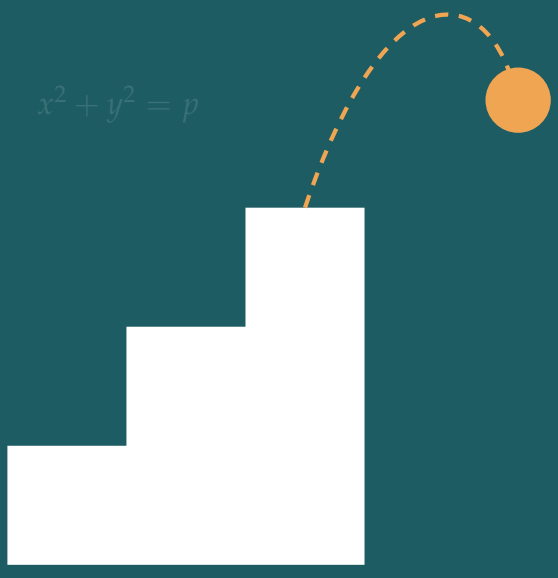
STEP 備試 · 大學數學

# Beyond STEP

## STEP I, II 及 III

1999 年詳細解答



KEITH McNULTY

# Beyond STEP

以歷屆 STEP 試題為門
走進大學數學

**STEP I, II 及 III**

## 1999 年詳細解答

Keith McNulty

**Beyond STEP：以歷屆 STEP 試題為門，走進大學數學**

**1999 年詳細解答**

1999 年 STEP 試卷（STEP I, II 及 III）全部 42 題的詳細解答。自樣本試卷以來每一道 STEP 題目的詳細解答、*Beyond STEP* 全書以及各章的單行本，均可在 `beyondstep.keithmcnulty.org` 免費閱讀及下載。

閱讀解答之前，請先認真嘗試解題。每個答案均已用數值方法核對，但解答數量如此之多，難免會有疏漏：如果發現有可疑之處，請相信自己的演算，並仔細檢查。

由機器從英文翻譯而成；尚未經審閱。

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。

# 目錄

## STEP I


第 1 題 互質整數的計數 純數學 3

第 2 題 軌跡與形心 純數學 6

第 3 題 輪換方程組 純數學 9

第 4 題 含絕對值的區域 純數學 13

第 5 題 小角近似 純數學 18

第 6 題 二次函數的極值 純數學 22

第 7 題 二階微分方程 純數學 26

第 8 題 調和級數的界限 純數學 30

第 9 題 勻速率運動學 力學 34

第 10 題 簡諧運動 力學 38

第 11 題 向心力作用下的圓周運動 力學 42

第 12 題 灌鉛骰子的不等式 概率與統計 46

第 13 題 指示變量 概率與統計 50

第 14 題 條件概率 概率與統計 53


## STEP II


第 1 題 大數的比較 純數學 56

第 2 題 二次方程的判別式 純數學 59

第 3 題 數學歸納法 純數學 63

第 4 題 二項式係數恆等式 純數學 66

第 5 題 三角方程 純數學 69

第 6 題 代換積分法 純數學 72

第 7 題 曲線描繪 純數學 75

第 8 題 三角級數 純數學 80

第 9 題 變力作用下的運動 力學 . . . 83
第 10 題 連續碰撞 力學 . . . 87
第 11 題 拋體與移動目標 力學 . . . 91
第 12 題 貝葉斯定理 概率與統計 . . . 95
第 13 題 連續隨機變量 概率與統計 . . . 98
第 14 題 最優停止 概率與統計 . . . 101

## STEP III

第 1 題 三次方程的根 純數學 . . . 105
第 2 題 實根的計數 純數學 . . . 108
第 3 題 黎曼和與極限 純數學 . . . 112
第 4 題 歐拉公式與多面體 純數學 . . . 116
第 5 題 斐波那契恆等式與數學歸納法 純數學 . . . 120
第 6 題 參數曲線與面積 純數學 . . . 123
第 7 題 群與同構 純數學 . . . 127
第 8 題 分段微分方程 純數學 . . . 130
第 9 題 牛頓殼層定理 力學 . . . 134
第 10 題 墜落鏈條的衝量 力學 . . . 138
第 11 題 剛體轉動 力學 . . . 141
第 12 題 離散聯合分佈 概率與統計 . . . 144
第 13 題 連續先驗分佈的貝葉斯定理 概率與統計 . . . 147
第 14 題 條件化求期望值 概率與統計 . . . 150

詳細解答

# STEP I 1999，第 1 題

純數學 • 互質整數的計數

> **STEP I 1999，第 1 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/99-s1-q1

## 關鍵思路

能否被 2 和 5 整除（divisibility）只取決於末位數字，而能否被 3 和 7 整除只取決於模（modulo）21 的餘數（remainder）。因此，我們要找的整數會按固定的規律一段接一段地重複出現，要數它們的個數，只需數出一段中有多少個，再乘以段數。至於平均值，竅門在於對稱（symmetry）：若 $n$ 在集合內，它的「拍檔」$N-n$ 也在集合內，於是這些數便圍繞中點兩兩配對。

## 解答

### 第 1 部分：小於一百萬且不能被 2 或 5 整除的整數

「不能被 2 或 5 整除」的意思是兩者都不能整除。一個整數能被 2 或 5 整除，當且僅當它的末位數字是 $0, 2, 4, 5, 6$ 或 $8$。所以我們要找的是末位數字為 1、3、7 或 9 的整數。

整數 $0, 1, \ldots, 999\,999$ 可分成 100 000 段，每段是十個連續整數 $10k, 10k+1, \ldots, 10k+9$，而每段恰好包含四個我們要找的整數。因此個數為

$$4 \times 100\,000 = 400\,000.$$

（用容斥原理（inclusion–exclusion）也得到相同結果：$10^6 - \frac{1}{2}10^6 - \frac{1}{5}10^6 + \frac{1}{10}10^6 = 400\,000$。）

求平均值時，把每個這樣的 $n$ 與 $10^6 - n$ 配對。若 $n$ 的末位數字是 $1, 3, 7, 9$，則 $10^6 - n$ 的末位數字分別是 $9, 7, 3, 1$，而且因為 $n \neq 0$，所以 $0 < 10^6 - n < 10^6$。因此 $n \mapsto 10^6 - n$ 把這個集合映射到它自身，而且這個映射是它自己的逆映射（inverse）。它沒有不動點（fixed point），因為 $n = 500\,000$ 是偶數。於是這個集合分成 200 000 對，每對之和為

$10^6$。總和是 $200\,000 \times 10^6$，平均值是

$$\frac{200\,000 \times 10^6}{400\,000} = 500\,000.$$

**第 2 部分：小於 4179 且不能被 3 或 7 整除的整數**

4179 這個數是精心挑選的：

$$4179 = 3 \times 1393 = 3 \times 7 \times 199 = 21 \times 199.$$

$n$ 能否被 3 或 7 整除，只取決於 $n$ 除以 21 所得的餘數。在餘數 $0, 1, \ldots, 20$ 之中，3 的倍數有 $0, 3, \ldots, 18$（共七個），7 的倍數有 $0, 7, 14$（共三個），其中 0 在兩邊都數了一次。所以有 $21 - 7 - 3 + 1 = 12$ 個餘數不能被兩者中任何一個整除：

$$1,\ 2,\ 4,\ 5,\ 8,\ 10,\ 11,\ 13,\ 16,\ 17,\ 19,\ 20.$$

整數 $0, 1, \ldots, 4178$ 恰好組成 199 個完整的段，每段長 21，所以個數為

$$12 \times 199 = 2388.$$

作為驗算，用容斥原理：$4179 - 1393 - 597 + 199 = 2388$。

**想一想** 這裏的數與十進制數字毫無關係，為甚麼對稱的竅門仍然行得通？

關鍵在於 21 整除 4179。於是 $4179 - n \equiv -n \pmod 3$ 以及 $\pmod 7$，所以 $3 \mid 4179 - n$ 當且僅當 $3 \mid n$，對 7 亦然。此外，$n = 0$ 已被排除（它能被任何數整除），所以集合中每個 $n$ 都滿足 $1 \le n \le 4178$，從而亦有 $1 \le 4179 - n \le 4178$。因此 $n \mapsto 4179 - n$ 把這個集合映射到它自身。它沒有不動點，因為 4179 是奇數。於是這 2388 個整數組成 1194 對，每對之和為 4179，平均值是

$$\frac{1194 \times 4179}{2388} = \frac{4179}{2} = 2089\tfrac{1}{2}.$$

## 驗算

用電腦把所有整數逐一檢查，可確認個數 400 000 和 2388 以及平均值 500 000 和 2089.5 完全正確。你也可以親手驗算一個小例子：在小於 21 的範圍內，上面列出的十二個餘數之和為 $126 = 12 \times \frac{21}{2}$，所以它們的平均值是 $10.5 = \frac{21}{2}$，與配對論證所預測的完全一致。

## 常見錯誤

- 把「不能被 2 或 5 整除」理解為「不能同時被兩者整除」，即不能被 10 整除。這樣會得出 900 000 個整數。

- 使用容斥原理時，忘記把 21 的倍數（或 10 的倍數）加回去，因為它們被減了兩次。
- 沒有檢查集合是否對稱，就假定平均值必定是範圍的中點。這裏之所以行得通，只因為 10 整除 $10^6$，而 21 整除 4179。

## STEP 之外

在 $1, \ldots, m$ 之中與 $m$ 互質（coprime）的整數的個數，就是歐拉函數（Euler's totient function）$\varphi(m)$。這裏 $\varphi(10) = 4$ 及 $\varphi(21) = 12$，而所求的個數就是 $\varphi(10^6) = 400\,000$ 和 $199\,\varphi(21)$。配對 $n \leftrightarrow m - n$ 說明，當 $m > 2$ 時，與 $m$ 互質的整數之和總是 $\frac{1}{2} m\,\varphi(m)$，這個事實是踏入模 $m$ 整數算術的第一步。想進一步閱讀，達文波特（Davenport）的 *The Higher Arithmetic* 是極佳的入門書。

詳細解答

# STEP I 1999，第 2 題

純數學 • 軌跡與形心

> **STEP I 1999，第 2 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/99-s1-q2

## 關鍵思路

寫下條件並展開，便會得到 $x^2$ 和 $y^2$ 的係數同為 3，而且沒有 $xy$ 項：這正是圓的特徵。配方法（completing the square）顯示出圓心的位置，而形心（centroid）亦自然出現，因為一次項中出現的正是 $x_1 + x_2 + x_3$。至於對 $a^2$ 的限制，只要問一問「半徑的平方」何時可以是負數，答案便呼之欲出。

## 解答

### 軌跡

設 $P = (x, y)$ 為軌跡（locus）上的一點，並記 $A_i = (x_i, y_i)$，其中 $i = 1, 2, 3$。條件是

$$\sum_{i=1}^{3} \left[ (x - x_i)^2 + (y - y_i)^2 \right] = a^2.$$

展開，並利用 $x_1 + x_2 + x_3 = 3\bar{x}$ 及 $y_1 + y_2 + y_3 = 3\bar{y}$，得

$$3x^2 - 6\bar{x}\,x + 3y^2 - 6\bar{y}\,y + \sum_i (x_i^2 + y_i^2) = a^2.$$

這就是軌跡的方程。要看清它是甚麼，把方程除以 3 再配方：

$$(x - \bar{x})^2 + (y - \bar{y})^2 = \frac{a^2 - \sum_i (x_i^2 + y_i^2) + 3\bar{x}^2 + 3\bar{y}^2}{3}.$$

分子可以整理得相當簡潔。由於 $\sum_i x_i = 3\bar{x}$，

$$\sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - 2\bar{x} \cdot 3\bar{x} + 3\bar{x}^2 = \sum_i x_i^2 - 3\bar{x}^2,$$

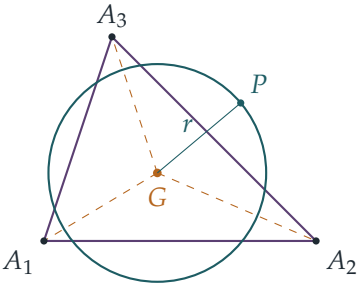
$y$ 的情況亦相同。因此，若記

$$S = \sum_{i=1}^{3}\Big[(x_i - \bar{x})^2 + (y_i - \bar{y})^2\Big]$$

為三點與其形心 $G = (\bar{x}, \bar{y})$ 的距離的平方和，則軌跡為

$$\boxed{(x - \bar{x})^2 + (y - \bar{y})^2 = \frac{a^2 - S}{3}}.$$

**詮釋**。當 $a^2 > S$ 時，這是一個圓，圓心是三點的形心 $G$，半徑為 $\sqrt{(a^2 - S)/3}$。隨着 $a^2$ 變化，我們得到一族以 $G$ 為圓心的同心圓（concentric circles）。在臨界情況 $a^2 = S$ 時，這個「圓」收縮成單獨一點 $G$。




### 為何 $a^2$ 不能小於 $S$

> **想一想** 方框內方程的右邊是負數，這代表甚麼？

左邊 $(x - \bar{x})^2 + (y - \bar{y})^2$ 是實數的平方和，所以永不為負。若 $a^2 < S$，右邊是負數，沒有任何實點 $(x, y)$ 滿足方程：軌跡是空集。因此，要有點能夠按題目所述移動，必須有 $a^2 \geq S$。

把同一番運算倒過來看，會很有啟發。我們實際上證明了的，是以下對任意點 $P$ 均成立的恆等式（identity）：

$$PA_1^2 + PA_2^2 + PA_3^2 = 3\,PG^2 + S.$$

由於 $PG^2 \geq 0$，和 $PA_1^2 + PA_2^2 + PA_3^2$ 必定不小於 $S$，而等號僅在 $P = G$ 時成立。用文字來說：形心就是令到三點的距離平方和最小的那一點，而最小值為 $S$。一個比這最小值還要小的常數和 $a^2$ 是不可能的。

## 驗算

- 取 $A_1 = (0,0)$、$A_2 = (4,0)$、$A_3 = (1,3)$，則 $G = (\frac{5}{3}, 1)$，且 $S = \frac{34}{9} + \frac{58}{9} + \frac{40}{9} = \frac{44}{3}$。在 $P = G$ 時，距離平方和恰好是 $\frac{44}{3} \approx 14.67$；在 $P = (0,0)$ 時，它是 $0 + 16 + 10 =$

26，而 $3PG^2 + S = 3 \cdot \frac{34}{9} + \frac{44}{3} = 26$，與恆等式所預測的一致。

- 另一個有用的形式：$S = \frac{1}{3}\left(A_1A_2^2 + A_2A_3^2 + A_3A_1^2\right)$。在上述例子中，各邊長的平方為 16、18 和 10，同樣得出 $\frac{44}{3}$。
- 我們以分數作精確運算，在 2000 組隨機抽取的三點上檢驗了這個恆等式，並以數值方法確認了圓上的點給出的和為 $a^2$，以及該和的最小值為 $S$，在 $G$ 處取得。

## 常見錯誤

- 把半徑留作 $\sqrt{\left(a^2 - \sum(x_i^2 + y_i^2)\right)/3 + \bar{x}^2 + \bar{y}^2}$，而沒有認出它可以用 $S$ 表示；這樣一來，最後一部分便很難看得出來。
- 忘記退化情況（degenerate case）$a^2 = S$，此時軌跡是單獨一點，而不是圓。

## STEP 之外

恆等式 $\sum PA_i^2 = 3PG^2 + \sum GA_i^2$ 是轉動慣量（moment of inertia）的平行軸定理（parallel axis theorem）的一種形式；在統計學中，它就是把平方和分解為「組間」（between）與「組內」（within）兩部分：平均值正是令離差平方和最小的那個值，而這就是最小二乘法（least squares）擬合之所以行得通的根本原因。在線性代數中，它是喬裝了的畢氏定理（Pythagoras' theorem），因為由 $G$ 到 $P$ 的向量與各離差 $A_i - G$（在適當的意義下）正交（orthogonal），而這些離差之和為零。Strang 的 *Introduction to Linear Algebra*（《線性代數導論》）闡述了這種以投影（projection）看待最小二乘法的觀點。

詳細解答

# STEP I 1999，第 3 題

純數學 • 輪換方程組

**STEP I 1999，第 3 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/99-s1-q3

## 關鍵思路

每條方程都是 $x_k = f(x_{k+1})$ 的形式，其中 $f(x) = 1 + 1/x$，而整個方程組首尾相接，形成一個環。相鄰兩項之差遵從一個簡單的規律：每個差等於下一個差除以兩個大於 1 的數之積。因此沿着環走一圈，這些差只會愈來愈小。既然最後回到起點，它們必定全部為零。

以下一律把下標按循環（cyclic）方式理解：$x_{n+1}$ 即 $x_1$，$x_{n+2}$ 即 $x_2$。這樣，方程組中每條方程都可寫成同一形式

$$x_k = 1 + \frac{1}{x_{k+1}}, \qquad k = 1, 2, \ldots, n.$$

## 解答

### 第 (i) 部分

所有 $x_k$ 皆為正數，所以對每個 $k$ 都有 $1/x_{k+1} > 0$。因此

$$x_k = 1 + \frac{1}{x_{k+1}} > 1 \qquad \text{對於} k = 1, 2, \ldots, n,$$

而由於下標是循環的，這已涵蓋 $x_1, \ldots, x_n$ 中的每一個（最後一個是 $x_n = 1 + 1/x_1$）。所以 $x_1, x_2, \ldots, x_n > 1$。

### 第 (ii) 部分

把第一條方程減去第二條方程：

$$x_1 - x_2 = \left(1 + \frac{1}{x_2}\right) - \left(1 + \frac{1}{x_3}\right) = \frac{1}{x_2} - \frac{1}{x_3} = \frac{x_3 - x_2}{x_2 x_3} = -\frac{x_2 - x_3}{x_2 x_3},$$

這正是所求的結果。我們只用了 $x_1 = 1 + 1/x_2$ 和 $x_2 = 1 + 1/x_3$（以及 $n \geq 3$，使得 $x_3$ 是另一個變量）。對任意兩條相鄰的方程作完全相同的計算，便得到對每個 $k$ 都有

$$x_k - x_{k+1} = -\frac{x_{k+1} - x_{k+2}}{x_{k+1} x_{k+2}} \qquad \text{(下標按循環方式理解).}$$

### 第 (iii) 部分

記相鄰兩項之差為 $d_k = x_k - x_{k+1}$，故 $d_n = x_n - x_1$。(ii) 的一般形式就是 $d_k = -d_{k+1}/(x_{k+1}x_{k+2})$。取絕對值（absolute value），並利用第 (i) 部分所給出的 $x_{k+1}x_{k+2} > 1$，

$$|d_k| = \frac{|d_{k+1}|}{x_{k+1}x_{k+2}} \leq |d_{k+1}|,$$

而且只要 $d_{k+1} \neq 0$，不等式（inequality）就是嚴格成立的。把這些不等式沿着環逐一串連起來，

$$|d_1| \leq |d_2| \leq |d_3| \leq \cdots \leq |d_n| \leq |d_1|.$$

兩端是同一個數，所以這串不等式中每一個都是等式。特別地，$|d_1| = |d_2|$。但假如 $d_2 \neq 0$，我們便會有嚴格的 $|d_1| < |d_2|$，所以 $d_2 = 0$；於是所有 $|d_k|$ 既然都等於 $|d_2|$，也就全部為零。

> ? **想一想** 我們究竟在哪一步用到了方程組首尾相接，即 $x_n = 1 + 1/x_1$？

只在這串不等式的最後一環 $|d_n| \leq |d_1|$ 用到，它來自 $x_n = 1 + 1/x_1$。沒有這一環，這些差大可以隨 $k$ 增加而愈來愈大；事實上，一條不閉合的鏈 $x_1, \ldots, x_n$ 不一定是常數。正是因為要回到起點，所謂的縮小才被迫變成完全沒有縮小。

所以對每個 $k$ 都有 $x_k = x_{k+1}$，即 $x_1 = x_2 = \cdots = x_n$。

### $x_1$ 的值

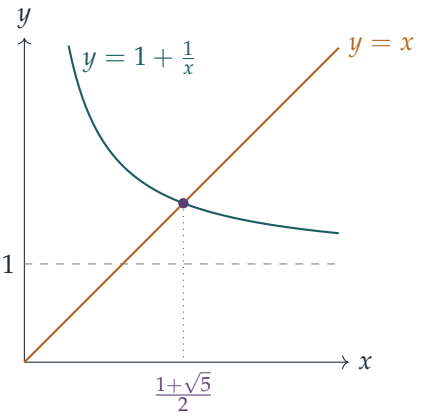
把這個共同的值稱為 $x$。現在任何一條方程都變成 $x = 1 + 1/x$，所以

$$x^2 - x - 1 = 0, \qquad x = \frac{1 \pm \sqrt{5}}{2}.$$

由第 (i) 部分知 $x > 1$，故負根須捨去，所以

$$x_1 = \frac{1 + \sqrt{5}}{2}.$$

反過來，令每個 $x_k$ 都等於這個值，確實滿足全部 $n$ 條方程，所以方程組的確有這個解，而由 (iii) 可知再沒有其他解。從幾何上看，$x$ 就是曲線 $y = 1 + 1/x$ 與直線 $y = x$ 相交之處：




## 驗算

- 取 $x = \frac{1+\sqrt{5}}{2} \approx 1.6180$，得 $1 + 1/x \approx 1 + 0.6180 = 1.6180$，符合要求。
- (ii) 中的恆等式（identity）可以用任何一條不閉合的鏈來檢驗：取 $x_3 = 2$，則 $x_2 = \frac{3}{2}$，$x_1 = \frac{5}{3}$。於是 $x_1 - x_2 = \frac{1}{6}$，而 $-(x_2 - x_3)/(x_2 x_3) = \frac{1/2}{3} = \frac{1}{6}$。
- 沿着環走完一整圈，得 $x_1 = f^n(x_1)$，其中 $f(x) = 1 + 1/x$。對 $n = 3, \ldots, 11$，在 $(0, 20)$ 上以數值方法掃描 $f^n(x) - x$，只找到一次變號，位於 $x \approx 1.618$，由此確認沒有其他解。

## 常見錯誤

- 單憑 $|d_1| \le |d_2|$ 就去證明 (iii)。一步只說明這些差不會變大；你必須沿着環走完一整圈，回到 $|d_1|$。
- 先假設各個 $x_k$ 相等，然後求 $x_1$。第 (iii) 部分才是這樣做的依據，而它需要第 (i) 部分的嚴格不等式 $x_{k+1} x_{k+2} > 1$。
- 忘記捨去 $\frac{1-\sqrt{5}}{2}$，它是負數。

## STEP 之外

這道題的核心在於：在 $x > 1$ 上，$f(x) = 1 + 1/x$ 會把點拉得更近，因為 $|f(a) - f(b)| = |a - b|/(ab)$。這樣的映射稱為壓縮映射（contraction），而巴拿赫不動點定理（Banach fixed-point theorem）指出，完備空間（complete space）上的壓縮映射恰有一個不動點，而且從任何起點開始迭代（iteration）都會收斂（converge）到這個不動點。在這裏從 $x = 1$ 開始迭代，得到的是相鄰斐波那契數（Fibonacci numbers）之比 $1, 2, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \ldots$，以及黃金比（golden ratio）的連分數（continued fraction）$1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}$。分析方面，可讀 Abbott 的 *Understanding Analysis*。連分數方面，可讀 Davenport 的 *The Higher Arithmetic*。

詳細解答

# STEP I 1999，第 4 題

純數學 • 含絕對值的區域

> **STEP I 1999，第 4 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/99-s1-q4

## 關鍵思路

絕對值（modulus）符號把平面分成若干區域，在每個區域內各括號的正負號固定，於是條件就變成普通的線性不等式（最後一部分則是雙曲線型的不等式）。另一個大大節省時間的方法是利用對稱（symmetry）和平移（translation）：$|x|$ 和 $|y|$ 令集合對兩條坐標軸都對稱，所以只需在一個象限內計算，然後作反射；而把 $x$ 換成 $x-1$，只是把整個圖形向右平移一個單位。

## 解答

### 第 (i) 部分

把 $x$ 換成 $-x$ 或把 $y$ 換成 $-y$，集合都不變，所以它對兩條坐標軸都對稱。在第一象限（$x \geq 0$，$y \geq 0$）內，條件就是 $x+y \leq 1$：即頂點為 $(0,0)$、$(1,0)$、$(0,1)$ 的三角形。對兩條坐標軸作反射，便得到四個三角形，合起來是一個以一角豎立的正方形，頂點為

$$(1,0), \quad (0,1), \quad (-1,0), \quad (0,-1).$$

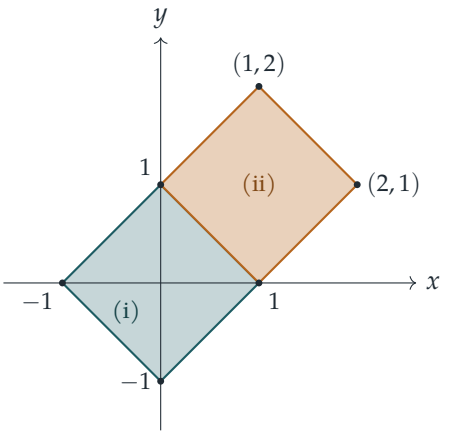
換句話說，它是四個半平面（half-plane）$\pm x \pm y \leq 1$ 的交集。邊界包括在內（不等式是 $\leq$），面積為 2。

