算較高的序列。

回顧

- **選對狀態。**只有最後兩次投擲才有影響，所以四個狀態、四條方程便概括了整個遊戲。
- **首步分析（first-step analysis）。**每條方程都是以下一次投擲為條件：從某狀態出發的概率，就是下一次投擲所到達的各個狀態的概率的平均值。
- **找出決定性的事件。**在第 3 階段，「第一個 HH 之前是 T」這一點無須列方程便決定了勝負。

背後的數學

為甚麼有些序列較好

在任何一個位置，所有三次投擲的序列出現的機會都相同，而每個序列的期望等待時間（expected waiting time）也不難計算。但各序列的等待時間並不相同：TTH 平均要等 8 次投擲，TTT 卻要等 14 次，因為與自身重疊的序列（例如 TTT，只差一步的 TTH 會令已累積的兩次投擲前功盡廢）出現得較慢。而在兩個序列的競賽中，關鍵在於每個序列與對方如何重疊。THH 勝 HHT，是因為 HHT 每次出現（除了在最開頭之外），前面都有一個 T，這表示 THH 剛剛已經出現。後選一方的竅門總是一樣：取對手序列的頭兩個符號，再在前面加上與對手中間符號相反的那個符號。

康威公式（Conway's formula）

約翰·康威發現了一條計算勝算的絕妙公式。對於長度為 n 的兩個序列 X 和 Y，定義相關數（correlation）XY 為一個 n 位的二進制數：由左至右，若 X 中由第 i 個符號開始的

部分與 Y 開頭相同長度的部分相同，則第 i 位為 1。那麼 Y 比 X 先出現的勝算是

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

取 X = HHT 及 Y = THH：$XX = 100_2 = 4$（只有整個 HHT 與自身吻合）；$XY = 001_2 = 1$（只有 HHT 末尾的 T 與 THH 的開頭吻合）；$YY = 100_2 = 4$；而 $YX = 011_2 = 3$（THH 末尾的 HH 和 H 都與 HHT 的開頭吻合）。所以 THH 先出現的勝算是 $(4-1):(4-3) = 3:1$，即第 3 階段的 $\frac{3}{4}$。康威公式適用於任何長度、任何字母表的序列，其證明用到鞅（martingale）的理論：想像有一間賭場，賭客就序列的出現下注，再利用公平遊戲的期望利潤為零這一事實。

Penney 與加德納（Gardner）

Walter Penney 於 1969 年在 *Journal of Recreational Mathematics* 的一篇短文中介紹了這個遊戲。馬丁·加德納於 1974 年在他的 *Scientific American* 專欄中介紹了它，使它廣為人知，文中還包括康威公式，這條公式是康威發現的，但他一直沒有發表。這類非傳遞關係（埃弗龍骰子（Efron's dice）是另一個例子：骰子 A 勝 B，B 勝 C，C 勝 D，而 D 勝 A）一再令人驚訝，在投票理論中也很重要，因為多數人的偏好可以形成循環：這就是孔多塞悖論（Condorcet's paradox）。

思維技巧

首步分析

若一個過程在各狀態之間轉移，要求出與它有關的概率或期望值（expectation），可以第一步爲條件。這樣每個狀態都給出一條線性方程。

這是處理馬可夫鏈（Markov chain）的基本技巧，你會在 Section 12.16 以及大學概率課程中一再遇到。同樣的方法可以求出先到達某一狀態而非另一狀態的概率、被吸收前的期望時間，以及賭博問題中的破產概率。

延伸練習

1 **四人同場。**試做 STEP II 2015，第 12 題（`step.maths.org/questions/15-s2-q12`）：四名玩家同時參與的 Penney 遊戲。

2 **等待時間。**證明直至 HH 首次出現的期望投擲次數是 6，而直至 HT 首次出現的則是 4。（使用首步分析。）

3 **最佳對策。**對於八個長度為三的序列，逐一找出以最大概率勝過它的序列。哪個序列最容易被擊敗？

4 **康威公式。**用康威公式驗證 HHT 勝 HTT 的概率是 $\frac{2}{3}$。

延伸閱讀

- **Martin Gardner, *Time Travel and Other Mathematical Bewilderments***（W. H. Freeman）。書中包括 Penney 遊戲和康威公式。
- **Ronald Graham, Donald Knuth & Oren Patashnik, *Concrete Mathematics***（Addison-Wesley），第 8.4 節。以生成函數（generating function）處理 Penney 遊戲。
- **David Williams, *Probability with Martingales***（Cambridge University Press）。以鞅給出的證明，留待日後再讀。

12.9 用不公平的硬幣作公平的決定

題目

STEP I 2019，第 11 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2019 年 STEP 試卷（zip 檔案）內的 STEP I 試卷：

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

引子

你要作出一個公平（fair）的決定，但手上唯一的硬幣是彎的，而且你不知道它偏差（bias）有多大。這樣還能公平地作出選擇嗎？馮紐曼（John von Neumann）在 1951 年給出的答案是：擲兩次。若結果是 HT，就選酒店；若是 TH，就選茶室；若是 HH 或 TT，就再擲兩次。由於 HT 和 TH 的概率（probability）都是 pq，所以不論 p 是多少，這個決定都完全公平。

本題的程序剛好相反：它在出現 HH 和 TT 時作決定，這樣做很快，卻不公平（選酒店的概率是 $\frac{p^2}{p^2+q^2}$）。本題要證明的是，不論硬幣如何，這個程序都不會拖得很久。比較這兩個程序，可以看到產生隨機性時的一個基本取捨（trade-off）：速度與公平之間的取捨。

動筆之前

想一想 在一輪之中，作出決定的概率是多少？沒有決定的概率又是多少？$2pq$ 最大可以取甚麼值？

若兩枚硬幣結果相同，便作出決定，概率是 p^2+q^2；否則他們繼續下去，概率是 $2pq$。由於 $pq = p(1-p) \le \frac{1}{4}$（當 $p = \frac{1}{2}$ 時等號成立），所以 $2pq \le \frac{1}{2}$。

逐步解題

階段 1 在第 n 輪作出決定

各輪之間互相獨立（independent）。要在第 n 輪作出決定，須先有 $n-1$ 輪結果不同，然後一輪結果相同：

$$\mathrm{P}(\text{作出決定的輪次為}n) = (2pq)^{n-1}(p^2+q^2). \qquad \blacksquare$$

階段 2 在第 n 輪或之前作出決定

? **想一想** 在首 n 輪中沒有作出決定的概率是多少？

n 輪之內沒有決定的概率是 $(2pq)^n \le \left(\frac{1}{2}\right)^n$，所以不論 p 是多少，在第 n 輪或之前作出決定的概率至少是 $1-\frac{1}{2^n}$。 $\blacksquare$

階段 3 三個人

? **想一想** 第一輪出現兩個正面和一個反面之後，第三個人把反面翻成正面。第二輪要發生甚麼事才會有決定？如果第一輪是兩個反面和一個正面呢？

第一輪作出決定的概率是 p^3+q^3。否則，有兩枚硬幣結果相同，另一枚不同。若多數是正面（概率為 $3p^2q$），那枚不同的硬幣會被翻成正面，而第二輪要作出決定，就需要另外兩人都擲出正面：概率是 p^2（此時已不可能全是反面）。同理，若多數是反面（概率為 $3pq^2$），要作出決定就需要兩個反面：q^2。所以

$$\mathrm{P}(\text{作出決定的輪次不遲於}2) = p^3+q^3+3p^2q\cdot p^2+3pq^2\cdot q^2.$$

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

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

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

因為 $s \le \frac{1}{4}$。 $\blacksquare$

回顧

- **補事件（complement）。**「在第 n 輪或之前至少發生一次」最容易的算法是 1−（「n 輪之內從未發生」）。
- **為最壞情況定界。**未知的 p 是靠不等式（inequality）$pq \le \frac{1}{4}$ 來處理的。
- **對稱函數（symmetric functions）。**對 p 和 q 對稱的式子，用 $p+q=1$ 和 pq 來表示時便會簡化。

背後的數學

馮紐曼的妙計

題目中的程序很快，但有偏差。馮紐曼的程序則是公平的：擲兩次，出現 HT 或 TH 就作決定，出現 HH 或 TT 就重複。每次嘗試成功的概率是 $2pq$，所以平均需要 $\frac{1}{2pq}$ 次嘗試；當硬幣偏差很大時，這會很慢（圖 12.9.1）。題目的程序每輪作出決定的概率是 $p^2+q^2 \geq \frac{1}{2}$，但當 $p > \frac{1}{2}$ 時，它的結果偏向酒店。

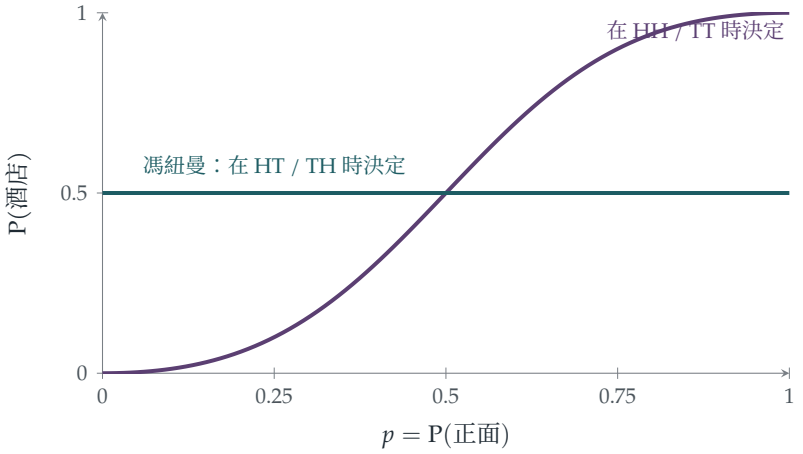


圖 12.9.1 選中酒店的概率，對應硬幣的偏差。在兩枚硬幣結果相同時作決定（紫紅色）很快，但偏向較可能出現的一面。馮紐曼的程序（藍綠色）對每一個 p 都完全公平，代價是當硬幣偏差很大時需時較長。

提取隨機性

馮紐曼的妙計是第一個隨機性提取器（randomness extractor）：一種把不完美的隨機性變成完美隨機性的程序。它浪費很多（每兩次投擲最多只得到一個公平的位元，而且往往一個也得不到）；更好的方法會把被棄掉的 HH 和 TT 重新利用，從而提取硬幣所含的幾乎全部隨機性，而這隨機性的多少是用硬幣的熵（entropy）來量度的，即每次投擲 $-p\log_2 p - q\log_2 q$ 個位元。提取器用於電腦的硬件隨機數產生器：這些產生器收集不可預測但有偏差的物理雜訊，並必須把它變成無偏差的位元，供密碼學（cryptography）使用。

馮紐曼與隨機數

馮紐曼在 1951 年的論文「Various techniques used in connection with random digits」中描述了這個妙計。當時他和同事正在為設計核武器而進行一些最早期的電腦模擬（即

「蒙地卡羅方法」（Monte Carlo method））。在同一篇論文中，他就用算術方法產生隨機數說了一句名言：「任何人若考慮用算術方法來產生隨機數字，當然就是身處罪惡之中。」

思維技巧

利用對稱性消去未知量

當某個參數未知時，找出那些不論參數取何值、概率都相等的事件，並以它們作爲決定的依據。

HT 和 TH 有相同的概率 pq，所以比較這兩者就把 p 完全消去了。在大學裏，同樣的想法在統計學中以樞軸量（pivotal quantities）的形式出現，也出現於實驗設計之中：在那裏，隨機化（randomisation）和對稱性會把未知的效應抵消掉。

延伸練習

1 **輪數的期望值。**求第 (i) 部分的程序所需輪數的期望值（expected number），以及馮紐曼的程序所需輪數的期望值。對哪些 p 而言，馮紐曼的程序較慢？

2 **公平的三選一。**利用一枚有偏差的硬幣，設計一個程序，從三個選項中選出一個，使每個選項的概率都恰好是 $\frac{1}{3}$。

3 **用公平硬幣造出有偏差的骰子。**利用一枚公平的硬幣，設計一個程序，模擬一枚擲出正面的概率恰好是 $\frac{1}{3}$ 的硬幣。所需投擲次數的期望值是多少？

4 **再看三個人。**在第 (ii) 部分中，他們去酒店的概率是多少？請以 p 的函數表示。當 $p = \frac{1}{2}$ 時，這個程序公平嗎？

延伸閱讀

- **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，論隨機數。
- **Michael Mitzenmacher & Eli Upfal，*Probability and Computing***（Cambridge University

Press），論提取隨機性與熵。

12.10 把繩頭繫起來

題目

STEP III 2008，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/08-s3-q13`

引子

把一百條繩子放進袋中，只讓繩頭露在外面。隨機選兩個繩頭綁在一起，如此重複，直至再沒有鬆開的繩頭為止。最後你會得到一堆糾纏的繩圈。究竟有多少個？一百條繩子原則上可以結成一百個小環，也可以結成一個巨大的圈。出人意表的是，繩環數目的期望值（expected number）只有約 3.3；即使有四萬條繩子，也不過約 6。

只要看出一個關鍵之處，問題便幾乎不費吹灰之力：每打一個結之後，不論剛才發生了甚麼事，箱中都只是少了一條繩子。這個過程會「忘記」自己的歷史，於是繩環的數目就變成一連串獨立（independent）擲硬幣結果之和，只是每次的概率（probability）各有不同。

動筆之前

想一想 把兩個繩頭綁在一起之後，還剩下多少條繩子（結成環的不再算作一「條」）？這是否取決於有沒有結成環？

如果兩個繩頭屬於同一條繩子，這條繩子便成為一個環，餘下 $n-1$ 條。如果兩個繩頭屬於不同的繩子，這兩條繩子便接成較長的一條，同樣餘下 $n-1$ 條。無論是哪一種情況，我們都剩下 $n-1$ 條繩子，每條各有兩個繩頭。

逐步解題

階段 1　第一個結

繩頭共有 $2n$ 個。先選定一個；與它相綁的繩頭等可能地是其餘 $2n-1$ 個繩頭中的任何一個，而當中恰好有一個屬於同一條繩子。因此第一個結結成環的概率是 $\frac{1}{2n-1}$，而第一步所產生的繩環數目的期望值就是 $\frac{1}{2n-1}$。

階段 2　所有的結

> **想一想**　根據上面的觀察，打第二個結之前的情況是怎樣的？第三個呢？設 I_k 為指示變量，表示在剩下 k 條繩子時所打的結是否結成環。各個 $\mathrm{E}(I_k)$ 是多少？

每次打結之前，箱中有 k 條繩子，其中 $k = n, n-1, \dots, 1$，而這個結結成環的概率是 $\frac{1}{2k-1}$。（最後一個結在只剩一條繩子時打，必定結成環。）繩環的總數是 $R = I_n + I_{n-1} + \cdots + I_1$，所以

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

階段 3　方差

> **想一想**　各個指示變量（indicator）I_k 是否獨立？（某一步結成環的概率，是否取決於之前發生過的事？）

在每一步，不論之前發生過甚麼事，結成環的概率都是 $\frac{1}{2k-1}$，因為當時的格局永遠是「k 條繩子」。所以各個 I_k 互相獨立，方差（variance）可以相加：

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

階段 4　四萬條繩子

> **想一想**　試用調和數（harmonic numbers）表示直至 $2n-1$ 為止所有奇數的倒數之和。

把直至 $2n$ 為止所有正整數的倒數之和，減去其中偶數的倒數之和，便得

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

當 $n = 40\,000$ 時，$4n = 160\,000 = 20^4$，所以 $\mathrm{E}(R) \approx \frac{1}{2} \cdot 4 \ln 20 \approx 6$。 ■（準確值是 6.28；兩者之差來自歐拉常數（Euler's constant），因為 $H_n \approx \ln n + 0.577$。）

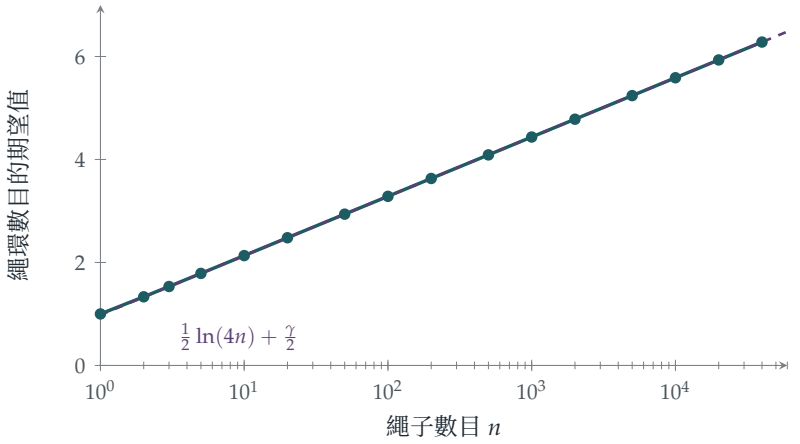


圖 12.10.1 繩環數目的期望值與繩子數目的關係，橫軸採用對數尺度。準確值（藍綠色）落在直線 $\frac{1}{2}\ln(4n) + \frac{\gamma}{2}$（紫紅色虛線）上，其中 $\gamma \approx 0.577$ 是歐拉常數。把繩子數目乘以 10，繩環只會多出約 1.15 個。

回顧

- **找出過程忘記了甚麼。**每打一個結之後，不論過去的經過如何，箱中都是「k 條繩子」。
- **又是指示變量。**繩環的數目是指示變量之和，每個結對應一個。
- **有獨立性便有方差。**由於每一步的概率都不取決於過去，各個指示變量互相獨立。

背後的數學

對數增長

繩環數目的期望值以 $\frac{1}{2}\ln n$ 的速度增長，而方差與平均值相當接近（兩者都是一些很小的概率之和）。根據中心極限定理（central limit theorem）（Section 12.24），當 n 很大時，繩環數目近似服從正態分佈（normal distribution），平均值和方差都約為 $\frac{1}{2}\ln(4n)$。若有 40 000 條繩子，你通常會得到 2 至 11 個環。

隨機置換的循環

同一個對數也出現在 n 件物件的隨機置換（random permutation）的循環（cycle）數目之中：平均有 $H_n \approx \ln n$ 個（繩子問題是它的近親，對應的是 $\frac{1}{2}\ln n$）。隨機置換的循環很少：通常有一個巨大的循環包含了大部分元素，而最大循環所佔元素的比例，平均約為 0.62，即 Golomb–Dickman 常數。隨機地把繩子綁在一起，情況也是一樣：大部分繩子最終都落在一兩個長圈之中。這些事實屬於隨機組合結構（random combinatorial structures）的研究範圍，而這門學問又與數論（number theory）有關（一個隨機整數的質因數，其表現就像一個隨機置換的循環）。

隨機紐結與聚合物

綁繩子問題是某些物理系統的簡化模型。溶液中的長鏈聚合物分子會首尾相接，形成環和鏈，而成環的統計規律也遵從類似的法則。在紐結理論（knot theory）中，數學家問的是一個更難的問題：一個隨機的閉合圈有多大機會打了結？對於很長的隨機圈，答案趨向 1，這個結果由 De Witt Sumners 和 Stuart Whittington 於 1988 年證明。

思維技巧

> **找出無記憶的結構**
>
> 如果每一步之後，系統所處的狀態只取決於已經走了多少步，那麼各步的表現便互相獨立，期望值和方差都可以直接相加。
>
> 這正是優惠券收集問題（coupon collector）（Section 9.11）背後的想法，當中每個階段都是一次全新的幾何等待；它也在馬可夫鏈（Markov chains）的理論中再次出現，在那裏只有當前的狀態才起作用。

延伸練習

1. **小的情況。**對於 $n = 2$，列出所有可能情況，並驗證繩環數目的期望值是 $\frac{4}{3}$。

2. **一個大圈。**證明整個過程最終得到單一個包含所有繩子的圈的概率是 $\prod_{k=2}^{n}\left(1 - \frac{1}{2k-1}\right)$，並證明當 n 很大時，這個概率約為 $\sqrt{\frac{\pi}{4n}}$。（提示：找出這個乘積與 $\binom{2n}{n}$ 的關係。）

3. **置換的循環。**證明 n 件物件的隨機置換的循環數目的期望值是 H_n，方法是逐個元素地構造這個置換（即「中國餐館」構造法）。

4 **方差。**證明 $\mathrm{Var}(R) = \mathrm{E}(R) - \sum_{k=1}^{n} \frac{1}{(2k-1)^2}$，並證明被減去的那個和趨向 $\frac{\pi^2}{8}$。

延伸閱讀

- **Peter Winkler, *Mathematical Puzzles: A Connoisseur's Collection***（A K Peters），書中收錄了這道綁繩子謎題。
- **Richard Arratia, Andrew Barbour & Simon Tavaré, *Logarithmic Combinatorial Structures***（European Mathematical Society）。講述循環、質數與隨機結構，可留待日後閱讀。
- **Colin Adams, *The Knot Book***（American Mathematical Society）。紐結理論入門，內容包括隨機紐結。

12.11 郵票、杯子與 e

題目

STEP I 1996，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/96-s1-q13`

引子

有這樣一種古怪的賣郵票方法：買家不斷猜郵票藏在哪一個杯子下面，每一輪多加一個杯子；到他第一次猜錯時，桌上有多少個杯子，他就要付多少英鎊。平均而言，他要付 £e，約為 £2.72。

e 這個數在概率（probability）中出現之頻繁，簡直不可思議：沒有人拿回自己襯衫的機會（Section 9.8）、祕書問題（secretary problem，Section 12.12）、泊松分佈（Poisson distribution）。本題也許是它眾多現身之中最簡單的一次，而且原來它與另一道著名謎題有着完全相同的分佈（distribution）：要把多少個介乎 0 與 1 之間的隨機數相加，總和才會超過 1？平均而言，是 e 個。

動筆之前

想一想 若 $N = 4$，必須發生甚麼事？概率是多少？

Jim 必須在有 2 個杯子時猜中（概率為 $\frac{1}{2}$），在有 3 個杯子時也猜中（$\frac{1}{3}$），然後在有 4 個杯子時猜錯（$\frac{3}{4}$）。所以 $\mathrm{P}(N = 4) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{3}{4} = \frac{3}{4!} = \frac{1}{8}$。

逐步解題

階段 1 分佈

若 $N = k$（其中 $k \geq 2$），Jim 在有 $2, 3, \ldots, k-1$ 個杯子時都猜中，而在有 k 個杯子時猜錯：

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

（驗算：$\frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \cdots = \sum_{k \geq 2} \left(\frac{1}{(k-1)!} - \frac{1}{k!} \right) = 1$，這是一個裂項和（telescoping sum）。）