### 第 (ii) 部分

設 $X = x-1$，$Y = y-1$。條件於是變成 $|X|+|Y| \leq 1$，也就是 $(X,Y)$ 坐標下的第 (i) 部分。因此這個集合就是把第 (i) 部分的正方形平移 $(1,1)$ 所得：中心為 $(1,1)$，頂點為

$$(2,1), \quad (1,2), \quad (0,1), \quad (1,0),$$

由直線 $x+y=1$、$x+y=3$、$y=x+1$ 和 $y=x-1$ 圍成。它與兩條坐標軸相接，接點分別是 $(1,0)$ 和 $(0,1)$。




**第 (iii) 部分**

同樣先作平移：設 $X=x-1$，$Y=y+1$，條件變成

$$|X|\le 1+|Y|.$$

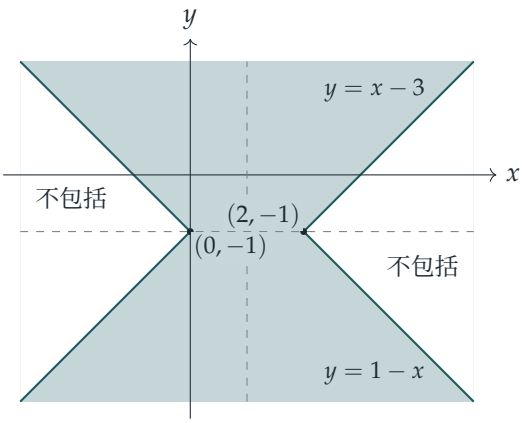
這個條件對 $X$ 軸和 $Y$ 軸都對稱，即對直線 $x=1$ 和 $y=-1$ 對稱。在象限 $X\ge 0$、$Y\ge 0$ 內，條件是 $X\le 1+Y$：即直線 $X=1+Y$ 左方的所有點，這條直線通過 $(1,0)$，斜率為 1。作反射之後，這個集合包含平面上所有的點，只除去兩個開的楔形區域：

- 右邊是 $X>1+|Y|$：一個直角楔形，頂點為 $(X,Y)=(1,0)$，向右張開，邊界為 $Y=\pm(X-1)$；
- 左邊是 $X<-1-|Y|$：它的鏡像，頂點為 $(-1,0)$，向左張開。

**想一想** 這個集合是否像 (i) 和 (ii) 那樣有界（bounded）？

不是。那個負號改變了一切：$|Y|$ 越大，條件反而越容易滿足，所以對每個固定的 $Y$，容許的 $X$ 構成區間 $-1-|Y|\le X\le 1+|Y|$，而這個區間隨 $|Y|$ 增大而變闊。特別地，整條垂直線 $X=0$（即 $x=1$）都在集合之內。

回到 $x,y$ 坐標，兩個頂點是 $(2,-1)$ 和 $(0,-1)$。被排除的右楔形是 $x>2+|y+1|$，邊界是 $x\ge 2$ 時的兩條半直線 $y=x-3$ 和 $y=1-x$；被排除的左楔形是 $x<-|y+1|$，邊界是 $x\le 0$ 時的 $y=x-1$ 和 $y=-1-x$。邊界線本身屬於這個集合。在直線 $y=-1$ 上，集合只是線段 $0\le x\le 2$，那是它最窄的地方。




虛線 $x = 1$ 和 $y = -1$ 是對稱軸。

**第 (iv) 部分**

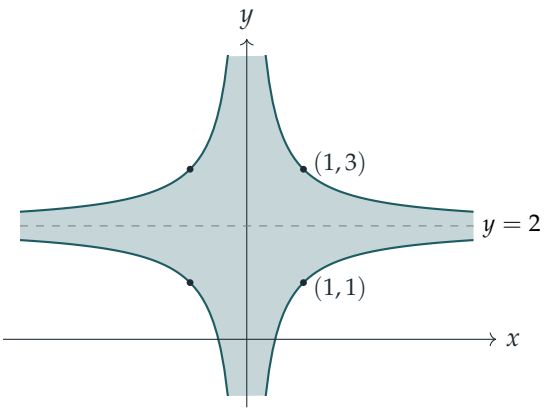
設 $Y = y - 2$，條件變成 $|x|\,|Y| \le 1$，它對 $x$ 軸方向和直線 $Y = 0$（即 $y = 2$）都對稱。在象限 $x \ge 0$、$Y \ge 0$ 內：

- 若 $x = 0$，條件對任何 $Y$ 都成立，所以整條 $y$ 軸都在集合之內；
- 若 $x > 0$，我們可以除以 $x$，條件變成 $0 \le Y \le 1/x$：即雙曲線（hyperbola）$Y = 1/x$ 下方的區域。

對兩條對稱軸作反射，這個集合就是

$$x = 0, \quad \text{或} \quad 2 - \frac{1}{|x|} \le y \le 2 + \frac{1}{|x|} \ \ (x \ne 0),$$

即夾在雙曲線 $(y-2)x = \pm 1$ 四條分支之間的區域。它包含整條 $y$ 軸和整條直線 $y = 2$，而這兩條直線正是邊界曲線的漸近線（asymptote）。它看起來像一個以 $(0, 2)$ 為中心的粗身加號：「正方形」$|x| \le 1$、$|y-2| \le 1$ 的四個角落在邊界上，位於 $(\pm 1, 1)$ 和 $(\pm 1, 3)$；當 $|x| \to \infty$ 時，橫臂向 $y = 2$ 收窄；當 $|y| \to \infty$ 時，直臂向 $y$ 軸收窄。這個集合是無界的（不過有趣的是，它的面積是無限大）。




## 驗算

- 取幾個點代入原來的不等式檢驗。在 (iii) 中，$(1, 100)$ 給出 $0 - 101 \leq 1$，所以它在集合之內（證實集合是無界的），而 $(2.01, -1)$ 給出 $1.01 > 1$，剛好在右邊頂點之外。在 (iv) 中，$(2, 2.5)$ 給出 $2 \times 0.5 = 1$，在邊界上，而 $(2, 2.51)$ 則剛好在外面。
- 在 (i) 中，四個三角形每個面積為 $\frac{1}{2}$，所以正方形的面積是 2；用格點數算得到 2.002。
- 我們為每一部分在 $[-6, 6]^2$ 內測試了 400,000 個隨機點，把原來的不等式與我們的描述（半平面、被排除的楔形，或 $|y - 2| \leq 1/|x|$）互相比較；結果完全一致。

## 常見錯誤

- 在 (ii) 和 (iii) 中，平移方向弄錯：$|x - 1|$ 把圖形移到 $x = +1$，而 $|y + 1|$ 把圖形移到 $y = -1$。
- 在 (iii) 中，仿照 (i) 畫出一個有界的菱形。負號把 $|y + 1|$ 的作用倒轉了，集合其實是無界的。
- 在 (iv) 中，沒有另外處理 $x = 0$ 就除以 $|x|$，因而遺漏了整條 $y$ 軸都屬於這個集合這一事實。

## STEP 之外

(i) 中的集合是「計程車」範數（taxicab norm）$\|(x, y)\|_1 = |x| + |y|$ 的單位球（unit ball），而 (ii) 則是以 $(1, 1)$ 為中心、半徑為 1 的球。不同的範數給出形狀不同的單位球（通常的距離給出圓形，$\max(|x|, |y|)$ 給出邊為水平的正方形），而分析或度量空間（metric space）

的入門課程會證明，在 $\mathbb{R}^2$ 上它們所定義的收斂（convergence）和連續（continuity）概念完全相同。Bartle 與 Sherbert 合著的 *Introduction to Real Analysis*（實分析入門）是認識這些概念的好起點。

詳細解答

# STEP I 1999，第 5 題

純數學 • 小角近似

**STEP I 1999，第 5 題**

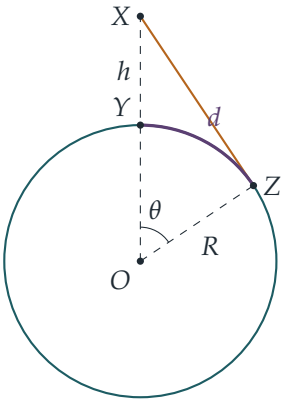
繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：

step.maths.org/questions/99-s1-q5

## 關鍵思路

「剛好在地平線上看見」的意思是，由 $Z$ 望向 $X$ 的視線剛好擦過地球表面，所以這條視線是地球在 $Z$ 處的切線（tangent），並與半徑 $OZ$ 垂直。這樣便得到一個以地心 $O$ 為頂點之一的直角三角形，而弧（arc）$YZ$ 的長度等於 $R$ 乘以 $O$ 處的角。至於那兩個近似式（approximation），則分別來自考慮這個角很小的情況，以及這個角接近直角的情況。

## 解答




### 第 (i) 部分

設 $O$ 為地球的中心，並設 $\theta = \angle YOZ$。由於 $X$ 位於 $Y$ 的正上方，$O$、$Y$、$X$ 三點共線（collinear），且 $OX = R + h$。由於從 $Z$ 剛好可以看見 $X$，直線 $ZX$ 與地球相切於 $Z$，所

以它與半徑 $OZ$ 垂直。在直角三角形 $OZX$ 中，

$$\cos\theta = \frac{OZ}{OX} = \frac{R}{R+h}.$$

在半徑為 $R$ 的圓上，對應圓心角 $\theta$ 的弧 $YZ$ 的長度為 $R\theta$，所以

$$d = R\arccos\left(\frac{R}{R+h}\right).$$

這裏 $0 < \theta < \frac{\pi}{2}$，因為 $0 < R/(R+h) < 1$。

**第 (ii) 部分**

當 $h$ 相對於 $R$ 很小時，$\cos\theta = R/(R+h)$ 接近 1，所以 $\theta$ 很小，我們可以使用 $\cos\theta \approx 1 - \frac{1}{2}\theta^2$。此外，

$$\frac{R}{R+h} = 1 - \frac{h}{R+h} \approx 1 - \frac{h}{R},$$

因為 $R + h \approx R$。令兩者相等，

$$1 - \tfrac{1}{2}\theta^2 \approx 1 - \frac{h}{R} \quad\Longrightarrow\quad \theta^2 \approx \frac{2h}{R} \quad\Longrightarrow\quad \theta \approx \sqrt{\frac{2h}{R}},$$

其中因 $\theta > 0$ 而取正根。因此

$$d = R\theta \approx R\sqrt{\frac{2h}{R}} = \sqrt{2}\,(Rh)^{1/2},$$

這正是所要求的形式，其中 $k = \sqrt{2}$。

**第 (iii) 部分**

現在 $R/(R+h)$ 很小，所以 $\cos\theta$ 接近 0，而 $\theta$ 接近 $\frac{\pi}{2}$，並不是小角。題目給出的近似式只適用於小角，所以我們改為量度那個*確實*很小的角。

> ? **想一想** 這裏哪一個角是小角？哪一個近似式適用於它？

記 $\theta = \frac{\pi}{2} - \phi$，其中 $\phi = \angle OXZ$ 是衛星處的角，當衛星離得很遠時，這個角很小。於是 $\cos\theta = \sin\phi$，所以

$$\sin\phi = \frac{R}{R+h}.$$

我們需要把這個式子準確至 $R/h$ 的二階，因為答案中有一項 $R^2/h$。由二項式展開（binomial expansion），因 $R/h < 1$ 而成立，

$$\frac{R}{R+h} = \frac{R}{h}\left(1 + \frac{R}{h}\right)^{-1} = \frac{R}{h} - \frac{R^2}{h^2} + \cdots.$$

利用 $\sin\phi \approx \phi$，

$$\phi \approx \frac{R}{h} - \frac{R^2}{h^2}.$$

因此

$$\begin{aligned} d = R\theta &= R\left(\frac{\pi}{2} - \phi\right) \\ &\approx \frac{\pi}{2}R - \frac{R^2}{h} + \frac{R^3}{h^2}. \end{aligned}$$

最後一項比 $R^2/h$ 還要再小一個因子 $R/h$，所以我們把它捨去，得到

$$d \approx \frac{\pi}{2}R - \frac{R^2}{h},$$

這正是所要求的形式，其中 $a = \frac{\pi}{2}$ 及 $b = -1$。

在這個階數下使用 $\sin\phi \approx \phi$ 是穩妥的：它的誤差約為 $\frac{1}{6}\phi^3$，屬於 $R^3/h^3$ 的階，所以只會在 $R^4/h^3$ 這一階才影響 $d$。

## 驗算

- 極限情況合理。當 $h \to 0$ 時，$d \to 0$。當 $h \to \infty$ 時，$d \to \frac{\pi}{2}R$，即繞地球四分之一圈：一顆非常遙遠的衛星幾乎可以看見整個半球。$-R^2/h$ 這一項是負數，所以 $d$ 從下方趨近 $\frac{\pi}{2}R$，這是必然的，因為 $\theta < \frac{\pi}{2}$。
- 數值上，取 $R = 1$ 及 $h = 10^{-4}$，$d$ 的準確值除以 $\sqrt{2Rh}$ 得 0.99996。取 $h = 1$ 及 $R = 0.01$，$d$ 的準確值與 $\frac{\pi}{2}R - R^2/h$ 相差 $9.9 \times 10^{-7}$，約為 $R^3/h^2$，與預測相符。
- 若衛星位於半徑為 6370 km 的地球上空 300 km 處，$d$ 的準確值約為 1918 km，而 $\sqrt{2Rh}$ 給出 1955 km。

## 常見錯誤

- 在第 (iii) 部分使用 $\cos\theta \approx 1 - \frac{1}{2}\theta^2$，但那裏 $\theta$ 接近 $\frac{\pi}{2}$，並不是小角。
- 在第 (iii) 部分只把 $R/(R+h)$ 展開至一階，結果失去了 $R^2/h$ 這一項，得出錯誤的 $b$。

## STEP 之外

這兩部分都是漸近展開（asymptotic expansion）：把 $d$ 的準確值按一個小參數（small parameter）$h/R$ 或 $R/h$ 的冪展開，保留至指定階數的各項，同時估計誤差的上界。利

用大 $O$ 記號和帶餘項的泰勒定理（Taylor's theorem with remainder）把「≈」說得精確，是大學一年級數學分析的課題；可參閱 Abbott 的 *Understanding Analysis*（數學分析入門教材）。Siklos 的 *Advanced Problems in Mathematics*（進階數學難題集）中有更多這類近似法的 STEP 練習。

詳細解答

# STEP I 1999，第 6 題

純數學 • 二次函數的極值

> **STEP I 1999，第 6 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/99-s1-q6

## 關鍵思路

直線在閉區間（closed interval）上的極值必定出現在端點，而哪一端取最大值、哪一端取最小值，只取決於斜率（gradient）的正負。對於 $c > 0$ 的二次函數（quadratic），圖像是一條 $\cup$ 形的拋物線（parabola）：最大值仍然在端點取得，但最小值則當頂點位於區間內時在頂點（vertex）取得，否則在較近的一端取得。因此，整道題歸根究底就是要問：頂點 $x = -b/(2c)$ 相對於 $[-10, 10]$ 位於何處。

## 解答

### 第 (i) 部分

設 $f(x) = bx + a$，則 $f(10) = a + 10b$ 及 $f(-10) = a - 10b$。

- 若 $b > 0$，$f$ 嚴格遞增（strictly increasing），故最小值為 $f(-10) = a - 10b$（在 $x = -10$ 取得），最大值為 $f(10) = a + 10b$（在 $x = 10$ 取得）。
- 若 $b = 0$，則對所有 $x$ 均有 $f(x) = a$，故最大值和最小值都是 $a$。
- 若 $b < 0$，$f$ 嚴格遞減（strictly decreasing），故最大值為 $f(-10) = a - 10b$（在 $x = -10$ 取得），最小值為 $f(10) = a + 10b$（在 $x = 10$ 取得）。

三種情況可以歸納為

$$\text{最大值} = a + 10|b|, \qquad \text{最小值} = a - 10|b|.$$

**第 (ii) 部分**

設 $g(x) = cx^2 + bx + a$，其中 $c \geq 0$。若 $c = 0$，這就是第 (i) 部分，答案如上所述為 $a \pm 10|b|$。以下取 $c > 0$，並配方（complete the square）：

$$g(x) = c\left(x + \frac{b}{2c}\right)^2 + a - \frac{b^2}{4c}.$$

圖像是一條 $\cup$ 形的拋物線，頂點在 $x_0 = -\dfrac{b}{2c}$，$g$ 在該處取得所有實數 $x$ 之中的最小值。此外，$g$ 在 $x < x_0$ 時遞減，在 $x > x_0$ 時遞增。兩個端點的值為

$$g(10) = a + 100c + 10b, \qquad g(-10) = a + 100c - 10b.$$

**最大值**。

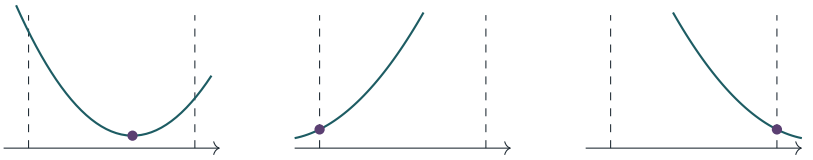
> **想一想** 為甚麼無論頂點在哪裏，$g$ 在這個區間上的最大值都必定在其中一個端點取得？

由於 $c > 0$，$g(x)$ 只透過距離 $|x - x_0|$ 而取決於 $x$，並且隨這個距離增大而增大。在 $[-10, 10]$ 上，距離 $|x - x_0|$ 在其中一個端點最大（即離 $x_0$ 較遠的那一端）。所以最大值是 $g(10)$ 與 $g(-10)$ 之中較大的一個，而兩者之差為 $g(10) - g(-10) = 20b$。因此：

- 若 $b > 0$，最大值為 $a + 100c + 10b$，在 $x = 10$ 取得；
- 若 $b < 0$，最大值為 $a + 100c - 10b$，在 $x = -10$ 取得；
- 若 $b = 0$，最大值為 $a + 100c$，在兩個端點同時取得。

在所有情況下，最大值 $= a + 100c + 10|b|$。這個結果在 $c = 0$ 時同樣成立。

**最小值**。 這時 $x_0$ 的位置就至關重要了。頂點位於區間內，當且僅當 $-10 \leq -\frac{b}{2c} \leq 10$，即（各項乘以 $2c > 0$）$-20c \leq b \leq 20c$，亦即 $|b| \leq 20c$。




虛線為 $x = \pm 10$，圓點標示函數在區間上的最小值。

- $|b| \leq 20c$（**頂點在區間內**）。最小值就是全局最小值（global minimum），

$$a - \frac{b^2}{4c}, \quad \text{於} x = -\frac{b}{2c}.$$

- $b > 20c$（**頂點在左方**，$x_0 < -10$）。在 $[-10, 10]$ 上有 $x > x_0$，故 $g$ 在該處遞增，最小值為 $g(-10) = a + 100c - 10b$，在 $x = -10$ 取得。
- $b < -20c$（**頂點在右方**，$x_0 > 10$）。這時 $g$ 在整個區間上遞減，故最小值為 $g(10) = a + 100c + 10b$，在 $x = 10$ 取得。

頂點在區間外的兩種情況可以合併為：當 $|b| > 20c$ 時，最小值 $= a + 100c - 10|b|$。

> ? **想一想** 在邊界 $|b| = 20c$ 上會怎樣？兩條公式在那裏會否互相矛盾？

不會。若 $b = 20c$，則 $a - \frac{b^2}{4c} = a - 100c$，而 $a + 100c - 10b = a - 100c$ 亦然，因為這時頂點恰好在 $x = -10$。$b = -20c$ 的情況同理。所以當 $b$ 越過 $\pm 20c$ 時，答案是連續地變化的，邊界歸入哪一種情況都沒有分別。

**總結**。 對於 $-10 \le x \le 10$ 及 $c \ge 0$：

$$\text{最大值} = a + 100c + 10|b|, \qquad \text{最小值} = \begin{cases} a - \dfrac{b^2}{4c}, & c > 0 \text{ 及} |b| \le 20c, \\ a + 100c - 10|b|, & |b| > 20c, \\ a, & b = c = 0. \end{cases}$$

$c = 0$、$b \neq 0$ 的情況屬於 $|b| > 20c = 0$，所得結果正是第 (i) 部分的答案。

## 驗算

- 取 $g(x) = x^2 + 4x + 1$：這裏 $|b| = 4 \le 20$，故最小值為 $1 - 4 = -3$，在 $x = -2$ 取得，而最大值為 $1 + 100 + 40 = 141$，在 $x = 10$ 取得。直接代入計算，結果相符。
- 取 $g(x) = x^2 + 30x$：這裏 $b = 30 > 20$，故頂點 $x = -15$ 在區間之外，最小值為 $g(-10) = 100 - 300 = -200$。事實上 $g(-10) = -200$，而且 $g$ 在 $[-10, 10]$ 上遞增。
- 另外用電腦作了簡短的檢驗：對 1000 組隨機的 $(a, b, c)$（包括 $c = 0$ 和邊界 $|b| = 20c$ 的情況），以精確的有理數運算，把公式與在 $[-10, 10]$ 內 401 個點上逐點搜尋的結果比較。所有情況均相符。

## 常見錯誤

- 沒有檢查頂點是否位於 $[-10, 10]$ 內，就直接把 $a - \frac{b^2}{4c}$ 當作最小值。
- 遺漏了 $c = 0$（沒有頂點，而且不能除以 $c$）和 $b = c = 0$ 這兩種情況。

▶ 頂點在區間外時，弄錯哪一端取最小值：應該是較接近頂點的那一端。

## STEP 之外

這道題背後的事實是極值定理（Extreme Value Theorem）：在有界閉區間上的連續函數（continuous function）必定取得最大值和最小值，而這些值不是在端點取得，就是在區間內的駐點（stationary point）取得。像 $cx^2 + bx + a$（其中 $c \geq 0$）這樣的凸函數（convex function）在端點取得最大值，這個觀察是凸優化（convex optimisation）的第一步，而凸優化正是現代應用數學許多範疇的基礎。Abbott 的 *Understanding Analysis* 對極值定理有嚴謹的證明。

詳細解答

# STEP I 1999，第 7 題

純數學 • 二階微分方程

> **STEP I 1999，第 7 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/99-s1-q7

## 關鍵思路

函數 $\sin(k\sin^{-1}x)$ 看來很難求兩次導數，但只要設 $x=\sin\theta$，使 $y=\sin k\theta$，它就變得易於處理。先把一階導數（first derivative）乘以 $\sqrt{1-x^2}$ 再求導，可令代數運算保持簡短。至於最後一部分，重點在於：一個二階微分方程（second-order differential equation）若在某一點給定了函數值和斜率，就只有一個解，所以我們求得的多項式必定是 $\sin(3\sin^{-1}x)$。

## 解答

### 微分方程

設 $y=\sin(k\sin^{-1}x)$，其中 $-1<x<1$。根據鏈式法則（chain rule），並利用 $\dfrac{\mathrm{d}}{\mathrm{d}x}\sin^{-1}x=\dfrac{1}{\sqrt{1-x^2}}$，

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{k\cos(k\sin^{-1}x)}{\sqrt{1-x^2}},\qquad\text{所以}\qquad\sqrt{1-x^2}\,\frac{\mathrm{d}y}{\mathrm{d}x}=k\cos(k\sin^{-1}x).$$

把第二個式子對 $x$ 求導（左邊用積法則（product rule），右邊用鏈式法則）：

$$\sqrt{1-x^2}\,\frac{\mathrm{d}^2y}{\mathrm{d}x^2}-\frac{x}{\sqrt{1-x^2}}\,\frac{\mathrm{d}y}{\mathrm{d}x}=-\frac{k^2\sin(k\sin^{-1}x)}{\sqrt{1-x^2}}=-\frac{k^2y}{\sqrt{1-x^2}}.$$

全式乘以 $\sqrt{1-x^2}$（當 $-1<x<1$ 時它不等於零），得

$$(1-x^2)\frac{\mathrm{d}^2y}{\mathrm{d}x^2}-x\frac{\mathrm{d}y}{\mathrm{d}x}+k^2y=0,$$

這正是所要證明的 $(*)$。

### $k = 3$ 時的三次多項式解

設 $y = Ax^3 + Bx^2 + Cx + D$。由條件「當 $x = 0$ 時 $y = 0$」得 $D = 0$，而由「當 $x = 0$ 時 $\frac{\mathrm{d}y}{\mathrm{d}x} = 3$」得 $C = 3$。現在

$$y' = 3Ax^2 + 2Bx + C, \qquad y'' = 6Ax + 2B,$$

代入 $(*)$，其中 $k^2 = 9$：

$$(1 - x^2)(6Ax + 2B) - x(3Ax^2 + 2Bx + C) + 9(Ax^3 + Bx^2 + Cx + D) = 0.$$

此式須對所有 $x$ 成立，所以我們比較係數（compare coefficients）：

$$\begin{aligned}
x^3: &\quad -6A - 3A + 9A = 0 \text{ （對任何 } A \text{ 都成立），}\\
x^2: &\quad -2B - 2B + 9B = 5B = 0,\\
x^1: &\quad 6A - C + 9C = 6A + 8C = 0,\\
x^0: &\quad 2B + 9D = 0.
\end{aligned}$$

因此 $B = 0$（與 $D = 0$ 一致），而 $6A = -8C = -24$，得 $A = -4$。所求的解是

$$y = 3x - 4x^3.$$

### $\sin 3\theta$ 的表達式

> **想一想** 同一個方程，我們有兩個解：$\sin(3\sin^{-1} x)$ 和 $3x - 4x^3$。為甚麼它們必定是同一個函數？

兩者都滿足 $k = 3$ 時的 $(*)$，而且在 $x = 0$ 處滿足相同的條件。對於多項式，這是我們在構造時已經定下的。對於 $y = \sin(3\sin^{-1} x)$，我們有 $y(0) = \sin 0 = 0$，並且由上面的一階導數得 $y'(0) = 3\cos 0/1 = 3$。

在 $-1 < x < 1$ 上，我們可以把 $(*)$ 除以 $1 - x^2 \neq 0$。這樣得到 $y'' = \dfrac{xy' - 9y}{1 - x^2}$，是一個係數連續（continuous）的二階線性方程，而這類方程在 $x = 0$ 處給定 $y$ 和 $y'$ 的值後，恰好只有一個解。（可以想像由 $x = 0$ 出發逐步向外走：方程會根據 $y$ 和 $y'$ 告訴你 $y''$，所以初始值決定了一切。）因此