階段 2　期望值

想一想　化簡 $k \cdot \frac{k-1}{k!}$。你會得到甚麼級數（series）？

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

階段 3　方差

想一想　求 $\mathrm{E}(N^2)$：可寫成 $k^2 = k(k-1) + k$，或直接化簡 $\frac{k^2(k-1)}{k!}$。

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

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

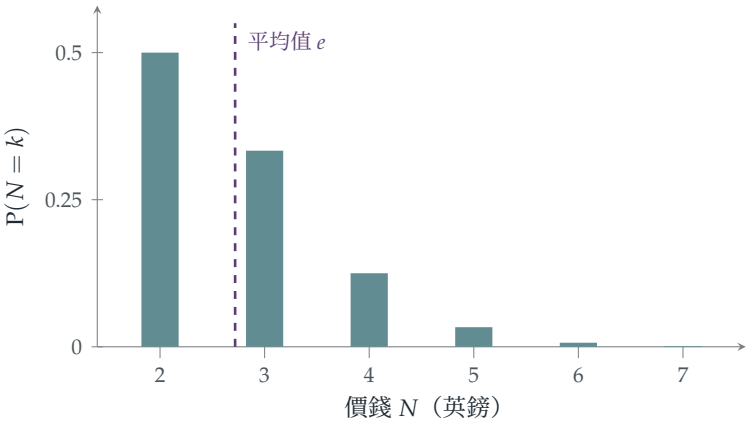


圖 12.11.1　價錢的分佈，$\mathrm{P}(N = k) = \frac{k-1}{k!}$。有一半機會 Jim 只需付 £2；那條又長又細的尾巴把平均值拉高至 £$e \approx$ £2.72。

回顧

► **概率的乘積。**一連串獨立（independent）的成功之後接着一次失敗，所得的乘積可

化簡為 $\frac{k-1}{k!}$。

- **約去階乘。**乘以 k 或 k^2 再約簡，那些和式就變成了指數級數（exponential series）。
- **以平移求 $\mathrm{E}(N^2)$。**以 $j = k-2$ 重新編排下標，和式便分拆成兩份 e。

背後的數學

換了裝扮的同一個分佈

從 $[0,1]$ 中均勻（uniformly）地隨機選取數 $U_1, U_2, \ldots$，並設 M 為令它們的總和超過 1 所需的個數。首 k 個數之和小於 1 的概率，就是區域 $\{u_1 + \cdots + u_k < 1, u_i \geq 0\}$ 的體積，即 $\frac{1}{k!}$（面積為 $\frac{1}{2}$ 的三角形、體積為 $\frac{1}{6}$ 的四面體，如此類推）。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 的分佈。在兩個問題中都有 $\mathrm{P}(N > k) = \frac{1}{k!}$：就杯子而言，這是連續猜中 $k-1$ 次的概率。利用 $\mathrm{E}(N) = \sum_{k \geq 0} \mathrm{P}(N > k)$（Section 9.9 中的一條公式），一行便得出 $\mathrm{E}(N) = \sum \frac{1}{k!} = e$。

為何 e 不斷出現

每當概率中含有 $\frac{1}{k!}$，不論它來自 k 件物件按某個特定次序排列、k 個概率依次為 $\frac{1}{2}, \frac{1}{3}, \ldots$ 的相繼事件，還是單純形（simplex）的體積，對 k 求和都會產生指數級數。錯排（derangements，Section 9.8）給出 $\sum \frac{(-1)^k}{k!} = e^{-1}$；這裏我們則有 $\sum \frac{1}{k!} = e$。至於形如 $\left(1 - \frac{1}{n}\right)^n \to e^{-1}$ 的極限（limit），它們源自大量獨立而不大可能發生的事件，正好解釋了祕書問題和泊松分佈中的 e。

蒙地卡羅 e

「總和超過 1」這個問題提供了一個以實驗估計 e 的方法：產生隨機數，數一數需要多少個才超過 1，重複多次，然後取平均值。它是蒲豐投針（Buffon's needle，Section 12.20）的近親，後者用來估計 π。這類蒙地卡羅（Monte Carlo）方法以賭場命名，利用隨機抽樣來估計難以直接計算的量；如今它們廣泛應用於科學的各個範疇，由粒子物理學到金融，主要用於計算高維積分。

思維技巧

認出級數

當期望值是一個含有階乘的和式時，把各項化簡，直至出現一個標準級數（指數級數、幾何級數、二項級數）。

在大學，生成函數(generating function)把這種做法系統化：概率生成函數(probability generating function) $\mathrm{E}(t^N)$ 把整個分佈包裝起來，而它在 $t = 1$ 處的導數則給出平均值和方差。

延伸練習

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

2 **尾部公式。**對於取非負整數值的隨機變量（random variable），證明 $\mathrm{E}(N) = \sum_{k\geq 0} \mathrm{P}(N > k)$，並用它一行求出 $\mathrm{E}(N)$。

3 **由一個杯子開始。**假如遊戲由一個杯子開始 (Jim 不可能猜錯)，分佈會有甚麼改變？價錢的期望值是多少？

4 **單純形。**證明 $\{u_1 + u_2 + u_3 < 1, u_i \geq 0\}$ 的體積為 $\frac{1}{6}$，並用歸納法（induction）證明 k 維的情況為 $\frac{1}{k!}$。

延伸閱讀

- **Eli Maor, *e: The Story of a Number***（Princeton University Press）。e 的歷史，包括它在概率中的各次現身。
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），關於生成函數。
- **Paul Nahin, *Digital Dice***（Princeton University Press）。以蒙地卡羅模擬解決的概率謎題。

12.12 何時停止尋找

題目

STEP I 2001，第 14 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/01-s1-q14`

引子

你正在逐一面試應徵者，而且必須即場決定：取錄眼前這一位，然後停止；還是永遠拒絕這一位，再見下一位。你想要的是最好的那一位。如果太早取錄，你很可能錯過更好的人；如果等得太久，最好的那一位很可能已經走了。你應該怎樣做？

答案是概率（probability）論中最著名的結果之一：先拒絕最初 37% 的應徵者，然後取錄第一位比這些人都好的應徵者。這個做法大約有 37% 的機會成功，而 37% 正是 $\frac{1}{e}$。本題計算其中的基本概率，而與它配對的 STEP II 2006 題目（見練習）則分析這個策略（strategy）最簡單的版本：只拒絕第一位應徵者。

動筆之前

想一想 在隨機次序之下，全體之中最好的應徵者落在最先面試的 n 位之中的概率是多少？

最好的應徵者落在 N 個位置中任何一個的可能性均等，所以概率是 $\frac{n}{N}$。

逐步解題

階段 1 最先 n 位中的最佳者就是全體最佳者

最先 n 位中的最佳者是全體最佳者，當且僅當全體最佳者在最先 n 位之中：概率為 $\frac{n}{N}$。

■

階段 2　最佳或次佳

想一想　最先 n 位中的最佳者是全體次佳者，當且僅當次佳者在最先 n 位之中而最佳者不在其中。這個概率是多少？

次佳者在最先 n 位之中的概率是 $\frac{n}{N}$；在此條件下，最佳者在其餘 $N-1$ 個位置的其中一個，當中有 $N-n$ 個位置排在最先 n 位之後。所以

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

當 $N=4$，$n=2$ 時：$\frac{1}{2}+\frac{1}{2}\cdot\frac{2}{3}=\frac{5}{6}$。把最好的兩位應徵者在四個位置中的排法逐一列出（全部 12 種有序排法的可能性均等），除了兩位頂尖應徵者同時落在第 3 和第 4 位的 2 種排法之外，最先兩位中的最佳者都是頂尖兩位之一：$\frac{10}{12}=\frac{5}{6}$。

階段 3　最優策略

想一想　策略：先拒絕最初 r 位應徵者，然後取錄第一位比他們全部都好的應徵者。如果最佳應徵者在位置 $k>r$，這個策略在甚麼情況下會選中此人？

設最佳應徵者在位置 k，這個策略選中此人，當且僅當位置 $r+1,\dots,k-1$ 上沒有任何應徵者勝過最初 r 位全部的人，即是說，最先 $k-1$ 位應徵者中的最佳者在最初 r 位之中。根據第 1 階段（把 N 換成 $k-1$），這件事的概率是 $\frac{r}{k-1}$。最佳應徵者落在每個位置 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　當 N 很大時，記 $x=\frac{r}{N}$。這個和大約等於 $\ln\frac{N}{r}$。x 取甚麼值時 $x\ln\frac{1}{x}$ 達到最大？

成功的概率大約是 $x\ln\frac{1}{x}$。求導可得，$\ln\frac{1}{x}-1=0$ 於 $x=\frac{1}{e}$ 時成立，而最大值是 $\frac{1}{e}\approx 0.368$（圖 12.12.1）。

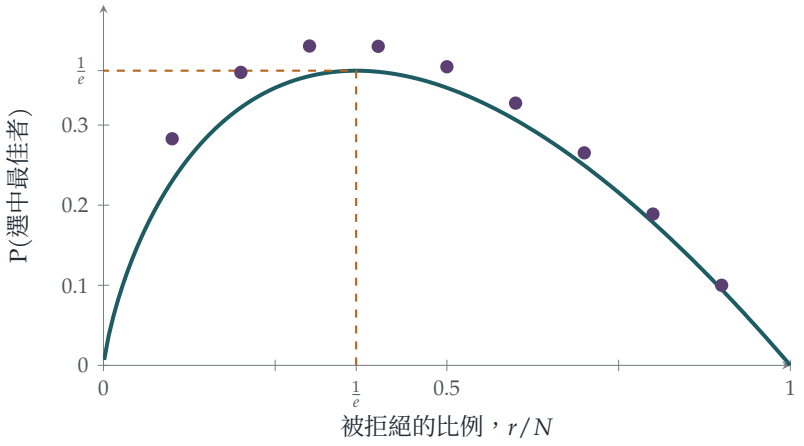


圖 12.12.1 若先拒絕佔比例 r/N 的最初一批應徵者，選中最佳應徵者的概率：極限曲線 $x \ln \frac{1}{x}$（藍綠色），以及 $N = 10$ 時的精確值（紫紅色）。最佳策略是拒絕大約 $\frac{1}{e} \approx 37\%$ 的應徵者，成功的概率大約是 $\frac{1}{e}$，不論 N 有多大。

回顧

- **位置的對稱性。**在隨機次序之下，每位應徵者落在任何位置的可能性均等，由此立即得出 $\frac{n}{N}$。
- **以最佳者的位置為條件。**計算策略的成功概率時，我們以最佳應徵者所在的位置為條件分情況處理。
- **求和變成積分。**當 N 很大時，調和和（harmonic sum）變成了對數，再經一次求最大值便得出 $\frac{1}{e}$。

背後的數學

最優停止

祕書問題（secretary problem）是最優停止（optimal stopping）的奠基例子：根據隨時間陸續到來的資訊，決定何時採取行動，以使期望回報最大。可以證明，在所有策略之中（而不只是「先拒絕 r 位，然後取錄下一位至今最佳者」這一類），這種閾值規則（threshold rule）是最優的，論證的方法是由最後一位應徵者開始向前逆推。同樣的數學也決定了何時賣出股票、何時行使金融期權，以及應該花多久去找泊車位。一個值得注意的特點是，在極限之下答案與 N 無關：不論應徵者有一百位還是一百萬位，你成功的概率都是 $\frac{1}{e}$。

各種變化

如果你只要選中頂尖兩位應徵者的其中一位便滿意，最佳策略大約有 57% 的機會成功。如果你想做到的不是使選中最佳者的機會最大，而是使所選應徵者的期望*名次*最優，最優策略便複雜得多，而期望名次趨向大約 3.87（Chow、Moriguti、Robbins 和 Samuels，1964）。如果你有一定的概率可以回頭召回已被拒絕的應徵者，又或者你知道應徵者分數的分佈（即「完全資訊」問題），答案又會不同。練習中的雪糕題是最簡單的情形：只拒絕一個，然後取下一個比它好的。

是誰解決的？

這個問題在 1950 年代流傳於數學家之間，馬丁·加德納（Martin Gardner）於 1960 年以「googol 遊戲」之名把它普及開來（在紙條上寫下數字；你的對手逐張翻開，並且必須在最大的那個數出現時停下）。Thomas Ferguson 的文章「Who solved the secretary problem?」(1989)追溯了它錯綜複雜的歷史；有一種說法指，天文學家開普勒(Johannes Kepler）在 1611 年挑選第二任妻子時，會見了十一位人選，最後向第五位求婚。

思維技巧

把策略參數化，然後求最優

當最佳策略未知時，考慮一族帶有參數的自然策略，把成功概率寫成該參數的函數，然後求最優。

在大學裏，這種方法用於決策論、博弈論和統計學。要證明這一族之中的最佳策略就是全體之中的最佳策略，是更困難也更有趣的一步，通常要用*逆向歸納法*（backward induction）來完成。

延伸練習

1. **雪糕。**試做 STEP II 2006，第 13 題（`step.maths.org/questions/06-s2-q13`）：這是祕書問題的小型版本，規則很簡單。拒絕第一個，然後取第一個勝過它的。

2. **較小的 N。**當 $N = 4$ 時，求 r 的最佳值和成功的概率。用列舉法驗算。

3. **最佳或次佳。**利用第 (ii) 部分，論證「最先 n 位中的最佳者是頂尖兩位之一」的可能性遠高於「最先 n 位中的最佳者就是最佳者」。這對於選中頂尖兩位之一的策略有甚麼啟示？

4 **精確的最大值。**當 $N = 10$ 時，從圖中驗證 $r = 3$ 是最優的。當 $N = 100$ 時，證明 $r = 37$ 是最優的（準確至四位小數，$r = 36$、37、38 的成功概率分別是 0.3710、0.3710、0.3708）。

延伸閱讀

- **Thomas Ferguson，「Who solved the secretary problem?」**，*Statistical Science* 4 (1989)。歷史與各種變化；可在網上免費閱讀。
- **Brian Christian & Tom Griffiths，*Algorithms to Live By***（William Collins）。把最優停止及其他算法應用於日常決策。
- **Martin Gardner，*New Mathematical Diversions***（Mathematical Association of America）。googol 遊戲。

12.13 三人決鬥

題目

STEP II 1995，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/95-s2-q12`

引子

三個人用餐包決鬥。Cuthbert 百發百中，Bertie 每五次擲中三次，而 Algernon 準頭最差，每五次只擲中兩次。Algernon 先擲。他應該瞄準誰？答案出人意表：誰也不瞄準。他應該故意把餐包擲偏。

原因在於策略。只要三人都仍在場，Bertie 和 Cuthbert 這兩個危險的對手就會互相瞄準。假如 Algernon 淘汰了其中一人，餘下的那個便會轉而對付他。他故意擲失，就能讓那兩人先行分出勝負，然後由他向倖存者先擲一次。事實上，只要採用最佳方案，準頭最差的人反而最有可能勝出。這場「三人決鬥」（truel）是博弈論（game theory）中常被引用的例子，說明弱勢有時反而是優勢。

動筆之前

想一想 在一場二人決鬥中，Percy 先擲。他要麼一擲即中，要麼兩人都擲失，於是局面完全重演。試寫出一條關於 Percy 勝出概率（probability）的方程。

設 P 為 Percy 勝出的概率。他一擲即中的概率是 p；否則（概率為 $1-p$）輪到 Rodney 擲，而 Rodney 擲失的概率是 $1-r$，此時局面回到起點。所以 $P = p + (1-p)(1-r)P$。

逐步解題

階段 1　二人決鬥

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 ($p=\frac{2}{5}$) 對 Bertie ($r=\frac{3}{5}$)，由 Algernon 先擲，這個概率是 $\frac{2/5}{2/5+3/5-6/25}=\frac{10}{19}$；若由 Bertie 先擲，Algernon 勝出的概率是 $\left(1-\frac{3}{5}\right)\cdot\frac{10}{19}=\frac{4}{19}$。

階段 2 Algernon 的三個選擇

想一想 在每種情況下，還剩下誰？下一個輪到誰擲？

(i) *Algernon* 擲中 *Cuthbert*。接着由 Bertie 擲，目標是 Algernon。Algernon 要在 Bertie 先擲的二人決鬥中勝出：概率為 $\frac{4}{19}$。

(ii) *Algernon* 擲中 *Bertie*。接着由 Cuthbert 擲，目標是 Algernon，而他從不失手。Algernon 的勝出機會是 0。

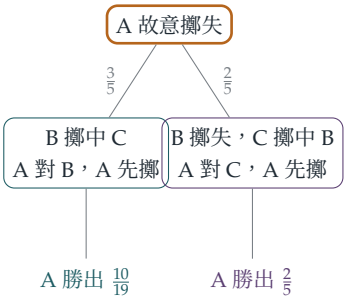
(iii) *Algernon* 擲失。Bertie 向 Cuthbert 擲。他有 $\frac{3}{5}$ 的概率擲中，剩下 Algernon 和 Bertie，並由 Algernon 先擲：Algernon 勝出的概率是 $\frac{10}{19}$。Bertie 有 $\frac{2}{5}$ 的概率擲失，然後 Cuthbert 擲中 Bertie，Algernon 便要面對 Cuthbert，而且只有一次出手機會：他必須擲中才能勝出，概率為 $\frac{2}{5}$。合計為

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

階段 3 建議

想一想 如果 Algernon 瞄準 Cuthbert，他可能擲中（概率為 $\frac{2}{5}$），也可能擲失。他整體的勝出機會是多少？試與故意擲失比較。

瞄準 Cuthbert 的勝出概率是 $\frac{2}{5}\cdot\frac{4}{19}+\frac{3}{5}\cdot\frac{226}{475}=\frac{200+678}{2375}=\frac{878}{2375}\approx 0.370$。瞄準 Bertie 則更差。故意擲失的勝出概率是 $\frac{226}{475}\approx 0.476$，是三者中最高的。所以 Algernon 應該把餐包擲偏，這樣他勝出的概率是 $\frac{226}{475}$。 $\blacksquare$



採用這個方案：
A 勝出 $\frac{226}{475} \approx 0.48$
B 勝出 $\frac{135}{475} \approx 0.28$
C 勝出 $\frac{114}{475} \approx 0.24$

圖 12.13.1 Algernon 的最佳方案。他故意擲失，讓兩個較強的對手先行交鋒，並確保自己能向倖存者先擲一次。結果準頭最差的人最有可能勝出，而百發百中的人反而最不可能勝出。

回顧

- **重複出現的局面可以列出方程。**兩人都擲失後，決鬥「重新開始」，由此得出一條關於勝出概率的線性方程（linear equation）。
- **考慮對手的反應。**Algernon 的每個選擇，都會改變 Bertie 和 Cuthbert 接下來的行動。
- **比較所有選擇，包括甚麼都不做。**最好的一着，正是看來最差的那一着。

背後的數學

策略與博弈論

三人決鬥展示了博弈論的核心思想：最佳行動取決於其他參與者會如何回應。擲中最危險的對手 Cuthbert 看似理所當然，但這樣做會改變局勢，反而對 Algernon 不利。在本題中，Bertie 和 Cuthbert 的策略（strategy）是固定的（他們互相瞄準）；在完整的博弈論分析中，每個參與者都會自行選擇策略，並且知道其他人也在這樣做。這類分析始於馮·諾伊曼（John von Neumann）與摩根斯坦（Oskar Morgenstern）在 1944 年的工作，以及納殊（John Nash）在 1950 年的工作，如今已成為經濟學、演化生物學和政治學的基礎。

強大反成弱點

三人決鬥中誰會勝出？若 Algernon 採用最佳方案，Algernon 勝出的概率是 $\frac{226}{475}$，Bertie 是 $\frac{135}{475}$，Cuthbert 是 $\frac{114}{475}$。準頭最好的人最不可能倖存，因為凡是可以選擇目標的人都會瞄準他。同樣的規律也見於有多名候選人的選舉、企業之間的競爭，以及自然界：最強

的競爭者可能成為首個被攻擊的目標。Martin Shubik 在 1950 年代把三人決鬥引入博弈論，並藉此探問：最適者是否必然生存？

思維技巧

多想一步

選擇行動之前，先弄清楚它會怎樣改變其他各人面對的局面，以及他們會如何回應。

到了大學，這會發展成博弈論中的逆向歸納法（backward induction）：先分析博弈的終局，再往回推，從而決定最佳的第一步。

延伸練習

1 **Bertie 與 Cuthbert。**驗證：若 Algernon 故意擲失，Bertie 勝出的概率是 $\frac{135}{475}$，Cuthbert 勝出的概率是 $\frac{114}{475}$。

2 **準頭較好的投手。**假如 Algernon 擲中的概率是 a 而不是 $\frac{2}{5}$，當 a 取哪些值時，故意擲失仍然是他最佳的第一步？

3 **Bertie 應該瞄準 Cuthbert 嗎？**已知 Algernon 故意擲失，Bertie 瞄準 Cuthbert 而不瞄準 Algernon，是否對他自己有利？（假設 Cuthbert 的行為不變。）

4 **投擲次數有限的決鬥。**在一場每人只可擲兩次的二人決鬥中，求 Percy 勝出的概率，並與不限次數的決鬥比較。

延伸閱讀

- **D. Marc Kilgour & Steven Brams，「The truel」**，*Mathematics Magazine* 70 (1997)。一篇淺白易讀的三人決鬥綜述。
- **Avinash Dixit & Barry Nalebuff，*Thinking Strategically***（Norton）。寫給一般讀者的博弈論讀物，當中也談到三人決鬥。
- **Ken Binmore，*Game Theory: A Very Short Introduction***（Oxford University Press）。

12.14 混合策略

題目

STEP I 1989，第 16 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/89-s1-q16`

引子

玩「包剪揼」時，任何固定的出法都會被對手利用：每次都出揼（石頭），對手很快便學會出包（布）。唯一穩妥的策略就是隨機出手。本題是一個更精細的版本：各種結果的得失並不相等，怎樣隨機化才對也並不明顯。結果發現，只要以概率 $\frac{1}{4}$、$\frac{1}{2}$、$\frac{1}{4}$ 分別猜 1、2、3，不論對手怎樣做，玩家都可以保證長遠而言不會輸錢。

這讓我們初次窺見馮紐曼（John von Neumann）在 1928 年提出的**極小極大定理**（minimax theorem），它是博弈論（game theory）的奠基結果：在任何一方所得即另一方所失的二人博弈中，雙方各有一個隨機化策略（randomised strategy），能保證在對手發揮得完美無瑕時，仍取得可能範圍內的最佳結果。