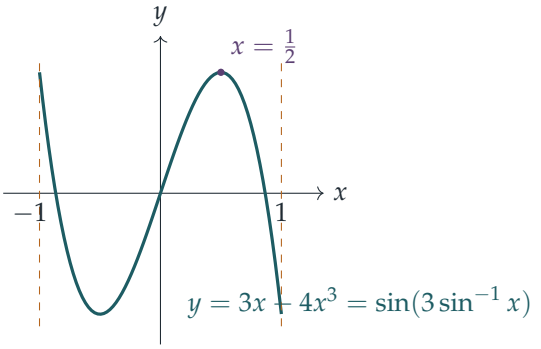
$$\sin(3\sin^{-1} x) = 3x - 4x^3 \qquad \text{對於 } -1 < x < 1,$$

而根據連續性，在 $x = \pm 1$ 處同樣成立。

現在設 $x = \sin\theta$，其中 $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$，於是 $\sin^{-1} x = \theta$：

$$\boxed{\sin 3\theta = 3\sin\theta - 4\sin^3\theta.}$$

這個恆等式（identity）對所有 $\theta$ 都成立，而不僅限於 $|\theta| \le \frac{\pi}{2}$。兩邊的周期（period）都是 $2\pi$。若 $\frac{\pi}{2} \le \theta \le \frac{3\pi}{2}$，設 $\varphi = \pi - \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$。那麼 $\sin\varphi = \sin\theta$，且 $\sin 3\varphi = \sin(3\pi - 3\theta) = \sin 3\theta$，所以由 $\varphi$ 的恆等式即可得出 $\theta$ 的恆等式。




這個圖像就是大家熟悉的曲線 $\sin 3\theta$ 透過 $x = \sin\theta$ 所見的樣子：當 $\theta$ 由 $-\frac{\pi}{2}$ 變到 $\frac{\pi}{2}$ 時，$3\theta$ 走過一個半振盪，因此在 $x = \pm\frac{1}{2}$ 處（即 $3\theta = \pm\frac{\pi}{2}$）有兩個轉向點（turning point）。

## 驗算

- $\theta = \frac{\pi}{2}$：$\sin\frac{3\pi}{2} = -1$，而 $3 - 4 = -1$。　$\theta = \frac{\pi}{6}$：$\sin\frac{\pi}{2} = 1$，而 $\frac{3}{2} - \frac{1}{2} = 1$。
- 數值上，當 $\theta = 0.7$：$\sin 2.1 = 0.86321$，而 $3(0.64422) - 4(0.64422)^3 = 0.86321$。
- 同一個恆等式也可以由 $\sin(2\theta + \theta)$ 和二倍角公式（double-angle formulae）推出，這是一個很好的獨立驗算。

## 常見錯誤

- 直接用商法則（quotient rule）對 $\dfrac{k\cos(k\sin^{-1}x)}{\sqrt{1-x^2}}$ 求導：做法正確，但十分繁複。先消去平方根會簡潔得多。
- 忘記說明爲甚麼那個三次多項式等於 $\sin(3\sin^{-1}x)$。單是滿足同一個方程並不足夠：兩個解還必須有相同的 $y(0)$ 和 $y'(0)$。

## STEP 之外

方程 $(*)$ 是切比雪夫（Chebyshev）方程。當 $k = n$ 是整數時，它有一個多項式解，而 $\cos(n\cos^{-1}x)$ 就是切比雪夫多項式（Chebyshev polynomial）$T_n(x)$，在數值逼近（numerical approximation）中佔有核心地位。我們用到的唯一性定理（uniqueness theorem，

即 Picard–Lindelöf 定理）會在大學一年級的微分方程課程中證明。處理這類方程的冪級數（power series）方法，可參閱 Boyce and DiPrima, *Elementary Differential Equations*（一本微分方程入門教科書）。

詳細解答

# STEP I 1999，第 8 題

純數學 • 調和級數的界限

**STEP I 1999，第 8 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/99-s1-q8

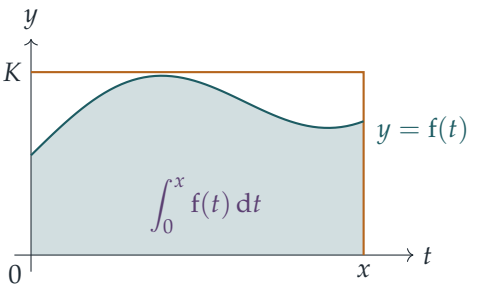
## 關鍵思路

積分就是面積，所以把曲線困在一個長方形之內，它的面積便被夾在 0 與長方形的面積之間。題目建議的被積函數（integrand）可用部分分式（partial fractions）拆開，其積分恰好等於 $\ln\left(\frac{n}{n-1}\right) - \frac{1}{n}$，而被積函數的上下界便給出所求的不等式。之後一切都會裂項相消（telescope）：對 $n$ 求和，一連串的對數化為 $\ln N$，一連串的倒數之差則幾乎全部消去。

## 解答

### 估計積分的上下界

繪出 $y = \mathrm{f}(t)$ 在 $0 \le t \le x$ 的圖像。由於 $0 \le \mathrm{f}(t) \le K$，整條曲線位於以 $(0,0)$ 和 $(x,K)$ 為頂點的長方形之內，並且在 $t$ 軸上或其上方。




積分就是曲線下方陰影部分的面積。這面積不可能是負數，因為曲線從不跌到軸的下

方；它也不可能超過長方形的面積 $Kx$，因為該區域完全位於長方形之內。所以

$$0 \le \int_0^x \mathrm{f}(t)\,\mathrm{d}t \le Kx.$$

**關於 $\ln\left(\frac{n}{n-1}\right)$ 的不等式**

取 $n > 1$。當 $0 \le t \le 1$ 時，$n - t \ge n - 1 > 0$，所以被積函數有定義。把它拆成部分分式，

$$\frac{t}{n(n-t)} = \frac{1}{n} \cdot \frac{n-(n-t)}{n-t} = \frac{1}{n-t} - \frac{1}{n},$$

所以

$$\int_0^1 \frac{t}{n(n-t)}\,\mathrm{d}t = \Big[-\ln(n-t) - \frac{t}{n}\Big]_0^1 = \ln n - \ln(n-1) - \frac{1}{n} = \ln\Big(\frac{n}{n-1}\Big) - \frac{1}{n}.$$

現在估計被積函數在 $[0,1]$ 上的上下界。分子 $t$ 非負，分母 $n(n-t)$ 為正，所以被積函數至少為 0。當 $t$ 由 0 增加至 1 時，分子增大而正的分母減小，所以被積函數是遞增的，其最大值在 $t = 1$ 處取得：

$$0 \le \frac{t}{n(n-t)} \le \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}.$$

在第一個結果中取 $x = 1$ 及 $K = \frac{1}{n-1} - \frac{1}{n}$，得

$$0 \le \ln\Big(\frac{n}{n-1}\Big) - \frac{1}{n} \le \frac{1}{n-1} - \frac{1}{n},$$

這正是所求。

**由 $n = 2$ 求和至 $N$**

把 $n = 2, 3, \ldots, N$（其中 $N \ge 2$）的不等式相加。對數部分會裂項相消，因為

$$\sum_{n=2}^{N} \ln\Big(\frac{n}{n-1}\Big) = \ln\Big(\frac{2}{1} \cdot \frac{3}{2} \cdots \frac{N}{N-1}\Big) = \ln N,$$

右邊各項亦然：

$$\sum_{n=2}^{N}\Big(\frac{1}{n-1} - \frac{1}{n}\Big) = \Big(1 - \frac{1}{2}\Big) + \Big(\frac{1}{2} - \frac{1}{3}\Big) + \cdots + \Big(\frac{1}{N-1} - \frac{1}{N}\Big) = 1 - \frac{1}{N}.$$

因此

$$0 \le \ln N - \sum_{n=2}^{N} \frac{1}{n} \le 1 - \frac{1}{N} \le 1.$$

（當 $N = 1$ 時，和式為空，而 $\ln 1 = 0$，所以命題此時同樣成立。）

**想一想** 這個雙重不等式的哪一半告訴我們調和級數（harmonic series）發散（diverge），哪一半給出上界？

左半說明該和至多為 $\ln N$；右半說明它至少為 $\ln N - 1$。要證明發散，我們需要下界，所以用右半。

### 發散

由 $\ln N - \sum_{n=2}^{N} \frac{1}{n} \le 1$ 可得 $\sum_{n=2}^{N} \frac{1}{n} \ge \ln N - 1$，再加上 $n = 1$ 的一項，

$$\sum_{n=1}^{N} \frac{1}{n} \ge \ln N.$$

由於當 $N \to \infty$ 時 $\ln N \to \infty$，該和亦趨向無窮大。

### $N < 10^{30}$ 時的上界

現在用左半：$\sum_{n=2}^{N} \frac{1}{n} \le \ln N$，所以

$$\sum_{n=1}^{N} \frac{1}{n} \le 1 + \ln N.$$

由於 $2^{10} = 1024 > 1000 = 10^3$，兩邊取十次方得 $10^{30} < 2^{100}$。所以若 $N < 10^{30}$，則 $N < 2^{100}$，而 ln 是遞增函數，故

$$\ln N < 100 \ln 2 < 100,$$

這裏用到 $\ln 2 < 1$，它成立是因為 $2 < \mathrm{e}$。因此

$$\sum_{n=1}^{N} \frac{1}{n} \le 1 + \ln N < 101.$$

## 驗算

- 對多個 $n > 1$ 的值，以辛普森法則（Simpson's rule）計算 $\int_0^1 \frac{t}{n(n-t)}\,\mathrm{d}t$，結果都與 $\ln\left(\frac{n}{n-1}\right) - \frac{1}{n}$ 吻合。例如當 $n = 2$ 時，兩者都給出 $\ln 2 - \frac{1}{2} \approx 0.193$，它介乎 0 與 $\frac{1}{2}$ 之間。
- 對每個不超過 $2 \times 10^5$ 的 $N$，$\ln N - \sum_{n=2}^{N} \frac{1}{n}$ 都落在 $[0, 1]$ 內。當 $N = 2 \times 10^5$ 時，它約為 0.4228，接近 $1 - \gamma \approx 0.4228$，其中 $\gamma$ 是歐拉常數（Euler's constant）。
- 這個上界相當寬鬆：$1 + 30 \ln 10 \approx 70.1$，而事實上 $\sum_{n=1}^{10^{30}} \frac{1}{n} \approx 69.65$。調和級數確實發散，但發散得極其緩慢。

## 常見錯誤

- 用錯雙重不等式的一半：證明發散需要該和的下界，而最後一部分需要上界。
- 遺漏 $n = 1$ 的一項：不等式所涉及的是 $\sum_{n=2}^{N}$，而結論所涉及的是 $\sum_{n=1}^{N}$。
- 沒有說明為何被積函數的最大值在 $t = 1$ 處取得。只需指出它是遞增的（或直接指出 $t \leq 1$ 及 $n - t \geq n - 1$）便已足夠。

## STEP 之外

$\sum_{n=1}^{N} \frac{1}{n} - \ln N$ 這個量是遞減的，並且有下界，所以它收斂（converge）。它的極限是歐拉常數 $\gamma \approx 0.5772$，而至今仍無人知道 $\gamma$ 是否有理數。像我們在這裏所做的那樣，把和式與積分作比較，就是積分判別法（integral test），它是大學一年級分析課程中處理級數的首批工具之一。Abbott 的 *Understanding Analysis*（《理解分析》）是了解其嚴格論證的好去處。

詳細解答

# STEP I 1999，第 9 題

力學 • 勻速率運動學

**STEP I 1999，第 9 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：

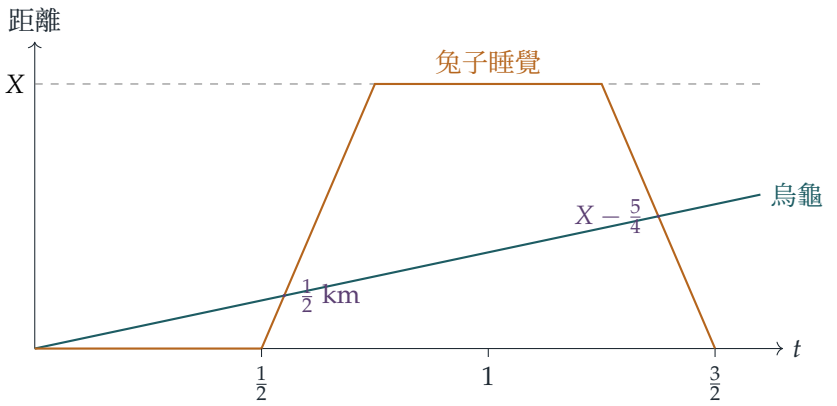
step.maths.org/questions/99-s1-q9

## 關鍵思路

所有運動都以恆定速率進行，所以每段路程所需的時間就是距離除以速率。兩隻動物每次相遇時，都是在同一時刻身處同一位置，因此我們分別寫出牠們各自在該刻的時鐘時間，再令兩者相等。兩次相遇給出兩條關於 $v$ 和 $V$ 的方程。改用 $1/v$ 和 $1/V$（每公里所需的小時數）來運算，方程便保持線性（linear）。

## 解答

我們把兩隻動物都看成沿同一條直線路徑以恆定速率（constant speed）運動，並忽略轉身所需的時間。時間 $t$ 以小時為單位，由起步一刻開始計算。下面的距離—時間圖（distance–time graph）（按最後求得的 $X$ 值繪畫）顯示了我們用到的兩個交點。



### 第一次相遇

烏龜在時間 $\dfrac{1/2}{v}$ 到達距離起點 $\frac{1}{2}\,\mathrm{km}$ 的位置。兔子在 $t=\frac{1}{2}$ 起步，並在時間 $\frac{1}{2}+\dfrac{1/2}{V}$ 到達同一位置。牠們同時身處該處，所以

$$\frac{1}{2v}=\frac{1}{2}+\frac{1}{2V},\qquad \text{即}\qquad \frac{1}{v}=1+\frac{1}{V}. \tag{1}$$

### 第二次相遇

兔子在時間 $\frac{1}{2}+\dfrac{X}{V}$ 到達菜園，睡了半小時，然後在 $1+\dfrac{X}{V}$ 離開。接着牠往回跑了 $\frac{5}{4}\,\mathrm{km}$，所以牠與烏龜相遇的時間是

$$1+\frac{X}{V}+\frac{5}{4V}=1+\frac{X+\frac{5}{4}}{V}.$$

烏龜此時仍在去程途中，距離起點 $X-\frac{5}{4}\,\mathrm{km}$，所以牠已經走了 $\dfrac{X-\frac{5}{4}}{v}$ 小時。令兩個時間相等：

$$\frac{X-\frac{5}{4}}{v}=1+\frac{X+\frac{5}{4}}{V}. \tag{2}$$

**想一想** 我們有兩條方程和兩個未知數，但 (2) 式中 $1/v$ 與 $X$ 相乘。怎樣才可以單獨求出 $V$？

我們把 (1) 代入 (2)。這樣 $X/V$ 的項便會互相抵消，$V$ 的結果因而如此簡潔：

$$\begin{aligned}\left(X-\tfrac{5}{4}\right)\left(1+\frac{1}{V}\right)&=1+\frac{X+\frac{5}{4}}{V}\\ X-\tfrac{5}{4}+\frac{X-\frac{5}{4}}{V}&=1+\frac{X+\frac{5}{4}}{V}\\ X-\tfrac{9}{4}&=\frac{5/2}{V}.\end{aligned}$$

因此

$$V=\frac{5/2}{X-\frac{9}{4}}=\frac{10}{4X-9},$$

正如所求。(要使這個速率為正數，我們需要 $X>\frac{9}{4}$，在最後一部分要記住這一點。)

然後由 (1) 得

$$\frac{1}{v}=1+\frac{4X-9}{10}=\frac{4X+1}{10},\qquad \text{所以}\qquad v=\frac{10}{4X+1}.$$

**求 $X$**

兔子在 $1+\dfrac{X}{V}$ 離開菜園，再用 $\dfrac{X}{V}$ 小時回到家，所以牠在 $1+\dfrac{2X}{V}$ 完成全程。令這個時間等於 $\frac{3}{2}$，得 $\dfrac{2X}{V}=\frac{1}{2}$，所以 $V=4X$。於是

$$\begin{aligned} 4X &= \frac{10}{4X-9} \\ 16X^2-36X-10 &= 0 \\ 8X^2-18X-5 &= 0 \\ (2X-5)(4X+1) &= 0. \end{aligned}$$

根 $X=-\frac{1}{4}$ 不可能是距離（況且我們需要 $X>\frac{9}{4}$），所以

$$X=\tfrac{5}{2}\text{ km}.$$

取這個值時，兔子的速率是 $V=10\,\text{km/h}$，烏龜的速率是 $v=\frac{10}{11}\,\text{km/h}$。

## 驗算

把 $X=\frac{5}{2}$、$V=10$、$v=\frac{10}{11}$ 代回故事之中。烏龜在 $t=0.55$ 到達 $\frac{1}{2}\,\text{km}$ 處，而兔子在 $t=0.5$ 起步，需要 0.05 h 才到達該處：牠們相遇。兔子在 $t=0.75$ 到達菜園，並在 $t=1.25$ 離開。牠在 $t=1.375$ 到達距離菜園 1.25 km 的位置（亦即距離起點 1.25 km），而烏龜到達 1.25 km 處的時間是 $1.25\times1.1=1.375$：牠們再次相遇。兔子回到家的時間是 $1.25+0.25=1.5$。把任意選取的 $X>\frac{9}{4}$ 的值代入 $V$ 和 $v$ 的公式，再重演整場比賽，同樣會得到相遇位置完全正確的結果。

## 常見錯誤

- 寫兔子的時鐘時間時，漏掉了牠其中一次半小時的睡眠，尤其是在菜園的那一次。
- 把第二個相遇點當成距離起點 $\frac{5}{4}\,\text{km}$，而不是距離菜園。
- 把二次方程（quadratic equation）的兩個根都保留：負的距離必須捨去，並要說明理由。

## STEP 之外

像上面那樣的距離—時間圖，把一道文字題轉化為直線在何處相交的問題，這正是任何力學課程中運動學（kinematics）的起點。這裏改用速率的倒數（配速，pace）來運算，

使方程變成線性；選取能把問題線性化（linearise）的變量，是應用數學中反覆出現的技巧。若想看看由這些想法發展出來的大學一年級程度運動理論，可參閱 Kleppner and Kolenkow, *An Introduction to Mechanics*（力學入門）。

詳細解答

# STEP I 1999，第 10 題

力學 • 簡諧運動

**STEP I 1999，第 10 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/99-s1-q10

## 關鍵思路

質點把彈簧分成兩條彈簧 $AP$ 和 $PB$，兩者的彈性模量（modulus of elasticity）都與整條彈簧相同，同為 $\lambda$（模量是物料本身的性質），但自然長度（natural length）分別是 $a$ 和 $b$。只用一個坐標，即質點與上端的距離，對每一段寫出虎克定律（Hooke's law），便會發現運動方程（equation of motion）對這個坐標而言是線性的。這個坐標的係數決定了周期（period），我們只需看看它取決於甚麼、不取決於甚麼。

## 解答

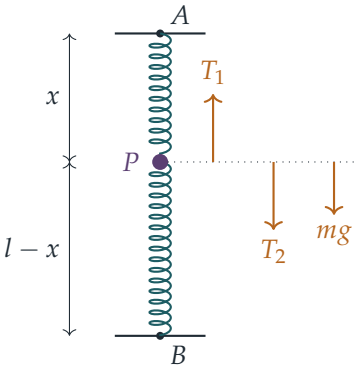
### 模型

設質點的質量為 $m$。彈簧是輕的，所以每一段上各處的張力（tension）都相同，而且它在伸長和壓縮時都服從虎克定律（它是彈簧而不是彈性繩，所以既能拉也能推）。先假設 $A$ 是上端。設在時間 $t$ 時質點位於 $A$ 下方距離 $x$ 處，於是 $AP = x$，$PB = l - x$。

由於 $AP$ 的自然長度是 $a$，$PB$ 的自然長度是 $b$，由虎克定律得

$$T_1 = \frac{\lambda(x-a)}{a}, \qquad T_2 = \frac{\lambda(l-x-b)}{b},$$

其中 $T_1$ 是 $AP$ 的張力，$T_2$ 是 $PB$ 的張力。張力為正時，該段把質點拉向它的固定端；數值為負則表示該段受壓縮，把質點推開。無論哪種情況，這些公式都給出正確的帶符號的力，所以我們根本不必事先知道哪一段被拉長。




**簡諧運動**

對質點沿向下方向分解力，作用於它的力有重量 $mg$、向上的張力 $T_1$ 和向下的張力 $T_2$：

$$\begin{aligned} m\ddot{x} &= mg - \frac{\lambda(x-a)}{a} + \frac{\lambda(l-x-b)}{b} \\ &= mg + \lambda - \frac{\lambda x}{a} + \frac{\lambda l}{b} - \frac{\lambda x}{b} - \lambda \\ &= mg + \frac{\lambda l}{b} - \lambda\left(\frac{1}{a} + \frac{1}{b}\right)x. \end{aligned}$$

常數 $\pm\lambda$ 互相抵消，右邊是 $x$ 的線性函數，斜率為負。記

$$\omega^2 = \frac{\lambda}{m}\left(\frac{1}{a} + \frac{1}{b}\right) = \frac{\lambda(a+b)}{mab}, \qquad x_0 = \frac{mg + \lambda l/b}{\lambda(a+b)/(ab)} = \frac{a(mgb+\lambda l)}{\lambda(a+b)}.$$

於是運動方程變成

$$\ddot{x} = -\omega^2(x - x_0),$$

所以，設 $X = x - x_0$（即相對於平衡位置（equilibrium position）$x_0$ 的位移），便有 $\ddot{X} = -\omega^2 X$。這正是以 $x = x_0$ 為中心的簡諧運動（simple harmonic motion）的方程，周期為

$$\boxed{\frac{2\pi}{\omega} = 2\pi\sqrt{\frac{mab}{\lambda(a+b)}}}$$

這就證明了第一個結果。

**想一想** 一旦其中一段受到壓縮，運動不就不再是簡諧運動了嗎？

如果是彈性繩，確實如此，因為它的張力公式只在繩被拉長時才成立。彈簧在受壓縮時仍服從同一條虎克定律公式，所以上面那一條方程在整個運動過程中都成立（只要彈簧圈沒有壓至互相緊貼），因此無須限制 $l$ 或振幅（amplitude）。

**與 $l$ 及擺放方向無關**

長度 $l$ 只在 $x_0$ 中出現：改變 $l$ 會移動振動的中心，但 $\omega$ 不變，因而周期也不變。

如果改為 $B$ 在最上方，同樣的推導仍然適用，只是兩段的角色互換：上面一段的自然長度現在是 $b$，下面一段是 $a$。這只不過是把 $a$ 和 $b$ 全部對調。周期只透過 $ab$ 和 $a+b$ 取決於 $a$ 和 $b$，而這兩個式子都是對稱的，所以周期不變。（平衡位置則確實會改變，因為重力現在拉長的是自然長度為 $b$ 的那一段，而不是自然長度為 $a$ 的那一段。）

所以，無論 $l$ 取甚麼值，也無論哪一端在最上方，周期都是 $2\pi\sqrt{mab/(\lambda(a+b))}$，這正是所要證明的。

## 驗算

- 質量兩側的兩條彈簧相當於並聯（in parallel）：勁度（stiffness）$\lambda/a$ 和 $\lambda/b$ 相加得 $\lambda(a+b)/(ab)$，與上面的 $\omega^2$ 一致。
- 極限情況 $b \to \infty$：下面一段變得非常弱，我們便回到一個質量掛在一條自然長度為 $a$ 的彈簧上的情形，周期為 $2\pi\sqrt{ma/\lambda}$。
- 單位：$\lambda$ 是力，所以 $ma/\lambda$ 的單位是 $\mathrm{kg\,m/N} = \mathrm{s}^2$。
- 數值上，取 $m=1$、$\lambda=50$、$a=1$、$b=2$，公式給出周期 0.72552。對 $l=2$、3 和 4.5（包括兩段都受壓縮的一種情況），並分別以兩端為最上方，用數值方法逐步求解完整的運動方程，所得周期在五位有效數字內相同；另外隨機選取多組常數，結果的相差也都在 $10^{-5}$ 以內。

## 常見錯誤

- 對各段使用 $\lambda a/(a+b)$ 或類似的模量。每一段的模量仍然是 $\lambda$；改變的是勁度 $\lambda/$(自然長度)。
- 弄錯下方張力的符號。被拉長的 $PB$ 把質點向下拉，即拉向 $B$。
- 忘記論證擺放方向的部分：你需要說明兩端對調就是把 $a$ 和 $b$ 對調，而周期對於它們是對稱的。

## STEP 之外

把一個系統拆成若干條彈簧再把勁度相加，是邁向耦合振子（coupled oscillators）的第一步：若彈簧上有多個質量，運動方程便組成一個線性方程組，而可能出現的頻率就是某個矩陣的特徵值（eigenvalues），即所謂「簡正模式」（normal modes）。當質量數

目極多而每個都極小時，取極限便得到彈簧或桿中縱波的波動方程（wave equation）。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教材）討論了振動和簡正模式。

詳細解答

# STEP I 1999，第 11 題

力學 • 引力作用下的圓周運動

**STEP I 1999，第 11 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/99-s1-q11

本題在 *Beyond STEP* 第 10.8 節排成三角形的三顆恆星中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch10/three-stars.html

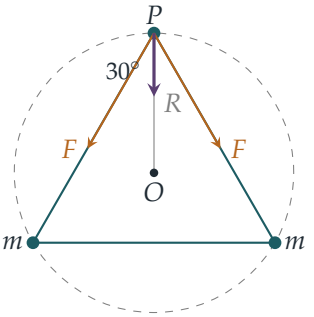
## 關鍵思路

先猜出運動的形式，再加以檢驗。如果三角形繞着中心作剛性轉動（rigid rotation），每顆星都沿圓周運動，因此它所受的淨力（net force）必須恰好是 $m\omega^2 R$，並且指向中心。根據對稱性，其餘各星對每顆星的引力（gravitational pull）正好指向中心。所以我們只需令這個引力的大小等於 $m\omega^2 R$，便可讀出 $\omega$。在中心放置一顆星，只是多加一個大小已知、指向中心的引力。