動筆之前

想一想 寫出一個表，列出每一對猜測之下 B 贏得的金額（B 的猜測列於左側，A 的猜測列於頂端）。

按規則逐一推算（例如，若 B 說 1 而 A 說 3，兩數相差 2，說出較小數的 B 要付 £2）：

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

這個表是**反對稱**（antisymmetric）的（把兩位玩家對調，正負號便會改變），所以這個博

弈是公平的，意思是雙方都沒有與生俱來的優勢。

逐步解題

階段 1　對付固定的猜測

若 A 每次都說 1，B 應該說 3，每局贏 £2。若 A 每次都說 2，B 說 1，贏 £1。若 A 每次都說 3，B 說 2，贏 £1。任何固定的猜測都會被對手利用。

階段 2　對付隨機的猜測

> **想一想**　若 A 以均勻的概率隨機猜測，當 B 說 j 時，B 的期望收益（expected gain）是多少？（求表中第 j 行的平均值。）

各行的平均值分別是：$j=1$ 時為 $\frac{1}{3}(0+1-2)=-\frac{1}{3}$，$j=2$ 時為 0，$j=3$ 時為 $\frac{1}{3}(2-1+0)=\frac{1}{3}$。所以 B 每局贏得金額的期望值是

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

當 $b_3=1$ 時取得最大值：每次都說 3，平均每局贏 $\frac{1}{3}$ 鎊。均勻的隨機化並不能保護 A。

階段 3　一般情況

> **想一想**　現在設 A 以概率 a_j 說 j。把 B 贏得金額的期望值按 A 的猜測歸項：a_1 的係數（即第 1 列以 b_j 為權的加權平均）是甚麼？

沿着表中各列由上而下讀，

$$\text{B 贏得金額的期望值}=(-b_2+2b_3)a_1+(b_1-b_3)a_2+(-2b_1+b_2)a_3,$$

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

> **想一想**　你能令三者同時為零嗎？（共有三條方程，但它們並非獨立；為甚麼？）

它們並非獨立：$X+2Y+Z=0$ 恆成立（請自行驗證）。所以由 $Y=0$ 得 $b_1=b_3$，由 $X=0$ 得 $b_2=2b_3$，兩者合起來便迫使 $Z=0$ 亦成立。再配合 $b_1+b_2+b_3=1$：

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

於是 $X=Y=Z=0$。這樣，不論 a_j 取甚麼值，B 贏得金額的期望值都是 0，B 長遠而言不會有任何損失。 ■

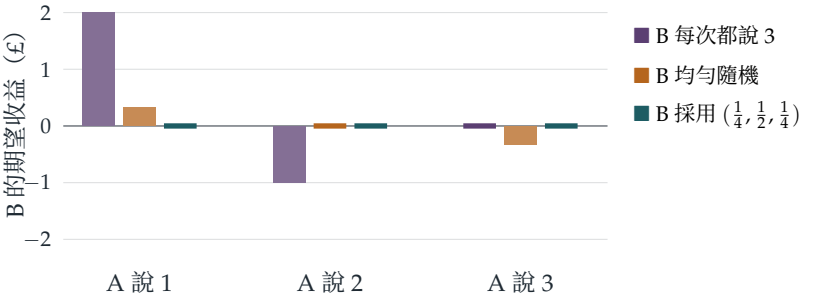


圖 12.14.1 B 採用三種不同策略時，面對 A 每一種選擇的期望收益。每次都說 3，遇上 A 說 1 會贏，遇上 A 說 2 卻會輸；均勻隨機猜測，遇上 A 說 3 會輸。策略 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ 則不論 A 怎樣做，收益都恰好是 0（圖中以幼條表示），所以永遠不會被對手利用。

回顧

- **寫出得失表。**所有推理都圍繞一個 3×3 矩陣的行和列。
- **純策略（pure strategy）會被對手利用。**面對任何固定的猜測，甚至面對均勻隨機的猜測，對手都有一個有利可圖的應對方法。
- **拉平。**穩妥的策略令對手選甚麼都沒有分別：A 的每一種選擇都帶來相同的期望結果。

背後的數學

極小極大定理

在一個二人零和（zero-sum）博弈中，設得失矩陣（payoff matrix）為 M，B 選取概率 $\mathbf{b}$，A 選取概率 $\mathbf{a}$，則 B 贏得金額的期望值是 $\mathbf{b}^{\mathsf{T}} M \mathbf{a}$。馮紐曼的極小極大定理指出

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

用文字說：B 預先公開自己的隨機化策略而仍能保證取得的金額，等於 A 預先公開自己的策略而能把 B 壓制到的金額。這個共同的數值稱為博弈的值（value）。在本題中，由反對稱性可知博弈的值是 0，而策略 $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ 對雙方都是最優的。這個定理可用線性代數（linear algebra）和凸性（convexity）證明（它等價於線性規劃（linear programming）的對偶定理），而它推廣到多人博弈的版本，即納殊均衡（Nash equilibrium），為納殊（John Nash）贏得諾貝爾經濟學獎。

為甚麼要隨機化？

這裏的隨機性無關運氣，而是關乎資訊。如果 B 的選擇可以被預測，A 便能加以利用；B 以恰當的概率隨機選擇，既令自己的選擇無從預測，又確保任何預測都幫不了對手。現實中的例子包括足球的十二碼罰球——對職業球員的研究發現，主射球員和守門員在左右兩邊之間隨機選擇，比例與博弈論的最優值十分接近——以及網球的發球。

馮紐曼 1928 年的論文

馮紐曼在 1928 年證明了極小極大定理，論文題為「On the theory of games of strategy」（論策略博弈的理論），研究動機來自撲克，因為撲克中的「偷雞」（bluffing，即虛張聲勢）正是一種隨機化。1944 年，他與經濟學家 Oskar Morgenstern 出版了 *Theory of Games and Economic Behavior*（博弈論與經濟行為），博弈論自此成為一門學科。馮紐曼後來憶述，他一直視極小極大定理為關鍵結果，沒有它，就沒有值得發表的博弈理論。

思維技巧

以隨機化消除對手的優勢

當任何固定的選擇都會被對手利用時，就隨機選擇，而所用的概率要令對手無論選哪一項都沒有分別。

在大學裏，博弈論、線性規劃和演算法理論都會用到這個想法。例如，隨機演算法（randomised algorithm）會作出任何對手都無法預測的隨機選擇，藉此防範最壞情況的輸入。

延伸練習

1. **A 的最佳應對。**若 B 採用 $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$，證明 A 的每一種策略所得的期望值都是 0。若 B 改用 $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$，求 A 的最佳應對。

2. **包剪揼。**寫出「包剪揼」的得失矩陣，並求最優策略。為甚麼這裏的最優策略是均勻的，而本題的卻不是？

3. **唯一性。**證明 $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$ 是 B 能保證不輸的唯一策略。（提示：它必須令 X、Y、Z 全部非負；利用 $X + 2Y + Z = 0$。）

4. **改動後的博弈。**若相差 2 時的罰款由 £2 改為 £3，求最優策略。

延伸閱讀

- **Ken Binmore, *Playing for Real***（Oxford University Press）。一本詳盡而平易近人的博弈論入門書。
- **Avinash Dixit & Barry Nalebuff, *The Art of Strategy***（Norton）。以體育和商業的例子講解博弈論。
- **William Poundstone, *Prisoner's Dilemma***（Anchor）。馮紐曼、博弈論與冷戰。

12.15 滅絕還是爆發

題目

STEP II 2003，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/03-s2-q12`

引子

1873 年，高爾頓（Francis Galton）在《*Educational Times*》上提出一個問題：為甚麼貴族的姓氏那麼容易失傳？每個男人所生的兒子數目是隨機的，而每個兒子都會把姓氏傳下去。即使每個男人平均有多於一個兒子，這個姓氏就能永遠流傳下去嗎？這個問題開創了分支過程（branching process）的理論，如今這套理論用來模擬核連鎖反應、傳染病的擴散、細胞羣體的增長，以及突變能否存活下去。

本題建立的是最簡單的模型：每個粒子分裂成 0、1 或 2 個粒子。粒子總數的期望值（expectation）按幾何級數增長或縮減。但期望值並不是故事的全部。即使期望的粒子總數急劇膨脹，整個族系完全滅絕的概率仍然是正數，而一個漂亮的不動點（fixed point）論證就能把這個概率求出來。

動筆之前

? **想一想** 如果某一階段有兩個粒子，下一階段粒子數目的分佈是甚麼？

兩個粒子各自獨立地產生後代，後代數目服從 $\mathrm{B}(2, p)$，所以總數服從 $\mathrm{B}(4, p)$：共有四次獨立的「機會」，每次成功的概率為 p。

逐步解題

階段 1　先是兩個，然後又是兩個

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

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

階段 2　逆向推理

> **想一想**　$X_2 = 2$ 可以有多少種方式發生？（X_1 可以是甚麼？）

要使 $X_2 = 2$，可以是 $X_1 = 1$（概率為 $2pq$），其後產生 2 個後代（p^2）；也可以是 $X_1 = 2$，其後由 $\mathrm{B}(4, p)$ 得出 2 個。（$X_1 = 0$ 則甚麼也不剩。）所以 $\mathrm{P}(X_2 = 2) = 2p^3q + 6p^4q^2$，而

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

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

階段 3　粒子總數的期望值

> **想一想**　如果在第 $n-1$ 階段有 k 個粒子，第 n 階段粒子數目的期望值是多少？然後對 k 取平均。

每個粒子平均有 $\mathrm{E}(X) = 2p$ 個後代，所以 k 個粒子平均有 $2pk$ 個後代：$\mathrm{E}(X_n \mid X_{n-1} = k) = 2pk$。對 k 取平均，得 $\mathrm{E}(X_n) = 2p\,\mathrm{E}(X_{n-1})$。由於 $\mathrm{E}(X_0) = 1$，

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

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

回顧

- **獨立二項變量之和。**k 個粒子各自給出 $\mathrm{B}(2, p)$，合起來就是 $\mathrm{B}(2k, p)$。
- **時間上的貝葉斯定理（Bayes' theorem）。**$\mathrm{P}(X_1 = 2 \mid X_2 = 2)$ 是由較後的觀察結果逆向推斷較早的狀態。
- **條件期望值（conditional expectation）。**以上一代的情況為條件，期望值就化成一個等比數列（geometric sequence）。

背後的數學

滅絕的概率

當 $p = \frac{3}{4}$ 時，粒子總數的期望值無止境地增長。這是否表示羣體一定能存活下去？不一定：如果第一個粒子沒有後代（概率為 $q^2 = \frac{1}{16}$），這個族系便立即終結。設 e 為最終滅絕的概率。第一個粒子有 k 個後代，而每個後代的族系各自獨立地以概率 e 滅絕，所以

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

其中 $\mathrm{G}(s) = (q + ps)^2$ 是後代數目的概率生成函數（probability generating function）。滅絕概率是 G 的一個不動點：事實上是最小的非負不動點。當 $p = \frac{3}{4}$ 時，由 $(1+3e)^2 = 16e$ 得 $9e^2 - 10e + 1 = 0$，所以 $e = \frac{1}{9}$ 或 1，而滅絕概率是 $\frac{1}{9}$。當 $p = \frac{1}{4}$ 時，在 $[0,1]$ 內唯一的不動點是 $e = 1$：滅絕是必然的（圖 12.15.1）。

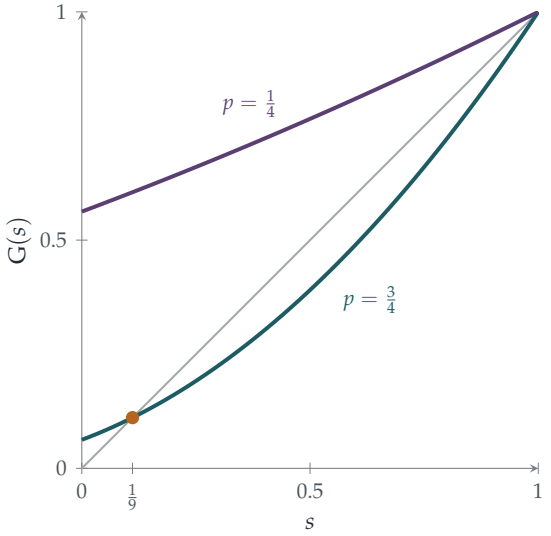


圖 12.15.1　後代數目的生成函數 $\mathrm{G}(s) = (q + ps)^2$ 與直線 $y = s$。滅絕概率就是兩者相交的最小一點。當 $p = \frac{1}{4}$（紫紅色，平均值 $\frac{1}{2}$）時，該點是 $s = 1$：滅絕是必然的。當 $p = \frac{3}{4}$（藍綠色，平均值 $\frac{3}{2}$）時，該點是 $s = \frac{1}{9}$（琥珀色）：羣體以概率 $\frac{8}{9}$ 永遠存活下去。

臨界門檻

一般而言，如果後代數目的平均值小於 1，分支過程必然滅絕；平均值等於 1 時也是如此，唯一的例外是每個個體都恰好有一個後代的平凡情況（這時族系只是一個接一個地永遠延續下去）；如果平均值大於 1，則分支過程以正概率存活下去。後代數目的平均值就是流行病學中的繁殖數（reproduction number）R：$R > 1$ 的疫症有可能大規模爆發，但即使如此，早期的爆發仍可能純粹因為偶然而無疾而終，其概率正好由上面的不動點方程給出。在核物理學中，如果每次裂變平均恰好引發另一次裂變，連鎖反應就處於臨

界（critical）狀態，而操作反應堆的學問，就在於讓它一直保持在這個狀態。

高爾頓、Watson 與 Bienaymé

高爾頓的問題在 1874 年由 Henry William Watson 牧師作答，但答案有部分是錯的：他找到了不動點方程，卻漏掉了第二個根，因而斷定每個姓氏最終都會失傳。其實正確的分析早在 1845 年已由法國統計學家 Irénée-Jules Bienaymé 提出，只是他的研究一直被人遺忘，直到 1970 年代才重見天日。這個模型現在稱為*高爾頓—Watson 過程*（Galton–Watson process，有時稱為 Bienaymé–Galton–Watson 過程），是每一門隨機過程（random processes）課程中的核心例子。

思維技巧

以第一代為條件

一個會自我更新的過程（每個後代都開始一個獨立的、與整個過程相同的複本），必定滿足一條以第一步爲條件而得出的方程。

方程 $e = \mathrm{G}(e)$ 是一個不動點問題，與 Section 2.2 和 Section 8.7 中遇到的屬同一類。在大學裏，生成函數和不動點是研究分支過程、隨機樹和隨機網絡的主要工具。

延伸練習

1 **一般情況下的滅絕。**設後代數目的分佈為 $\mathrm{B}(2, p)$，證明當 $p \le \frac{1}{2}$ 時滅絕概率為 1，而當 $p > \frac{1}{2}$ 時滅絕概率為 $\left(\frac{q}{p}\right)^2$。

2 **方差（variance）。**證明 $\mathrm{Var}(X_n)$ 滿足 $\mathrm{Var}(X_n) = (2p)^2\mathrm{Var}(X_{n-1}) + 2pq\,\mathrm{E}(X_{n-1})$，並求 $p = \frac{1}{2}$ 時的方差。

3 **在第 n 代或之前滅絕。**設 $e_n = \mathrm{P}(X_n = 0)$。證明 $e_n = \mathrm{G}(e_{n-1})$，並就 $p = \frac{3}{4}$ 計算 e_1、e_2、e_3。它們是否趨近 $\frac{1}{9}$？

4 **姓氏。**如果每個男人有 0、1、2 或 3 個兒子的概率各為 $\frac{1}{4}$，他的姓氏最終失傳的概率是多少？

延伸閱讀

- **Geoffrey Grimmett & David Stirzaker, *Probability and Random Processes***（Oxford University Press），論分支過程的一節。
- **Adam Kucharski, *The Rules of Contagion***（Profile）。講述繁殖數與疫症爆發，適合一般讀者。
- **David Kendall, "Branching processes since 1873"**，*Journal of the London Mathematical Society* 41 (1966)。講述這段歷史，包括 Bienaymé 的貢獻。

12.16 順向與逆向的馬可夫鏈

題目

STEP III 1996，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/96-s3-q12`

引子

Ecks 教授的定理有一個規律：她下一個證明甚麼，只取決於她上一個證明了甚麼。一個過程的未來若只取決於現在而與過去無關，就稱為**馬可夫鏈**（*Markov chain*）。馬可夫鏈無處不在：天氣模型、蛇梯棋這種圖版遊戲、DNA 模型、Google 最初的 PageRank 演算法，以至統計學家用來模擬複雜概率分佈的方法，全都用得上它。

本題要求出各類定理在長遠而言所佔的比例，即**平穩分佈**（*stationary distribution*）：長遠來說，她的定理每六個之中有一個是新而正確的，而有一半是錯的。然後題目把這條鏈在時間上倒過來運行，並問：已知今天的定理，昨天的定理是哪一類？

動筆之前

想一想 「每一行（row）之和為 1」這個條件是甚麼意思？各列（column）之和是否也是 1？

第 i 行列出的是：已知現時的類型為 i，下一個定理屬於各類型的概率；三者必有其一發生，所以每一行之和為 1。各列則不一定：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。

逐步解題

階段 1 **一步**

根據全概率定律（law of total probability），以現時定理的類型 i 為條件，

$$b_j = \sum_i \mathrm{P}(\text{現時}i)\,\mathrm{P}(\text{下一個}j \mid \text{現時}i) = a_1p_{1j} + a_2p_{2j} + a_3p_{3j}. \qquad \blacksquare$$

用矩陣的語言來說，行向量（row vector）$\mathbf{b} = \mathbf{a}P$。

階段 2　平穩分佈

想一想　我們要求 $\mathbf{a} = \mathbf{a}P$，且 $a_1 + a_2 + a_3 = 1$。把各方程寫出來。哪些最簡單？

各方程為

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

（它們並非獨立：把三式相加，兩邊都得 $a_1 + a_2 + a_3$。）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 $8a_1 = \frac{4}{3}$，$a_1 = \frac{1}{6}$，$a_3 = \frac{1}{2}$：

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

階段 3　把鏈倒過來運行

想一想　$q_{ij} = \mathrm{P}(\text{上一個}i \mid \text{現時}j)$。運用貝葉斯定理（Bayes' theorem），上一個定理和現時的定理都採用平穩分佈。

根據貝葉斯定理，

$$q_{ij} = \frac{\mathrm{P}(\text{上一個}i)\,\mathrm{P}(\text{現時}j \mid \text{上一個}i)}{\mathrm{P}(\text{現時}j)} = \frac{a_ip_{ij}}{a_j}.$$

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

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

現在是每一列之和為 1，因為第 j 列列出的是上一個定理各種可能的類型。

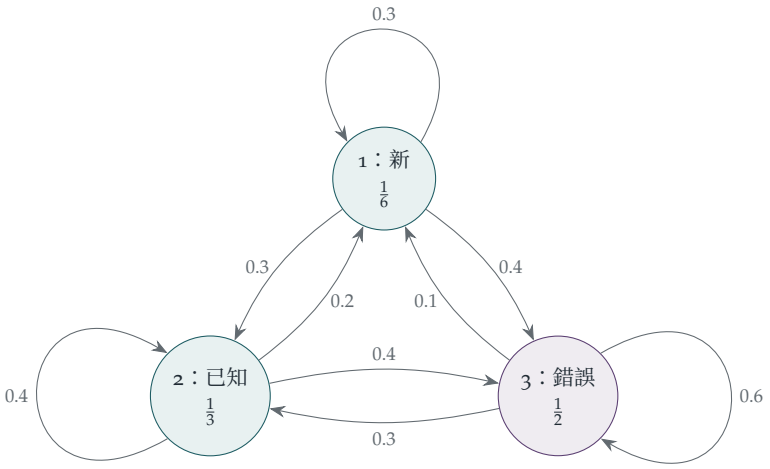


圖 12.16.1　把 Ecks 教授的定理看成一條馬可夫鏈。箭頭顯示下一個定理屬於各類型的概率。圓圈內的分數是平穩分佈：長遠來說，每六個定理有一個是新而正確的，每三個有一個是已知的，而有一半是錯的。

回顧

- **以矩陣表達全概率定律。**鏈每走一步，就是把一個行向量乘以矩陣 P。
- **平穩即特徵向量（eigenvector）。**$\mathbf{a} = \mathbf{a}P$ 說明 $\mathbf{a}$ 是 P 的左特徵向量，對應的特徵值（eigenvalue）為 1。
- **用貝葉斯定理倒推。**逆向概率構成另一個隨機矩陣（stochastic matrix），由正向的隨機矩陣和平穩分佈構造而成。

背後的數學

收斂至平衡

讓 Ecks 教授從任何一個定理開始；再證明 n 個之後，類型的分佈是 $\mathbf{a}_0 P^n$。對於像這樣的一條鏈，即每種類型都可以緊接在任何一種類型之後，不論起點為何，P^n 都會收斂至一個每一行均等於平穩分佈 $(\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$ 的矩陣。收斂的速度由 P 的其他特徵值（這裏是 0.2 和 0.1）決定，它們以幾何方式縮小：只需幾個定理之後，她的過去就被遺忘了。這個關於馬可夫鏈的*遍歷定理*（*ergodic theorem*）是大學二年級概率課程的核心結果，馬可夫鏈蒙地卡羅（Markov chain Monte Carlo）方法之所以行得通，靠的正是它：要從一個複雜的分佈中抽樣，就設計一條以該分佈為平穩分佈的馬可夫鏈，然後讓它運行一段長時

間。

可逆性

逆向鏈 (q_{ij}) 與正向鏈 (p_{ij}) 並不相同：從她的定理的一份長紀錄中，你可以分辨出時間的方向。若一條鏈的逆向版本與正向版本一致，即對所有 i、j 都有 $a_i p_{ij} = a_j p_{ji}$，就稱為*可逆*（*reversible*）；網絡上的隨機遊走（random walk）以及處於熱平衡的物理系統的模型都屬此類。可逆性是一個威力強大的性質，因為它令平穩分佈很容易求出。

馬可夫與普希金

馬可夫（Andrey Markov）在 1906 年引入他的鏈，部分原因是要證明大數定律（law of large numbers）並不需要獨立性；當時他正與數學家 Pavel Nekrasov 爭論，後者聲稱獨立性是必需的。1913 年，馬可夫把他的想法應用於普希金（Pushkin）的詩體小說《葉甫蓋尼·奧涅金》（*Eugene Onegin*）的首 20 000 個字母，把每個字母分為元音或輔音，並證明每個字母的類型取決於前一個字母的類型，而這種關係可用一條兩個狀態的鏈很好地描述。這是史上第一次對文本作統計分析，也是所有預測輸入系統的始祖。