## 解答

我們把各星看成質點（particles），它們只在彼此的引力作用下運動。設 $O$ 為質心（centre of mass）。對三個相等的質量而言，質心就是三角形的形心（centroid），它與每個頂點的距離是

$$R = \frac{2}{3} \times \frac{\sqrt{3}}{2} a = \frac{a}{\sqrt{3}}.$$




**三顆星**

考慮位於頂部頂點的星 $P$。其餘兩顆星各以大小為

$$F = \frac{\gamma m^2}{a^2},$$

的力吸引它，方向沿着三角形的一條邊。經過 $P$ 的每條邊都與 $PO$ 成 $30°$ 角，因為 $PO$ 平分位於 $P$ 的 $60°$ 角。沿 $PO$ 方向及垂直於 $PO$ 方向分解：

- 在垂直於 $PO$ 的方向，兩個分量大小相等、方向相反，因此互相抵消；
- 在沿 $PO$ 的方向，合力（resultant）為 $2F\cos 30° = \sqrt{3}\,\dfrac{\gamma m^2}{a^2}$，指向 $O$。

根據對稱性，其餘兩顆星的情況相同。

現在假設三角形以角速度（angular velocity）$\omega$ 繞 $O$ 作剛性轉動。於是每顆星都沿半徑為 $R$ 的圓周以恆定速率 $R\omega$ 運動，因此它需要一個指向 $O$ 的加速度 $R\omega^2$。在指向 $O$ 的方向應用牛頓第二定律（Newton's second law），得

$$m\,\frac{a}{\sqrt{3}}\,\omega^2 = \sqrt{3}\,\frac{\gamma m^2}{a^2} \quad\Longrightarrow\quad \omega^2 = \frac{3\gamma m}{a^3}.$$

**想一想** 為甚麼這樣就證明了這種運動是可能的，而不僅僅是各力在某一瞬間恰好平衡？

讓各星從頂點出發，速度大小為 $R\omega$，方向垂直於 $OP$，並且全部朝同一方向轉動。只要整個排列仍然是邊長為 $a$、中心在 $O$ 的等邊三角形，每顆星所受的力，其大小和方向都恰好符合以角速度 $\omega$ 作圓周運動的要求。所以這個剛性轉動的三角形，在每一刻都滿足每顆星的運動方程（equation of motion）。因此它是這個系統的一個真正的運動：它就是由這些初始條件（initial conditions）所決定的解。總動量（momentum）為零，所以 $O$ 保持不動，正如所料。因此每顆星都可以以角速度 $(3\gamma m/a^3)^{1/2}$ 繞質心公轉。

### 在中心加上第四顆星

在 $O$ 放置一顆質量為 $\lambda m$ 的星。質心仍然是 $O$。外圍三顆星對中心星的引力大小相等，彼此成 120° 角，所以合力為零。因此，只要外圍三顆星保持對稱的排列，中心星便會在 $O$ 保持靜止。

現在每顆外圍星都多受一個指向 $O$ 的引力，大小為

$$\frac{\gamma(\lambda m)m}{R^2} = \frac{3\gamma\lambda m^2}{a^2},$$

這是因為 $R^2 = a^2/3$。這個額外的力沿 $PO$ 作用，所以它只是加在向內的合力之上。指向 $O$ 方向的運動方程變成

$$m\,\frac{a}{\sqrt{3}}\,\omega^2 = \sqrt{3}\,\frac{\gamma m^2}{a^2} + 3\lambda\,\frac{\gamma m^2}{a^2}$$

$$\omega^2 = \frac{3\gamma m}{a^3} + \frac{3\sqrt{3}\,\lambda\gamma m}{a^3} = \frac{3\gamma m}{a^3}\left(1+\sqrt{3}\,\lambda\right).$$

用與先前相同的論證，可知這個剛性轉動是一個真正的運動。所以，當第四顆星靜止於中心時，外圍三顆星可以繞着它公轉，角速度為

$$\omega = \left(\frac{3\gamma m\,(1+\sqrt{3}\,\lambda)}{a^3}\right)^{1/2}.$$

## 驗算

- 代入 $\lambda = 0$ 便得回第一個答案，這是必然的。
- 量綱（dimensions）：$\gamma m/r^2$ 是加速度，所以 $\gamma m/a^3$ 的單位是 時間$^{-2}$，正好與 $\omega^2$ 相符。
- 如果 $\lambda$ 非常大，外圍各星幾乎感受不到彼此的引力。這時它們只繞着中心的質量運行，有 $\omega^2 \approx \gamma\lambda m/R^3 = 3\sqrt{3}\,\gamma\lambda m/a^3$，與主導項吻合。
- 我們以數值方法對完整的運動方程進行積分（取 $\gamma = m = a = 1$）。從預測的圓周運動速度出發，各星在公轉三周後回到起始位置，相對誤差小於 $10^{-7}$，對 $\lambda = 0$、1 和 2.5 皆然。例如，$\lambda = 1$ 給出 $\omega^2 = 3(1+\sqrt{3}) \approx 8.196$。

## 常見錯誤

- 以 $a$ 而不是 $R = a/\sqrt{3}$ 作為圓的半徑，或作為與中心星的距離。
- 把兩個大小為 $F$ 的引力直接相加成 $2F$，忘記了它們與 $PO$ 成 30° 角。
- 忘記說明中心星為何保持不動：作用於它的三個引力互相抵消。

## STEP 之外

這是拉格朗日（Lagrange）三體問題（three-body problem）等邊三角形解的一個特殊情況；該解對位於旋轉等邊三角形頂點的任意三個質量都成立。同樣的想法可以得出拉格朗日點（Lagrange points）$L_4$ 和 $L_5$，特洛伊小行星（Trojan asteroids）就在那裏與木星共用同一軌道。這類排列是否穩定（stable），是一個更深入的問題，要把運動方程在旋轉解附近線性化（linearising）才能解答。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學入門教材）在這個程度上對有心力（central forces）和軌道有嚴謹的論述。

詳細解答

# STEP I 1999，第 12 題

概率與統計 • 灌鉛骰子的不等式

**STEP I 1999，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/99-s1-q12

本題在 *Beyond STEP* 第 9.13 節做了手腳的骰子中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch09/loaded-dice.html

## 關鍵思路

第一行的不等式（inequality）是整道題的原動力。假如每個總和出現的機會均等，那麼 2 和 12 這兩個「角落」總和各自會迫使某個概率乘積等於 $\frac{1}{11}$，而總和 7 就會包含兩個交叉項，其和的形式為 $c\,(x + x^{-1})$。該不等式令這個和變得太大。第 (ii) 部分是同一個想法換了另一副面貌：當所有項都相等時，平方和最小。

## 解答

### 第 (i) 部分

**不等式**。當 $x > 0$ 時可以取平方根，而

$$x + x^{-1} - 2 = \left(\sqrt{x} - \frac{1}{\sqrt{x}}\right)^2 \geq 0,$$

所以 $x + x^{-1} \geq 2$，而且只有當 $\sqrt{x} = 1/\sqrt{x}$，即 $x = 1$ 時等號才成立。（等價的做法是：把 $(x-1)^2 \geq 0$ 乘以正數 $x^{-1}$。）

**骰子**。設 $p_s$ 為總和等於 $s$ 的概率（probability），其中 $2 \leq s \leq 12$。由於兩顆骰子互相獨立（independent），出現數對（第一顆擲得 $i$，第二顆擲得 $j$）的概率是 $q_i r_j$。下面的方格表顯示哪些數對給出我們所需的總和。

| | $q_1$ | $q_2$ | $q_3$ | $q_4$ | $q_5$ | $q_6$ |
|---|---|---|---|---|---|---|
| $r_1$ | 2 | 3 | 4 | 5 | 6 | 7 |
| $r_2$ | 3 | 4 | 5 | 6 | 7 | 8 |
| $r_3$ | 4 | 5 | 6 | 7 | 8 | 9 |
| $r_4$ | 5 | 6 | 7 | 8 | 9 | 10 |
| $r_5$ | 6 | 7 | 8 | 9 | 10 | 11 |
| $r_6$ | 7 | 8 | 9 | 10 | 11 | 12 |

總和 2：一格

總和 12：一格

總和 7：六格

總和 2 只能來自 $(1,1)$，總和 12 只能來自 $(6,6)$，所以

$$p_2 = q_1 r_1, \qquad p_{12} = q_6 r_6.$$

總和 7 來自六個數對，當中包括 $(1,6)$ 和 $(6,1)$。這六項全部非負，所以

$$p_7 = q_1 r_6 + q_2 r_5 + \cdots + q_6 r_1 \ \geq \ q_1 r_6 + q_6 r_1.$$

現在用反證法（proof by contradiction），假設全部十一個總和出現的機會均等。它們已窮盡所有可能，所以每個總和的概率都是 $\frac{1}{11}$。於是 $q_1 r_1 = \frac{1}{11}$ 及 $q_6 r_6 = \frac{1}{11}$。特別地，$q_1, r_1, q_6, r_6$ 均不為零，所以可以寫成 $r_1 = \dfrac{1}{11 q_1}$ 及 $r_6 = \dfrac{1}{11 q_6}$。代入後，

$$\begin{aligned} p_7 \ \geq \ q_1 r_6 + q_6 r_1 &= \frac{q_1}{11 q_6} + \frac{q_6}{11 q_1} = \frac{1}{11}\left(x + x^{-1}\right), \qquad \text{其中} x = \frac{q_1}{q_6} > 0, \\ &\geq \frac{2}{11} \end{aligned}$$

這是根據剛才證明的不等式。但 $p_7 = \frac{1}{11} < \frac{2}{11}$，矛盾。因此，無論兩顆骰子怎樣灌鉛，總和 $2, 3, \ldots, 12$ 都不可能全部機會均等。

**想一想** 為甚麼必須指出 $q_1$ 和 $q_6$ 不為零？

我們曾除以 $q_1$ 和 $q_6$，而不等式 $x + x^{-1} \geq 2$ 需要 $x > 0$。這兩點都有保證，因為乘積 $q_1 r_1$ 和 $q_6 r_6$ 各自等於 $\frac{1}{11} \neq 0$。少了這一步，論證就會有漏洞：一般而言，一顆永遠擲不出 1 的骰子是完全容許的。

### 第 (ii) 部分

設 $S$ 為第一顆骰子兩次擲得相同點數的事件（event）。兩次投擲互相獨立，所以兩次都擲得 $j$ 的概率是 $q_j^2$，而這六個事件互不相交（disjoint）。因此

$$\mathrm{P}(S) = q_1^2 + q_2^2 + \cdots + q_6^2.$$

已知 $q_j \geq 0$ 及 $\sum q_j = 1$，我們要求這個值最小可以是多少。把一個量度骰子偏離公平程度的平方和展開：

$$0 \leq \sum_{j=1}^{6} \left(q_j - \tfrac{1}{6}\right)^2 = \sum_{j=1}^{6} q_j^2 - \tfrac{1}{3}\sum_{j=1}^{6} q_j + 6 \cdot \tfrac{1}{36}$$
$$= \sum_{j=1}^{6} q_j^2 - \tfrac{1}{3} + \tfrac{1}{6} = \mathrm{P}(S) - \tfrac{1}{6}.$$

所以 $\mathrm{P}(S) \geq \frac{1}{6}$，而且等號恰好在每個 $q_j = \frac{1}{6}$，即骰子公平時成立。

**想一想** 1/36 從何而來？為甚麼說它是個煙幕？

$\frac{1}{36}$ 是用公平骰子擲出某一個特定數對（例如一對六）的概率。「兩個點數相同」涵蓋六個這樣的數對，所以對公平骰子來說，其概率是 $6 \times \frac{1}{36} = \frac{1}{6}$，而把骰子灌鉛只會令它增加。

所以答案是**不可能**：無論 $q_j$ 取甚麼值，兩個點數相同的概率至少是 $\frac{1}{6}$，遠大於 $\frac{1}{36}$。

## 驗算

- 公平骰子：$p_2 = p_{12} = \frac{1}{36}$，但 $p_7 = \frac{6}{36}$，明顯不相等，與結果所預測的一致。
- 以數值方法搜尋 20,000 對隨機灌鉛的骰子，從未出現全部十一個總和的概率與 $\frac{1}{11}$ 相差都在 0.05 以內的情況；用分數作精確檢查，亦證實每次都有 $p_7 \geq q_1 r_6 + q_6 r_1$。
- 對於 $(q_1, \ldots, q_6) = (0.1, 0.2, 0.3, 0.15, 0.15, 0.1)$ 的骰子，我們得到 $\mathrm{P}(S) = 0.195 \geq \frac{1}{6}$；模擬 200,000 次雙擲的結果是 0.194。

## 常見錯誤

- 假設兩顆骰子完全相同（$q_i = r_i$）。題目容許兩顆骰子的灌鉛方式不同，而上面的論證並不需要兩者一致。
- 寫成 $p_7 = q_1 r_6 + q_6 r_1$。其實還有另外四項；我們只需要一個不等式，所以可以把它們捨去，但要說明原因（$q_i r_j \geq 0$）。
- 在第 (ii) 部分，以 $\sum q_j^2 \geq \frac{1}{36}$ 作答，或者忘記「相同」是指六個結果中的任何一個。

## STEP 之外

第 (ii) 部分所說的就是 $\sum q_j^2 \geq \frac{1}{n}\left(\sum q_j\right)^2$，這是柯西－施瓦茲不等式（Cauchy–Schwarz inequality）的特殊情況；$\sum q_j^2$ 這個量稱為*碰撞概率*（collision probability），在信息論（information theory）和雜湊（hashing）的分析中都會出現。第 (i) 部分是關於生成函數（generating function）的經典結果：兩顆骰子總和的概率生成函數（probability generating function）是 $Q(t)R(t)$，而有一個巧妙的另證指出，若總和均勻分佈，$1 + t + \cdots + t^{10}$ 就必須等於一個常數乘以兩個 5 次實係數多項式的積，而這兩個多項式各有一個實根，然而 $1 + t + \cdots + t^{10} = (t^{11} - 1)/(t - 1)$ 卻沒有實根。這類生成函數的理論可參閱 Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門教材）。

詳細解答

# STEP I 1999，第 13 題

概率與統計 • 指示變量

> **STEP I 1999，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/99-s1-q13

## 關鍵思路

不要直接數單元的數目，改為數間隙：單元的數目等於相鄰磁鐵彼此分開的位置數目再加一。某處會否出現間隙，只取決於相鄰兩塊磁鐵的指向是否相同，而這 $N-1$ 個事件原來是互相獨立（independent）的公平擲幣。因此間隙的數目服從二項分佈（binomial distribution），期望值（expectation）和方差（variance）便可立即得出。

## 解答

### 建立模型

我們把「隨機擺放」理解為：每塊磁鐵指向任何一方的可能性均等，且各磁鐵互相獨立。若一塊磁鐵的北極在右邊（SN），稱它為 $A$ 型；若北極在左邊（NS），稱它為 $B$ 型。設 $X_1, \ldots, X_N$ 為由左至右各磁鐵的類型。

考慮相鄰的兩塊磁鐵。互相面對的兩端，是第一塊磁鐵的右端和第二塊磁鐵的左端。

- 類型相同（先 A 後 A）：相對的兩端是 N 和 S，屬異極，所以它們會**相連**。（先 B 後 B 的情況相同，只是換成 S 和 N。）
- 類型不同（先 A 後 B）：相對的兩端是 N 和 N，所以它們會**分開**。（先 B 後 A 則是 S 和 S。）



因此，第 $i$ 塊與第 $i+1$ 塊磁鐵之間出現間隙，當且僅當 $X_i \neq X_{i+1}$。對 $i = 1, \ldots, N-1$，定義

$$G_i = \begin{cases} 1 & \text{若} X_i \neq X_{i+1}\text{（有間隙）}, \\ 0 & \text{若} X_i = X_{i+1}\text{（相連）}. \end{cases}$$

每一個間隙都把整列磁鐵多分出一段，所以獨立單元的數目是

$$U = 1 + G_1 + G_2 + \cdots + G_{N-1}.$$

### 間隙的分佈

每個 $G_i$ 取值 1 的概率都是 $\frac{1}{2}$：不論 $X_i$ 是甚麼，$X_{i+1}$ 與它不同的概率都是 $\frac{1}{2}$。

> **想一想** 「在 $i$ 處有間隙」和「在 $i+1$ 處有間隙」這兩個事件都牽涉第 $i+1$ 塊磁鐵。各個 $G_i$ 真的互相獨立嗎？

是的，儘管相鄰的 $G$ 共用同一塊磁鐵。先固定第一塊磁鐵的類型 $X_1$。這樣，數列 $G_1, \ldots, G_{N-1}$ 便逐一決定了 $X_2, X_3, \ldots, X_N$（若 $G_i = 0$，$X_{i+1}$ 等於 $X_i$；若 $G_i = 1$，則為另一類型）；反過來，各磁鐵的類型也決定了所有的 $G$。所以，不論 $X_1$ 取哪一個值，$(X_2, \ldots, X_N)$ 的 $2^{N-1}$ 種等可能的選擇，與 $(G_1, \ldots, G_{N-1})$ 的 $2^{N-1}$ 個可能數列一一對應（one-to-one）。因此每一個間隙數列出現的概率都是 $1/2^{N-1}$，這正好表示各個 $G_i$ 互相獨立，並且各以概率 $\frac{1}{2}$ 取值 0 或 1。

由此可得 $G_1 + \cdots + G_{N-1} \sim \mathrm{B}\left(N-1, \frac{1}{2}\right)$。

### 期望值與方差

服從 $\mathrm{B}(n, p)$ 的變量，其平均值為 $np$，方差為 $np(1-p)$。取 $n = N-1$、$p = \frac{1}{2}$，並留意加上常數 1 只會令平均值平移，不會改變方差，於是

$$\mathrm{E}(U) = 1 + \tfrac{1}{2}(N-1) = \frac{N+1}{2},$$
$$\mathrm{Var}(U) = (N-1) \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = \frac{N-1}{4}.$$

如果你不想直接引用二項分佈的結果：$\mathrm{E}(G_i) = \frac{1}{2}$，而且由於 $G_i^2 = G_i$，可得 $\mathrm{Var}(G_i) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$。由期望值的線性性質（linearity）可得平均值，而獨立性則讓我們可以把 $N-1$ 個方差相加。

## 驗算

- $N = 1$：一塊磁鐵必定是一個單元，而公式給出平均值 1 和方差 0。

- $N = 2$：兩塊磁鐵組成 1 個或 2 個單元，概率各為 $\frac{1}{2}$，所以平均值是 $\frac{3}{2}$，方差是 $\frac{5}{2} - \frac{9}{4} = \frac{1}{4}$，與公式相符。
- 對所有 $2^N$ 種排列逐一窮舉，在 $N \leq 12$ 的情況下所得結果與 $\frac{N+1}{2}$ 和 $\frac{N-1}{4}$ 完全一致；另外以 $N = 20$ 進行模擬，得到的平均值約為 10.5，方差約為 4.73（預測值為 10.5 和 4.75）。

## 常見錯誤

- 以為兩塊指向相同的磁鐵會互相排斥。其實它們相對的是異極（N 對 S），所以會互相吸引並相連。
- 沒有說明獨立性的理由便把方差相加。這裏的獨立性確實成立，但並不明顯，因為相鄰的間隙共用同一塊磁鐵。
- 把間隙數目當作 $N$ 個而非 $N - 1$ 個，因而寫出 $\mathrm{Var}(U) = \frac{N}{4}$；又或者在平均值中漏掉額外的 1。

## STEP 之外

藉着數單元之間的邊界來數單元，這種方法就是指示變量（indicator variables）法，是概率論中最有用的工具之一：把一個計數寫成若干個 0–1 變量之和，然後運用期望值的線性性質，而這完全不需要獨立性。這裏的 $U$ 所數的，是由兩種符號組成的隨機序列中的游程（runs）數目。游程數目是統計學中 Wald–Wolfowitz 游程檢驗（runs test）的基礎，這個檢驗用來判斷一個序列看來是否隨機。關於指示變量及其方差的更多內容，可參閱 Blitzstein 與 Hwang 的 *Introduction to Probability*（概率論入門）。

詳細解答

# STEP I 1999，第 14 題

概率與統計 • 條件概率

> **STEP I 1999，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/99-s1-q14

## 關鍵思路

這個連續隨機變量（continuous random variable）只是用來引出一個離散（discrete）隨機變量。只要把密度函數（density）化成得 4、3、2、1 分的四個概率，第 (i) 部分就只是一行的期望值（expectation）計算。第 (ii) 部分是條件概率（conditional probability）；由於總分必須剛好等於 3，獲勝的方式寥寥可數，我們可以把它們全部列出。

## 解答

### 第 (i) 部分

對於 $0 \le a \le b \le 1$，

$$\mathrm{P}(a \le X < b) = \int_a^b 2x \,\mathrm{d}x = b^2 - a^2.$$

（對連續隨機變量而言，包括或不包括端點並無分別。）設 $S$ 為擲一支飛鏢所得的分數，則

$$\mathrm{P}(S=4) = \left(\tfrac{1}{4}\right)^2 - 0 = \tfrac{1}{16}, \qquad \mathrm{P}(S=3) = \left(\tfrac{1}{2}\right)^2 - \left(\tfrac{1}{4}\right)^2 = \tfrac{3}{16},$$
$$\mathrm{P}(S=2) = \left(\tfrac{3}{4}\right)^2 - \left(\tfrac{1}{2}\right)^2 = \tfrac{5}{16}, \qquad \mathrm{P}(S=1) = 1 - \left(\tfrac{3}{4}\right)^2 = \tfrac{7}{16}.$$

這四個概率之和為 1，正如所料。留意外圍的環較容易擲中：密度 $2x$ 隨 $x$ 增大，因為離中心愈遠，面積愈大。

所以

$$\mathrm{E}(S) = 4 \cdot \tfrac{1}{16} + 3 \cdot \tfrac{3}{16} + 2 \cdot \tfrac{5}{16} + 1 \cdot \tfrac{7}{16} = \frac{4+9+10+7}{16} = \frac{30}{16} = \frac{15}{8},$$

即為所求。

### 第 (ii) 部分

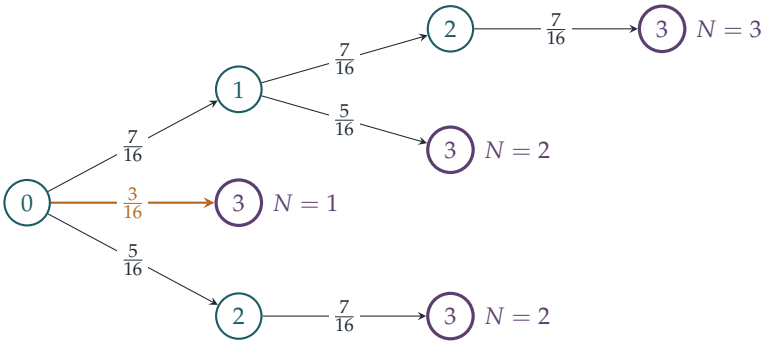
假設各次投擲互相獨立（independent），每次的分數都依從上述 $S$ 的分佈。設 $W$ 為我獲勝的事件，$N$ 為這局遊戲的投擲次數。

> **想一想** 哪些得分序列會導致獲勝？

我獲勝，當且僅當累積總分達到 3 而之前沒有超過它。每次得分至少為 1，所以總分每次至少增加 1，遊戲最多只會擲三次。一局獲勝的遊戲，就是一個由正數得分組成、總和剛好為 3 的序列。（只有當總分低於 3 時我才會再擲，所以每個這樣的序列確實是一局完整的遊戲。）這樣的序列只有四個：

$$(3), \qquad (1,2), \qquad (2,1), \qquad (1,1,1).$$

得分 4 永遠不會出現，因為它會令總分立即超過 3。




樹形圖顯示每條以獲勝告終的路線上的累積總分（獲勝狀態以紫色圓圈標示），箭頭上標示每次投擲的概率。

根據獨立性，每個序列的概率等於其中各個得分的概率之積：

$$\begin{aligned}
\mathrm{P}(W \cap N = 1) &= \mathrm{P}(3) = \tfrac{3}{16} = \tfrac{768}{4096}, \\
\mathrm{P}(W \cap N = 2) &= \mathrm{P}(1,2) + \mathrm{P}(2,1) = 2 \cdot \tfrac{7}{16} \cdot \tfrac{5}{16} = \tfrac{70}{256} = \tfrac{1120}{4096}, \\
\mathrm{P}(W \cap N = 3) &= \mathrm{P}(1,1,1) = \left(\tfrac{7}{16}\right)^3 = \tfrac{343}{4096}.
\end{aligned}$$

這三個事件互斥（disjoint），合起來就是 $W$，所以

$$\mathrm{P}(W) = \frac{768 + 1120 + 343}{4096} = \frac{2231}{4096}.$$

因此

$$\mathrm{P}(N = 3 \mid W) = \frac{\mathrm{P}(W \cap N = 3)}{\mathrm{P}(W)} = \frac{343/4096}{2231/4096} = \frac{343}{2231},$$

即為所求。

## 驗算

- 四個得分概率 $\frac{1}{16}, \frac{3}{16}, \frac{5}{16}, \frac{7}{16}$ 之和為 1。它們都是奇數除以 16，因為 $(k+1)^2 - k^2 = 2k+1$。
- 目標答案中的 $343 = 7^3$ 一看便知，擲三次而獲勝的情況必定是 $(1,1,1)$，這印證了我們的設定。
- 以電腦模擬 $10^6$ 局遊戲，得出 $\mathrm{P}(W) \approx 0.5452$（準確值為 0.5447）及 $\mathrm{P}(N=3 \mid W) \approx 0.1535$（準確值為 $343/2231 \approx 0.1537$）。

## 常見錯誤

- 把各個得分區當作等可能。概率來自對密度函數求積分，而不是來自區間的闊度。
- 計算了 $(1,2)$ 卻忘記 $(2,1)$，結果得出 $\mathrm{P}(W) = \frac{1671}{4096}$。
- 答案寫成 $\mathrm{P}(N=3 \text{ 及} W) = \frac{343}{4096}$，而沒有除以 $\mathrm{P}(W)$。

## STEP 之外

第 (ii) 部分其實是一條小型的馬可夫鏈（Markov chain）：狀態就是累積總分，而 3 和「爆分」是吸收態（absorbing states）。目標分數較大時，逐一列出路線便不再可行。這時我們改為以 $w_t$ 表示由總分 $t$ 出發而獲勝的概率，從而得到遞推關係（recurrence）$w_t = \sum_s \mathrm{P}(S=s)\, w_{t+s}$。這種首步分析（first-step analysis）是大學一年級概率論的重要工具。Grimmett 與 Welsh 所著的 *Probability: An Introduction*（概率論入門教科書）在討論隨機漫步（random walks）和馬可夫鏈的部分對此有所講解。

詳細解答

# STEP II 1999，第 1 題

純數學 • 大數的比較

**STEP II 1999，第 1 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/99-s2-q1

## 關鍵思路

這些數大得根本無法直接比較，就連它們的對數（logarithm）也大得無從入手。突破口是取 $\log_{10}$ 兩次。由於 $x$、$y$ 和 $z$ 各自都是 10 的冪，每個雙重對數（double logarithm）都會化為 100、$x$ 和 $y$ 的簡短的和，而這些和很容易比較。因為 $\log_{10}$ 是遞增的，比較雙重對數就等於比較原來的數。

## 解答

全文以 log 表示 $\log_{10}$，並記

$$L(a) = \log(\log a).$$

函數 $t \mapsto \log t$ 是嚴格遞增（strictly increasing）的，所以對於 $a, b > 10$ 的數（此時 $\log a, \log b > 1 > 0$），$a < b$ 當且僅當 $L(a) < L(b)$。我們用到的基本事實是

$$\log x = 100, \qquad \log y = x = 10^{100}, \qquad \log z = y = 10^{x}.$$

### 第 (i) 部分

在斯特靈近似公式（Stirling's approximation）中取 $n = x$，再取 $\log_{10}$，得

$$\begin{aligned}\log a_1 &\approx \log\sqrt{2\pi} + \left(x + \tfrac{1}{2}\right)\log x - x\log \mathrm{e} \\ &= 100x - x\log \mathrm{e} + 50 + \log\sqrt{2\pi} \\ &= x\,(100 - \log \mathrm{e}) + 50.4\ldots.\end{aligned}$$

現在 $\log \mathrm{e} \approx 0.434$，而常數 $50.4\ldots$ 與 $x = 10^{100}$ 相比完全可以忽略。所以

$$\log a_1 \approx 99.566 \times 10^{100}, \qquad L(a_1) \approx 100 + \log 99.566 \approx 100 + 1.998 = 101.998\ldots \approx 102,$$

正如所求。

**第 (ii) 部分**

我們對每個數先計算 $\log a_i$，再計算 $L(a_i)$。

- $a_2 = x^y$： $\log a_2 = y \log x = 100y$，所以 $L(a_2) = 2 + \log y = x + 2$。
- $a_3 = y^x$： $\log a_3 = x \log y = x^2 = 10^{200}$，所以 $L(a_3) = 200$。
- $a_4 = z^x$： $\log a_4 = x \log z = xy$，所以 $L(a_4) = \log x + \log y = x + 100$。
- $a_5 = \mathrm{e}^{xyz}$： $\log a_5 = xyz \log \mathrm{e}$，所以

$$L(a_5) = 100 + x + y + \log(\log \mathrm{e}) \approx y + x + 99.64,$$

  因為 $\log(\log \mathrm{e}) = \log 0.434\ldots \approx -0.36$。
- $a_6 = z^{1/y}$： $\log a_6 = \frac{1}{y} \log z = 1$，所以 $a_6 = 10$，這是精確值。
- $a_7 = y^{z/x}$： $\log a_7 = \frac{z}{x} \log y = \frac{z}{x} \cdot x = z$，所以 $L(a_7) = \log z = y$。

連同第 (i) 部分得到的 $L(a_1) \approx 102$，各個雙重對數為

$$\underbrace{0}_{a_6},\ \underbrace{102}_{a_1},\ \underbrace{200}_{a_3},\ \underbrace{x+2}_{a_2},\ \underbrace{x+100}_{a_4},\ \underbrace{y}_{a_7},\ \underbrace{y+x+99.6}_{a_5}.$$

? **想一想** 第 (i) 部分只給出了 $a_1$ 的近似值。這樣足以確定它的位置嗎？

足夠，而且我們甚至不需要用到斯特靈公式。$x!$ 的 $x$ 個因子每個都不超過 $x$，所以 $x! \le x^x$，並且 $\log a_1 \le x \log x = 10^{102}$，由此得 $L(a_1) \le 102 < 200 = L(a_3)$。另外，$a_1 = x! > 10 = a_6$ 是顯然的。所以 $a_6 < a_1 < a_3$ 嚴格成立。

現在我們逐一說明餘下的比較。

- $a_3 < a_2$： $200 < 10^{100} + 2$。
- $a_2 < a_4$： $x + 2 < x + 100$。
- $a_4 < a_7$：我們需要 $x + 100 < y = 10^x$。這個不等式成立，而且相差極其懸殊，因為對於每個 $t \ge 3$ 都有 $10^t > t + 100$，而這裏 $t = x = 10^{100}$。
- $a_7 < a_5$： $y < y + x + 99.6$，因為 $x + 99.6 > 0$。（直接來看：$\log a_5 = z \cdot xy \log \mathrm{e}$ 而 $xy \log \mathrm{e} > 1$，所以 $\log a_5 > z = \log a_7$。）

除 $a_6$ 以外，所有數都大於 10，所以 $L$ 的單調性（monotonicity）適用，而大小次序為

$$a_6 < a_1 < a_3 < a_2 < a_4 < a_7 < a_5.$$

## 驗算

對於第 (i) 部分，我們取 $n = 10^2, 10^3, 10^5, 10^8$，把斯特靈公式與 $\log_{10} n!$ 的精確值（透過對數伽瑪函數（log-gamma function）求得）作比較。$a_1$ 的雙重對數公式與精確值相差不超過 $10^{-6}$，而當 $n = 10^{100}$ 時，公式給出 101.998。對於第 (ii) 部分，我們取 $x = 10^k$，其中 $k = 2, 3, 4$，把整個計算重做一遍，這時 $y = 10^x$ 可以作為整數精確地儲存。每次都得出相同的次序 $a_6 < a_1 < a_3 < a_2 < a_4 < a_7 < a_5$，所以這個次序並不取決於 100 這個數的大小。

## 常見錯誤

- 只比較單重對數，然後迷失方向：$\log a_7 = z$ 和 $\log a_4 = xy$ 仍然太大，無法憑肉眼比較。要再取一次對數。
- 把 $\mathrm{e}^{xyz}$ 轉換為以 10 為底時，忘記了因子 $\log \mathrm{e}$。在這裏它幾乎沒有影響，但確實會令 $L(a_5)$ 偏移約 0.36。
- 沒有留意到 $a_6$ 其實就是 10，是所有數中最小的。

## STEP 之外

迭代對數（iterated logarithm）和指數塔在分析學和計算機科學中用來量度增長。例如，迭代對數 $\log^* n$ 出現在某些算法的運行時間中，而「塔」函數則在拉姆齊理論（Ramsey theory）中用作結果的界。斯特靈公式本身是一個漸近級數（asymptotic series）的首項，而嚴謹地證明它是大學一年級分析課的經典習題。想熟習當中涉及的極限和不等式，Abbott 的 *Understanding Analysis* 是很好的入門讀物。

詳細解答

# STEP II 1999，第 2 題

純數學 • 二次方程的判別式

**STEP II 1999，第 2 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/99-s2-q2

## 關鍵思路

一切取決於 $(*)$ 的判別式（discriminant）。由於 $x$ 的係數帶有因子 2，四分之一判別式 $\frac{1}{4}(b^2-4ac)$ 特別簡潔，而且它原來是 $n$ 的二次式（quadratic）。於是每一部分都變成關於另一個二次式的正負號的問題，這次以 $n$ 為變量。

## 解答

全篇我們都假設 $pn^2+q\geq 0$，使平方根為實數，而 $(*)$ 是一個係數為實數的普通二次方程。（在第 (i) 及第 (iii) 部分，由於 $q>0$ 或 $q\geq 0$，這個條件自動成立。）取 $a=n$、$b=2\sqrt{pn^2+q}$ 及 $c=rn+s$，方程有實根（real roots）當且僅當 $b^2-4ac\geq 0$，即 $\frac{1}{4}b^2-ac\geq 0$。現在

$$\begin{aligned}\tfrac{1}{4}b^2-ac&=(pn^2+q)-n(rn+s)\\&=(p-r)n^2-sn+q.\end{aligned}$$

設 $f(n)=(p-r)n^2-sn+q$。則

$$(*)\text{ 沒有實根}\iff f(n)<0.$$

### 第 (i) 部分

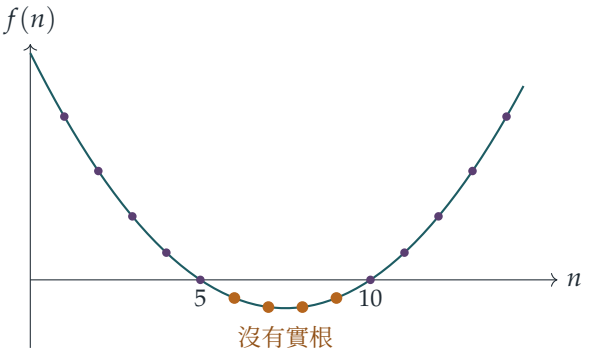
這裏 $p-r=1$，所以

$$f(n)=n^2-15n+50=(n-5)(n-10).$$

它為負數當且僅當 $5<n<10$。由於 $n$ 是正整數，$(*)$ 沒有實根的情況恰好是

$$n\in\{6,7,8,9\}.$$

當 $n = 5$ 及 $n = 10$ 時，$f(n) = 0$，方程有一個實數重根（repeated root），所以這兩個值不包括在內。




**第 (ii) 部分**

現在 $p < r$，所以 $f$ 的首項係數（leading coefficient）$p - r$ 為負數。把 $f$ 看成實變量 $t$ 的二次式：

$$f(t) = (p - r)t^2 - st + q.$$

它的判別式是

$$s^2 - 4(p - r)q,$$

根據假設 $4q(p - r) > s^2$，這個判別式為負數。

**想一想** 為甚麼 $f$ 的判別式為負數，能告訴我們 $f(n)$ 的正負號？

判別式為負數的二次式沒有實零點，所以它的圖像永不穿過橫軸，對所有實數 $t$ 都保持同一個正負號。首項係數為負數時，圖像是一條開口向下的拋物線（parabola），所以這個符號是負號。具體來說，用配方法（completing the square），

$$f(t) = (p - r)\Big(t - \frac{s}{2(p - r)}\Big)^2 + \frac{4q(p - r) - s^2}{4(p - r)},$$

兩項都是負數或零：第一項是因為 $p - r < 0$，第二項是因為其分子為正數而分母為負數。第二項嚴格小於零，所以對每一個實數 $t$ 都有 $f(t) < 0$。

特別地，對每一個正整數 $n$ 都有 $f(n) < 0$，所以無論 $n$ 取甚麼值，$(*)$ 都沒有實根，這正是所要證明的。

留意這些假設必然導致 $q < 0$：由於 $p - r < 0$ 而 $4q(p - r) > s^2 \geq 0$，$q$ 必定是負數。因此當 $n$ 較小時，$pn^2 + q$ 確實可以是負數，而我們一直沿用的「平方根為實數」這個假設在這裏就很重要。（假如容許係數為虛數，$x = 0$ 便有可能是根：試試 $p = 1$、$r = 2$、$s = -2$、$q = -10$、$n = 1$。）

### 第 (iii) 部分

現在 $n=1$、$p-r=1$ 及 $q=s^2/8\geq 0$，所以 $p+q>0$，各係數均為實數。這裏

$$f(1)=1-s+\frac{s^2}{8}=\frac{s^2-8s+8}{8}.$$

方程有實根當且僅當 $f(1)\geq 0$，亦即當且僅當

$$s^2-8s+8\geq 0.$$

$s^2-8s+8=0$ 的根是

$$s=\frac{8\pm\sqrt{64-32}}{2}=4\pm 2\sqrt{2},$$

所以 $s^2-8s+8=\big(s-(4-2\sqrt{2})\big)\big(s-(4+2\sqrt{2})\big)$。這個乘積非負，當且僅當 $s$ 位於兩根之間的開區間（open interval）之外。因此 $(*)$ 有實根當且僅當

$$s\leq 4-2\sqrt{2}\quad\text{或}\quad s\geq 4+2\sqrt{2},$$

這正是所要證明的。每一步都是可逆的，「當且僅當」的兩個方向正是由此而來。在兩個端點處，方程的根是重根。

## 驗算

- 在 (i) 中，取 $n=7$：$\frac{1}{4}b^2=3\cdot 49+50=197$ 而 $ac=7\cdot 29=203$，所以 $\frac{1}{4}b^2-ac=-6<0$。取 $n=11$：$413-11\cdot 37=6>0$。
- 在 (iii) 中，$4-2\sqrt{2}\approx 1.1716$ 而 $4+2\sqrt{2}\approx 6.8284$。取 $s=4$，得 $f(1)=1-4+2=-1<0$，它位於兩根之間的空隙內，與預期相符。
- 我們用電腦就 (i) 搜索了由 1 至 199 的 $n$，就 (ii) 測試了約 5 百萬個隨機情況，並就 (iii) 測試了 $10^5$ 個隨機的 $s$ 值。全部結果都與上文一致。

## 常見錯誤

- 在 (i) 中把 $n=5$ 及 $n=10$ 也包括在內。這時判別式為零，所以方程有一個（重複的）實根。
- 在 (ii) 中只驗證 $n$ 很大時 $f(n)<0$。判別式條件的要點，正在於 $f$ 處處為負數，包括 $n$ 較小的時候。
- 忘記了 $x$ 的係數中的因子 2，結果在判別式中把 $4(pn^2+q)$ 放錯了位置。

## STEP 之外

這裏的做法，是把參數（parameter）當作新的變量，再考慮以它為變量的二次式的正負號；這正是研究多項式的根如何隨係數變化的起點。在大學，這會引伸到高次多項式的判別式，以及這樣的一個概念：二次型（quadratic form）$at^2 + bt + c$ 是負定（negative definite）的，當且僅當 $a < 0$ 且 $b^2 < 4ac$。同一個判別法用於兩個變量時，可以為多元函數的駐點（stationary points）分類。Siklos 的 *Advanced Problems in Mathematics*（《高等數學難題》）中還有很多建基於嚴謹判別式論證的 STEP 題目。

詳細解答

# STEP II 1999，第 3 題

純數學 • 數學歸納法

**STEP II 1999，第 3 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/99-s2-q3

## 關鍵思路

不要一次又一次地從頭對 $\mathrm{e}^{-x^3}$ 求導。應該把第 $n$ 階導數（derivative）寫成 $\mathrm{e}^{-x^3}S_n(x)$，然後對這個式子再求導一次。這樣便得到一個把 $S_{n+1}$ 與 $S_n$ 及 $S_n'$ 連繫起來的遞推關係（recurrence），而題目的每一部分都可由這一個關係推出：各個多項式（polynomial）的具體形式、歸納法（induction）證明、次數（degree）與首項係數（leading coefficient），以及最後的不等式。

## 解答

### 遞推關係

根據定義，

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}\left(\mathrm{e}^{-x^3}\right) = \mathrm{e}^{-x^3}S_n(x).$$

利用積法則（product rule）再求導一次，

$$\frac{\mathrm{d}^{n+1}}{\mathrm{d}x^{n+1}}\left(\mathrm{e}^{-x^3}\right) = \mathrm{e}^{-x^3}S_n'(x) - 3x^2\mathrm{e}^{-x^3}S_n(x),$$

再乘以 $\mathrm{e}^{x^3}$，得

$$\boxed{S_{n+1}(x) = S_n'(x) - 3x^2S_n(x).}$$

這對所有 $n \geq 0$ 都成立，其中 $S_0(x) = 1$（第零階導數就是函數本身）。

**$S_2$ 與 $S_3$**

由 $S_0 = 1$ 出發，遞推關係給出 $S_1 = 0 - 3x^2 = -3x^2$，這也可以直接看出。然後

$$S_2 = S_1' - 3x^2 S_1 = -6x - 3x^2(-3x^2) = 9x^4 - 6x,$$

正如題目所求。再用一次遞推關係，

$$\begin{aligned} S_3 &= S_2' - 3x^2 S_2 = (36x^3 - 6) - 3x^2(9x^4 - 6x) \\ &= 36x^3 - 6 - 27x^6 + 18x^3, \end{aligned}$$

所以

$$S_3(x) = -27x^6 + 54x^3 - 6.$$

**歸納法證明**

我們證明一個更強的命題，因為這樣做不用多花工夫，還能一併回答下一問：

$P(n)$：$S_n(x)$ 是一個次數爲 $2n$ 的多項式，其首項係數爲 $(-3)^n$。

**基礎情況**。$S_1(x) = -3x^2$ 的次數為 2，首項係數為 $-3 = (-3)^1$，所以 $P(1)$ 成立。（$P(0)$ 亦成立，因為 $S_0 = 1$。）

**歸納步驟**。假設 $P(k)$ 成立，即

$$S_k(x) = (-3)^k x^{2k} + (\text{其餘各項，次數均低於} 2k).$$

多項式的導數仍是多項式，$3x^2 S_k$ 也是多項式，所以 $S_{k+1} = S_k' - 3x^2 S_k$ 是多項式。至於它的次數，把兩部分分開來看：

- $S_k'$ 的次數至多為 $2k - 1$；
- $-3x^2 S_k(x) = -3(-3)^k x^{2k+2} + (\text{較低次項}) = (-3)^{k+1} x^{2k+2} + \cdots$。

> **想一想** $S_k'$ 會不會把 $-3x^2 S_k$ 的最高次項抵消掉？

不會：$S_k'$ 的次數至多為 $2k - 1$，遠低於 $2k + 2$，所以它碰不到 $x^{2k+2}$ 這一項。因此 $S_{k+1}$ 的次數恰好是 $2k + 2 = 2(k + 1)$，首項係數為 $(-3)^{k+1}$（非零）。這正是 $P(k + 1)$。

由歸納法，$P(n)$ 對所有 $n \geq 1$ 成立。所以 $S_n(x)$ 是一個階（即次數）為 $2n$ 的多項式，而 $x^{2n}$ 的係數為 $(-3)^n$。這與 $S_2$（$9 = (-3)^2$，次數為 4）及 $S_3$（$-27 = (-3)^3$，次數為 6）吻合。

### 最後一部分

假設 $S_n'(a) = 0$。在遞推關係中代入 $x = a$，導數項便消失：

$$S_{n+1}(a) = S_n'(a) - 3a^2 S_n(a) = -3a^2 S_n(a).$$

因此

$$S_n(a)\, S_{n+1}(a) = -3a^2 \big(S_n(a)\big)^2 \le 0,$$

因為 $a^2 \ge 0$ 且 $\big(S_n(a)\big)^2 \ge 0$。等號成立當且僅當 $a = 0$ 或 $S_n(a) = 0$。證明至此完成。

## 驗算

- 首項：$S_3$ 應以 $(-3)^3 x^6 = -27x^6$ 開首，事實確是如此。
- $S_3$ 的常數項：在 $x = 0$ 附近，$\mathrm{e}^{-x^3} = 1 - x^3 + \cdots$，所以它在 0 處的第三階導數是 $-3! = -6$。這與 $S_3(0) = -6$ 相符。
- 一個駐點（stationary point）：$S_3'(x) = -162x^5 + 162x^2$ 在 $x = 1$ 處為零。在該點 $S_3(1) = 21$ 而 $S_4(1) = -3 \cdot 21 = -63$，所以乘積為 $-1323 \le 0$。
- 我們亦在三十個隨機有理數點 $a$ 上，把 $S_0, \ldots, S_9$ 與 $\mathrm{e}^{-(a+h)^3}$ 的精確泰勒級數（Taylor series）作了對照，並在 $[-4, 4]$ 內 $S_1, \ldots, S_8$ 的每一個駐點上檢驗了該不等式。

## 常見錯誤

- 忘記說明為何次數恰好是 $2n + 2$。你需要首項係數非零，並且 $S_k'$ 無法把它抵消，這就是為甚麼值得把首項係數一併放進歸納假設之中。
- 只證明「$S_n$ 是多項式」，然後憑規律猜出次數。題目要求你從歸納論證本身得出次數。

## STEP 之外

這裏的多項式是埃爾米特多項式（Hermite polynomials）的近親，後者是用同一技巧對 $\mathrm{e}^{-x^2}$ 定義的（即所謂「羅德里格斯公式」（Rodrigues formula））。它們滿足與我們這裏相似的遞推關係，並出現於概率論和量子諧振子（quantum harmonic oscillator）之中。最後的不等式讓我們初次窺見，這類遞推關係如何支配一整族多項式的零點和轉向點（turning point）的位置。Boyce 與 DiPrima 的 *Elementary Differential Equations*（常微分方程入門教材）在討論級數解的一章中會遇到埃爾米特方程。

詳細解答

# STEP II 1999，第 4 題

純數學 • 二項式係數恆等式

> **STEP II 1999，第 4 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/99-s2-q4

## 關鍵思路

當兩個多項式（polynomial）恆等時，它們各次 $x$ 冪的係數（coefficient）都相同。因此我們選取兩個二項展開式（binomial expansion）的乘積，使其 $x^n$ 的係數正好是所求的和，然後從該乘積較簡單的形式再計算同一個係數。唯一真正要決定的，是選哪兩個因式相乘：把其中一個因式的 $x$ 換成 $-x$，便得到第 (i) 部分所需的正負號；把其中一個因式求導，便得到第 (ii) 部分多出來的 $t$。

## 解答

### 第一個結果

根據二項式定理（binomial theorem），

$$(1+x)^n = \sum_{r=0}^{n} \binom{n}{r} x^r.$$

把兩個這樣的展開式相乘。$x^n$ 項是由第一個括號取 $x^s$、第二個括號取 $x^{n-s}$ 而得，其中 $s$ 滿足 $0 \le s \le n$。所以 $(1+x)^n(1+x)^n$ 中 $x^n$ 的係數是

$$\sum_{s=0}^{n} \binom{n}{s}\binom{n}{n-s} = \sum_{s=0}^{n} \binom{n}{s}^2,$$

這裏用了對稱性 $\binom{n}{n-s} = \binom{n}{s}$，由階乘（factorial）公式可清楚看出。在恆等式的另一邊，$(1+x)^{2n}$ 中 $x^n$ 的係數是 $\binom{2n}{n}$。兩個多項式恆等，所以它們 $x^n$ 的係數相等：

$$\sum_{s=0}^{n} \binom{n}{s}^2 = \binom{2n}{n}.$$

### 第 (i) 部分

我們需要正負交替的符號，所以把第一個因式換成 $(1-x)^n = \sum_r \binom{n}{r}(-1)^r x^r$，並使用恆等式（identity）

$$(1-x)^n(1+x)^n \equiv (1-x^2)^n.$$

與之前一樣，左邊的 $x^n$ 項來自第一個括號的 $x^s$ 和第二個括號的 $x^{n-s}$，所以 $x^n$ 的係數是

$$\sum_{s=0}^{n}(-1)^s\binom{n}{s}\binom{n}{n-s} = \sum_{s=0}^{n}(-1)^s\binom{n}{s}^2.$$

在右邊，

$$(1-x^2)^n = \sum_{k=0}^{n}\binom{n}{k}(-1)^k x^{2k}$$

只含 $x$ 的偶數次冪。由於 $n$ 是偶數，$x^n$ 確實出現，即當 $2k=n$，亦即 $k=n/2$ 時，其係數為 $(-1)^{n/2}\binom{n}{n/2}$。比較係數，得

$$\sum_{s=0}^{n}(-1)^s\binom{n}{s}^2 = (-1)^{n/2}\binom{n}{n/2},$$

正是所求。（同樣的論證顯示，當 $n$ 是奇數時這個和等於 0，因為此時 $(1-x^2)^n$ 沒有 $x^n$ 項。）

### 第 (ii) 部分

> **想一想** 怎樣才能令 $\binom{n}{t}$ 旁邊出現一個因子 $t$？

求導（differentiation）便可做到：$\dfrac{\mathrm{d}}{\mathrm{d}x}x^t = t\,x^{t-1}$。把二項展開式求導，得

$$n(1+x)^{n-1} = \sum_{t=1}^{n} t\binom{n}{t}x^{t-1}.$$

現在乘以 $(1+x)^n = \sum_u \binom{n}{u}x^u$，並使用恆等式

$$n(1+x)^{n-1}(1+x)^n \equiv n(1+x)^{2n-1}.$$

在左邊，$x^{n-1}$ 項來自第一個括號的 $t\binom{n}{t}x^{t-1}$ 乘以第二個括號的 $\binom{n}{n-t}x^{n-t}$，所以 $x^{n-1}$ 的係數是

$$\sum_{t=1}^{n} t\binom{n}{t}\binom{n}{n-t} = \sum_{t=1}^{n} t\binom{n}{t}^2.$$

在右邊，$n(1+x)^{2n-1}$ 中 $x^{n-1}$ 的係數是 $n\binom{2n-1}{n-1}$。最後，

$$\binom{2n}{n} = \frac{(2n)!}{n!\,n!} = \frac{2n}{n}\cdot\frac{(2n-1)!}{(n-1)!\,n!} = 2\binom{2n-1}{n-1},$$