思維技巧

概率作向量，轉移作矩陣

當一個系統以固定的概率在各狀態之間轉移時，把分佈寫成向量，把轉移寫成矩陣。這樣，長遠行爲就變成一個關於特徵向量的問題。

在大學裏，這樣做會把第 7 章的線性代數帶進概率論。佩龍－弗羅賓尼斯定理（Perron–Frobenius theorem）保證平穩分佈存在，而其他特徵值則量度到達平穩分佈有多快。

延伸練習

1 **兔子窩。**試做 STEP III 2003，第 13 題（`step.maths.org/questions/03-s3-q13`）：在一個正方形的四個頂點上的隨機遊走，每個頂點各有一個出口，以首步方程求解。

2 **兩步。**若 Ecks 教授上一個定理是新而正確的，求再下一個定理屬於三種類型的概率。

3 **特徵值。**證明 P 的特徵值為 1、0.2 和 0.1。

4 **可逆嗎？**檢查 Ecks 教授的鏈是否滿足 $a_i p_{ij} = a_j p_{ji}$。答案告訴你甚麼？

延伸閱讀

- **James Norris, *Markov Chains***（Cambridge University Press）。標準的大學本科教科書。
- **Olle Häggström, *Finite Markov Chains and Algorithmic Applications***（Cambridge University Press）。篇幅短小，講馬可夫鏈蒙地卡羅尤其出色。
- **Brian Hayes,「First links in the Markov chain」**，*American Scientist* 101 (2013)。重述馬可夫對普希金作品的分析。

12.17 巴拿赫的火柴盒

題目

STEP II 1992，第 16 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/92-s2-q16`

引子

相傳數學家巴拿赫（Stefan Banach）煙癮很大，兩邊口袋各放一盒火柴。每次要用火柴，他就隨機選一個口袋。有一天，他發現其中一盒空了。另一盒還剩多少根火柴？這就是巴拿赫火柴盒問題（Banach's matchbox problem），由斯坦豪斯（Hugo Steinhaus）為向巴拿赫致意而設計；本題則是的士司機的版本，把火柴換成了糖果。

微妙之處在於「發現」一詞。司機取走一包糖的最後一顆時，並不會察覺那包已經空了，要到下一次伸手去拿那包糖時才知道。所以這個事件是：她伸手去拿太妃糖，而全部 n 顆太妃糖早已吃完。計算這件事在第 N 次伸手時發生的方法數目，是二項分佈（binomial distribution）的一道巧妙練習；而巴拿赫問題的答案有一個出人意表的特點：另一盒最可能剩下的火柴數目是 0（與 1 並列）。

動筆之前

想一想 若司機在第 N 次伸手時發現太妃糖已經吃完，第 N 次伸手必須是甚麼？之前的 $N-1$ 次伸手又必須包含甚麼？

第 N 次伸手必須是去拿太妃糖，而之前的 $N-1$ 次伸手必須剛好取走了 n 顆太妃糖（即全部）和 $N-1-n$ 顆薄荷糖。此外，她在此之前不可以已經發現薄荷糖空了。

逐步解題

階段 1　發現太妃糖吃完了

> **想一想**　頭 $N-1$ 次伸手可以按甚麼次序出現？$N-1-n$ 要符合甚麼條件，才能確保她從未發現那包薄荷糖是空的？

那 n 次拿太妃糖的伸手，可以是頭 $N-1$ 次伸手中的任何 n 次。只要她最多取了 m 顆薄荷糖，即 $N-1-n \le m$，她就從未在薄荷糖吃完之後再伸手去拿薄荷糖。因此，對於 $n+1 \le N \le n+m+1$，

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

其中 T_N 是事件「她在第 N 次伸手時發現太妃糖吃完了」（因子 p^{n+1} 已包括第 N 次伸手本身）。

階段 2　發現薄荷糖吃完了

把兩者的角色對調，用同樣的論證可得，對於 $m+1 \le N \le n+m+1$，

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

階段 3　條件概率

> **想一想**　已知在第 N 次伸手時有所發現。那次拿的是太妃糖的概率是多少？

事件 T_N 與 M_N 互斥（exclusive），所以

$$\mathrm{P}(\text{太妃糖} \mid \text{有所發現的伸手次數為} N) = \frac{\binom{N-1}{n} p^{n+1} q^{N-1-n}}{\binom{N-1}{n} p^{n+1} q^{N-1-n} + \binom{N-1}{m} q^{m+1} p^{N-1-m}},$$

其中二項式係數（binomial coefficient）在上述範圍以外一律視為 0。（若 $N \le m$，則只有太妃糖可能已經吃完，概率為 1；同樣，若 $N \le n$ 但 $N > m$，概率為 0。）∎

> **停一停，試一試**　當 $n = m$ 且 $p = q = \frac{1}{2}$ 時，驗算一下這條公式。根據對稱性，對每一個可能的 N，答案都應該是 $\frac{1}{2}$。是否如此？

回顧

- **準確理解事件。**「發現」是指伸手去拿一包已空的糖，而不是取走最後一顆糖。

- **固定最後一步，再數其餘。**第 N 次伸手已經固定；頭 $N-1$ 次可以按任何次序出現，於是得出一個二項式係數。
- **檢查邊界。**限制條件 $N-1-n \le m$ 和 $N-1-m \le n$ 排除了不可能出現的過程。

背後的數學

巴拿赫的火柴盒

在巴拿赫的版本中，每盒起初有 n 根火柴，選中每個口袋的機會均等。當他首次發現其中一盒空了，設 K 為另一盒的火柴數目。若 $K = k$，他必定已從空的那盒取了 n 根，從另一盒取了 $n-k$ 根，次序不拘，然後再一次伸手去拿空的那盒；而兩盒之中任何一盒都可以是空的那盒：

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

這些概率隨 k 增大而遞減，唯一例外是 $\mathrm{P}(K=0) = \mathrm{P}(K=1)$（圖 12.17.1）。剩餘火柴數目的期望值（expected value）約為 $2\sqrt{n/\pi} - 1$：若每盒有 50 根火柴，約為 7。這個 $\sqrt{n}$ 是經過 $2n$ 次隨機選擇之後，兩盒之間差距的典型大小，與生日問題（Section 9.9）和隨機漫步（random walk）中的平方根定律如出一轍。

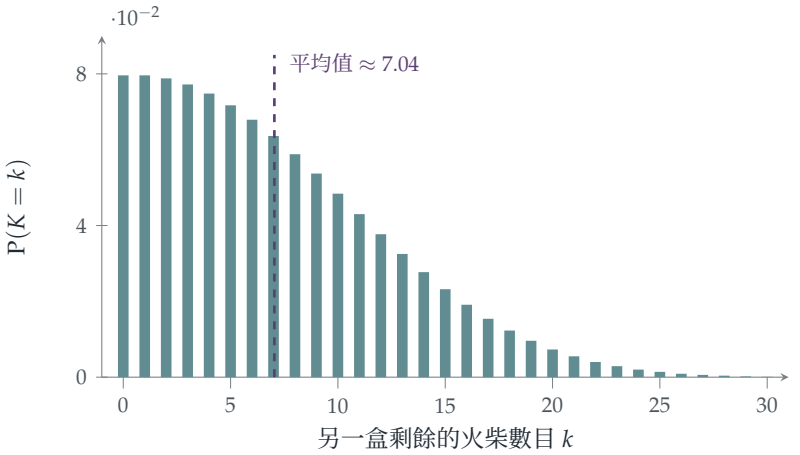


圖 12.17.1 每盒有 50 根火柴的巴拿赫火柴盒問題：首次發現其中一盒空了的時候，另一盒剩餘火柴數目的分佈。最可能的數值是 0 和 1，兩者概率相同，但平均值約為 7。

負二項分佈

在一連串獨立試驗（independent trials）中，每次試驗成功的概率為 p，則取得第 r 次成功所需的試驗次數服從負二項分佈（negative binomial distribution）：$\mathrm{P}(N=k) =$

$\binom{k-1}{r-1}p^r q^{k-r}$。的士司機的問題，是兩個負二項過程（太妃糖和薄荷糖）之間的競賽，每個過程所需的「成功」次數都比該包糖的數目多一次。負二項分佈見於保險（索償宗數）、生態學（動物聚集分佈的數目）和遺傳學，往往用來代替泊松分佈（Poisson distribution），因為它更有彈性。

巴拿赫與蘇格蘭咖啡館

巴拿赫是泛函分析（functional analysis）的奠基人之一，他的數學研究有不少是在 Lwów（今稱 Lviv）的蘇格蘭咖啡館（Scottish Café）裏完成的。他和同事把問題寫在一本由侍應保管的筆記簿上，這就是《蘇格蘭書》（Scottish Book），解題獎品從一杯啤酒到一隻活鵝不等。斯坦豪斯是巴拿赫的恩師，他為了一篇向巴拿赫的吸煙習慣致意的演說，設計了這道火柴盒問題。最先在書刊中分析這個問題的，是費勒（William Feller）1950 年出版的著名概率論教科書。

思維技巧

固定最終事件，數出通往它的路徑

要求一個過程在某一時刻以某種方式結束的概率，可先固定最後一步發生的事，再數出通往這一步的所有過程。

在大學裏，這是負二項分佈、隨機漫步的首達時間（first-passage time），以及 Section 9.10 中反射原理（reflection principle）的基礎。

延伸練習

1. **小規模情況。**設 $n = m = 1$ 及 $p = \frac{1}{2}$，列出所有可能的序列，並求 N 的分佈。

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

3. **並列最高。**證明在巴拿赫問題中，對每一個 n 都有 $\mathrm{P}(K = 0) = \mathrm{P}(K = 1)$。

4. **哪一包先吃完？**就的士司機的問題，當 $n = m = 2$ 及 $p = \frac{2}{3}$ 時，求太妃糖是首先被發現空了的那一包（不論何時發現）的概率。

延伸閱讀

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

第 VI 章，正是這本書把巴拿赫問題介紹給全世界。

- **R. Daniel Mauldin (ed.), *The Scottish Book***（Birkhäuser）。收錄蘇格蘭咖啡館的問題，並附評註。
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），內容涉及負二項分佈。

12.18 無記憶性

題目

STEP III 2021，第 11 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2021 年 STEP 試卷（zip 檔案）內的 STEP III 試卷：

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

引子

一顆放射性原子在實驗室裏待了一百萬年，一直沒有衰變。它是否「快要」衰變了？不是：它在下一秒內衰變的機會，與一顆剛剛產生的原子完全相同。放射性原子不會老化。直至衰變為止的等候時間服從指數分佈（*exponential distribution*），而這個分佈的決定性特徵，就是它沒有記憶。

這道題以一種令人印象深刻的形式展示了這種無記憶性（memorylessness）。把一個指數等候時間拆成整數個單位和小數部分。原來這兩部分是獨立（*independent*）的：知道你等了多少個完整的小時，對於分鐘數毫無提示。這道題還會在練習中解釋，為甚麼你等巴士的時間，可以顯得比兩班巴士之間的間隔還要長。

動筆之前

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

$\mathrm{P}(X > x) = \int_x^\infty \lambda e^{-\lambda u}\,\mathrm{d}u = e^{-\lambda x}$。於是 $\mathrm{P}(X > s+t \mid X > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathrm{P}(X > t)$：已經等了 s 之後，餘下的等候時間與從頭開始等候的分佈相同。這就是無記憶性質（*memoryless property*）。

逐步解題

階段 1　整數部分

$Y = n$ 當且僅當 $n \le X < n+1$：

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

所以 Y 服從幾何分佈（geometric distribution）：每一個完整單位都以概率 $e^{-\lambda}$ 被「捱過」。

階段 2 小數部分

想一想 $Z < z$ 表示 X 落在其中一個區間 $[n, n+z)$ 內。把這些區間的概率加起來。

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

階段 3 Z 的平均值

想一想 對於取值於 $[0,1]$ 的隨機變量（random variable），$\mathrm{E}(Z) = \int_0^1 \mathrm{P}(Z > z)\,\mathrm{d}z$。運用這條公式。

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

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

（當 λ 很小，即等候時間很長時，這個值趨向 $\frac{1}{2}$：小數部分幾乎是均勻分佈的。當 λ 很大時，它接近 $\frac{1}{\lambda}$，因為 X 幾乎總是小於 1。）

階段 4 獨立性

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

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

由 (i) 及 (ii)，$\mathrm{P}(Y = n)\mathrm{P}(z_1 < Z < z_2) = (1 - e^{-\lambda})e^{-n\lambda} \cdot \frac{e^{-\lambda z_1} - e^{-\lambda z_2}}{1 - e^{-\lambda}}$，兩者相同。所以 Y 與 Z 獨立。 $\blacksquare$

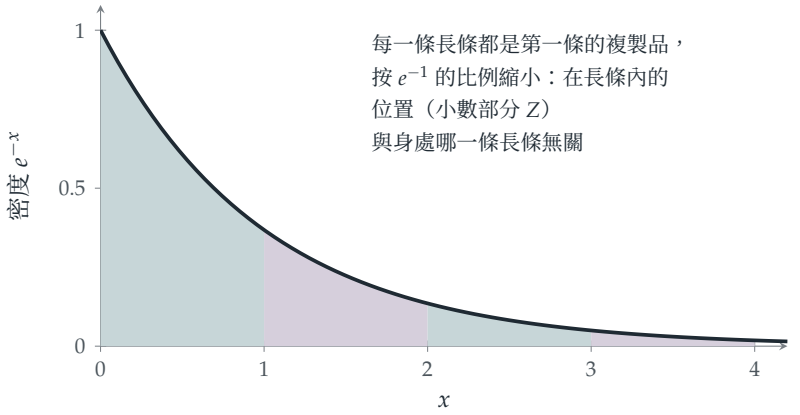


圖 12.18.1 $\lambda = 1$ 的指數密度，分割成單位寬度的長條。由於 $e^{-(n+z)} = e^{-n}e^{-z}$，每一條長條的形狀都相同。$X$ 落在哪一條長條（整數部分 Y），與它落在長條內的甚麼位置（小數部分 Z），兩者是獨立的。

回顧

- **逐段求和。**Z 的分佈和聯合概率（joint probability），都是把各區間 $[n, n+1)$ 上的概率相加而得的。
- **尾部公式。**$\mathrm{E}(Z) = \int \mathrm{P}(Z > z)\,\mathrm{d}z$ 讓我們不必求出密度。
- **能夠分解即是獨立。**聯合概率分拆成一個 n 的函數乘以一個 z_1、z_2 的函數。

背後的數學

指數分佈為何特別

Y 與 Z 的獨立性直接來自 $e^{-\lambda(n+z)} = e^{-\lambda n}e^{-\lambda z}$，而這正是無記憶性質的另一副面貌。事實上，指數分佈是唯一沒有記憶的連續分佈：若對所有 $s, t \geq 0$ 都有 $\mathrm{P}(X > s+t) = \mathrm{P}(X > s)\mathrm{P}(X > t)$，則 $g(t) = \ln \mathrm{P}(X > t)$ 滿足 $g(s+t) = g(s) + g(t)$，而（對於單調（monotonic）的 g）這就迫使 $g(t) = -\lambda t$。正因如此，指數分佈可用來描述那些「隨機」發生、沒有任何累積過程的事件的等候時間：放射性衰變、顧客的到達、不會損耗的電子元件的故障。

最小值與等候時間悖論

若 X_1 與 X_2 是獨立的指數隨機變量，率分別為 λ_1 和 λ_2，則 $\mathrm{P}(\min > x) = e^{-\lambda_1 x}e^{-\lambda_2 x}$，所以最小值服從率為 $\lambda_1 + \lambda_2$ 的指數分佈。等候兩個獨立隨機事件中最先發生的一個，

就如同以合併後的率等候單一事件（STEP III 2002 Q13，見練習）。無記憶性也解釋了等候時間悖論（*waiting-time paradox*）：若巴士隨機到站，平均間隔為 10 分鐘，則一位在隨機時刻到達的乘客平均要等 10 分鐘，而不是 5 分鐘，因為餘下的等候時間是無記憶的；而乘客到達時所處的那個間隔，平均長達 20 分鐘，因為較長的間隔更容易被碰上。

盧瑟福（Rutherford）與泊松過程（Poisson process）

1910 年，歐內斯特·盧瑟福與漢斯·蓋革（Hans Geiger）以 7.5 秒為一個時段，數算一個放射源所放出的阿爾法粒子（alpha particles），發現這些計數服從泊松分佈（Poisson distribution）；數學家 Harry Bateman 證明了，只要各次放射以恆定的率獨立發生，結果就理應如此。這樣一來，相鄰兩次放射之間的時間便服從指數分佈。泊松過程，即間隔服從指數分佈的一連串事件，如今已成為描述時間上隨機性的基本模型，應用於電話通話、保險索償以及神經元的放電；下一節 Section 12.19 會把它推廣到空間。

思維技巧

以因式分解檢驗獨立性

若兩個隨機變量的聯合概率可分解爲其中一個變量的函數乘以另一個變量的函數，則這兩個隨機變量獨立。

在大學裏，聯合密度（joint density）能否分解是檢驗獨立性的標準方法，你亦會藉此在多變量問題中辨認出獨立的分量，例如 Section 12.22 中服從正態分佈（normal distribution）的點的半徑與角度。

延伸練習

1. **兩條巴士路線。**試做 STEP III 2002，第 13 題（`step.maths.org/questions/02-s3-q13`）。這道題利用兩個指數隨機變量的最小值，求出等候下一班巴士的時間。

2. **離散的無記憶性。**證明幾何分佈是無記憶的：$\mathrm{P}(Y \geq m+n \mid Y \geq m) = \mathrm{P}(Y \geq n)$。

3. **等候時間悖論。**巴士的到站是一個率為 λ 的泊松過程。證明包含某固定時刻 t（t 很大）的那個間隔，其期望長度為 $\frac{2}{\lambda}$，即平均間隔的兩倍。

4. **Z 的密度。**求 Z 的概率密度（probability density），並就 $\lambda = 1$ 和 $\lambda = 5$ 繪出其草圖。

延伸閱讀

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），論及指數分佈與泊松過程。
- **Sheldon Ross, *Introduction to Probability Models***（Academic Press）。指數分佈、泊松過程與排隊。
- **David Spiegelhalter, *The Art of Uncertainty***（Pelican）。日常生活中的隨機性與等候。

12.19 平面上的隨機點

題目

STEP II 2012，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/12-s2-q13`

引子

假如超級市場、樹木、恆星或流動電話發射站「隨機」散佈在某個區域內，你要走多遠才到達最近的一個？答案只取決於密度 k（每單位面積的平均數目），而且有一個獨特的形狀，就是瑞利分佈（*Rayleigh distribution*）。平均而言，最近的一個位於距離 $\frac{1}{2\sqrt{k}}$ 處：若每平方公里有一間超級市場，這距離就是 500 米。

這個模型稱為空間泊松過程（*spatial Poisson process*），即時間上隨機到達（Section 12.18）的二維版本。計算過程簡短而優美，但它真正的價值在於充當基準：生態學家把實際的最近鄰距離與這條公式比較，以判斷樹木或鳥巢究竟是聚集成羣、隨機散佈，還是刻意彼此隔開。

動筆之前

想一想 一個平均值為 μ 的泊松隨機變量（Poisson random variable），取值 0 的概率是多少？

$\mathrm{P}(0) = e^{-\mu}$。半徑為 y 的圓面積是 πy^2，所以圓內超級市場數目的平均值為 $k\pi y^2$。

逐步解題

階段 1　附近沒有超級市場

在距離 y 以內沒有超級市場的概率是 $e^{-\pi k y^2}$。

階段 2 最近距離的分佈

? **想一想** $Y \geq y$ 表示在距離 y 以內沒有超級市場。那麼 $\mathrm{P}(Y < y)$ 是甚麼？

$\mathrm{P}(Y < y) = 1 - e^{-\pi k y^2}$，求導後便得到密度函數（density）

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

階段 3 平均值

? **想一想** 可以用分部積分法（integration by parts）計算 $y\,\mathrm{f}(y)$ 的積分，也可以利用 $\mathrm{E}(Y) = \int_0^\infty \mathrm{P}(Y > y)\,\mathrm{d}y$。然後作代換，以便使用題目給出的高斯積分（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 $\mathrm{E}(Y) = \frac{1}{\sqrt{2\pi k}}\sqrt{\frac{\pi}{2}} = \frac{1}{2\sqrt{k}}$。 $\blacksquare$

階段 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 $\mathrm{d}u = 2\pi k y\,\mathrm{d}y$），上式變成 $\int_0^\infty \frac{u}{\pi k} e^{-u}\,\mathrm{d}u = \frac{1}{\pi k}$。因此

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

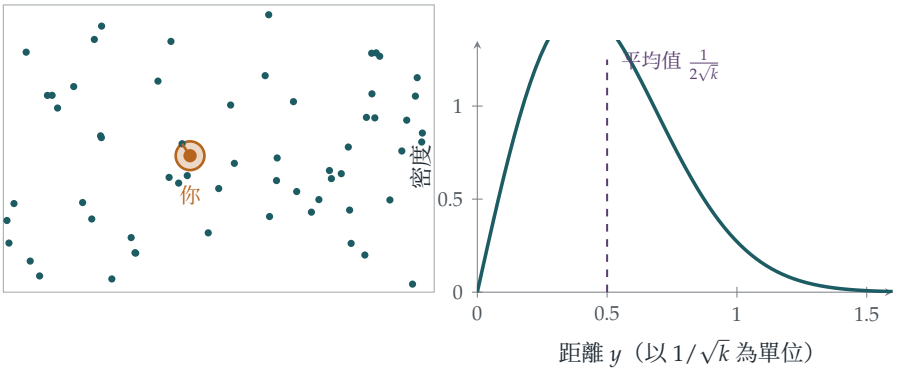


圖 12.19.1　左：六十個點隨機散佈在面積為 24 的區域內（所以 $k = 2.5$），以及由選定的一點到其中最近一點的距離（琥珀色）。右：該距離的瑞利密度函數 $2\pi k y e^{-\pi k y^2}$，距離以 $1/\sqrt{k}$ 為單位。其平均值為 $\frac{1}{2\sqrt{k}}$。