所以比較係數後乘以二，得

$$\sum_{t=1}^{n} 2t\binom{n}{t}^2 = 2n\binom{2n-1}{n-1} = n\binom{2n}{n}.$$

*另一方法*：有一個巧妙的對稱論證。把 $t$ 換成 $n-t$，並利用 $\binom{n}{n-t} = \binom{n}{t}$，可見 $\sum_{t=0}^{n} t\binom{n}{t}^2 = \sum_{t=0}^{n}(n-t)\binom{n}{t}^2$。把同一個和的這兩個表達式相加，得

$$\sum_{t=0}^{n} 2t\binom{n}{t}^2 = n\sum_{t=0}^{n}\binom{n}{t}^2 = n\binom{2n}{n}$$

這是根據第一個結果，而 $t=0$ 的一項為零。

## 驗算

取 $n=4$，此時帕斯卡三角形（Pascal's triangle）的相應一行是 $1,4,6,4,1$。於是 $1+16+36+16+1 = 70 = \binom{8}{4}$；交錯和是 $1-16+36-16+1 = 6 = (-1)^2\binom{4}{2}$；而 $2(1\cdot 16 + 2\cdot 36 + 3\cdot 16 + 4\cdot 1) = 280 = 4\times 70$。我們亦已對 39 或以下的每一個 $n$ 精確驗證過全部三個恆等式。

## 常見錯誤

- 忘記在第 (i) 部分中右邊只有偶數次冪，所以正正因為 $n$ 是偶數，$x^n$ 才會出現。
- 在第 (ii) 部分比較了錯誤的冪：求導之後指數減一，所以我們需要的是 $x^{n-1}$，而不是 $x^n$。

## STEP 之外

第一個恆等式是范德蒙恆等式（Vandermonde's identity）$\sum_s \binom{m}{s}\binom{n}{k-s} = \binom{m+n}{k}$ 的特殊情況，後者亦有一個計數證明：從 $m$ 名女士和 $n$ 名男士中選出 $k$ 人，方法是先決定選多少名女士。用多項式作為記錄數列的工具，就是*生成函數*（generating functions）的方法，它是組合數學（combinatorics）和概率論中的核心工具。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率論入門）全書都使用生成函數，而 Siklos 的 *Advanced Problems in Mathematics*（高等數學難題）則有更多這種類型的題目。

詳細解答

# STEP II 1999，第 5 題

純數學 • 三角方程

**STEP II 1999，第 5 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/99-s2-q5

## 關鍵思路

有兩條自然的路線。一是以 $\cos\alpha$ 為未知數：先把 $\sin\alpha$ 分離出來，兩邊平方，再利用 $\sin^2+\cos^2=1$ 得出一條關於 $\cos\alpha$ 的二次方程（quadratic equation）。二是把 $5\cos x+12\sin x$ 寫成單一餘弦 $13\cos(x-\beta)$，這個寫法同時說明二次方程的兩個根確實都會出現。至於最後一部分，由於餘弦在 $[0,\pi]$ 上遞減，$\alpha<\frac{3}{4}\pi$ 其實是一個關於 $\cos\alpha$ 的命題。於是問題歸結為比較兩個根式（surd）的大小。

## 解答

### $\cos\alpha$ 的兩個值

設 $c=\cos\alpha$。由方程得 $12\sin\alpha=7-5c$。兩邊平方，並利用 $\sin^2\alpha=1-c^2$，

$$144(1-c^2)=49-70c+25c^2,$$
$$169c^2-70c-95=0.$$

由二次公式（quadratic formula）得

$$c=\frac{70\pm\sqrt{70^2+4\cdot169\cdot95}}{2\cdot169}=\frac{70\pm\sqrt{69120}}{338}.$$

由於 $69120=576\times120=24^2\times120$，故 $\sqrt{69120}=24\sqrt{120}$。因此

$$\cos\alpha=\frac{35-12\sqrt{120}}{169}\qquad\text{或}\qquad\cos\alpha=\frac{35+12\sqrt{120}}{169},$$

這就是所求的結果，而另一個值是 $\dfrac{35+12\sqrt{120}}{169}$。

**想一想** 平方可能會產生增根。這兩個值是否真的都來自原方程的解？

是的，而單一餘弦的形式正好說明原因。設 $\beta$ 為滿足 $\cos\beta = \frac{5}{13}$ 及 $\sin\beta = \frac{12}{13}$ 的銳角（acute angle）。則

$$5\cos x + 12\sin x = 13\cos(x-\beta),$$

所以方程即是 $\cos(x-\beta) = \frac{7}{13}$。設 $\gamma$ 為滿足 $\cos\gamma = \frac{7}{13}$ 的銳角，於是 $\sin\gamma = \frac{\sqrt{169-49}}{13} = \frac{\sqrt{120}}{13}$。方程的解為 $x = \beta \pm \gamma + 2k\pi$，而

$$\cos(\beta\pm\gamma) = \cos\beta\cos\gamma \mp \sin\beta\sin\gamma = \frac{35 \mp 12\sqrt{120}}{169}.$$

所以兩個符號各自給出一個真正的解：$\alpha = \beta + \gamma$ 對應負號，$\alpha = \beta - \gamma$ 對應正號。

### 第二象限的解

假設 $\frac{1}{2}\pi < \alpha < \pi$。則 $\cos\alpha < 0$。由於 $\sqrt{120} > 10$，可知 $35 + 12\sqrt{120} > 0$，所以這個根被排除。另外 $12\sqrt{120} > 120 > 35$，所以另一個根是負數。因此

$$\cos\alpha = \frac{35 - 12\sqrt{120}}{169}.$$

餘弦在 $[0,\pi]$ 上嚴格遞減（strictly decreasing）。$\alpha$ 和 $\frac{3}{4}\pi$ 都在這個區間內，所以

$$\alpha < \tfrac{3}{4}\pi \iff \cos\alpha > \cos\tfrac{3}{4}\pi = -\tfrac{1}{\sqrt{2}}.$$

因此我們須證明 $\dfrac{12\sqrt{120}-35}{169} < \dfrac{1}{\sqrt{2}}$，即 $12\sqrt{120} - 35 < \dfrac{169}{\sqrt{2}}$。

**想一想** 不用計算機，又要避免把可能是負數的量平方，我們怎樣比較這些根式的大小？

我們先粗略地估計 $\sqrt{120}$ 的上界（bound）。因為 $120 < 121$，所以 $\sqrt{120} < 11$，於是

$$12\sqrt{120} - 35 < 132 - 35 = 97.$$

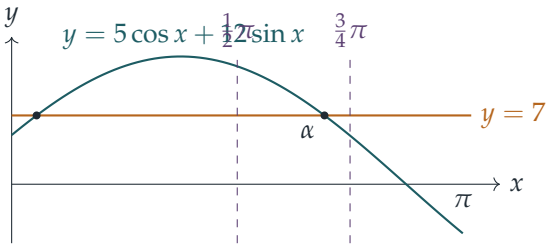
現在只需證明 $97 < \dfrac{169}{\sqrt{2}}$，即 $97\sqrt{2} < 169$。兩邊皆為正數，所以這等價於兩邊平方後的不等式（inequality）：

$$2\times 97^2 = 18\,818 < 28\,561 = 169^2,$$

而這是成立的。綜合以上各步，

$$\cos\alpha = -\frac{12\sqrt{120}-35}{169} > -\frac{97}{169} > -\frac{1}{\sqrt{2}} = \cos\tfrac{3}{4}\pi,$$

又因為餘弦在 $[0,\pi]$ 上遞減，所以 $\alpha < \frac{3}{4}\pi$，證明完畢。




## 驗算

- 以數值計算，在 $[0, 2\pi)$ 內的根為 $x \approx 0.1738$ 和 $x \approx 2.1782$。它們的餘弦值分別是 0.98493 和 $-0.57073$，與 $(35 \pm 12\sqrt{120})/169$ 吻合。
- 第二象限的根是 2.1782，小於 $\frac{3}{4}\pi \approx 2.3562$，正如圖像所示。兩者相差的餘裕相當充足：$\cos\alpha \approx -0.571$，而 $-1/\sqrt{2} \approx -0.707$。這就是粗略的估計 $\sqrt{120} < 11$ 已經足夠的原因。
- 作為對二次方程的合理性檢查，兩根之積為 $-\frac{95}{169}$，兩根之和為 $\frac{70}{169}$。

## 常見錯誤

- 沒有檢查正負號就把不等式兩邊平方。這裏 $12\sqrt{120} - 35$ 和 $169/\sqrt{2}$ 都是正數，但理由應當寫明。
- 忘記餘弦在 $[0, \pi]$ 上是遞減的，因而不等號要反轉：$\alpha < \frac{3}{4}\pi$ 對應的是 $\cos\alpha > -\frac{1}{\sqrt{2}}$。
- 沒有解釋為何在第二象限中要排除正根。

## STEP 之外

最後一部分是嚴格估計的一個小練習。你把一個難以處理的量換成一個較簡單的界，再檢查餘裕是否仍然足夠；分析學（analysis）中許多不等式正是這樣證明的。由餘弦值推斷角的大小這一步，用到了嚴格單調函數（strictly monotonic function）保持（或反轉）次序這個事實，而這正是大學第一門分析課中反函數（inverse function）背後的想法。如想進一步了解如何嚴謹地處理不等式的證明，可參閱 Abbott 的 *Understanding Analysis*（分析學入門教材）。

詳細解答

# STEP II 1999，第 6 題

純數學 • 代換積分法

**STEP II 1999，第 6 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/99-s2-q6

## 關鍵思路

分式 $\frac{ax+b}{cx+d}$ 的導數（derivative）是一個常數除以 $(cx+d)^2$。因此，只要被積函數（integrand）是 $\frac{1}{(x+3)^2}$ 乘以 $\frac{x+1}{x+3}$ 或 $\frac{x+2}{x+3}$ 的函數，把該分式代換為新變數，就能把 $\frac{1}{(x+3)^2}$ 完全吸收，餘下的最多只是 $\int \ln y \, \mathrm{d}y$。其餘兩個積分則可由對數定律（laws of logarithms）推出。

## 解答

### 導數

由商法則（quotient rule），

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{a(cx+d) - c(ax+b)}{(cx+d)^2} = \frac{ad-bc}{(cx+d)^2}.$$

分子中含 $x$ 的項互相抵消，只餘下一個常數。整道題之所以行得通，關鍵正在於此。

### 第一個積分

設 $y = \dfrac{x+1}{x+3}$。這符合 $(*)$ 的形式，其中 $a=1$，$b=1$，$c=1$，$d=3$，所以

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{3-1}{(x+3)^2} = \frac{2}{(x+3)^2}, \qquad \text{即} \qquad \frac{\mathrm{d}x}{(x+3)^2} = \tfrac{1}{2}\,\mathrm{d}y.$$

當 $x=0$ 時，$y=\frac{1}{3}$；當 $x=1$ 時，$y=\frac{1}{2}$。利用 $\int \ln y\,\mathrm{d}y = y\ln y - y$（以 $u=\ln y$ 作分部積分法（integration by parts）），可得：

$$\begin{aligned}\int_0^1 \frac{1}{(x+3)^2}\ln\left(\frac{x+1}{x+3}\right)\mathrm{d}x &= \frac{1}{2}\int_{1/3}^{1/2}\ln y\,\mathrm{d}y = \frac{1}{2}\Big[y\ln y - y\Big]_{1/3}^{1/2}\\ &= \frac{1}{2}\left(-\tfrac{1}{2}\ln 2 - \tfrac{1}{2} + \tfrac{1}{3}\ln 3 + \tfrac{1}{3}\right)\\ &= \frac{1}{6}\ln 3 - \frac{1}{4}\ln 2 - \frac{1}{12},\end{aligned}$$

其中用到 $\ln\frac{1}{2} = -\ln 2$ 和 $\ln\frac{1}{3} = -\ln 3$。這正是題目要求的結果。把這個積分記為 $I_1$。

### 一個相伴的積分

**想一想** $x^2+3x+2$ 與我們剛才用過的代換有甚麼關係？

它可因式分解為 $(x+1)(x+2)$，所以

$$\frac{x^2+3x+2}{(x+3)^2} = \frac{x+1}{x+3}\cdot\frac{x+2}{x+3} \qquad 及 \qquad \frac{x+1}{x+2} = \frac{x+1}{x+3}\Big/\frac{x+2}{x+3}.$$

因此，最後兩個積分中的所有對數，都是 $\ln\frac{x+1}{x+3}$ 與 $\ln\frac{x+2}{x+3}$ 的和與差（當 $0\le x\le 1$ 時，這些分式全部為正）。所以我們還需要多計算一個積分：

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

設 $y=\dfrac{x+2}{x+3}$。這次 $a=1$，$b=2$，$c=1$，$d=3$，所以 $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{1}{(x+3)^2}$，而積分上下限變為 $y=\frac{2}{3}$ 和 $y=\frac{3}{4}$：

$$\begin{aligned}J &= \int_{2/3}^{3/4}\ln y\,\mathrm{d}y = \Big[y\ln y - y\Big]_{2/3}^{3/4}\\ &= \tfrac{3}{4}(\ln 3 - 2\ln 2) - \tfrac{3}{4} - \tfrac{2}{3}(\ln 2 - \ln 3) + \tfrac{2}{3}\\ &= \frac{17}{12}\ln 3 - \frac{13}{6}\ln 2 - \frac{1}{12}.\end{aligned}$$

### 餘下的兩個積分

由於 $\ln\frac{x^2+3x+2}{(x+3)^2} = \ln\frac{x+1}{x+3} + \ln\frac{x+2}{x+3}$，第二個積分是

$$\begin{aligned}I_1 + J &= \left(\tfrac{1}{6}+\tfrac{17}{12}\right)\ln 3 - \left(\tfrac{1}{4}+\tfrac{13}{6}\right)\ln 2 - \tfrac{1}{12} - \tfrac{1}{12}\\ &= \frac{19}{12}\ln 3 - \frac{29}{12}\ln 2 - \frac{1}{6}.\end{aligned}$$

由於 $\ln\frac{x+1}{x+2} = \ln\frac{x+1}{x+3} - \ln\frac{x+2}{x+3}$，第三個積分是

$$\begin{aligned} I_1 - J &= \left(\tfrac{1}{6} - \tfrac{17}{12}\right)\ln 3 - \left(\tfrac{1}{4} - \tfrac{13}{6}\right)\ln 2 \\ &= \frac{23}{12}\ln 2 - \frac{5}{4}\ln 3, \end{aligned}$$

其中常數項 $-\frac{1}{12}$ 已互相抵消。

## 驗算

- 在 $[0,1]$ 上使用辛普森法則（Simpson's rule），得 $I_1 \approx -0.07352$，$J \approx -0.02878$，而第二和第三個積分分別 $\approx -0.10230$ 和 $-0.04473$，全部與精確形式吻合至小數點後十位。
- 粗略估計上下界：$\int_0^1 (x+3)^{-2}\,\mathrm{d}x = \frac{1}{12}$，而 $\ln\frac{x+1}{x+2}$ 介乎 $\ln\frac{1}{2} \approx -0.69$ 與 $\ln\frac{2}{3} \approx -0.41$ 之間。所以第三個積分必定介乎約 $-0.058$ 與 $-0.034$ 之間，而 $-0.0447$ 確實在此範圍內。
- 每個被積函數在 $[0,1]$ 上都是負數（對數內的每個分式都小於 1），所以三個答案都應該是負數，而事實亦是如此。

## 常見錯誤

- 在第一個代換中漏掉因子 $\frac{1}{2}$：該處 $ad - bc = 2$，但在第二個代換中則是 1。
- 沒有更改積分上下限，或在代換後仍沿用 $x$ 的上下限 0 和 1。
- 為 $\frac{x+1}{x+2}$ 另作一個新的代換：這樣做是可行的，但 $(x+3)^{-2}$ 以 $y$ 表示時會變得相當繁複。把對數拆開要快得多。

## STEP 之外

形如 $x \mapsto \frac{ax+b}{cx+d}$ 而 $ad - bc \neq 0$ 的映射，稱為莫比烏斯變換（*Möbius transformations*）。把兩個這樣的變換複合，相當於把 $2\times 2$ 矩陣 $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 相乘，而 $ad - bc$ 就是行列式（determinant），這正是它出現在導數中的原因。在複分析（complex analysis）中，它們恰好就是擴充複平面（extended plane）上的全部共形雙射（conformal bijections），並把圓和直線映射為圓和直線。Needham 的 *Visual Complex Analysis*（《複分析：可視化方法》）中有一章對此有十分精彩的論述。

詳細解答

# STEP II 1999，第 7 題

純數學 • 曲線描繪

**STEP II 1999，第 7 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/99-s2-q7

## 關鍵思路

先求導，再把所有項通分到同一個公分母上：分子幾乎完全消去——所有二次項互相抵消，只剩下一個一次因式，駐點（stationary point）的位置隨即確定。至於描繪曲線，則把根號內的式子配方（complete the square）。這樣便知道曲線在哪裏有定義、當 $|x|$ 很大時有何表現，以及當 $a = 1$ 時平方根會變成絕對值（modulus），令曲線變成兩段等軸雙曲線（rectangular hyperbola）。

## 解答

設 $Q(x) = x^2 - 2x + a = (x-1)^2 + (a-1)$。

### 當 $a > 1$ 時恰有一個駐點

若 $a > 1$，則對任意 $x$ 均有 $Q(x) \geq a - 1 > 0$，所以曲線對所有實數 $x$ 都有定義，而且可導。由 $y = xQ^{-1/2}$ 及 $Q'(x) = 2x - 2$，用乘積法則（product rule）可得

$$\begin{aligned}\frac{\mathrm{d}y}{\mathrm{d}x} &= Q^{-1/2} - \tfrac{1}{2}x(2x-2)Q^{-3/2} = \frac{Q - x(x-1)}{Q^{3/2}} \\ &= \frac{x^2 - 2x + a - x^2 + x}{Q^{3/2}} = \frac{a - x}{(x^2 - 2x + a)^{3/2}}.\end{aligned}$$

對任意 $x$，分母都是正數而且有限，所以 $\mathrm{d}y/\mathrm{d}x = 0$ 當且僅當 $x = a$。因此 $C$ 恰有一個駐點，位於

$$x = a, \qquad y = \frac{a}{\sqrt{a^2 - a}} = \sqrt{\frac{a}{a-1}},$$

正如題目所求。由於 $\mathrm{d}y/\mathrm{d}x$ 的正負號與 $a-x$ 相同，曲線在 $x<a$ 時上升，在 $x>a$ 時下降，所以這一點是極大點（maximum）。

**想一想** 條件 $a>1$ 用在哪裏？

用了兩次：其一，它保證 $Q>0$ 處處成立，所以曲線沒有缺口，也沒有導數不存在的點；其二，它保證 $x=a$ 確實是曲線上的一點，因為 $Q(a)=a^2-a>0$。當 $a=1$ 時，原本應是駐點的 $x=a=1$ 正好是 $Q=0$ 之處，第 (ii) 部分出問題的地方就在這裏。

另外有兩項事實對描繪兩條曲線都有幫助。當 $|x|$ 很大時，$\sqrt{Q}=|x|\sqrt{1-2/x+a/x^2}\approx |x|$，所以

$$y\to 1 \text{ 當} x\to+\infty, \qquad y\to -1 \text{ 當} x\to-\infty.$$

此外，$y$ 的正負號與 $x$ 相同，而且曲線通過原點，在該處的斜率（gradient）為 $a/a^{3/2}=1/\sqrt{a}$。

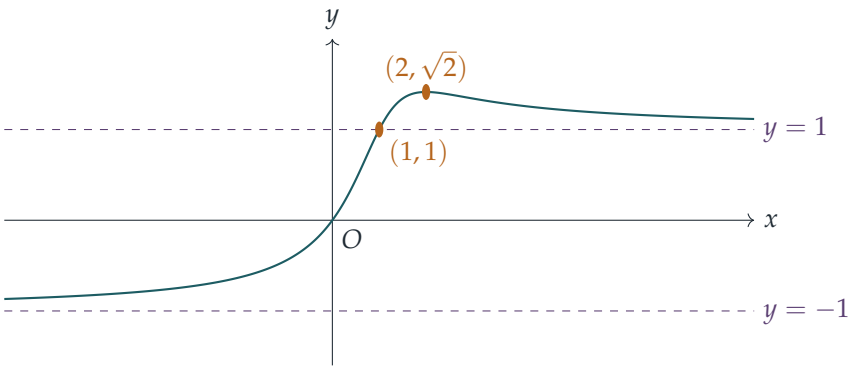
**第 (i) 部分：$a=2$**

這時 $y=\dfrac{x}{\sqrt{(x-1)^2+1}}$，對所有 $x$ 都有定義，並且

- 在 $(2,\sqrt{2})$ 有極大點，在 $x<2$ 時遞增，在 $x>2$ 時遞減；
- 原點在曲線上，該處的斜率為 $1/\sqrt{2}$；
- 有水平漸近線（asymptote）$y=1$（當 $x\to\infty$）和 $y=-1$（當 $x\to-\infty$）。

曲線在每條漸近線的哪一側？設 $y=\pm1$ 並把兩邊平方，得 $x^2=x^2-2x+2$，所以 $x=1$，而且只有 $y=+1$ 符合（因為 $y$ 的正負號與 $x$ 相同）。因此曲線與 $y=1$ 只相交一次，交點為 $(1,1)$：對所有 $x>1$，曲線都在 $y=1$ 之上，過了極大點之後從上方向漸近線下降。曲線永不與 $y=-1$ 相交，所以它一直在 $y=-1$ 之上，並隨着 $x$ 由 $-\infty$ 增大而上升，逐漸離開這條漸近線。

（如果想知道更多細節：再求導一次可得 $\mathrm{d}^2y/\mathrm{d}x^2=(2x^2-7x+4)/Q^{5/2}$，所以拐點（point of inflection）位於 $x=\frac{1}{4}(7\pm\sqrt{17})$，大約是 $x\approx0.72$ 和 $x\approx2.78$。描繪草圖時用不着這些，但它們可以解釋曲線的形狀。）




**第 (ii) 部分：$a = 1$**

這時 $Q = (x-1)^2$，所以要排除 $x = 1$，而且由於平方根取正值，

$$\sqrt{(x-1)^2} = |x-1|.$$

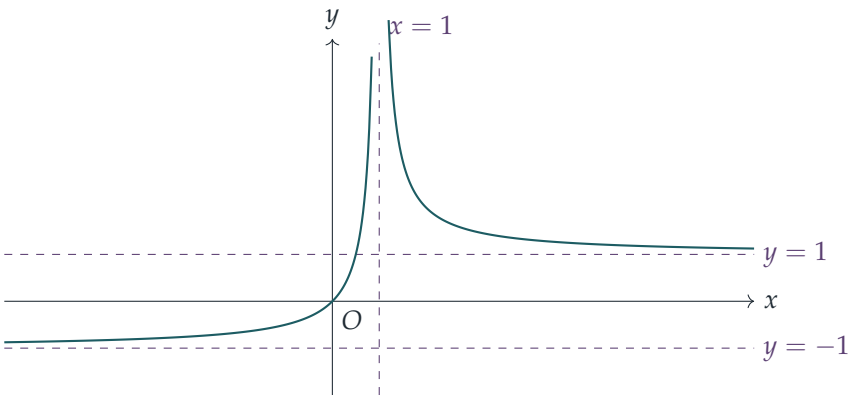
因此

$$y = \frac{x}{|x-1|} = \begin{cases} \dfrac{x}{1-x} = -1 + \dfrac{1}{1-x}, & x < 1, \\ \dfrac{x}{x-1} = 1 + \dfrac{1}{x-1}, & x > 1. \end{cases}$$

每一支都是以 $x = 1$ 為中心的等軸雙曲線的一部分。

- 當 $x < 1$：$y$ 遞增（其導數為 $1/(1-x)^2 > 0$），由 $-1$（當 $x \to -\infty$ 時從上方趨近）出發，以斜率 1 通過原點，再在 $x \to 1^-$ 時趨向 $+\infty$。
- 當 $x > 1$：$y$ 遞減（導數為 $-1/(x-1)^2 < 0$），由 $x \to 1^+$ 時的 $+\infty$ 下降，趨向 1（當 $x \to \infty$ 時從上方趨近）。

直線 $x = 1$ 是鉛垂漸近線，而且在兩側都有 $y \to +\infty$（分子接近 1，分母是很小的正數）。曲線沒有駐點：$a > 1$ 的曲線的極大點高度為 $\sqrt{a/(a-1)}$，當 $a \to 1^+$ 時，這個極大點已被推到無窮遠處。一般公式亦與此吻合，因為 $\mathrm{d}y/\mathrm{d}x = (1-x)/|x-1|^3$ 永不為零。




## 驗算

- 以數值方法檢驗：在數千個隨機點上，導數公式 $(a-x)/Q^{3/2}$ 與有限差分（finite difference）求得的導數吻合；而對若干個 $a>1$ 的值逐點檢查正負號，都恰好找到一次變號。
- 當 $a=2$：$y(2)=2/\sqrt{2}=\sqrt{2}\approx 1.414$，$y(1)=1$，而 $y(10^6)\approx 1.000001$，$y(-10^6)\approx -0.999999$，證實了曲線從哪一側趨近各漸近線。
- 當 $a=1$：$y(0.999)\approx 999$ 及 $y(1.001)\approx 1001$，兩者都是很大的正數，與上文所述一致。
- 當 $a\to\infty$ 時，極大點的高度 $\sqrt{a/(a-1)}$ 趨向 1，這是合理的：當 $a$ 極大時，曲線在原點附近近似於 $x/\sqrt{a}$，隆起的部分隨之變得平坦。

## 常見錯誤

- 在第 (ii) 部分寫成 $\sqrt{(x-1)^2}=x-1$。這樣會處處得到 $y=x/(x-1)$，左邊的一支因而出錯：它會在 $x\to-\infty$ 時趨向 $+1$，並在 $x\to 1^-$ 時趨向 $-\infty$。
- 把曲線兩端的漸近線都當作 $y=1$：由於 $\sqrt{x^2}=|x|$，當 $x\to-\infty$ 時極限是 $-1$。
- 忘記說明當 $a>1$ 時分母永不為零；缺少這一步，「分子只在 $x=a$ 處為零」尚不足以證明恰有一個駐點。

## STEP 之外

這個曲線族的性質在 $a=1$ 處突然改變：一條只有一個極大點的光滑曲線，變成了兩支分離的曲線，並帶有一條鉛垂漸近線。研究一個曲線族的定性形狀如何隨參數而變，正

是分歧理論（bifurcation theory）的起點；而處理 $\sqrt{x^2} = |x|$ 以及 $x \to \pm\infty$ 時的極限所需的那份謹慎，正是大學一年級分析課要嚴格闡明的。Abbott 的 *Understanding Analysis* 平易近人，很適合用來了解極限與連續性的嚴謹處理方法。

詳細解答

# STEP II 1999，第 8 題

純數學 • 三角級數

**STEP II 1999，第 8 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/99-s2-q8

## 關鍵思路

把 $\sin k\theta$ 乘以 $2\sin\frac{1}{2}\theta$，利用積化和差公式（product-to-sum formula）便可將它化成兩個餘弦之*差*，於是整個和式可以裂項相消（telescope）。之後兩部分的關鍵都在於選取合適的 $\theta$：取 $\theta = \pi/n$ 時，那個難纏的餘弦 $\cos(n+\frac{1}{2})\theta$ 會變成 $\cos\frac{1}{2}\theta$ 的簡單倍數。至於 (ii)，半角公式（half-angle formula）可把 $\sin^2$ 化為餘弦，而和式 $\sum k\cos k\theta$ 正好就是對 $(*)$ 求導所得的結果。

## 解答

**證明** $(*)$

假設 $\sin\frac{1}{2}\theta \neq 0$，使右邊有意義。在恆等式（identity）$2\sin A\sin B = \cos(A-B) - \cos(A+B)$ 中取 $A = k\theta$ 及 $B = \frac{1}{2}\theta$，得

$$2\sin\tfrac{1}{2}\theta\,\sin k\theta = \cos\left(k-\tfrac{1}{2}\right)\theta - \cos\left(k+\tfrac{1}{2}\right)\theta.$$

由 $k=0$ 加至 $n$，第 $k$ 行所減去的 $\cos(k+\frac{1}{2})\theta$ 會在第 $k+1$ 行加回來，因此只剩下首項和末項：

$$2\sin\tfrac{1}{2}\theta\sum_{k=0}^{n}\sin k\theta = \cos\left(-\tfrac{1}{2}\theta\right) - \cos\left(n+\tfrac{1}{2}\right)\theta = \cos\tfrac{1}{2}\theta - \cos\left(n+\tfrac{1}{2}\right)\theta,$$

因為餘弦是偶函數（even function）。兩邊除以 $2\sin\frac{1}{2}\theta$ 即得 $(*)$。

### 第 (i) 部分

取 $\theta = \pi/n$（其中 $n \geq 1$），則 $\sin\frac{1}{2}\theta = \sin\frac{\pi}{2n} \neq 0$。於是

$$\cos\left(n+\tfrac{1}{2}\right)\theta = \cos\left(\pi + \tfrac{\pi}{2n}\right) = -\cos\tfrac{\pi}{2n},$$

所以 $(*)$ 的分子是 $2\cos\frac{\pi}{2n}$，而

$$\sum_{k=0}^{n} \sin\left(\frac{k\pi}{n}\right) = \frac{2\cos\frac{\pi}{2n}}{2\sin\frac{\pi}{2n}} = \cot\frac{\pi}{2n}.$$

這是精確的結果。當 $n$ 很大時，$\frac{\pi}{2n}$ 很小，故 $\cos\frac{\pi}{2n} \approx 1$ 而 $\sin\frac{\pi}{2n} \approx \frac{\pi}{2n}$，由此得

$$\sum_{k=0}^{n} \sin\left(\frac{k\pi}{n}\right) \approx \frac{1}{\pi/(2n)} = \frac{2n}{\pi}.$$

### 第 (ii) 部分

由倍角公式（double-angle formula），$\sin^2\frac{k\pi}{2n} = \frac{1}{2}\left(1 - \cos\frac{k\pi}{n}\right)$，所以取 $\theta = \pi/n$ 時

$$\sum_{k=0}^{n} k\sin^2\left(\frac{k\pi}{2n}\right) = \frac{1}{2}\sum_{k=0}^{n} k - \frac{1}{2}\sum_{k=0}^{n} k\cos k\theta = \frac{n(n+1)}{4} - \frac{1}{2}\sum_{k=0}^{n} k\cos k\theta. \qquad (1)$$

要求出 $\sum k\cos k\theta$，我們把 $(*)$ 對 $\theta$ 求導；左邊正好變成這個和式。先把右邊拆成

$$\frac{1}{2}\cot\tfrac{1}{2}\theta - \frac{\cos(n+\frac{1}{2})\theta}{2\sin\frac{1}{2}\theta}.$$

記 $S = \sin\frac{1}{2}\theta$ 及 $c = \cos\frac{1}{2}\theta$。第一部分求導後得 $-\frac{1}{4}\operatorname{cosec}^2\frac{1}{2}\theta = -\frac{1}{4S^2}$。根據商法則（quotient rule），第二部分求導後得

$$\frac{(n+\frac{1}{2})\sin(n+\frac{1}{2})\theta \cdot 2S + \cos(n+\frac{1}{2})\theta \cdot c}{4S^2}.$$

因此

$$\sum_{k=0}^{n} k\cos k\theta = \frac{-1 + (2n+1)S\sin(n+\frac{1}{2})\theta + c\cos(n+\frac{1}{2})\theta}{4S^2}.$$

**想一想** 為甚麼要在求導之後才代入 $\theta = \pi/n$，而不是之前？

恆等式 $(*)$ 對所有 $\theta$（只要 $\sin\frac{1}{2}\theta \neq 0$）都成立，所以我們可以對它求導。假如先固定 $\theta = \pi/n$，得到的只是一個數，根本無從求導。一旦得到以 $\theta$ 表示的導數恆等式，我們便可以代入任何想要的值。

現在取 $\theta = \pi/n$。與 (i) 一樣，$(n+\frac{1}{2})\theta = \pi + \frac{1}{2}\theta$，所以 $\sin(n+\frac{1}{2})\theta = -S$ 而 $\cos(n+\frac{1}{2})\theta = -c$。因此

$$\sum_{k=0}^{n} k\cos\frac{k\pi}{n} = \frac{-1-(2n+1)S^2 - c^2}{4S^2} = \frac{-2-2nS^2}{4S^2} = -\frac{1}{2S^2} - \frac{n}{2},$$

其中用了 $c^2 = 1 - S^2$。代入 (1)：

$$\sum_{k=0}^{n} k\sin^2\left(\frac{k\pi}{2n}\right) = \frac{n(n+1)}{4} + \frac{n}{4} + \frac{1}{4S^2} = \frac{n(n+2)}{4} + \frac{1}{4\sin^2\frac{\pi}{2n}}.$$

這是精確的結果。當 $n$ 很大時，$\sin\frac{\pi}{2n} \approx \frac{\pi}{2n}$，故 $\frac{1}{4\sin^2(\pi/2n)} \approx \frac{n^2}{\pi^2}$；而 $\frac{n(n+2)}{4} = \frac{n^2}{4} + \frac{n}{2}$，其中第二項與 $n^2$ 相比可以忽略。所以

$$\sum_{k=0}^{n} k\sin^2\left(\frac{k\pi}{2n}\right) \approx \left(\frac{1}{4} + \frac{1}{\pi^2}\right) n^2,$$

正如題目所求。

## 驗算

- 在 (ii) 中取 $n = 1$：和為 $0 + 1\cdot\sin^2\frac{\pi}{2} = 1$，而精確公式給出 $\frac{3}{4} + \frac{1}{4} = 1$。
- 在 (i) 中取 $n = 2$：和為 $0 + 1 + 0 = 1$，而 $\cot\frac{\pi}{4} = 1$。
- 數值上，當 $n = 1000$ 時，(i) 的和是 $636.6192\ldots$，而 $2n/\pi = 636.6198\ldots$；(ii) 的和是 $351\,821.3$，與 $(\frac{1}{4} + \frac{1}{\pi^2})n^2$ 相差不超過 $0.15\%$。(ii) 的相對誤差（relative error）以 $1/n$ 的速度縮小，原因是略去了 $\frac{n}{2}$ 這一項。

## 常見錯誤

- 寫成 $\cos(\pi + x) = \cos x$：正確的值是 $-\cos x$，而正是這個負號令 (i) 的分子變成兩倍。
- 在 (ii) 中漏掉 $\frac{1}{2}\sum k = \frac{n(n+1)}{4}$ 這一項，而 $\frac{1}{4}n^2$ 正是由它而來。

## STEP 之外

第 (i) 部分其實是喬裝了的黎曼和（Riemann sum）：$\frac{\pi}{n}\sum\sin\frac{k\pi}{n} \to \int_0^{\pi}\sin x\,\mathrm{d}x = 2$，這正正就是 $\sum \approx 2n/\pi$ 的意思。同樣地，(ii) 所說的是 $\frac{1}{n^2}\sum k\sin^2\frac{k\pi}{2n} \to \int_0^1 x\sin^2\frac{\pi x}{2}\,\mathrm{d}x = \frac{1}{4} + \frac{1}{\pi^2}$。把同一個裂項技巧用於 $\sum\cos k\theta$，便會得到傅立葉分析（Fourier analysis）中的狄利克雷核（Dirichlet kernel）。Abbott 的 *Understanding Analysis*（數學分析入門教材）對黎曼和與傅立葉級數（Fourier series）有嚴謹的討論。

詳細解答

# STEP II 1999，第 9 題

力學 • 變力作用下的運動

> **STEP II 1999，第 9 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/99-s2-q9

## 關鍵思路

加速度（acceleration）只取決於距離 $x$，與時間無關，所以最合適的工具是 $\ddot{x} = v\dfrac{\mathrm{d}v}{\mathrm{d}x}$。這樣一來，運動方程（equation of motion）就變成一條可以直接積分的方程，從而得出 $v^2$ 作為 $x$ 的函數（其實就是一條喬裝了的能量方程（energy equation））。餘下的工作，都是在問 $v^2$ 在哪裏降至零：碎片就是在那裏停下，然後掉頭返回。

## 解答

我們把每塊碎片看成一個質點（particle），沿着一條通過定點 $G$ 的直線運動，以 $x$ 表示它與 $G$ 的距離，$v = \dot{x}$ 表示它的速度，並規定遠離 $G$ 的方向為正。起初 $x = 0$ 而 $v = u > 0$。由於加速度指向 $G$，在這個規定下它是負的。

### 第 (i) 部分

當 $x < R$ 時，$\ddot{x} = -k_1 x^3$，寫成 $\ddot{x} = v\dfrac{\mathrm{d}v}{\mathrm{d}x}$ 便得

$$v\frac{\mathrm{d}v}{\mathrm{d}x} = -k_1 x^3.$$

分離變數，並由起點（$x = 0$，$v = u$）開始積分：

$$\tfrac{1}{2}v^2 - \tfrac{1}{2}u^2 = -\tfrac{1}{4}k_1 x^4, \qquad \text{所以} \qquad v^2 = u^2 - \tfrac{1}{2}k_1 x^4 \quad (x < R).$$

只要碎片仍在向外運動，便有 $v = \sqrt{u^2 - \tfrac{1}{2}k_1 x^4}$。

**第 (ii) 部分**

由於 $2a^2 = k_1R^4$，(i) 的公式可寫成 $v^2 = u^2 - a^2(x/R)^4$。當 $x$ 由 0 開始增加時，$v^2$ 隨之遞減，並在

$$x_0 = \left(\frac{2u^2}{k_1}\right)^{1/4}.$$

處降至零。若 $u < a$，則 $2u^2 < 2a^2 = k_1R^4$，所以 $x_0 < R$。

**想一想** 在 $x_0$ 處速率為零，但為甚麼這就表示碎片不會走得更遠？

在 $x_0$ 處，碎片瞬間靜止，但它在該處的加速度是 $-k_1x_0^3 \neq 0$，方向指向 $G$。所以它並非處於平衡（equilibrium）狀態：它會開始向 $G$ 往回運動。事實上，$x$ 根本不可能超過 $x_0$，因為那樣的話 (i) 的公式便會給出 $v^2 < 0$。碎片最遠只能到達 $x_0 < R$，所以它永遠不會到達與 $G$ 相距 $R$ 之處，這正是所要證明的。

**第 (iii) 部分**

首先，$u \geq b$ 的碎片必定能到達 $R$。根據定義，

$$b^2 = \tfrac{1}{2}k_1R^4 + \frac{2k_2}{3R^3} = a^2 + \frac{2k_2}{3R^3} > a^2,$$

所以當 $x \leq R$ 時，由第 (i) 部分得

$$v^2 = u^2 - \tfrac{1}{2}k_1x^4 \geq u^2 - a^2 \geq b^2 - a^2 = \frac{2k_2}{3R^3} > 0.$$

特別地，碎片到達 $x = R$ 時，$v_R^2 = u^2 - \tfrac{1}{2}k_1R^4 = u^2 - a^2$。

當 $x \geq R$ 時，運動方程是 $v\dfrac{\mathrm{d}v}{\mathrm{d}x} = -k_2x^{-4}$。由 $x = R$ 開始積分：

$$\tfrac{1}{2}v^2 - \tfrac{1}{2}v_R^2 = \frac{k_2}{3}\left(\frac{1}{x^3} - \frac{1}{R^3}\right),$$

所以

$$\begin{aligned} v^2 &= u^2 - \tfrac{1}{2}k_1R^4 - \frac{2k_2}{3R^3} + \frac{2k_2}{3x^3} \\ &= u^2 - b^2 + \frac{2k_2}{3x^3} \qquad (x \geq R). \end{aligned}$$

這個組合正好說明了 $b$ 為何要這樣定義：$b^2$ 就是由 $G$ 逃逸到無窮遠處，以 $\tfrac{1}{2}v^2$ 計算所須付出的總「代價」。

若 $u \geq b$，則 $u^2 - b^2 \geq 0$ 而 $\frac{2k_2}{3x^3} > 0$，所以對每一個 $x \geq R$ 都有 $v^2 > 0$（這包括臨界情況 $u = b$）。連同 $x \leq R$ 時的下界，可知在碎片可能到達的每一個距離上，$v^2$ 都嚴格為正。

速度 $v$ 對時間連續，而且起初為 $u > 0$。它要變成負數，就必須經過 0，即在某個位置上有 $v^2 = 0$，而我們剛剛證明了這不會發生。所以自始至終 $v > 0$：由爆炸的一刻起，碎片一直在遠離 $G$。（由於 $v^2 \geq 2k_2/(3x^3)$，它會一直走下去，且 $x \to \infty$。）

### 第 (iv) 部分

現在 $a < u < b$。由於 $u > a$，(iii) 的論證表明在 $[0, R]$ 上 $v^2 \geq u^2 - a^2 > 0$，所以碎片到達 $R$ 時仍在向外運動。越過 $R$ 之後，

$$v^2 = \frac{2k_2}{3x^3} - (b^2 - u^2),$$

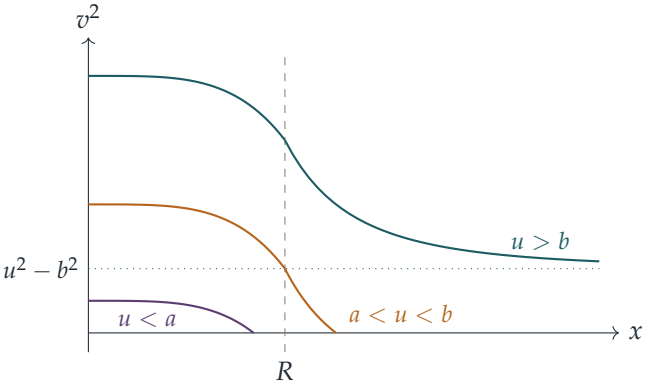
而此時 $b^2 - u^2 > 0$。這個式子隨 $x$ 增加而遞減，並在

$$x^3 = \frac{2k_2}{3(b^2 - u^2)}.$$

時等於零。這一點確實位於 $x \geq R$ 的區域內：由於 $u > a$，$b^2 - u^2 < b^2 - a^2 = \frac{2k_2}{3R^3}$，所以 $x^3 > R^3$。與 (ii) 一樣，碎片在該處停下，但仍有指向 $G$ 的非零加速度 $-k_2 x^{-4}$，而且它不可能走得更遠，否則 $v^2$ 便會是負數。與 $G$ 的最大距離是

$$\boxed{x_{\max} = \left(\frac{2k_2}{3(b^2 - u^2)}\right)^{1/3}}$$

下圖顯示當 $k_1 = 2$、$k_2 = 3$、$R = 1$（因此 $a^2 = 1$、$b^2 = 3$）時，三類初速率下 $v^2$ 對 $x$ 的圖像。每塊碎片都會一直前進，直至它的曲線與橫軸相交為止。




## 驗算

- 取 $k_1 = 2$、$k_2 = 3$、$R = 1$，得 $a = 1$、$b = \sqrt{3}$；當 $u = \sqrt{2}$ 時，公式給出 $x_{\max} = (2 \cdot 3/(3 \cdot 1))^{1/3} = 2^{1/3} \approx 1.26$，剛好超過 $R$，與預期相符。

- 當 $u \to a^+$ 時，$b^2 - u^2 \to 2k_2/(3R^3)$ 而 $x_{\max} \to R$；當 $u \to b^-$ 時，$x_{\max} \to \infty$。兩者在邊界上都與第 (ii) 和第 (iii) 部分吻合。
- 以數值方法逐步計算這個運動（龍格－庫塔法（Runge–Kutta），隨機選取十二組 $k_1$、$k_2$、$R$），結果重現了 $v^2$ 的公式、$u < a$ 時的折返點 $(2u^2/k_1)^{1/4}$、(iv) 中 $x_{\max}$ 的值（誤差在 0.1% 以內），以及 $u \geq b$ 時完全沒有折返。

## 常見錯誤

- 第二階段由 $x = 0$ 開始積分，而不是由 $x = R$ 並以 $v_R^2 = u^2 - \frac{1}{2}k_1R^4$ 開始：力的規律在 $R$ 處改變，所以積分常數也隨之改變。
- 弄丟正負號：加速度指向 $G$，所以 $v\,\mathrm{d}v/\mathrm{d}x$ 是負的 $k_1x^3$ 或 $k_2x^{-4}$。
- 在 (iii) 中只檢查 $x \geq R$，卻忘記在 $R$ 以內 $v$ 也必須保持為正。

## STEP 之外

設 $V'(x)$ 等於加速度的相反數，則對任何只取決於位置的力，$\frac{1}{2}v^2 + V(x)$ 這個量都是守恆的。畫出 $V$ 的圖像（一個「勢阱」（potential well）），便可一眼看出哪些運動是有界的、哪些會逃逸，而 $b$ 不過是這個勢阱的逃逸速率（escape speed）。同樣的推理可得出真實重力下的逃逸速度，也是對非線性振子（nonlinear oscillator）作相平面分析（phase-plane analysis）的起點。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics* 詳細闡述了這類能量圖。

詳細解答

# STEP II 1999，第 10 題

力學 • 連續碰撞

**STEP II 1999，第 10 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/99-s2-q10

## 關鍵思路

這條鏈上的每一次碰撞（impact）都是同一個模樣：一個質量為 $\mu$、以速率 $u$ 運動的粒子（particle），撞上一個質量為 $q\mu$ 的靜止粒子。因此，只要把第一次碰撞的計算按比例縮放，就能描述每一次碰撞。一旦知道每個粒子的最終速度，會否出現額外碰撞的問題，就歸結為比較相鄰粒子的最終速度；而當 $q = e$ 時，各個速度變得非常簡單，能量損失不過是一個等比級數（geometric series）。

## 解答

### 第一次碰撞

把各粒子視為沿直線運動的質點（point mass），並取由 $P_1$ 指向 $P_2$ 的方向為正。設 $v_1$ 和 $v_2$ 分別為 $P_1$ 和 $P_2$ 剛碰撞後的速度。由動量守恆（conservation of momentum）及牛頓恢復定律（Newton's law of restitution）得

$$mV = mv_1 + qm\,v_2,$$
$$v_2 - v_1 = eV.$$

把 $v_2 = v_1 + eV$ 代入第一條方程，得 $V = v_1 + qv_1 + qeV$，所以

$$v_1 = \left(\frac{1-eq}{1+q}\right)V, \qquad v_2 = v_1 + eV = \frac{1-eq+e+eq}{1+q}\,V = \left(\frac{1+e}{1+q}\right)V,$$

正如所求。

**每一次碰撞都是同一個模樣**

記

$$\alpha = \frac{1-eq}{1+q}, \qquad \beta = \frac{1+e}{1+q}.$$

上面的計算完全沒有用到 $m$ 的實際數值：一個質量為 $\mu$、以速度 $u$ 運動的粒子，撞上一個質量為 $q\mu$ 的靜止粒子後，自己以速度 $\alpha u$ 離開，並使對方以速度 $\beta u$ 彈出。由於每個 $P_{k+1}$ 的質量都是 $P_k$ 的 $q$ 倍，所以只要 $P_k$ 撞上 $P_{k+1}$ 時 $P_{k+1}$ 仍然靜止，這個結果便適用。

**當 $q \le e$ 時恰好有 $N-1$ 次碰撞**

假設 $q \le e$。那麼 $eq \le e^2 < 1$，所以 $\alpha > 0$；另外亦有 $\beta > 0$。

沿着這條鏈逐一追蹤。$P_1$ 撞上 $P_2$，碰撞後 $P_1$ 的速度為 $\alpha V$，$P_2$ 的速度為 $\beta V$。由於 $\beta V > 0$，$P_2$ 朝着靜止的 $P_3$ 運動並撞上它；碰撞後 $P_2$ 的速度為 $\alpha\beta V$，$P_3$ 的速度為 $\beta^2 V$。如此類推，$P_k$ 以速度 $u_k = \beta^{k-1}V > 0$ 到達靜止的 $P_{k+1}$，所以每一對相鄰粒子之間都有一次碰撞：即最少有 $N-1$ 次碰撞。這些碰撞之後，各粒子的速度為

$$w_k = \alpha\beta^{k-1}V \quad (1 \le k \le N-1), \qquad w_N = \beta^{N-1}V.$$

這些速度全部為正，所以所有粒子都朝同一方向運動。

? **想一想** 我們怎樣知道 $P_k$ 永遠不會再追上 $P_{k+1}$？

兩者碰撞之後，$P_k$ 從此以恆定速度 $w_k$ 運動，因為它往後唯一可能接觸到的只有 $P_{k+1}$ 或 $P_{k-1}$。與此同時，$P_{k+1}$ 在撞上 $P_{k+2}$ 之前以 $u_{k+1} = \beta^k V$ 運動，之後則以 $w_{k+1}$ 運動（若 $k+1 = N$，則永遠保持前者）。$P_k$ 與 $P_{k+1}$ 之間的距離只有在 $P_k$ 運動得較快時才會縮小，而：

- $u_{k+1} > w_k$，因為 $\beta^k V > \alpha\beta^{k-1}V$ 等價於 $\beta > \alpha$，即 $1+e > 1-eq$，而這是成立的；
- 對於 $k \le N-2$，有 $w_{k+1} \ge w_k$，因為 $w_{k+1}/w_k = \beta = \dfrac{1+e}{1+q} \ge 1$，而這正正是因為 $q \le e$；
- 對於 $k = N-1$，有 $w_N > w_{N-1}$，因為 $w_N/w_{N-1} = \beta/\alpha > 1$。

所以在兩者碰撞之後，$P_{k+1}$ 永遠不會比 $P_k$ 慢，兩者之間的距離永不縮小。（當 $q = e$ 時我們得到 $w_{k+1} = w_k$：距離保持不變，但不會縮小至零。）這對每一對相鄰粒子都成立，所以不會再有其他碰撞，碰撞次數恰好是 $N-1$ 次。

這就是 $q \le e$ 這個條件發揮作用的地方。假如 $q > e$，則 $\beta < 1$，每個粒子最終都比它後面的那個慢，這條鏈便會自己追上自己，產生更多碰撞。

### 當 $q = e$ 時損失的動能

現在 $q = e$，所以

$$\beta = \frac{1+e}{1+e} = 1, \qquad \alpha = \frac{1-e^2}{1+e} = 1-e.$$

每個粒子都是被速率為 $V$ 的粒子撞上；碰撞之後 $P_1, \ldots, P_{N-1}$ 以 $(1-e)V$ 運動，而 $P_N$ 以 $V$ 運動。由於 $P_k$ 的質量為 $e^{k-1}m$，最終的動能（kinetic energy）為

$$\begin{aligned} T &= \sum_{k=1}^{N-1} \tfrac{1}{2} e^{k-1} m (1-e)^2 V^2 + \tfrac{1}{2} e^{N-1} m V^2 \\ &= \tfrac{1}{2} m V^2 \left[ (1-e)^2 \cdot \frac{1-e^{N-1}}{1-e} + e^{N-1} \right] \\ &= \tfrac{1}{2} m V^2 \left[ (1-e)(1-e^{N-1}) + e^{N-1} \right], \end{aligned}$$

當中用了等比級數求和 $1 + e + \cdots + e^{N-2} = \dfrac{1-e^{N-1}}{1-e}$。最初的動能為 $\frac{1}{2}mV^2$，所以損失為

$$\begin{aligned} \tfrac{1}{2} m V^2 - T &= \tfrac{1}{2} m V^2 \left[ (1-e^{N-1}) - (1-e)(1-e^{N-1}) \right] \\ &= \tfrac{1}{2} m e \left( 1 - e^{N-1} \right) V^2, \end{aligned}$$