回顧

- **把距離問題化為計數問題。**「最近的一個比 y 更遠」就是「在 y 以內一個也沒有」。
- **先求累積分佈，再求密度。**$\mathrm{P}(Y < y)$ 可以直接寫出；求導後便得到密度函數。
- **以代換化為已知積分。**每個矩（moment）都化成了高斯積分或指數積分。

背後的數學

平面上的泊松過程

強度（intensity）為 k 的空間泊松過程，是一種隨機散佈的點集，滿足：互不相交的區域內的點數彼此獨立，而面積為 A 的區域內的點數服從平均值為 kA 的泊松分佈。把極大量的點獨立而均勻地投放到一片極大的區域上，並保持適當的平均密度，得到的就是這種過程：這正是「稀有事件定律」（law of rare events），它同樣解釋了錯排（derangement）的現象（Section 9.8）。它是完全空間隨機性（complete spatial randomness）的自然模型，實際的分佈形態都以它為比較標準。*Clark–Evans* 檢驗（*Clark–Evans test*）由兩位生態學家於 1954 年提出，把觀察所得的平均最近鄰距離與 $\frac{1}{2\sqrt{k}}$ 比較：小得多表示有聚集現象（種子落在母樹附近），大得多則表示間隔規則（有領域行為的鳥類，或植物之間爭奪水分）。

三維的情況

在三維空間中作同樣的計算，可得到最近的恆星、分子或星系的距離滿足 $\mathrm{P}(Y > y) = e^{-\frac{4}{3}\pi k y^3}$。錢德拉塞卡（Subrahmanyan Chandrasekhar）在 1943 年用它來研究一顆恆星受隨機分佈的鄰近恆星所施的引力，物理學家則用它來計算氣體分子的平均自由程（mean free path）。

瑞利

密度函數 $\frac{y}{\sigma^2} e^{-y^2/(2\sigma^2)}$ 以瑞利勳爵（Lord Rayleigh）命名，他在 1880 年研究聲學理論時遇上它：它描述了許多相位隨機的波疊加後的振幅。若一點的兩個坐標是獨立的正態隨機變量（normal random variable），它也給出該點與原點距離的分佈，這是理解它為何在這裏出現的另一個角度（練習 3）。今天，它用來模擬流動電話訊號的衰落，因為訊號是沿着許多相位隨機的路徑到達接收器的。

思維技巧

先求分佈函數

要求最小值、最大值或最近距離的分佈時，$\mathrm{P}(Y > y)$ 往往很容易直接寫出。求導便得到密度函數。

這個方法會在 Section 12.25 再次出現，用來處理許多隨機壽命中最長的一個。在大學裏，這是處理任何「順序統計量」（order statistic）的標準第一步。

延伸練習

1. **中位數。**求到最近超級市場的距離的中位數（median），並與平均值比較。

2. **第二近的。**求到第二近的超級市場的距離的分佈。（提示：$\mathrm{P}(Y_2 > y)$ 是在 y 以內最多只有一間超級市場的概率。）

3. **正態坐標。**若 U 和 V 是獨立的 $\mathrm{N}(0, \sigma^2)$ 隨機變量，證明 $\sqrt{U^2 + V^2}$ 的密度函數為 $\frac{r}{\sigma^2} e^{-r^2/(2\sigma^2)}$。$\sigma$ 取何值時會得到超級市場的分佈？

4. **Clark–Evans。**在一項調查中，一片面積 1 km^2 的樹林內有 400 棵樹，而一棵樹到其最近鄰的平均距離為 35 m。這與完全空間隨機性相符嗎？還是顯示有聚集或間隔的現象？

延伸閱讀

- **Peter Diggle, *Statistical Analysis of Spatial Point Patterns***（Arnold）。這方面的標準參考書，載有 Clark–Evans 檢驗及其後續發展的方法。
- **John Kingman, *Poisson Processes***（Oxford University Press）。一本優美的短篇著作，討論任意維數的泊松過程。
- **Geoffrey Grimmett & David Stirzaker, *Probability and Random Processes***（Oxford University Press），論泊松過程的一章。

12.20 蒲豐投針

題目

STEP I 1993，第 16 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/93-s1-q16`

引子

把一根針拋到鋪有平行地板條的地板上。針橫跨板縫的機會竟然與 π 有關，儘管眼前根本沒有任何圓形。拋很多次並數一數橫跨板縫的次數，就可以用實驗來估算 π。這就是*布豐投針問題*（*Buffon's needle*），由布豐伯爵喬治-路易·勒克萊爾（Georges-Louis Leclerc, Comte de Buffon）於 1733 年提出，並於 1777 年由他本人解決：它是*幾何概率*（*geometric probability*）的第一個問題，也是所有蒙地卡羅方法（Monte Carlo method）的始祖。

這個 π 來自角度：針的方向在半圈之內均勻分佈（uniform），把針橫跨平行線方向的「影子」長度對所有方向取平均，便得出因子 $\frac{2}{\pi}$。本題則以 $\cos^{-1}$ 的積分來完成這個計算。

動筆之前

想一想 針的中心與最近一條直線的距離為 x，而針與這些直線的垂直線所成的角為 y。針朝着該直線能伸展多遠？

針的每一半長度為 a，它在垂直於直線方向上的分量是 $a\cos y$。因此，針與最近的直線相交，當且僅當 $a\cos y \geq x$。（由於 $a < b$，針不可能觸及其他直線。）

逐步解題

階段 1 積分

設 $t = \cos y$，當 t 由 0 變到 1 時，$y = \cos^{-1} t$ 由 $\frac{\pi}{2}$ 變到 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：這個積分是 $\cos^{-1}$ 的圖像與兩條坐標軸之間的面積，它等於 $y = \cos x$ 由 0 到 $\frac{\pi}{2}$ 的曲線下面積，兩者只是沿 $y = x$ 反射而已。）

階段 2　給定位置

想一想　對於固定的 x，哪些角度 y 會使針與直線相交？這些角度的概率是多少？

若 $x \geq a$，針無法觸及直線。若 $x < a$，相交的條件是 $\cos y \geq \frac{x}{a}$，即 $y \leq \cos^{-1}\frac{x}{a}$。由於 Y 在 $[0, \frac{\pi}{2}]$ 上均勻分佈，

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

階段 3　對位置取平均

X 在 $[0, b]$ 上均勻分佈，其密度（density）為 $\frac{1}{b}$，所以

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

其中用了代換 $x = at$。

停一停，試一試　若針的長度恰好等於半間距（$a = \frac{b}{2}$，即針的長度 $2a$ 等於 b），相交的概率是 $\frac{1}{\pi}$。要把 π 估算至小數點後兩位，你需要拋多少次？

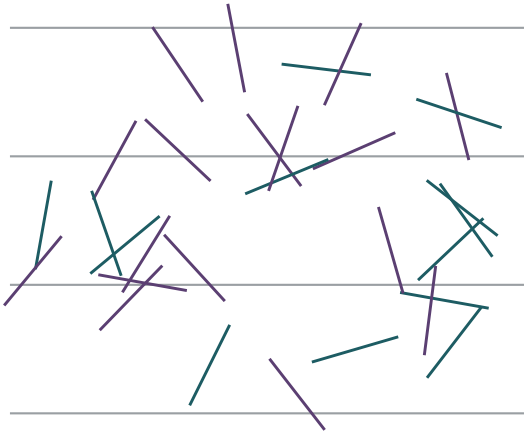


圖 12.20.1　三十根長度為 $2a = 1.4$ 的針，拋在相距 $2b = 2$ 的直線上。與直線相交的針以紫紅色表示。這裏有 17 根相交，而期望值是 13.4 根：小樣本只能給出粗略的估算，而布豐投針正是出名緩慢的 π 計算方法。

回顧

- **兩個獨立的坐標。**針的位置和角度互相獨立（independent）且均勻分佈，所以所求概率是一個二重積分（double integral），可以逐個變數計算。
- **先取條件，再求平均。**先固定 x，對角度取平均；然後再對 x 取平均。
- **代換顯露出常數。**以 $x = at$ 作縮放，答案便化約為單一個通用的積分。

背後的數學

布豐的麵條

以下的方法幾乎不用計算就能看出答案。把針屈曲成任意形狀，成為一根「麵條」，並設 C 為它與直線相交的次數。根據期望值的線性（linearity of expectation），一條曲線相交次數的期望值，等於它各小段的期望值之和，所以與長度成正比：$\mathrm{E}(C) = \kappa L$，其中 κ 是某個常數。要求出 κ，可取麵條為一個直徑等於直線間距 $2b$ 的圓：它總是恰好相交兩次，而其長度為 $2\pi b$。所以 $\kappa \cdot 2\pi b = 2$，得出 $\kappa = \frac{1}{\pi b}$。對於長度為 $2a < 2b$ 的直針，它最多相交一次，故 $\mathrm{P}(\text{相交}) = \mathrm{E}(C) = \frac{2a}{\pi b}$。這個論證出自 Joseph-Émile Barbier（1860 年），是概率推理中的瑰寶：那個 π 來自用來校準常數的圓。

估算 π，以及錯誤示範

若拋出 n 根針，其中 c 根相交，則 $\frac{c}{n} \approx \frac{2a}{\pi b}$，而 $\pi \approx \frac{2an}{bc}$。誤差只會按 $\frac{1}{\sqrt{n}}$ 的速度縮小，所以要把 π 算至小數點後三位，大約需要拋一百萬次。1901 年，意大利數學家 Mario Lazzarini 聲稱拋了 3408 次便得出 $\pi \approx \frac{355}{113} = 3.1415929\ldots$（準確至小數點後六位）。統計學家指出，這個結果好得令人難以置信：拋 3408 次，再配合一根精心挑選的針，恰好 1808 次相交便會得出 $\frac{355}{113}$，而只要在適當的時刻停止實驗，就能製造出這個結果。這是關於選擇性報告的經典警世故事。

布豐

布豐是一位博物學家，曾任巴黎皇家植物園園長，並著有一套 36 卷的自然史。他於 1733 年向法國科學院（Académie des Sciences）提出投針問題（以及一個使用方形地磚的版本，即「franc-carreau」遊戲），並於 1777 年發表解答。他的目的是要說明，概率中不單可以運用算術，也可以運用幾何。積分幾何（integral geometry）——研究隨機直線及其與曲線和曲面的交點的理論——正是由此發展而來，如今應用於體視學（stereology），即從二維切片估算三維結構（例如組織中的細胞）的大小和形狀的科學。

思維技巧

把概率看成區域上的積分

當結果取決於數個獨立的連續隨機變量時，一個事件的概率就是聯合密度（joint density）在該事件發生的區域上的積分。

在大學裏，這成為連續概率的基礎：聯合密度、邊際分佈（marginal distribution）和變數代換（change of variable），這些全都會在 Section 12.21 和 Section 12.22 再次出現。

延伸練習

1. **長針。**若 $a > b$，針可能與多於一條直線相交。證明相交次數的期望值仍然是 $\frac{2a}{\pi b}$。
2. **麵條。**把一根長度為 $2a$ 的針屈成直角（兩臂各長 a）。相交次數的期望值是多少？
3. **方形地磚。**把一枚半徑為 r 的硬幣拋到鋪有邊長為 $s > 2r$ 的方形地磚的地板上。求硬幣完全落在一塊地磚之內的概率。

4 **Lazzarini 的針。**設 $a = \frac{5}{6}b$（使得 $\frac{2a}{\pi b} = \frac{5}{3\pi}$），證明拋 3408 次而有 1808 次相交，會給出估算值 $\pi \approx \frac{355}{113}$。

延伸閱讀

- **Daniel Klain & Gian-Carlo Rota, *Introduction to Geometric Probability***（Cambridge University Press），此書由布豐投針問題講起。
- **Lee Badger, "Lazzarini's lucky approximation of π"**，*Mathematics Magazine* 67 (1994)。一個可疑實驗背後的統計學。
- **Paul Nahin, *Digital Dice***（Princeton University Press）。以蒙地卡羅方法處理概率問題。

12.21 折斷一根棒

題目

STEP III 1990，第 16 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/90-s3-q16`

引子

在一根木棒上隨機選兩點，把它折斷。三段木棒能否組成三角形？著名的答案是：概率（probability）為 $\frac{1}{4}$。如果能組成三角形，它是銳角（acute）三角形還是鈍角（obtuse）三角形？通常是鈍角三角形：所組成的三角形中，只有大約三分之一是銳角三角形。這類關於隨機圖形的問題，自十九世紀起便令概率學家樂此不疲，也是練習處理連續概率（continuous probability）的絕佳場地：區域內的一個隨機點、一個界定子區域的條件，以及一個計算其面積的積分。

本題採用兩種不同的折棒方法，並在每種情況下問出現鈍角的可能性有多大。答案 $3-2\sqrt{2}$ 和 $3-4\ln 2$ 都是出人意表的數，卻源自簡單的幾何。

動筆之前

想一想 三個長度能組成三角形，當且僅當每一個都小於其餘兩個之和。如果三個長度之和為 L，這對最長的一個有何含意？邊 c 所對的角何時是鈍角？

$c < a+b$ 連同 $a+b+c = L$，即 $c < \frac{L}{2}$；所以三段能組成三角形，正正是當每一段都短於總長度的一半。根據餘弦公式（cosine rule），c 所對的角是鈍角，正正是當 $c^2 > a^2+b^2$。

逐步解題

階段 1　先折一次，再對半分

三段的長度是 X、$\frac{1-X}{2}$、$\frac{1-X}{2}$。兩段相等的木棒必定短於 $\frac{1}{2}$，所以能組成三角形當且僅當 $X < \frac{1}{2}$。這個三角形是等腰（isosceles）三角形；兩個相等的角都是銳角，所以只有當 X 所對的角是鈍角時，它才是鈍角三角形：

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

已知 $X < \frac{1}{2}$ 時，X 在 $[0, \frac{1}{2}]$ 上均勻分佈（uniform），所以

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

階段 2　第二種構造方法

> **想一想**　設 (x, y) 位於以 $(1,0)$、$(1,1)$、$(0,1)$ 為頂點的三角形內，對邊長 x、y 和 $2 - x - y$ 檢查三條三角不等式（triangle inequality）。

該區域是 $\{x \le 1, y \le 1, x + y \ge 1\}$。三條不等式分別是：$x + y \ge 2 - x - y$，即 $x + y \ge 1$；$x + (2 - x - y) \ge y$，即 $y \le 1$；以及 $y + (2 - x - y) \ge x$，即 $x \le 1$。它們在區域內全部成立（在區域的邊上等號成立，此時三角形是退化（degenerate）的）。 $\blacksquare$

> **想一想**　B 處的角所對的邊是 $CA = y$。把條件 $y^2 > x^2 + (2 - x - y)^2$ 寫成關於 y 的不等式。

展開得 $y^2 > x^2 + 4 + x^2 + y^2 - 4x - 4y + 2xy$，所以 $4y - 2xy > 2x^2 - 4x + 4$，即 $2y(2 - x) > 2(x^2 - 2x + 2)$。由於 $2 - x > 0$，

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

階段 3　面積

> **想一想**　以 $2 - x$ 除 $x^2 - 2x + 2$。然後求區域中位於曲線上方那部分的面積，再除以區域的面積 $\frac{1}{2}$。

由於 $x^2 - 2x + 2 = -x(2 - x) + 2$，曲線是 $y = -x + \frac{2}{2-x}$。它由 $(0,1)$ 伸延至 $(1,1)$，而當 $0 < x < 1$ 時，它位於直線 $y = 1 - x$ 之上、$y = 1$ 之下（圖 12.21.1）。鈍角區域位於

曲線與 $y = 1$ 之間，面積為

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

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

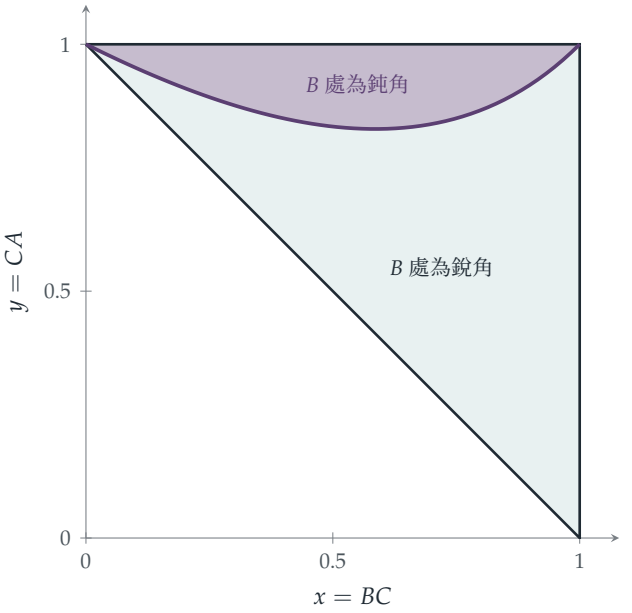


圖 12.21.1 第 (ii) 部分中 (x, y) 所有可能取值構成的三角形。在曲線 $y = \frac{x^2-2x+2}{2-x}$ 上方（紫色區域），B 處的角是鈍角；這個區域佔三角形的比例為 $3 - 4\ln 2 \approx 0.23$。

回顧

- **把幾何轉化為不等式。**「能組成三角形」和「是鈍角三角形」都化成了各長度之間的不等式。
- **連續變量的條件概率（conditional probability）。**已知 $X < \frac{1}{2}$ 時，X 在較小的區間上均勻分佈。
- **概率即面積之比。**對於區域內均勻分佈的點，概率等於面積除以總面積。

背後的數學

經典的折棒問題

在兩個獨立而均勻分佈的點 U 和 V 處把木棒折斷。數對 (U, V) 在單位正方形上均勻分佈，而三段能組成三角形，正正是當每一段都短於 $\frac{1}{2}$：這個條件切去四個小三角形，總面積為 $\frac{3}{4}$，餘下的概率是 $\frac{1}{4}$。已知能組成三角形時，它是銳角三角形的概率是 $\frac{1}{4}$ 當中的 $3\ln 2 - 2 \approx 0.079$，即大約 32%。這三段木棒是均勻隨機分割（uniform random partition）的一個例子，它們的聯合分佈（joint distribution）是狄利克雷分佈（Dirichlet distribution），這種分佈出現在貝葉斯統計學（Bayesian statistics）中，也出現在描述資源如何隨機分配的模型中。

怎樣折斷木棒，大有關係

請留意，題目中的兩種方法，對聽起來相似的問題給出了不同的答案 $3 - 2\sqrt{2}$ 和 $3 - 4\ln 2$，而兩者又都與經典方法的答案不同。正如貝特朗（Bertrand）的弦（Section 12.3）一樣，在指明具體程序之前，「隨機」一詞毫無意義。

隨機三角形

關於隨機三角形的問題，是維多利亞時代流行的趣題。路易斯·卡羅（Lewis Carroll）的 *Pillow Problems*（1893 年）問及「隨機」選取的三角形是鈍角三角形的概率，而他的答案，正如本節的答案一樣，完全取決於他所選擇的隨機性模型。關於隨機三角形的問題一直帶來驚喜：例如，在正方形內均勻隨機選取三點，它們組成鈍角三角形的概率約為 0.725，這個數值到二十世紀才有人準確算出。

思維技巧

樣本空間（sample space）的幾何

把隨機結果表示爲區域內的一點，把事件表示爲子區域，並以面積之比（或密度（*density*）的積分）計算概率。

在大學裏，這幅圖像是多元概率（multivariate probability）的基礎，也是變數代換公式（change-of-variables formula）的基礎；這條公式說明坐標改變時，密度如何隨之變換。

延伸練習

1. **經典問題。**證明：若在兩個獨立而均勻分佈的點處把木棒折斷，則各段能組成三角形的概率是 $\frac{1}{4}$。

2. **折斷較長的一段。**在一個均勻分佈的點處把木棒折斷，然後在較長的一段上取一個均勻分佈的點，再把它折斷。證明三段能組成三角形的概率是 $2\ln 2 - 1 \approx 0.386$。

3. **銳角還是鈍角。**在第 (ii) 部分中，求三角形 ABC 在 C 處（對邊為 $AB = 2 - x - y$）有鈍角的概率。由此求它有任何一個角是鈍角的概率。

4. **最長一段的期望長度。**對於經典的折棒問題，證明最長一段的期望（expected）長度是 $\frac{11}{18}$。

延伸閱讀

- **Martin Gardner, *The Colossal Book of Mathematics***（Norton），書中收錄了折棒問題。
- **Frederick Mosteller, *Fifty Challenging Problems in Probability***（Dover）。
- **Lewis Carroll, *Pillow Problems***（Dover，與 *A Tangled Tale* 合訂）。維多利亞時代的趣題，包括隨機三角形。

12.22 沒有平均值的分佈

題目

STEP III 1995，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/95-s3-q12`

引子

把許多次量度的結果取平均，隨機誤差便會互相抵消：平均值會穩定下來，趨向真實的數值。這條*大數定律*（*law of large numbers*）是整個統計學的基石。但它暗藏一個假設：這些量度必須有有限的平均值(mean)。本題構造了一個再自然不過的隨機變量(random variable)，即兩個獨立（independent）標準正態（standard normal）變量之比，而它根本沒有平均值。它的平均永遠不會穩定下來；事實上，一百萬個這樣的變量的平均，其分佈與單獨一個的分佈完全相同。

這就是*柯西分佈*（*Cauchy distribution*）。它的密度（density）看來像一個人畜無害的鐘形，但它的尾部實在太厚，偶然出現的極大數值足以左右任何平均。本題用一個漂亮的幾何論證把它找出來：兩個坐標之比就是一條直線的斜率，而徑向對稱令這條直線的角度呈均勻分佈。

動筆之前

想一想 若 (U, V) 是一個隨機點，其密度只取決於它與原點的距離，那麼由原點到該點的直線與坐標軸所成的角 Θ，你能說些甚麼？

把平面旋轉並不會改變密度，所以每個方向的可能性都相同：角 Θ 在 $[0, 2\pi)$ 上呈均勻（uniform）分佈。而 $\frac{U}{V}$ 只取決於方向，與距離無關。

逐步解題

階段 1　聯合密度（joint density）

由於 X 與 Y 獨立，它們的聯合密度就是各自密度的乘積：

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

它只透過 x^2+y^2（即與原點距離的平方）取決於 x 和 y：在以原點為圓心的圓上，密度是常數。這就是「徑向對稱」（radially symmetrical）的意思。

階段 2　運用對稱性

> **想一想**　$\frac{U}{V} < k$ 可以在兩種情況下發生，視乎 V 的正負號而定。這兩種情況與旋轉 180° 的變換 $(u,v) \mapsto (-u,-v)$ 有甚麼關係？

$\frac{U}{V} < k$ 的意思是：要麼 $V > 0$ 且 $U < kV$，要麼 $V < 0$ 且 $U > kV$。旋轉 $(u,v) \mapsto (-u,-v)$ 把第一個區域映射到第二個區域上，並且保持密度不變。所以兩者的概率相同，於是

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

階段 3　一個角

> **想一想**　在以 u 為橫軸、v 為縱軸的平面上，畫出區域 $\{V > 0, U < kV\}$。它是一個頂點在原點的楔形。它的角度是多少？

區域 $V > 0$ 是上半平面。通過原點的直線 $u = kv$ 與正 v 軸所成的角為 $\arctan k$（當 $k > 0$ 時偏向正 u 的一方）。上半平面中位於這條直線左方的部分，是一個角度為 $\frac{\pi}{2} + \arctan k$ 的楔形。根據徑向對稱，楔形的概率等於它的角度除以 2π，所以

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

對 k 求導，便得到密度

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

回顧

- **獨立的正態變量具有徑向對稱性。**兩個正態密度的乘積只取決於距離。
- **比值即方向。**$\frac{U}{V}$ 只取決於該點的角度，而這個角度呈均勻分佈。

► **楔形的概率。**在徑向對稱之下，楔形的概率與它的角度成正比。

背後的數學

沒有平均值

柯西分佈的平均值本應是 $\int_{-\infty}^{\infty} \frac{k}{\pi(1+k^2)}\,\mathrm{d}k$，但 $\int_0^K \frac{k}{1+k^2}\,\mathrm{d}k = \frac{1}{2}\ln(1+K^2) \to \infty$：這個積分並不收斂（converge），所以平均值不存在。（對稱性或會令人以為答案是 0，但正負兩部分都是無窮大，而 $\infty - \infty$ 並沒有值。）方差（variance）同樣是無窮大。

更糟的是，n 個獨立柯西變量的平均仍然是柯西變量，而且分佈完全相同。（其中一個看出這點的方法，是利用特徵函數（characteristic function）$\mathrm{E}(e^{itX}) = e^{-|t|}$。）所以取平均完全無補於事：一百萬次量度的平均，並不比單獨一次可靠。圖 12.22.1 顯示了實際情況：正態樣本的累計平均會穩定下來，而柯西樣本的累計平均卻永遠東歪西倒，不時被偶然出現的極大數值撞得偏離原位。

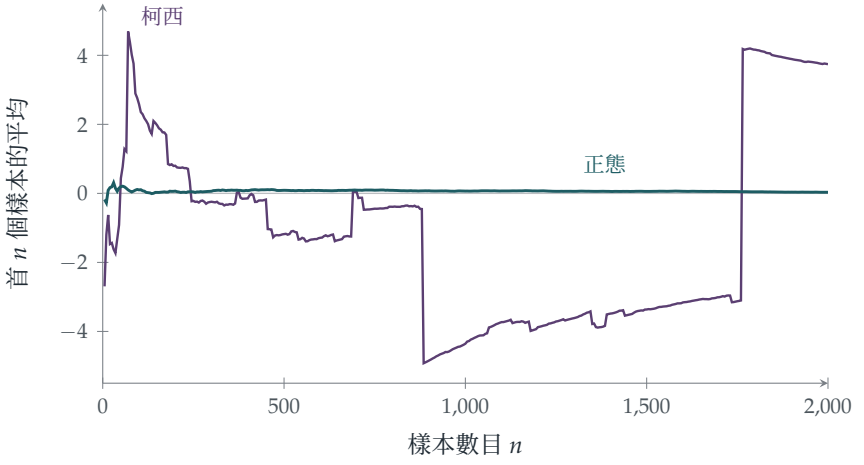


圖 12.22.1 2000 個獨立樣本的累計平均。對於標準正態樣本（藍綠色），平均值一如大數定律所預測，穩定下來並趨向 0。對於柯西樣本（紫紅色），它永遠不會穩定：每隔一段時間，單單一個極大的數值就會把它拋到另一個位置。

柯西變量從何而來

拿一支電筒，朝一個均勻隨機的方向照向一面牆：光點在牆上的位置便服從柯西分佈。在物理學中，同一個密度稱為勞侖茲分佈（*Lorentzian*），用來描述譜線和共振峯的形狀（比較 Section 11.1）。在金融學中，像柯西分佈這類厚尾（heavy-tailed）分佈，會用來為偶爾出現的股災建立模型，而正態模型往往嚴重低估這類事件。這對統計學是一個嚴肅的教訓：建基於平均值和方差的方法，遇上厚尾數據便會失效，這時就需要穩健的方

法，例如使用中位數（median）（柯西分佈的中位數完全沒有問題）。

泊松、柯西與一場論戰

這個分佈在 1824 年由泊松（Siméon Denis Poisson）研究過，他注意到它違反大數定律。柯西（Augustin-Louis Cauchy）則在 1853 年與 Irénée-Jules Bienaymé 就最小二乘法（method of least squares）公開論戰時遇上它：柯西用它來說明，當誤差屬於厚尾時，最小二乘法可以是一個很差的方法。這個分佈最終冠上了柯西的名字，正是 Stephen Stigler 所稱的「史蒂格勒命名定律」（Stigler's law of eponymy）的一例：沒有任何科學發現是以其最初發現者命名的。

思維技巧

檢查期望值是否存在

使用平均值或方差之前，先確定定義它的積分是收斂的。對稱性並不能令一個發散的積分變成零。

在大學裏，可積性條件是概率論中每一條定理的一部分：大數定律需要有限的平均值，中心極限定理（central limit theorem）（Section 12.24）需要有限的方差。當這些條件不成立時，柯西分佈就是標準的反例。

延伸練習

1 **電筒。**一支電筒與一面長而直的牆相距 1，所指的方向與垂直線成角 Θ，其中 Θ 在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 上呈均勻分佈。證明光點的位置 $\tan\Theta$ 服從柯西分佈。

2 **倒數。**證明若 C 服從柯西分佈，則 $\frac{1}{C}$ 亦然。（提示：$\frac{U}{V}$ 與 $\frac{V}{U}$。）

3 **中位數行得通。**證明柯西分佈的中位數是 0，而它的四分位數（quartiles）是 ± 1。

4 **半徑與角度。**設 X、Y 為獨立的標準正態變量，透過轉換成極坐標（polar coordinates），證明 $R^2 = X^2 + Y^2$ 服從平均值為 2 的指數分佈（exponential distribution），並且與角度獨立。（這是用來產生正態隨機數的 Box–Muller 方法的基礎。）

延伸閱讀

► **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction*** (Oxford University

Press），論隨機變量的變換。

- **Nassim Nicholas Taleb, *The Black Swan***（Penguin）。講述厚尾現象及其重要性，適合一般讀者。
- **Stephen Stigler, *The History of Statistics***（Harvard University Press）。書中記述了柯西與 Bienaymé 的論戰。

12.23 由概率得出斯特靈公式

題目

STEP II 2016，第 13 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/16-s2-q13`

引子

100! 有多大？要逐項乘出來得花點工夫：它有 158 位數字。斯特靈公式（*Stirling's formula*）$n! \approx \sqrt{2\pi n}\,(n/e)^n$ 只用一行，就能給出誤差在 0.1% 以內的答案。它是數學中最有用的近似式之一，在概率（probability）、統計物理和演算法分析中經常用到。

它有很多種證明（proof），可以用積分（integral）（Section 4.4），也可以用歐拉–麥克勞林公式（Euler–Maclaurin formula）。本題給出的是最出人意表的一種：用概率來證明。平均值（mean）為 n 的泊松分佈（Poisson distribution）近似於正態分佈（normal distribution），在峯頂比較兩者，$n!$ 的值就被確定下來。斯特靈公式中的 $\sqrt{2\pi}$，正正就是正態密度（normal density）中的 $\sqrt{2\pi}$。

動筆之前

? **想一想** 設離散隨機變量（discrete random variable）X 的平均值為 μ、方差（variance）為 σ^2，並且近似於正態分佈，那麼 $\mathrm{P}(X = \mu)$ 大約是多少？（試想一個寬度為 1、以 μ 為中心的區間的概率。）

按正態近似（normal approximation），$\mathrm{P}(X = \mu) \approx \mathrm{P}(\mu - \frac{1}{2} < X < \mu + \frac{1}{2})$，這大約等於寬度 1 乘以峯頂處的密度：$\frac{1}{\sigma\sqrt{2\pi}}$。

逐步解題

階段 1　十六枚硬幣

$X \sim \mathrm{B}(16, \tfrac{1}{2})$ 的平均值為 8，方差為 $16 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 4$，所以 $\sigma = 2$。由正態近似，

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

準確值是 $\binom{16}{8}/2^{16} = \tfrac{12870}{65536} \approx 0.1964$，兩者相差不到 2%。

階段 2　中央二項式係數

> **想一想**　$\mathrm{B}(2n, \tfrac{1}{2})$ 的平均值和標準差（standard deviation）是多少？分別寫出 $\mathrm{P}(X = n)$ 的準確值和近似值。

平均值是 n，方差是 $\tfrac{n}{2}$，所以 $\sigma = \sqrt{n/2}$。準確值是 $\mathrm{P}(X = n) = \binom{2n}{n}\tfrac{1}{2^{2n}}$；近似值是 $\tfrac{1}{\sqrt{n/2}\sqrt{2\pi}} = \tfrac{1}{\sqrt{\pi n}}$。令兩者相等，

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

階段 3　斯特靈公式

> **想一想**　參數為 n 的泊松分佈，其平均值和方差都等於 n。分別寫出 $\mathrm{P}(X = n)$ 的準確值和近似值。

準確值是 $\mathrm{P}(X = n) = e^{-n}\tfrac{n^n}{n!}$。當 n 很大時，泊松分佈近似於正態分佈（它是 n 個獨立的 Poisson(1) 變量之和），其中 $\sigma = \sqrt{n}$，所以 $\mathrm{P}(X = n) \approx \tfrac{1}{\sqrt{2\pi n}}$。令兩者相等，

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

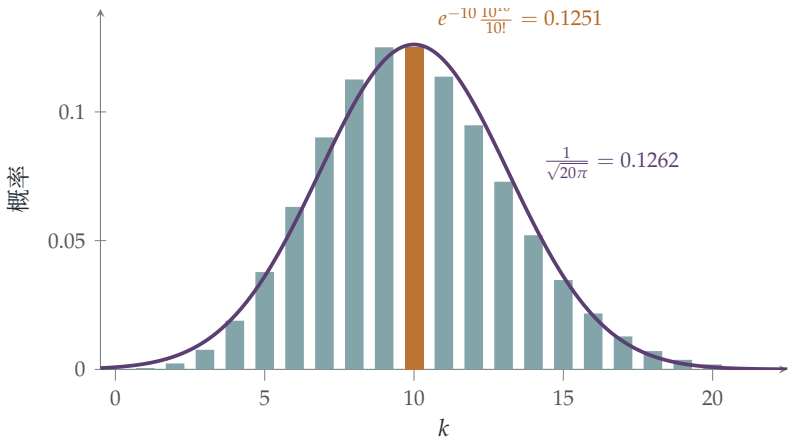


圖 12.23.1 平均值為 10 的泊松分佈（長條），以及平均值和方差與之相同的正態密度（紫色曲線）。在峯頂，泊松概率 $e^{-10}10^{10}/10!$（琥珀色）與正態密度 $\frac{1}{\sqrt{2\pi\cdot 10}}$ 十分接近。把這個近似等式移項整理，就是斯特靈公式。

回顧

- **一點上的概率 ≈ 密度乘以寬度。**對於格點分佈（lattice distribution），某一點上的正態密度近似於該點的概率。
- **準確公式遇上近似式。**二項分佈（binomial distribution）和泊松分佈的概率是準確的；令它們等於各自的正態近似，便得到階乘（factorial）的估計。
- **按公式的形狀選擇分佈。**$\binom{2n}{n}$ 令人想到 $\mathrm{B}(2n, \frac{1}{2})$；$\frac{n^n}{n!}$ 令人想到 $\mathrm{Poisson}(n)$。

背後的數學

這個近似有多準確？

這個論證只是啟發式（heuristic）的：它在單獨一點上使用正態近似，而一般的中心極限定理（central limit theorem）並不足以支持這樣做。它有一個嚴格的版本，稱為*局部中心極限定理*（*local central limit theorem*），所做的正是這件事，並使這個論證成為真正的證明。斯特靈公式其實出奇地準確：比值 $\frac{n!}{\sqrt{2\pi n}(n/e)^n}$ 大約是 $1 + \frac{1}{12n}$，所以即使 $n = 1$，它給出的也是 0.92 而不是 1；而當 $n = 10$ 時，它只偏小了 0.8%。完整的級數（series）$1 + \frac{1}{12n} + \frac{1}{288n^2} - \cdots$ 是*漸近級數*（*asymptotic series*）的一個例子：對於大的 n，它給出的近似值越來越好；但對於固定的 n，它最終會發散（diverge）。

階乘與正態分佈為何相連

斯特靈公式和正態密度中的 $\sqrt{2\pi}$ 是同一個常數，因為 $\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}$。棣莫弗（Abraham de Moivre）在 1733 年正是循這條路找到二項分佈的正態近似的：他需要在 n 很大時近似 $\binom{n}{k}$；他找到了 $n! \approx Cn^{n+1/2}e^{-n}$ 這個形式，卻沒有找到那個常數。斯特靈（James Stirling）證明了 $C = \sqrt{2\pi}$。正態分佈是從階乘的漸近性質中誕生的。

棣莫弗與斯特靈

棣莫弗是一位逃亡到倫敦的法國新教徒，靠當私人教師和為賭徒出謀獻策維生。他在 1718 年出版了 *The Doctrine of Chances*（《機會論》），並在 1733 年的一本小冊子中找到二項分佈的正態近似，這是正態曲線首次面世。蘇格蘭數學家斯特靈在他的 *Methodus Differentialis*（1730）中發表了該常數的值，棣莫弗亦大方地把功勞歸於他。階乘的近似至今仍是活躍的課題：組合數學（combinatorics）和密碼學（cryptography）都需要對極大的 n 計算 $n!$。

思維技巧

用概率去證明與概率無關的事情

把一個準確的概率與一個近似值作比較，可以得知公式中那些數的資訊。

這種「概率方法（probabilistic method）」是雙向的。這裏概率給出了階乘的估計；艾狄胥（Erdős）則用它來證明某些圖（graph）的存在（Section 9.2）。到了大學，你會見到概率被用來證明分析、數論和組合數學中的定理。

延伸練習

1. **檢驗公式。**用斯特靈公式估計 10! 和 20!，並與真確值 3 628 800 和 2 432 902 008 176 640 000 比較。

2. **由 (iii) 到 (ii)。**證明第 (iii) 部分的斯特靈公式可推出第 (ii) 部分的近似式。

3. **100! 的位數。**利用 $\log_{10} n! \approx n\log_{10}\frac{n}{e} + \frac{1}{2}\log_{10}(2\pi n)$ 證明 100! 有 158 位數字。

4. **擲硬幣打成平手。**利用第 (ii) 部分，證明擲一枚公平硬幣 1000 次而恰好得到 500 次正面的概率約為 0.025。

延伸閱讀

- **William Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1**（Wiley），第 II 章，當中證明了斯特靈公式。
- **Ronald Graham, Donald Knuth & Oren Patashnik, *Concrete Mathematics***（Addison-Wesley），第 9 章，講述漸近分析。
- **Stephen Stigler, *The History of Statistics***（Harvard University Press），講述棣莫弗與正態曲線的誕生。

12.24 中心極限定理

題目

STEP III 2008，第 12 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

`step.maths.org/questions/08-s3-q12`

引子

把許多微小而獨立的隨機效應加起來，不論每一個效應本身服從甚麼分佈，總和都會服從鐘形的正態曲線。這就是*中心極限定理*（central limit theorem），它也許是概率論中最重要的定理。它解釋了為何身高、量度誤差和考試分數都大致服從正態分佈（normal distribution），也解釋了為何統計學家即使對背後的分佈一無所知，仍能說出一個平均值有多可靠，只要該分佈的方差（variance）有限便可（試與 Section 12.22 比較）。

本題在一個特殊情況下證明這個定理。一隻青蛙隨機跳躍，每次跳躍服從雙邊指數分佈（two-sided exponential distribution），其密度函數（density）有一個尖峯，完全不像鐘形。跳了許多次之後，牠的總位移卻服從正態分佈。所用的工具是*矩母函數*（moment generating function），它把獨立變量之和化為乘積，再取對數，又把乘積化為和。

關於課程大綱的說明

矩母函數已於 2019 年從 STEP 考試大綱中刪除。隨機變量（random variable）X 的矩母函數是 $\mathrm{M}_X(\theta) = \mathrm{E}(e^{\theta X})$。它在 0 處的各階導數給出各階矩（moments）：$\mathrm{M}'(0) = \mathrm{E}(X)$，$\mathrm{M}''(0) = \mathrm{E}(X^2)$。若 X 與 Y 獨立（independent），則 $\mathrm{M}_{X+Y} = \mathrm{M}_X\mathrm{M}_Y$，而且 $\mathrm{M}_{cX}(\theta) = \mathrm{M}_X(c\theta)$。若矩母函數收斂，則（在適當的意義下）分佈也收斂。

動筆之前

想一想　密度函數 $\frac{1}{2}e^{-|x|}$ 是對稱的。它的圖像是甚麼樣子？它的平均值（mean）是多少？

它是一條帳篷形的曲線，在 0 處有一個高度為 $\frac{1}{2}$ 的尖峯，並向兩邊按指數衰減（圖 12.24.1）。由對稱性可知平均值為 0。

逐步解題

階段 1　矩母函數

當 $|\theta| < 1$ 時，把積分在 0 處分開，

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

展開得 $\mathrm{M}_X(\theta) = 1 + \theta^2 + \theta^4 + \cdots$，所以 $\mathrm{E}(X) = 0$，$\mathrm{E}(X^2) = 2$（即 $\frac{\theta^2}{2!}$ 的係數）。方差為 2。

階段 2　青蛙的位移

想一想　$Y = X_1 + \cdots + X_n$ 的矩母函數是甚麼？$Y/\sqrt{2n}$ 的又是甚麼？

由獨立性得 $\mathrm{M}_Y(\theta) = (1-\theta^2)^{-n}$，再作縮放，

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

（除以 Y 的標準差（standard deviation）$\sqrt{2n}$ 作縮放，令方差變為 1。）

階段 3　極限

想一想　取對數，並利用當 u 很小時 $\ln(1-u) \approx -u$。

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

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

這正是標準正態分佈的矩母函數。

階段 4　青蛙的機會

想一想 當 n 很大時，Y 近似服從 $\mathrm{N}(0, 2n)$。何時有 $\mathrm{P}(|Y| \geq 25) = 0.05$？

對於標準正態變量 Z，$\mathrm{P}(|Z| \geq 1.96) = 0.05$。因此我們需要 $\frac{25}{\sqrt{2n}} = 1.96$，由此得 $2n = \left(\frac{25}{1.96}\right)^2 \approx 162.7$，即 $n \approx 81.3$。由於落在遠處的機會隨 n 增大而增加，要有至少 5% 的機會，最少的跳躍次數約為 82。 ■

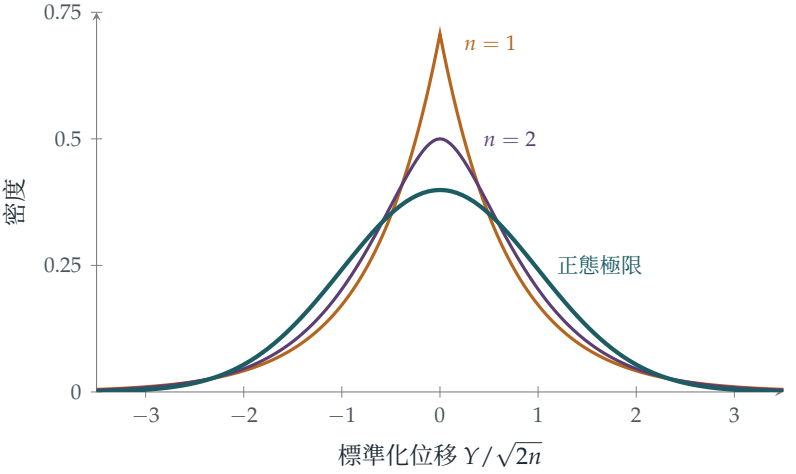


圖 12.24.1 青蛙跳了 $n = 1$ 次（琥珀色：一個拉普拉斯（Laplace）尖峯）和 $n = 2$ 次（紫紅色）之後，其標準化位移的密度函數，與標準正態密度函數（藍綠色）的比較。只跳了兩次，尖峯已經變得圓滑；中心極限定理指出，當 $n \to \infty$ 時，這些曲線收斂至正態曲線。

回顧

- **變換把和化為乘積。**獨立變量之和的矩母函數，等於各變量矩母函數的乘積。
- **對數把乘積化為和。**$\ln\left(1 - \frac{\theta^2}{2n}\right)^{-n}$ 變成了一個和，其極限很容易求得。
- **標準化（standardise）。**除以標準差 $\sqrt{2n}$ 令方差保持為 1，所以極限既不是 0，也不是無窮大。

背後的數學

一般的定理

同樣的論證適用於任何平均值為 0、方差 σ^2 有限的分佈（還須有矩母函數，不過這個條件可以避開）。這時 $\mathrm{M}(\theta) = 1 + \frac{1}{2}\sigma^2\theta^2 + o(\theta^2)$，而且

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

取極限之後只有方差留下來：原來分佈的形狀已被遺忘。這就是正態分佈具有普遍性的原因。在大學裏，這個定理是用特徵函數（characteristic function）$\mathrm{E}(e^{i\theta X})$ 來證明的，特徵函數必定存在，而分佈的收斂則由萊維連續性定理（Lévy's continuity theorem）賦予嚴格的意義。

有多快？

中心極限定理說明極限情況下會發生甚麼，卻沒有說明收斂得有多快。*Berry–Esseen* 定理（Berry–Esseen theorem）給出誤差的上界：一個常數乘以 $\frac{\mathrm{E}|X|^3}{\sigma^3\sqrt{n}}$。對於像青蛙跳躍那樣的對稱分佈，跳了幾十次之後近似已相當準確。對於偏態（skewed）分佈，或者在極端的尾部，即使 n 很大，近似也可能很差：正態近似會低估極罕見事件發生的機會，這個事實對保險和金融都有影響。

從棣美弗到 Lindeberg

棣美弗（De Moivre）在 1733 年發現了二項分佈（binomial distribution）的正態近似（Section 12.23），拉普拉斯在 1810 年把它推廣至更一般的變量之和，並首先提出量度誤差是許多微小效應之和這個想法。嚴格的證明要到很久之後才出現：李亞普諾夫（Aleksandr Lyapunov）在 1901 年給出證明，Jarl Lindeberg 則在 1922 年給出證明，並找到了基本上是令定理成立的最弱條件。「中心極限定理」這個名稱是波利亞（George Pólya）在 1920 年起的，意思是這個定理在概率論中處於中心地位。

思維技巧

變換、化簡、再變換回來

一個困難的運算（把獨立隨機變量相加，即對它們的密度函數作卷積（*convolution*）），經過變換之後變成一個容易的運算（乘法）。在那裏完成工作，然後詮釋所得的結果。

同樣的策略也是生成函數（generating functions）（Section 9.4）、分析學和工程學中的傅立葉（Fourier）變換和拉普拉斯變換，以及把乘法化為加法的對數背後

的原理。

延伸練習

1 **兩次跳躍的精確結果。**證明 $Y = X_1 + X_2$ 的密度函數為 $\frac{1}{4}(1 + |y|)e^{-|y|}$，並驗證其方差為 4。

2 **從二項到正態。**設 $X \sim \mathrm{B}(1, \frac{1}{2})$，並把它平移使平均值為 0（即取值 $\pm\frac{1}{2}$），求出 $\mathrm{M}_X(\theta) = \cosh\frac{\theta}{2}$，並證明 n 個這樣的變量的標準化和的矩母函數趨向 $e^{\theta^2/2}$。

3 **青蛙的精確結果。**利用電腦（或以 $n = 2$ 時的精確密度函數作檢驗），看看當 $n = 82$ 時正態近似有多準確。

4 **為何柯西不行？**柯西分佈（Cauchy distribution）（Section 12.22）沒有矩母函數。論證中哪一步失效？n 個柯西變量的平均值又會怎樣？

延伸閱讀

- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***（Oxford University Press），關於生成函數和極限定理的一章。
- **Hans Fischer, *A History of the Central Limit Theorem***（Springer）。從拉普拉斯到 Lindeberg。
- **David Spiegelhalter, *The Art of Statistics***（Pelican），講述為何正態分佈無處不在。

12.25 眾多之中的最大值

題目

STEP III 2025，第 11 題

繼續閱讀之前，請先看題目（原題為英文）。題目載於 OCR 提供下載的 2025 年 STEP 試卷（zip 檔案）內的 STEP III 試卷：

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

引子

下一個世紀最高的洪水會有多高？一千個燈泡之中，壽命最長的那個能用多久？在 n 名參賽者之中，跑得最快的馬拉松選手會有多快？這類關於極值（extremes）的問題，即許多隨機數值之中最大或最小的一個，在工程、保險和氣候科學中都極為重要，而極值的表現與平均值截然不同。

本題要找出 n 個指數分佈（exponential distribution）壽命中最大者的分佈。它的典型大小按 $\ln n$ 增長，雖然緩慢，卻沒有上限。然而它的離散程度卻完全不會增長：一千個壽命中的最大值，其可預測的程度與十個壽命中的最大值大致相同。在極限情況下，它的分佈形狀就是耿貝爾分佈（Gumbel distribution），也就是極值世界中與正態曲線相對應的角色。

動筆之前

想一想 $X_1, \ldots, X_n$ 中的最大者不超過 y，恰好等同於發生了甚麼事？

$Y \le y$ 恰好在每個 $X_i \le y$ 時成立。利用獨立性（independence），這個概率便化為一個乘積。

逐步解題

階段 1　最大值的分佈

每個 X_i 在 $y \ge 0$ 時都有 $\mathrm{F}(y) = 1 - e^{-\lambda y}$，由獨立性可得

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

階段 2　分位數

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

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

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

階段 3　n 很大的情況

> **想一想**　把 $\alpha^{1/n} = e^{(\ln \alpha)/n}$ 寫出來。當 n 很大時，指數很小。那麼 $1 - \alpha^{1/n}$ 近似於甚麼？

$\alpha^{1/n} = e^{(\ln\alpha)/n} \approx 1 + \frac{\ln \alpha}{n}$，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，$\lambda U(\alpha) \approx \ln n - \ln\!\left(\ln \frac{1}{1-\alpha}\right)$。

階段 4　增長與離散程度

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

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

當中的 $\ln n$ 項已互相抵消：區間的闊度與 n 無關。 $\blacksquare$

階段 5　一個 90% 區間

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

當 $\alpha = 0.05$ 時，$\ln 20 \approx 3$，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 $\ln\ln\frac{1}{1-\alpha} \approx \ln\frac{1}{20} \approx -3$。因此闊度約為 $\frac{1}{\lambda}(\ln 3 + 3) \approx \frac{4.1}{\lambda}$，而 Y 以概率 0.9 落在 $[L(0.05), U(0.05)]$ 之內，這個區間的闊度約為 $4\lambda^{-1}$。 $\blacksquare$

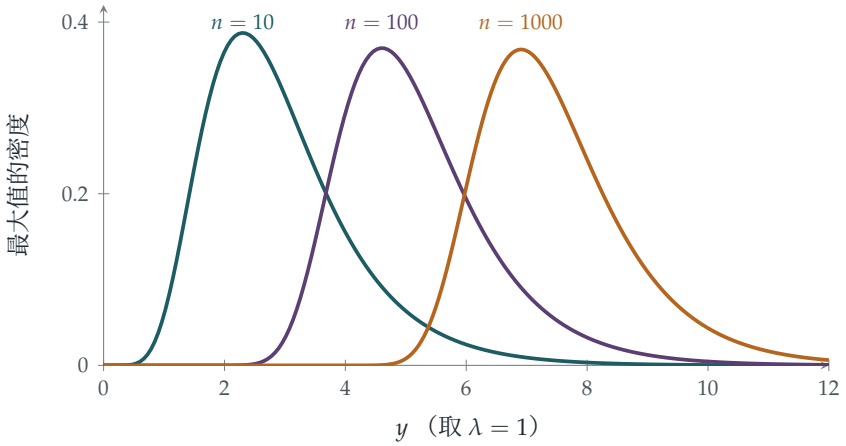


圖 12.25.1 n 個指數分佈壽命（平均值為 1）中最大者的密度，分別對應 $n = 10$、100 和 1000。n 每增至十倍，曲線便向右平移 $\ln 10 \approx 2.3$，但其形狀和闊度基本上保持不變：這就是平移了 $\ln n$ 的耿貝爾分佈。

回顧

- **所有值都不超過 y 時，最大值亦然。**獨立性使最大值的分佈函數成為一個冪。
- **以反函數求分位數。**解 $G(y) = \alpha$ 便得出區間兩端的精確公式。
- **先近似，再比較。**L 和 U 的近似式都含有 $\ln n$，在計算闊度時互相抵消。

背後的數學

耿貝爾分佈

由於 $\mathrm{P}(\lambda Y - \ln n \le x) = \left(1 - \frac{e^{-x}}{n}\right)^n \to \exp(-e^{-x})$，平移後的最大值 $\lambda Y - \ln n$ 在極限下服從耿貝爾分佈，其分佈函數（distribution function）為 $\exp(-e^{-x})$。它的平均值是歐拉常數（Euler's constant）$\gamma \approx 0.577$。你其實早已遇過它：集齊全部 n 款優惠券所需的時間（Section 9.11）就是 n 個（接近獨立的）等待時間中的最大值，這正是它約為 $n \ln n + \gamma n$ 的原因，而其波動屬耿貝爾型，數量級為 n。

極值理論

費雪（Fisher）、Tippett 和格涅堅科（Gnedenko）有一個非凡的定理（1928–1943）：許多獨立同分佈（independent identically distributed）的變量，其最大值經適當平移和縮放後，極限形狀只可能有三種：耿貝爾型（適用於輕尾分佈，例如指數分佈或正態分佈）、弗雷歇（Fréchet）型（適用於重尾分佈，例如柯西（Cauchy）分佈）和韋布爾（Weibull）型

（適用於有硬性上限的變量）。這是中心極限定理（central limit theorem）（Section 12.24）在極值方面的對應結果，所屬的學問稱為極值理論（extreme value theory）。工程師用它來設計防洪設施和海上平台，以抵禦「百年一遇的巨浪」；保險公司用它為巨災保障定價；氣候科學家則用它來估計熱浪正在怎樣變化。

耿貝爾與洪水

Emil Julius Gumbel（耿貝爾）是德國統計學家，直言不諱地反對納粹，1932 年被迫離開大學教席，最終逃往美國。他在 1930 和 1940 年代把極值理論應用於河流氾濫的研究。他的著作 *Statistics of Extremes*（1958）使這門學問成為實用的工具。荷蘭的三角洲工程（Delta Works）是在 1953 年北海洪災之後為保護荷蘭而興建的，設計時就運用了極值統計學，以抵禦預計約 10 000 年一遇的風暴潮。

思維技巧

最大值與最小值的分佈

對於獨立變量的最大值，$\mathrm{P}(\max \le y) = \prod \mathrm{P}(X_i \le y)$；對於最小值，$\mathrm{P}(\min > y) = \prod \mathrm{P}(X_i > y)$。

這一個想法就能處理最近的超級市場（Section 12.19）、第一班巴士（Section 12.18）和壽命最長的燈泡。在大學裏，它是次序統計量（order statistics）和極值理論的起點。

延伸練習

1. **最大值的期望值。**證明 $\mathrm{E}(Y) = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right)$。（提示：根據無記憶性（memorylessness），n 個壽命中的第一個結束之後，餘下的 $n-1$ 個會重新開始。）

2. **最小值。**證明 $X_1, \ldots, X_n$ 中的最小者服從參數為 $n\lambda$ 的指數分佈。

3. **中位數的精確值。**求 Y 的精確中位數，並在 $n = 10$ 時與近似值 $\frac{1}{\lambda}(\ln n - \ln \ln 2)$ 比較。

4. **均勻分佈的最大值。**若 $X_1, \ldots, X_n$ 在 $[0,1]$ 上均勻分佈（uniform），求它們的最大值 M 的分佈，並證明當 n 很大時，$n(1-M)$ 近似服從指數分佈。這屬於三種極值形狀中的哪一種？

延伸閱讀

- **Stuart Coles, *An Introduction to Statistical Modeling of Extreme Values***（Springer）。標準的入門書，附有洪水和海平面的真實數據。
- **Emil Julius Gumbel, *Statistics of Extremes***（Dover）。經典之作。
- **David Spiegelhalter, *The Art of Uncertainty***（Pelican），談罕見事件以及應該怎樣思考它們。

我們學到了甚麼

本章開首我們說過：不確定性是可以精確地推理的，但單憑直覺卻不行。我們提出了三個問題：題目問的究竟是甚麼？巧妙的切入點在哪裏？長遠而言會發生甚麼事？以下是本章二十五道題目所揭示的答案。

1. 題目問的究竟是甚麼？

首四節談的都是如何把問題弄清楚。「已知取出的是一塊有編號的圓片」與「已知取出的是編號為 3 的圓片」，兩者給出不同的答案（Section 12.1）。一種罕見疾病的檢測即使準確度達 90%，檢測呈陽性的病人真正患病的可能性也只有 9%，因為基本比率（base rate）舉足輕重（Section 12.2）。「隨機的弦」在我們說明它是如何選取之前，根本沒有期望長度可言（Section 12.3），而折斷的木棒（Section 12.21）說明的也是同一道理。此外，「獨立」（independent）原來有好幾種意思：事件可以兩兩獨立而並非全體獨立（Section 12.4），而不相關（uncorrelated）的變量也不一定獨立。在每一個例子中，解決之道都是確切地寫下有哪些可能的結果，以及它們是如何產生的。

2. 巧妙的切入點在哪裏？

許多題目都既有冗長的計算，也有簡短的洞見。

- **對稱性（symmetry）。**過程本身無法分辨的兩個座位（Section 12.5）；「正面較多」與「反面較多」是互補的鏡像（Section 12.6）；每一對選手在任何一場比賽中相遇的可能性均等（Section 12.7）；對任何硬幣而言，HT 與 TH 出現的可能性相同（Section 12.9）；徑向對稱的點，其方向服從均勻分佈（Section 12.22）。
- **指示變量（indicator）與線性（linearity）。**把配對（Section 12.4）、繩結（Section 12.10）和比賽場數（Section 12.7）寫成指示變量之和來點算，毋須求出完整的分佈，便可得到平均值和方差。
- **狀態與首步分析（first-step analysis）。**Penney 遊戲（Section 12.8）、兔子洞穴與 Ecks 教授（Section 12.16），以及姓氏的消亡（Section 12.15），全都是以下一步為條件來求解的。
- **策略。**三人決鬥（truel）中槍法最差的人應該故意射失（Section 12.13）；隨機化的策略無法被對手利用（Section 12.14）；先拒絕最初 $\frac{1}{e}$ 的應徵者是最優的做法（Sec-

tion 12.12）。

3. 長遠而言會發生甚麼事？

本章最後的部分追蹤隨機性如何隨時間演變，以及在大量試驗中的表現。分支過程（branching process）要麼消亡，要麼爆發式增長，而即使是爆發式增長的過程，也可能因偶然而消亡（Section 12.15）。馬可夫鏈（Markov chain）會趨於一個平穩分佈（stationary distribution），並忘卻自己的過去（Section 12.16）。服從指數分佈的等候時間則完全沒有記憶（Section 12.18）。平均值通常會穩定下來，但柯西（Cauchy）分佈卻不然，因為它沒有平均值（Section 12.22）。根據中心極限定理（central limit theorem），大量獨立效應之和會趨於正態分佈（Section 12.24），而這種正態性之中更隱藏着 $n!$ 的斯特靈（Stirling）公式（Section 12.23）。最大值的表現又截然不同：它們按 $\ln n$ 的速度增長，離散程度卻完全不會增長（Section 12.25）。而常數 e、π 和 γ 則不斷出現，在杯子與郵票之中，也在針與繩圈之中。

4. 概率的工具箱

方法	你在哪裏用過
寫下樣本空間（sample space）	條件概率（12.1）；貝特朗（Bertrand）（12.3）
貝葉斯定理（Bayes' theorem）與賠率（odds）	基本比率（12.2）；馬可夫逆推（12.16）
對稱性	12.5、12.6、12.7、12.22
指示變量與線性	12.4、12.7、12.10
首步分析	Penney（12.8）；兔子洞穴（12.16）；分支過程（12.15）
補事件（complement）與界	公平的決定（12.9）
以級數求期望值	郵票與 e（12.11）
參數化並求最優	祕書問題（12.12）
博弈論（game theory）式推理	三人決鬥（12.13）；混合策略（12.14）
點算通往終點的路徑	火柴盒（12.17）
先求分佈函數（distribution function）	12.18、12.19、12.25
把概率看成面積或積分	蒲豐（Buffon）（12.20）；折斷的木棒（12.21）
正態近似	斯特靈（12.23）；中心極限定理（12.24）

下一步

- 第 13 章由概率轉向統計：從數據估計未知的量，由 Section 13.1 開始。
- Section 9.8、Section 9.9 和 Section 9.11 的計數問題，是本章內容在組合數學方面的一面。

- 馬可夫鏈用到第 7 章的線性代數，而分支過程的不動點則用到 Section 8.7 的動力系統理論。

如果你想在入讀大學之前再進一步

- **Joseph Blitzstein & Jessica Hwang, *Introduction to Probability***：一流的入門課程，可在網上免費閱讀，並附有講課影片。
- **Geoffrey Grimmett & Dominic Welsh, *Probability: An Introduction***：簡潔而嚴謹，屬大學一年級程度。
- **Frederick Mosteller, *Fifty Challenging Problems in Probability***：經典謎題，附完整解答。
- **David Spiegelhalter, *The Art of Statistics* 及 *The Art of Uncertainty***：現實世界中的概率與風險。

拉普拉斯（Laplace）說過，概率就是化為計算的常識。本章說明了，之所以往往需要計算，正正是因為常識會失靈；而最好的計算不止於得出數字，它們還能解釋箇中道理。一種對稱性、一個指示變量、一個選得恰當的狀態，或一條極限定理，都能把一道謎題化為一個洞見，而大學的概率課程正是建基於這些洞見之上。

延伸練習答案

以下只是簡短的答案和提示，並非完整解答：它們給出結果和關鍵步驟，方便你核對自己的做法。請先認真做過每道練習，然後才看答案。

12.1 精確地取條件

1. 在 (iv) 和 (v) 中，其餘 $k-1$ 塊圓片都是從餘下的 $b+n-1$ 塊中隨機抽出的，其中 $n-1$ 塊有編號，所以兩者都等於 $\frac{(n-1)\binom{b}{k-2}}{\binom{b+n-1}{k-1}}$。(vi) 的答案是 $\frac{\binom{n}{2}\binom{b}{k-2}}{\binom{b+n}{k}-\binom{b}{k}}$。設 $N=b+n$，兩者只在 $(2N-nk)\binom{N-1}{k-1}=2k\binom{b}{k}$ 時才相等。這只會在個別特殊情況下發生，例如 $(b,n,k)=(2,4,3)$、$(3,3,4)$ 或 $(6,11,3)$，又或者是 $n=2$ 而所有圓片都被抽出這種平凡情況。對於題目所給的數值，等式並不成立。

2. 兩個孩子的（性別，星期幾）組合共有 $14^2=196$ 種，各種組合可能性相等，其中 $196-13^2=27$ 種包含至少一個星期二出生的男孩。在這些組合之中，$7^2-6^2=13$ 種是兩個男孩。所以概率是 $\frac{13}{27}$。

3. 你選了 1 號門。若汽車在 1 號門後（概率 $\frac{1}{3}$），換門會輸；若汽車在 2 號或 3 號門後，主持人必須打開另一扇後面是山羊的門，換門便會贏：概率為 $\frac{2}{3}$。若主持人隨機打開一扇門，而門後碰巧是山羊，那麼根據你所見，「汽車在你選的門後」和「汽車在另一扇關着的門後」這兩個結果的可能性相等：概率為 $\frac{1}{2}$。

4. $\mathrm{P}(\geq 2 \mid \geq 1)=\frac{1-\mathrm{P}(0)-\mathrm{P}(1)}{1-\mathrm{P}(0)}\approx 0.370$，但 $\mathrm{P}(\geq 2 \mid \text{葵扇 A})=1-\binom{48}{12}/\binom{51}{12}\approx 0.561$。知道是哪一張 A，會令手上有第二張 A 的可能性增加。

12.2 基本比率謬誤

1. 若屍體在那裏，兩者都說「有屍體」的概率是 pq；若不在，則是 $(1-p)(1-q)$。先驗勝算（prior odds）為 $1:1$，所以後驗概率（posterior probability）是 $\frac{pq}{pq+(1-p)(1-q)}=\frac{pq}{1-p-q+2pq}$。至於第二部分，在三種可能性相等的情況中，有兩種情況下「A 在那裏」為真，所以 $\mathrm{P}(\text{兩者都在} \mid \text{兩者都說}A)=\frac{pq}{2pq+(1-p)(1-q)}$。

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

3. 每一次 AL 測試呈陽性，都會把勝算乘以 $\frac{0.9}{1/11}=9.9$。由先驗勝算 $\frac{1}{99}$ 開始：一次陽性之後，勝算為 0.1（概率 $\frac{1}{11}$）；兩次之後，勝算為 0.99（概率約 0.497）；三次之後，勝算為 9.8（概率約 0.907）。

4. 在 100 名患者之中，99 人測試呈陽性；在 999 900 名沒有患病的人之中，約有 9999 人呈陽性。所以約有 10 100 人被告知可能患病，而當中只有 99 人（約 1%）真正患病。

12.3 「隨機」是甚麼意思？

1. 弦（chord）比內接等邊三角形的邊長 $\sqrt{3}a$ 長的條件分別是：它的遠端落在與近端相對的那三分之一圓周上（$\frac{1}{3}$）；它與 O 的距離小於 $\frac{a}{2}$（$\frac{1}{2}$）；它的中點落在半徑為 $\frac{a}{2}$ 的圓內（$\frac{1}{4}$）。

2. 以 C 為中點的弦垂直於 OC，所以它就是通過 C 而角度為 $t = \frac{\pi}{2}$ 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. 弦長為 $2\sqrt{a^2 - d^2}$，平均值為 $\frac{2}{a}\int_0^a \sqrt{a^2 - d^2}\,\mathrm{d}d = \frac{2}{a}\cdot\frac{\pi a^2}{4} = \frac{\pi a}{2}$。

4. 隨機端點：$\frac{2}{\pi\sqrt{4a^2 - \ell^2}}$。隨機距離：$\frac{\ell}{2a\sqrt{4a^2 - \ell^2}}$。隨機中點：$\frac{\ell}{2a^2}$。

12.4 兩兩獨立，卻非相互獨立

1. 展開後得 $\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = (a - pq)(x_1 - x_2)(y_1 - y_2)$，所以 $a = pq$；其餘三格隨之確定，於是兩者獨立。例子：在 $(0,0)$ 放概率 $\frac{1}{3}$，在 $(\pm1, \pm1)$ 各點放概率 $\frac{1}{6}$。兩個邊際分佈（marginal distribution）都是均勻的，而且 $\mathrm{E}(AB) = 0$，但 $\mathrm{P}(A = 0, B = 1) = 0 \neq \frac{1}{9}$。

2. 共有 21 對，每對相同的概率是 $\frac{1}{6}$，而各指示變量（indicator）兩兩獨立（pairwise independent）。所以平均值是 3.5，方差（variance）是 $21 \cdot \frac{1}{6} \cdot \frac{5}{6} \approx 2.92$，標準差約為 1.71。得分 7 大約比平均值高出兩個標準差。

3. 平均值是 $\binom{k}{2}\frac{1}{365}$。各指示變量兩兩獨立，與骰子的情況完全一樣，所以方差是 $\binom{k}{2}\frac{1}{365}\cdot\frac{364}{365}$。（它們並非相互獨立（mutually independent）：若 1,2 相同，2,3 也相同，則 1,3 亦相同。）

4. 取八個可能性相等的結果，$A = B = \{1,2,3,4\}$，$C = \{1,5,6,7\}$。那麼 $\mathrm{P}(A \cap B \cap C) = \frac{1}{8}$，但 $\mathrm{P}(A \cap B) = \frac{1}{2} \neq \frac{1}{4}$。

12.5 最後一位乘客的座位

1. 對所有 $n \geq 3$，答案都是 $\frac{1}{3}$。最後一個座位現在是 S_1、S_2 和 S_n 其中之一，而每一次隨機選擇對這三個座位都是對稱的。

2. 若 T_1 坐了 S_1，沒有人要讓位。若 T_1 坐了 S_3，則 T_1 和 T_3 都坐錯了位。若 T_1 坐了 S_2，便由 T_2 選擇：選 S_1 有 2 人坐錯位，選 S_3 則有 3 人。期望值（expectation）是 $\frac{1}{3}\left(0 + 2 + \frac{5}{2}\right) = \frac{3}{2}$，與公式給出的 $\frac{2}{3} + \frac{1}{3} + \frac{1}{2}$ 一致。

3. 當 $r \geq 2$ 時 $P_r = \frac{1}{2}$，於是 $\frac{1}{n}\left(1 + \frac{n-2}{2}\right) = \frac{1}{2}$，所以遞推關係（recurrence）與 $P_n = \frac{1}{2}$ 相容。

4. 最後一個空位必定是 S_1 或 S_n（任何其他 S_k 最遲在 T_k 登機時已被佔去）。所以若它不是 S_n，就是 S_1。

12.6 多一枚硬幣

1. 設 H' 為 Bob 首 $n+1$ 次擲幣的結果，X 為他最後一次擲幣的結果。若 $H' > H_A$，或者 $H' = H_A$ 而 X 是正面，Bob 便勝出。所以概率是 $\mathrm{P}(H' > H_A) + \frac{1}{2}\mathrm{P}(H' = H_A)$，而根據「多一枚硬幣」的結果，$\mathrm{P}(H' > H_A) = \frac{1}{2}$。

2. 不成立。當 $n = 0$ 時，Bob 勝出的概率是 p。那個巧妙的證明把正面和反面對調，而這只有對公平硬幣才是一種對稱。

3. $H_A = H_B$ 恰好在 $H_A + (\text{Bob 的反面數目}) = n$ 時成立，即一個 $\mathrm{B}(2n, \frac{1}{2})$ 變量等於 n。數值：$\frac{1}{2}$、$\frac{3}{8}$、$\frac{5}{16}$。

4. 不成立：擲出六點就像一枚 $p = \frac{1}{6}$ 的硬幣。當 $n = 0$ 時概率是 $\frac{1}{6}$，而一般而言概率低於 $\frac{1}{2}$。

12.7 淘汰賽中誰會遇上誰

1. 第 k 輪有 2^{n-k} 場比賽，而根據對稱性，每場比賽由 $\binom{2^n}{2}$ 對球員中任何一對對賽的可能性都相等。把整個淘汰賽的 $2^n - 1$ 場比賽加起來，得 $\frac{2^n - 1}{\binom{2^n}{2}} = \frac{2}{2^n}$。

2. 不會：在固定的對陣表中隨機編排位置，對稱性依然保持，同樣的計數方法仍然適用。

3. X 在第 k 輪的對手來自一組 2^{k-1} 名球員，每人從該組勝出的可能性相等。所以 Y 是該對手的概率為 $\frac{2^{k-1}}{2^n - 1} \cdot \frac{1}{2^{k-1}} = \frac{1}{2^n - 1}$，而 n 輪的總和是 $\frac{n}{2^n - 1}$。

4. 共有 $\binom{2n}{2} = n(2n-1)$ 場比賽；每一對都會相遇，所以概率是 1。

12.8 Penney 遊戲

1. HHT 只有在頭兩次擲出 HH 時才會勝出；否則在湊成 HH 之前已出現一個 T，THH 必定先出現。所以 HHT 勝出的概率是 $\frac{1}{4}$。四人一起玩時：HH 之後，HHT 必勝；TT 之後，TTH 必勝；由 HT 和 TH 開始，HTT 和 THH 的勝出機會分別是 $\frac{2}{3}$ 和 $\frac{1}{3}$（或倒過來），於是每名玩家的勝出概率是 $\frac{1}{4} \cdot 1$ 或 $\frac{1}{4}\left(\frac{2}{3} + \frac{1}{3}\right)$，即 $\frac{1}{4}$。TTH 擊敗 THH 的概率是 $\frac{2}{3}$（第 (i) 部分）。

2. 對於 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 $e_0 = 6$。對於 HT：一旦有了一個 H，便等待一個 T（平均擲兩次），而第一個 H 平均要擲兩次：共 4 次。

3. HHH 和 TTT 分別被 THH 和 HTT 擊敗，概率為 $\frac{7}{8}$；HHT 和 TTH 分別被 THH 和 HTT 擊敗，概率為 $\frac{3}{4}$；其餘四個（HTH、HTT、THH、THT）分別被 HHT、HHT、TTH、TTH 擊敗，概率為 $\frac{2}{3}$。HHH 和 TTT 最容易被擊敗。

4. 設 $A = \mathrm{HTT}$，$B = \mathrm{HHT}$，各相關數（correlation）為 $AA = 4$、$AB = 0$、$BB = 4$、$BA = 2$。B 勝出的勝算是 $(AA - AB) : (BB - BA) = 4 : 2$，即概率為 $\frac{2}{3}$。

12.9 用不公平的硬幣作公平的決定

1. 所需輪數服從幾何分佈（geometric distribution）：第 (i) 部分的平均值是 $\frac{1}{p^2+q^2}$，而馮·諾伊曼（von Neumann）的方法（HT 或 TH）的平均值是 $\frac{1}{2pq}$。由於 $p^2 + q^2 - 2pq = (p-q)^2$，對所有 $p \neq \frac{1}{2}$，馮·諾伊曼的方法都較慢。不過，公平的卻是它。

2. 用馮·諾伊曼的技巧製造公平的位元（bit）。每次取兩個：00、01 和 10 分別代表三個選項，11 則表示重新開始。

3. 把擲幣結果逐次與二進制展開式 $\frac{1}{3} = 0.010101\ldots$ 比較（把 H 讀作 1，T 讀作 0）。在第一次出現不同的那一擲，若擲幣結果小於該位數字（T 對着 1），便宣佈「正面」，否則宣佈「反面」。這樣得出正面的概率恰好是 $\frac{1}{3}$，平均用 2 次擲幣。

4. 在未能決定的一輪之後，其中一面有兩個，另一面有一個，而整個過程是這兩個狀態上的馬可夫鏈（Markov chain）。解之得 $\mathrm{P}(\text{酒店}) = \frac{p^3}{p^3+q^3}$，當 $p = \frac{1}{2}$ 時等於 $\frac{1}{2}$：這時是公平的，其他情況則不然。

12.10 把繩頭繫起來

1. 第一次打結形成一個環的概率是 $\frac{1}{3}$（其餘三個繩端之中，有一個屬於同一條繩），而最後一次打結則必定形成一個環。環的數目：2 個的概率是 $\frac{1}{3}$，1 個的概率是 $\frac{2}{3}$；平均值是 $\frac{4}{3}$。

2. 餘下 k 條繩時，一次打結不形成環的概率是 $1 - \frac{1}{2k-1}$。乘積 $\prod_{k=2}^{n} \frac{2k-2}{2k-1}$ 等於 $\frac{4^n}{2n\binom{2n}{n}}$，約為 $\frac{\sqrt{\pi n}}{2n} = \sqrt{\frac{\pi}{4n}}$。

3. 依次插入元素 $1, 2, \ldots, n$：元素 k 要麼開始一個新的循環（cycle）（概率 $\frac{1}{k}$），要麼放在之前 $k-1$ 個元素其中一個之後。把各指示變量加起來便得 H_n。

4. 各次打結的成環指示變量互相獨立，概率為 $\frac{1}{2k-1}$，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 郵票、杯子與 e

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

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

3. 完全沒有改變：只有一個杯時，Jim 必定「猜」中。期望價格仍然是 e。

4. $\int_0^1\frac{(1-u)^2}{2}\,\mathrm{d}u=\frac{1}{6}$。一般而言，在 $u_k=u$ 處的截面是一個按 $1-u$ 縮放的 $(k-1)$ 維單純形（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 何時停止尋找

1. 若最大的在第 $k\geq 2$ 位，你恰好在第一個雪糕是首 $k-1$ 個之中最大時才會得到它，概率為 $\frac{1}{k-1}$。對 k 取平均得 $\frac{1}{n}\sum_{k=2}^{n}\frac{1}{k-1}$，當 $n=4$ 時是 $\frac{11}{24}$。列出全部 24 種次序，可知得到第二、第三和第四大的概率分別是 $\frac{7}{24}$、$\frac{1}{6}$ 和 $\frac{1}{12}$。

2. 拒絕 $r=0,1,2,3$ 人分別給出 $\frac{1}{4}$、$\frac{11}{24}$、$\frac{5}{12}$、$\frac{1}{4}$。所以應拒絕一名候選人，成功概率為 $\frac{11}{24}$。

3. 首 n 人之中最好的一個位列頭兩名的概率是 $\frac{n}{N}+\frac{n}{N}\cdot\frac{N-n}{N-1}$，當 n 很小時接近 $\frac{n}{N}$ 的兩倍。若頭兩名任何一個都可以接受，你便可以少看一些候選人才作決定。（以此為目標的最佳策略，成功概率約為 0.57。）

4. 設 $\mathrm{P}(r)=\frac{r}{N}\sum_{k=r}^{N-1}\frac{1}{k}$：當 $N=10$ 時，$r=2,3,4$ 給出 0.366、0.399、0.398；當 $N=100$ 時，$r=36,37,38$ 給出 0.3710、0.3710、0.3708，其中 $r=37$ 僅僅領先。

12.13 三人決鬥

1. Algernon 擲失後，Bertie 擲向 Cuthbert。若他擲中 $(\frac{3}{5})$，便是 Algernon 對 Bertie，由 Algernon 先擲，Bertie 勝出的概率是 $1-\frac{2/5}{2/5+3/5-6/25}=\frac{9}{19}$。若他擲失，Cuthbert 會擲中 Bertie，然後除非 Algernon 先擲中 $(\frac{3}{5})$，否則 Cuthbert 勝出。所以 Bertie 勝出的概率是 $\frac{3}{5}\cdot\frac{9}{19}=\frac{135}{475}$，Cuthbert 則是 $\frac{2}{5}\cdot\frac{3}{5}=\frac{114}{475}$。

2. 對所有 $a<1$ 都成立。擲失給出 $\frac{3}{5}\cdot\frac{a}{a+\frac{3}{5}-\frac{3}{5}a}+\frac{2}{5}a$；擲中 Cuthbert 給出 $1-\frac{3/5}{3/5+a-\frac{3}{5}a}$（這時 Bertie 先向他擲），這個值總是較小；擲中 Bertie 則給出 0。

3. 不應該。若 Bertie 擲向 Algernon 而擲中，Cuthbert 隨即擲中 Bertie；若他擲失，Cuthbert 照樣擲中 Bertie。無論如何 Bertie 都一無所得，而擲向 Cuthbert 則有 $\frac{27}{95}$。

4. $p+(1-p)(1-r)p$，它小於 $\frac{p}{p+r-pr}$，即整個等比級數（geometric series）之和。

12.14 混合策略

1. 取 $b=(\frac{1}{4},\frac{1}{2},\frac{1}{4})$，則 $X=Y=Z=0$，所以無論對方怎樣做，B 都贏得 0。對着 $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$，A 猜 1、2 和 3 的期望贏款分別是 $-\frac{1}{3}$、0 和 $\frac{1}{3}$，所以 A 應該每次都猜 3。

2. 收益矩陣（payoff matrix）的對角線上是 0，其餘位置是 ± 1，把三個策略循環置換後矩陣不變，所以最優策略是 $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$。在原題中，相差 1 和相差 2 的罰則不同，所以各策略不能互換。

3. 若 $X、Y、Z \geq 0$ 且 $X + 2Y + Z = 0$，則三者都是 0。由此必有 $b_1 = b_3$ 及 $b_2 = 2b_3$。

4. 現在 $X = -b_2 + 3b_3、Y = b_1 - b_3$ 及 $Z = -3b_1 + b_2$，而 $X + 3Y + Z = 0$。令三者都等於零，得 $(\frac{1}{5}, \frac{3}{5}, \frac{1}{5})$。

12.15 滅絕還是爆發

1. 滅絕概率（extinction probability）是 $s = (q + ps)^2$ 在 $[0, 1]$ 內的最小根。兩個根是 1 和 $\left(\frac{q}{p}\right)^2$，而第二個根恰好在 $p > \frac{1}{2}$ 時小於 1。

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

3. 在第 n 代或之前滅絕，表示第一代每一顆粒子的後代都在 $n - 1$ 代之內消亡。取 $p = \frac{3}{4}$：$e_1 = 0.0625$，$e_2 \approx 0.0881$，$e_3 \approx 0.0999$，遞增並趨向 $\frac{1}{9} \approx 0.111$。

4. $s = \frac{1}{4}(1 + s + s^2 + s^3)$ 可因式分解為 $(s - 1)(s^2 + 2s - 1) = 0$，所以滅絕概率是 $\sqrt{2} - 1 \approx 0.414$。

12.16 順向與逆向的馬可夫鏈

1. 根據對稱性 $p_B = p_D$；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 出發，不經過 C 而在 A 離開：$x_A = \frac{1}{3} + \frac{2}{3}x_B$ 及 $x_B = \frac{1}{3}x_A$，所以 $x_A = \frac{3}{7}$。條件概率是 $\frac{3/7}{7/15} = \frac{45}{49}$。

2. 把第一行乘以 P：$(0.19, 0.33, 0.48)$。

3. 跡（trace）是 1.3，行列式（determinant）是 0.02，而 1 是一個特徵值（eigenvalue）（各行之和為 1）。另外兩個特徵值之和為 0.3，積為 0.02：它們是 0.2 和 0.1。

4. $a_1 p_{12} = \frac{1}{20}$ 但 $a_2 p_{21} = \frac{1}{15}$，所以這條鏈不是可逆的（reversible）：倒過來運行時，它的轉移概率並不相同（第 (iii) 部分的 q_{ij} 與 p_{ij} 不同）。

12.17 巴拿赫的火柴盒

1. 若第二次伸手要拿的與第一次是同一種，她當下便發現那包已空（$N = 2$）；否則她拿走另一顆糖果，並在第三次伸手時發現有一包已空。所以 $N = 2$ 或 3，概率各為 $\frac{1}{2}$。

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. 就餘下的拖肥糖和薄荷糖數目作首步遞推（first-step recursion），得 $\frac{64}{81}\approx 0.79$。

12.18 無記憶性

1. $\mathrm{P}(\min\ >\ t)\ =\ \mathrm{e}^{-\lambda_1 t}\mathrm{e}^{-\lambda_2 t}$。等候下一班巴士的時間服從指數分佈（exponential distribution），率為每分鐘 $\frac{1}{15}+\frac{1}{30}=\frac{1}{10}$，所以其中位數（median）是 $10\ln 2\approx 6.9$ 分鐘。
2. $\mathrm{P}(Y\geq k)=(1-p)^k$，所以比值是 $(1-p)^n$。
3. 包含 t 的那段間隔，由回溯至上一班巴士的時間和前望至下一班巴士的時間組成，兩者都服從平均值為 $\frac{1}{\lambda}$ 的指數分佈（當 t 很大時），所以間隔的期望長度是 $\frac{2}{\lambda}$。較長的間隔較有可能包含某個指定時刻。
4. 在 $[0,1]$ 上為 $\frac{\lambda \mathrm{e}^{-\lambda z}}{1-\mathrm{e}^{-\lambda}}$。當 $\lambda=1$ 時，它由約 1.58 緩緩下降至 0.58；當 $\lambda=5$ 時，它集中在 0 附近，由約 5.03 下降至 0.03。

12.19 平面上的隨機點

1. 由 $1-\mathrm{e}^{-\pi k y^2}=\frac{1}{2}$ 得中位數為 $\sqrt{\frac{\ln 2}{\pi k}}\approx\frac{0.470}{\sqrt{k}}$，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. 在極坐標下，R 的密度是 $\frac{1}{2\pi\sigma^2}\mathrm{e}^{-r^2/(2\sigma^2)}\cdot 2\pi r$。當 $\sigma^2=\frac{1}{2\pi k}$ 時，它與超級市場的分佈吻合。
4. 取 $k=4\times 10^{-4}$ 每 m^2，完全隨機的假設預測平均值為 $\frac{1}{2\sqrt{k}}=25\,\mathrm{m}$。觀察所得的 35 m 較大，這顯示樹木是互相疏開的（競爭），而不是聚集成叢。

12.20 蒲豐投針

1. 把針切成很多小段。相交次數的期望值具有可加性，而每一小段的貢獻與其長度成正比，所以無論 a 是多少，總和都是 $\frac{2a}{\pi b}$。
2. 根據同樣的可加性（蒲豐投麵條，Buffon's noodle），仍然是 $\frac{2a}{\pi b}$。
3. 圓心與每條邊的距離都必須至少為 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 折斷一根棒

1. 要組成三角形，每一段都必須短於 $\frac{1}{2}$。在由折斷點組成的單位正方形中，這個條件劃出一個面積為 $\frac{1}{4}$ 的區域。

2. 假設第一個折斷點在 $x < \frac{1}{2}$（另一情況是對稱的），而較長的一段 $1-x$ 再被均勻地折斷。新的兩段都必須短於 $\frac{1}{2}$，其概率為 $\frac{x}{1-x}$。所以答案是 $2\int_0^{1/2} \frac{x}{1-x}\,\mathrm{d}x = 2\ln 2 - 1$。

3. 當 $(2-x-y)^2 > x^2+y^2$，即 $y < \frac{2-2x}{2-x}$ 時，C 處的角是鈍角。這條曲線與 $x+y=1$ 之間的面積是 $\frac{3}{2} - 2\ln 2$，所以概率是 $3-4\ln 2$，與 B 處的相同。這並非巧合：三條邊各自最多為 1，總和為 2，並且是均勻分佈的，所以它們可以互換。因此 $\mathrm{P}(\text{鈍角三角形}) = 9 - 12\ln 2 \approx 0.68$。

4. 根據容斥原理（inclusion–exclusion），最長的一段 M 滿足 $\mathrm{P}(M > m) = 3(1-m)^2 - 3(1-2m)_+^2 + (1-3m)_+^2$，其中 $x_+ = \max(x,0)$。在 $[0,1]$ 上積分得 $\mathrm{E}(M) = 1 - \frac{1}{2} + \frac{1}{9} = \frac{11}{18}$。

12.22 沒有平均值的分佈

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

2. 把 U 和 V 對調，徑向對稱的密度保持不變，所以 $\frac{V}{U}$ 與 $\frac{U}{V}$ 有相同的分佈。

3. 密度是對稱的，所以中位數是 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 $s = r^2$ 後這就是 $\frac{1}{2\pi}\,\mathrm{d}\theta \cdot \frac{1}{2}\mathrm{e}^{-s/2}\,\mathrm{d}s$：角度是均勻分佈的，$R^2$ 服從平均值為 2 的指數分佈，而且兩者獨立。

12.23 由概率得出斯特靈公式

1. $\sqrt{20\pi}(10/\mathrm{e})^{10} \approx 3\,598\,696$（約低 0.8%）及 $\sqrt{40\pi}(20/\mathrm{e})^{20} \approx 2.4228 \times 10^{18}$（約低 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 位數字。

4. $\binom{1000}{500}2^{-1000} \approx \frac{1}{\sqrt{500\pi}} \approx 0.0252$。

12.24 中心極限定理

1. 作卷積（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. $\mathrm{E}(\mathrm{e}^{\theta X}) = \frac{1}{2}(\mathrm{e}^{\theta/2} + \mathrm{e}^{-\theta/2}) = \cosh\frac{\theta}{2}$。標準化後的和 $\frac{2}{\sqrt{n}}\sum X_i$ 的母函數（generating function）是 $\cosh^n \frac{\theta}{\sqrt{n}}$，其對數為 $n\ln\bigl(1 + \frac{\theta^2}{2n} + \cdots\bigr) \to \frac{\theta^2}{2}$。

3. 把 Y 寫成兩個獨立的 $\Gamma(82,1)$ 變量之差，再作數值積分，得 $\mathrm{P}(|Y| \geq 25) \approx 0.051$：正態近似（normal approximation）非常準確。

4. 它沒有矩母函數（moment generating function），甚至連平均值也沒有。事實上，n 個獨立柯西（Cauchy）變量的平均值本身也服從柯西分佈：取平均完全不會令它變得集中。

12.25 眾多之中的最大值

1. n 個壽命之中，最先完結的一個平均需時 $\frac{1}{n\lambda}$；之後餘下 $n-1$ 個，一切重新開始，如此類推。所以 $\mathrm{E}(Y) = \frac{1}{\lambda}\bigl(\frac{1}{n} + \frac{1}{n-1} + \cdots + 1\bigr)$。

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 $\lambda y = -\ln(1 - 2^{-1/n})$。當 $n = 10$ 時這是 2.704，而近似值為 $\ln 10 - \ln\ln 2 \approx 2.669$。

4. $\mathrm{P}(M \leq m) = m^n$，所以 $\mathrm{P}\bigl(n(1-M) > s\bigr) = \bigl(1 - \frac{s}{n}\bigr)^n \to \mathrm{e}^{-s}$。這屬於韋布爾（Weibull）型：各變量有上界，而最大值會擠向這個上界。

Beyond STEP

STEP 是大多數中學生一生中應考過最艱深的數學考試，也是最出色的考試之一：題目簡短、陌生而深刻，當中最精彩的，其實是喬裝了的大學數學小品。

本書精選最有趣的 STEP 試題，由 1987 年首次開考之前的樣本試卷（specimen papers），到 2025 年的試卷，並以每一道題目作為一扇門。每道題目都會仔細解答，沿途提出優秀學生會問的問題，然後帶你穿過這扇門：這個想法從何而來，到了大學會演變成甚麼，以及接下來該讀些甚麼。

本書為正在準備 STEP、有天分又有志投身數學的學生而寫。他們想要的不止是模擬試卷：他們想開始像數學系本科生那樣思考，並一探前方有甚麼正等着自己。

- 共 13 章，涵蓋純數學（pure mathematics）、力學（mechanics）、概率（probability）與統計（statistics）
- 136 道 STEP 歷屆試題，每題均附引導式解答
- 每道題目背後的大學數學
- 附答案的延伸練習，以及延伸閱讀

全書共 13 章，本冊為第 12 章。 全書各章及所有詳細解答均可免費閱讀、下載及分享，網址為

beyondstep.keithmcnulty.org

文字內容已貢獻予公有領域（CC0 1.0）。STEP 試題 © Cambridge University Press & Assessment。