這就是所求的結果。

## 驗算

- 逐次碰撞計算。當 $q = e$ 時，$P_k$ 撞上 $P_{k+1}$ 這次碰撞使 $P_k$ 的速度由 $V$ 變為 $(1-e)V$，使 $P_{k+1}$ 的速度由 0 變為 $V$，損失的動能為 $\frac{1}{2}e^{k-1}mV^2\left[1-(1-e)^2-e\right] = \frac{1}{2}m(1-e)e^kV^2$。對 $k = 1, \ldots, N-1$ 求和，得 $\frac{1}{2}m(1-e)V^2 \cdot e\,\dfrac{1-e^{N-1}}{1-e}$，答案相同。
- 極限情況。當 $e \to 0$ 時，損失趨向 0；事實上，當 $q = e \to 0$ 時，後面各粒子的質量小得可以忽略。取 $N = 2$，公式給出 $\frac{1}{2}me(1-e)V^2$，與直接計算單一次碰撞的結果吻合。
- 數值檢驗。我們隨機選取 $N \le 12$、隨機的間距和隨機的 $q \le e$，逐個事件地模擬整個運動，每次都恰好得到 $N-1$ 次碰撞。當 $N = 5$、$q = e = \frac{1}{2}$、$m = V = 1$ 時，模擬得出的損失為 $0.234375 = \frac{15}{64}$，與 $\frac{1}{2} \cdot \frac{1}{2}\left(1 - \frac{1}{16}\right)$ 一致。

## 常見錯誤

- 只檢查了首 $N-1$ 次碰撞確實發生。「恰好」二字還需要論證：往後沒有任何粒子會追上它前面的那一個。

- 把 $P_k$ 的最終速度與 $P_{k+1}$ 剛碰撞後的速度（$\beta u_k$）比較，而不是與它撞上 $P_{k+2}$ 並減慢之後的最終速度比較。
- 弄錯恢復定律中的正負號：等於 $e$ 乘以接近速率的，是分離速率 $v_2 - v_1$。

## STEP 之外

一串質量逐個把動量傳遞下去的粒子鏈，是衝擊波（shock）在顆粒物料中傳播的一維模型。質量穩定遞減的鏈，例如本題中 $q < 1$ 的情況，是物理學家研究的對象，用來分散和吸收衝擊。要重新整理能量的計算，有一個漂亮的方法，就是利用約化質量（reduced mass）$\mu = m_1 m_2/(m_1 + m_2)$：一次碰撞損失 $\frac{1}{2}\mu(1 - e^2)w^2$，其中 $w$ 是接近速率。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）在質心參考系（centre-of-mass frame）中討論碰撞，在那裏這類結果變得一目了然。

詳細解答

# STEP II 1999，第 11 題

力學 • 拋體與移動目標

**STEP II 1999，第 11 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/99-s2-q11

## 關鍵思路

目標始終留在圓上，所以炮彈必須落在水平面上與 $O$ 相距恰好 $R$ 之處；而炮總可以轉向，對準目標屆時所在的位置。因此一切歸結為同一個拋體（projectile）的兩項事實：它的射程（range）和飛行時間（time of flight）。轉過的角為 $\beta = \omega T$，再利用 $\sin^2\alpha + \cos^2\alpha = 1$ 消去 $\alpha$，便得到那個四次方程（quartic）。最後一部分則是韋達定理（Vieta's formulas）加上圓的一條弦（chord）。

## 解答

我們把炮彈視為只受重力作用的質點（particle）（不計空氣阻力），從圓所在水平面上的 $O$ 點發射，其中 $0 < \alpha < \frac{1}{2}\pi$。

**第 1 部分：當 $V^2 < gR$**

取坐標軸分別沿地面（在發射所在的平面內）及鉛垂向上。在時間 $t$，

$$x = Vt\cos\alpha, \qquad y = Vt\sin\alpha - \tfrac{1}{2}gt^2.$$

炮彈回到水平面（$y = 0$，$t > 0$）的時間為

$$T = \frac{2V\sin\alpha}{g}, \qquad \text{此時水平距離為} \qquad x(T) = \frac{2V^2\sin\alpha\cos\alpha}{g} = \frac{V^2\sin 2\alpha}{g}.$$

由於 $\sin 2\alpha \le 1$，無論在哪一個鉛垂面內發射，炮彈的落點與 $O$ 的距離都不超過 $V^2/g$。目標始終在水平面上與 $O$ 相距 $R$ 之處，所以要命中就必須有 $R \le V^2/g$，即 $V^2 \ge gR$。因此若 $V^2 < gR$，炮彈永遠不能擊中目標，這正是所要證明的。

**第 2 部分：$\beta$ 所滿足的方程**

現設 $V^2 > gR$。炮彈在時間 $T$、距離 $R$ 處擊中目標，所以

$$T = \frac{2V\sin\alpha}{g}, \qquad R = \frac{2V^2\sin\alpha\cos\alpha}{g}.$$

目標以恆定角速率 $\omega$ 轉動，所以 $\beta = \omega T$，由此得

$$\sin\alpha = \frac{g\beta}{2V\omega}.$$

把它代入射程的方程，

$$\cos\alpha = \frac{gR}{2V^2\sin\alpha} = \frac{gR}{2V^2}\cdot\frac{2V\omega}{g\beta} = \frac{R\omega}{V\beta}.$$

於是由 $\sin^2\alpha + \cos^2\alpha = 1$ 得

$$\frac{g^2\beta^2}{4V^2\omega^2} + \frac{R^2\omega^2}{V^2\beta^2} = 1,$$

全式乘以 $4V^2\omega^2\beta^2$，得

$$g^2\beta^4 - 4\omega^2V^2\beta^2 + 4R^2\omega^4 = 0,$$

這就是所求的方程。

? **想一想** 為甚麼恰好有兩個值，而不是最多四個，或者一個也沒有？

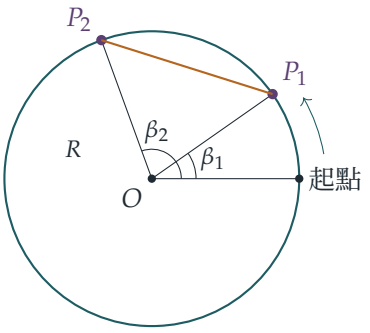
把它看成 $\beta^2$ 的二次方程，其判別式（discriminant）為

$$16\omega^4V^4 - 16g^2R^2\omega^4 = 16\omega^4\left(V^4 - g^2R^2\right) > 0,$$

因為 $V^2 > gR$。所以 $\beta^2$ 有兩個相異的實數值。它們的和 $4\omega^2V^2/g^2$ 與積 $4R^2\omega^4/g^2$ 都是正數，所以 $\beta^2$ 的兩個值都是正數。角 $\beta = \omega T$ 是正數，所以 $\beta^2$ 的每個值恰好對應一個 $\beta$，於是我們恰好有兩個候選值 $\beta_1 < \beta_2$。

每一個候選值都確實對應一次命中。對任何一個根，$g\beta/(2V\omega)$ 和 $R\omega/(V\beta)$ 這兩個數都是正數，而且（根據該方程）它們的平方和為 1，所以存在 $\alpha \in (0, \frac{1}{2}\pi)$，以這兩個數分別作為其正弦和餘弦。以這個仰角（elevation），在通過目標轉過 $\beta$ 後所到達之點的鉛垂面內發射，飛行時間便是 $\beta/\omega$，射程便是 $R$，所以炮彈與目標相遇。因此 $\beta$ 恰好有兩個可能的值。（它們對應於射程同為 $R$ 的兩個互餘仰角 $\alpha$ 和 $\frac{1}{2}\pi - \alpha$。）

**第 3 部分：距離 $P_1P_2$**




兩個彈着點都在圓上，從發射一刻目標所在的位置起按逆時針方向量度，所對的角分別為 $\beta_1$ 和 $\beta_2$。所以角 $P_1OP_2$ 為 $\beta_2-\beta_1$（如有需要可按模 $2\pi$ 化簡，這不影響弦長），而由等腰三角形 $OP_1P_2$ 得

$$P_1P_2 = 2R\left|\sin\frac{\beta_2-\beta_1}{2}\right|.$$

由 $\beta^2$ 的二次方程，

$$\beta_1^2+\beta_2^2 = \frac{4\omega^2V^2}{g^2}, \qquad \beta_1^2\beta_2^2 = \frac{4R^2\omega^4}{g^2},$$

又因為 $\beta_1,\beta_2>0$，可以取正平方根：$\beta_1\beta_2 = 2R\omega^2/g$。因此

$$\begin{aligned}(\beta_2-\beta_1)^2 &= \beta_1^2+\beta_2^2-2\beta_1\beta_2\\ &= \frac{4\omega^2V^2}{g^2}-\frac{4R\omega^2}{g} = \frac{4\omega^2}{g^2}\left(V^2-Rg\right),\end{aligned}$$

所以 $\beta_2-\beta_1 = \dfrac{2\omega}{g}\sqrt{V^2-Rg}$。故此

$$P_1P_2 = 2R\sin\left(\frac{\omega}{g}\sqrt{V^2-Rg}\right),$$

這正是所要證明的。（嚴格來說，這需要該正弦值非負，即 $\frac{\omega}{g}\sqrt{V^2-Rg}\le\pi$；否則距離是這個式子的絕對值。）

## 驗算

- 當 $V^2\to gR^+$ 時，兩個仰角都趨向 45°，所以 $P_1$ 和 $P_2$ 應該重合：公式的值確實趨向 0。
- 直接驗證 $\beta_1\beta_2 = 2R\omega^2/g$：$\beta_1\beta_2 = (2V\omega/g)^2\sin\alpha_1\sin\alpha_2$，而代入 $\alpha_2 = \frac{1}{2}\pi-\alpha_1$ 後，這等於 $(2V\omega/g)^2\sin\alpha_1\cos\alpha_1 = (4V^2\omega^2/g^2)(gR/2V^2)$，與預期相符。
- 數值上，取 $g=9.8$、$R=10$、$V=14$、$\omega=0.3$（國際單位制），逐步計算兩條軌跡，得出弦長約為 5.97 m，與 $20\sin\left(0.3\sqrt{98}/9.8\right)\approx 5.969$ 吻合。

## 常見錯誤

- 忘記了炮可以對準任何一個鉛垂面：第一部分只關乎射程，與目標碰巧身在何處無關。
- 對 $\beta_1^2\beta_2^2$ 開平方，卻沒有說明為甚麼 $\beta_1\beta_2$ 是正數。
- 寫成 $P_1P_2 = R(\beta_2 - \beta_1)$，這是弧長，而不是直線距離。

## STEP 之外

擊中移動中的目標，是一個小型的追蹤（pursuit）或攔截（interception）問題：先求出命中的時間，再求出瞄準的方向。在兩條方程之間消去一個參數而得到一個多項式，然後不必解它，直接從係數讀出所需的資料，這種技巧你在大學裏會經常用到。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）正是以這種嚴謹的態度處理拋體和轉動運動。

詳細解答

# STEP II 1999，第 12 題

概率與統計 • 貝葉斯定理

**STEP II 1999，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/99-s2-q12

## 關鍵思路

這是一道保留了參數的貝葉斯定理（Bayes' theorem）題目。只要用全概率公式（law of total probability）求出 $\mathrm{P}(S)$，兩個條件概率（conditional probability）都是 $p$ 的簡單有理函數（rational function），而且各自在容許的區間上都是單調（monotonic）的。因此，每位檢查員的做法中所謂的「最大值」，都落在區間 $0 \le p \le \frac{1}{3}$ 的端點上；而真正有趣的地方，在於這對單憑一次觀察作估計有甚麼啟示。

## 解答

### 條件概率

設 $A$、$B$、$C$ 分別為隨機抽出的一塊晶片由 $A$、$B$、$C$ 製造的事件，並設 $S$ 為該晶片完好的事件。這三個事件構成樣本空間（sample space）的一個分割（partition），其中

$$\mathrm{P}(A) = 2p, \quad \mathrm{P}(B) = p, \quad \mathrm{P}(C) = 1 - 3p,$$
$$\mathrm{P}(S \mid A) = \tfrac{7}{10}, \quad \mathrm{P}(S \mid B) = \tfrac{8}{10}, \quad \mathrm{P}(S \mid C) = \tfrac{9}{10}.$$

根據全概率公式，

$$\begin{aligned}\mathrm{P}(S) &= 2p \cdot \tfrac{7}{10} + p \cdot \tfrac{8}{10} + (1 - 3p) \cdot \tfrac{9}{10} \\ &= \tfrac{1}{10}\left(14p + 8p + 9 - 27p\right) = \tfrac{1}{10}(9 - 5p).\end{aligned}$$

留意在整個 $0 \le p \le \frac{1}{3}$ 的範圍內都有 $9 - 5p > 0$，所以可以用它作除數。根據貝葉斯定理，

$$\mathrm{P}(A \mid S) = \frac{\mathrm{P}(S \mid A)\,\mathrm{P}(A)}{\mathrm{P}(S)} = \frac{\frac{14p}{10}}{\frac{9-5p}{10}} = \frac{14p}{9 - 5p},$$

同理

$$\mathrm{P}(C \mid S) = \frac{\frac{9}{10}(1-3p)}{\frac{1}{10}(9-5p)} = \frac{9(1-3p)}{9-5p} = \frac{9-27p}{9-5p}.$$

### 第一位檢查員的估計

當 $p$ 由 0 增加至 $\frac{1}{3}$ 時，分子 $14p$ 遞增，而正的分母 $9-5p$ 遞減，所以 $\mathrm{P}(A \mid S)$ 嚴格遞增（strictly increasing）。（等價地說，它的導數是 $126/(9-5p)^2 > 0$。）因此它的最大值出現在右端點 $p = \frac{1}{3}$，其值為

$$\frac{14/3}{9-5/3} = \frac{14/3}{22/3} = \frac{7}{11}.$$

所以第一位檢查員的估計是 $\hat{p} = \frac{1}{3}$。

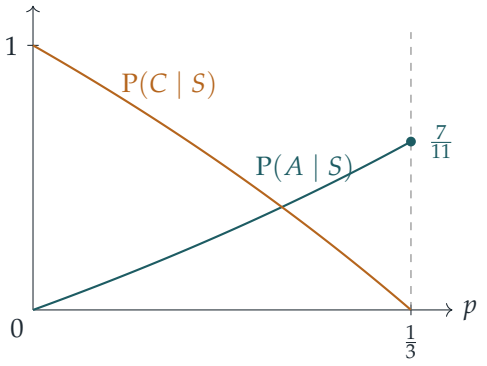
### 第二位檢查員的估計

**想一想** $\mathrm{P}(C \mid S)$ 是遞增還是遞減？隨着 $p$ 增大，它的分子和分母都在減小。

當分子和分母朝同一方向變化時，我們不能一眼看出答案，所以要用商法則（quotient rule）求導：

$$\frac{\mathrm{d}}{\mathrm{d}p}\frac{9-27p}{9-5p} = \frac{-27(9-5p)+5(9-27p)}{(9-5p)^2} = \frac{-243+135p+45-135p}{(9-5p)^2} = \frac{-198}{(9-5p)^2} < 0.$$

所以 $\mathrm{P}(C \mid S)$ 在 $[0, \frac{1}{3}]$ 上嚴格遞減，它的最大值出現在 $p = 0$，其值為 1。第二位檢查員的估計是 $\hat{p} = 0$。




### 評論

兩位檢查員用的是同一個方法，各自依據的單次觀察也同樣有效，卻得出容許範圍內兩個相反的極端：$\hat{p} = \frac{1}{3}$（完全沒有晶片來自 $C$）和 $\hat{p} = 0$（所有晶片都來自 $C$）。每個估

計其實只是「令我這塊晶片的製造商盡可能常見的 $p$ 值」，而只有一個樣本時，必然會得出這樣的結果。第二個估計甚至與第一位檢查員手上的晶片互相矛盾，因為 $p = 0$ 表示 $A$ 根本不製造晶片。兩個估計都不可信：一次觀察所含的信息實在太少，不足以估計 $p$，我們需要更大的樣本（連同來自每個製造商的晶片數目）。

還有一點值得留意：晶片是否完好，幾乎不起任何作用。「完好且來自 $A$」的似然（likelihood）是 $\frac{7}{10} \cdot 2p$，它取得最大值的位置與 $2p$ 本身相同；因子 $\frac{7}{10}$ 與 $p$ 無關。

## 驗算

- 當 $p = 0$ 時，所有晶片都來自 $C$，所以 $\mathrm{P}(C \mid S)$ 應為 1，而 $\mathrm{P}(A \mid S)$ 應為 0；兩條公式都與此吻合。
- 同理可得 $\mathrm{P}(B \mid S) = 8p/(9 - 5p)$，而三個條件概率之和為

$$\frac{14p + 8p + 9 - 27p}{9 - 5p} = 1.$$

- 當 $p = 0.2$ 時，公式給出 $\mathrm{P}(A \mid S) = 0.35$ 和 $\mathrm{P}(C \mid S) = 0.45$；以 400 000 塊晶片進行模擬，得出 0.3505 和 0.4505。

## 常見錯誤

- 令導數等於零，然後尋找轉向點（turning point）。這裏根本沒有轉向點：兩個最大值都在 $[0, \frac{1}{3}]$ 的端點上。
- 忘記 $\mathrm{P}(C \mid S)$ 需要仔細論證：分子和分母都在減小，所以單說「分子變小了」並不足夠。

## STEP 之外

兩位檢查員所做的，實際上就是最大似然估計（maximum likelihood estimation）：選出令他們所見的結果概率最大的參數值。只有一次觀察時，最大似然估計值往往落在參數空間（parameter space）的邊界上，這提醒我們：它需要大量數據才會表現良好。貝葉斯學派（Bayesian）的做法則是先為 $p$ 設定一個先驗分佈（prior distribution），然後每觀察一塊晶片就更新一次，這樣便不會單憑一次觀察而得出如此極端的結論。Blitzstein 與 Hwang 所著的 *Introduction to Probability*（概率論入門教材）對這兩種觀點都有討論。

詳細解答

# STEP II 1999，第 13 題

概率與統計 • 連續隨機變量

**STEP II 1999，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/99-s2-q13

## 關鍵思路

這道題雖然歸入力學（mechanics）一類，其實是一道關於連續隨機變量（continuous random variable）的題目。只要留意到較短的一段有簡單的分佈，一切便迎刃而解：若折斷點在棒上均勻分佈（uniform），則較短的一段在棒長的 $[0, \frac{1}{2}]$ 上均勻分佈。比值 $R$ 是較短一段的遞增函數，所以「$R < r$」可化為只涉及較短一段的命題。由此得出分佈函數（distribution function），再求導便得到密度（density）。

## 解答

### $R$ 的分佈

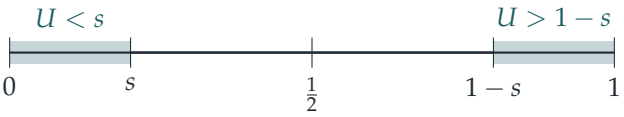
設棒長為 1；比值與所用單位無關。設折斷點與其中一端的距離為 $U$，其中 $U$ 在 $[0,1]$ 上均勻分佈。兩段的長度分別為 $U$ 和 $1-U$，所以較短一段的長度為

$$S = \min(U, 1-U),$$

而較長一段的長度為 $1-S$。於是 $0 \le S \le \frac{1}{2}$，並且

$$R = \frac{S}{1-S}.$$

對於 $0 \le s \le \frac{1}{2}$，事件 $S < s$ 恰好在 $U < s$ 或 $U > 1-s$ 時發生。這是兩個互不相交（disjoint）的區間，總長度為 $2s$，所以 $\mathrm{P}(S < s) = 2s$。換句話說，$S$ 在 $[0, \frac{1}{2}]$ 上均勻分佈。




現設 $0 \le r \le 1$。函數 $s \mapsto \dfrac{s}{1-s}$ 在 $[0, \frac{1}{2}]$ 上嚴格遞增（strictly increasing）（其導數為 $1/(1-s)^2 > 0$），所以我們可以把不等式反解出來：

$$R < r \iff \frac{S}{1-S} < r \iff S < r(1-S) \iff S < \frac{r}{1+r},$$

其中兩邊同乘 $1-S$ 是穩妥的，因為 $1 - S \ge \frac{1}{2} > 0$。當 $0 \le r \le 1$ 時，$\dfrac{r}{1+r}$ 的值落在 $[0, \frac{1}{2}]$ 之內，所以

$$\boxed{\mathrm{P}(R < r) = \frac{2r}{1+r} \qquad (0 \le r \le 1).}$$

由於較短的一段永遠不會比較長的一段長，$R$ 必定落在 $[0,1]$ 之內。所以當 $r \le 0$ 時 $\mathrm{P}(R < r) = 0$，當 $r \ge 1$ 時 $\mathrm{P}(R < r) = 1$，這與公式在 $r = 1$ 處取值 1 吻合。

> **想一想** 為甚麼不直接用 $U$ 來做，即取 $R = U/(1-U)$？

這只在 $U \le \frac{1}{2}$ 時才正確。當 $U > \frac{1}{2}$ 時，較短的一段是 $1-U$，而 $R = (1-U)/U$。你可以分成這兩種情況來處理，根據對稱性，每種情況各貢獻 $\frac{r}{1+r}$。改用 $S$ 則一次過把這些細節處理妥當。

### 密度、平均值與方差

在 $(0,1)$ 上對分佈函數 $F(r) = \dfrac{2r}{1+r} = 2 - \dfrac{2}{1+r}$ 求導，得到概率密度函數（probability density function）

$$f(r) = \frac{2}{(1+r)^2} \quad (0 \le r \le 1), \qquad f(r) = 0 \text{ 其他情況.}$$

驗算一下：$\int_0^1 f(r)\,\mathrm{d}r = F(1) - F(0) = 1$。

求平均值（mean）時，寫成 $r = (1+r) - 1$：

$$\begin{aligned}\mathrm{E}(R) &= \int_0^1 \frac{2r}{(1+r)^2}\,\mathrm{d}r = 2\int_0^1 \left(\frac{1}{1+r} - \frac{1}{(1+r)^2}\right)\mathrm{d}r \\ &= 2\Big[\ln(1+r) + \frac{1}{1+r}\Big]_0^1 = 2\left(\ln 2 + \tfrac{1}{2} - 1\right) \\ &= 2\ln 2 - 1 \approx 0.386.\end{aligned}$$

求二階矩（second moment）時，寫成 $r^2 = (1+r)^2 - 2(1+r) + 1$，於是

$$\frac{r^2}{(1+r)^2} = 1 - \frac{2}{1+r} + \frac{1}{(1+r)^2}.$$

那麼

$$\begin{aligned}\mathrm{E}(R^2) &= 2\int_0^1 \left(1 - \frac{2}{1+r} + \frac{1}{(1+r)^2}\right) \mathrm{d}r = 2\Big[r - 2\ln(1+r) - \frac{1}{1+r}\Big]_0^1 \\ &= 2\left(1 - 2\ln 2 - \tfrac{1}{2} + 1\right) = 3 - 4\ln 2.\end{aligned}$$

因此方差（variance）為

$$\begin{aligned}\mathrm{Var}(R) &= \mathrm{E}(R^2) - \big(\mathrm{E}(R)\big)^2 = 3 - 4\ln 2 - \big(4(\ln 2)^2 - 4\ln 2 + 1\big) \\ &= 2 - 4(\ln 2)^2 \approx 0.0782.\end{aligned}$$

含 $\ln 2$ 的項互相抵消，剩下這個簡潔的形式。

## 驗算

- 在 $r = 1$ 處概率為 1，理應如此。在 $r = \frac{1}{2}$ 處概率為 $\frac{2}{3}$：較長一段超過較短一段的兩倍，恰好是折斷點落在棒兩端各三分之一的範圍內，而這兩部分的總長度為 $\frac{2}{3}$。
- 平均值 0.386 介乎 0 與 1 之間，並且小於 $\frac{1}{2}$，這與密度由 $r = 0$ 處的 2 遞減至 $r = 1$ 處的 $\frac{1}{2}$ 相符。
- 模擬 $10^6$ 次隨機折斷，得到樣本平均值 0.3862 及樣本方差 0.0783，相比之下理論值為 $2\ln 2 - 1 = 0.38629$ 及 $2 - 4(\ln 2)^2 = 0.07819$。

## 常見錯誤

- 對所有 $U$ 都取 $R = U/(1-U)$。這個量可以大於 1，所以它並不是較短一段與較長一段之比。
- 忘記折斷點可以靠近任何一端，因而得出 $\mathrm{P}(R < r) = \frac{r}{1+r}$。
- 用較短一段的平均值 $\frac{1}{4}$ 來得出 $\mathrm{E}(R) = \frac{1/4}{3/4} = \frac{1}{3}$。一般而言 $\mathrm{E}\big(g(S)\big) \neq g\big(\mathrm{E}(S)\big)$。

## STEP 之外

這裏用的是變換隨機變量的標準方法，即「分佈函數法」。若 $Y = g(X)$ 而 $g$ 為遞增函數，則 $F_Y(y) = F_X\big(g^{-1}(y)\big)$，求導後便得到變量代換公式（change-of-variables formula）$f_Y(y) = f_X\big(g^{-1}(y)\big)\left|\mathrm{d}g^{-1}/\mathrm{d}y\right|$。同一想法可以引申到順序統計量（order statistics），以及經典的折棒問題，例如三段能構成三角形的概率。Grimmett 與 Welsh 所著的 *Probability: An Introduction*（概率論入門）是研讀連續隨機變量變換的好去處。

詳細解答

# STEP II 1999，第 14 題

概率與統計 • 最優停止

> **STEP II 1999，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/99-s2-q14

## 關鍵思路

骰子沒有記憶。每當你決定再擲一次，你就回到同一個遊戲的最開端，所以從那一刻起你所能期望的最好結果仍是同一個數 $V$，即整個遊戲的最佳期望得分。這樣一來，每次擲骰後的決定就變成一個簡單的比較：如果手上的點數勝過 $V$ 就保留，否則再擲。把這一點寫下來，便得到一條關於 $V$ 的方程，解出它就同時知道這個值和應採用的策略。

## 解答

### 設定

把擲出 1 的一擲稱為強制的：遊戲就此結束，得分為 1。擲出其他任何點數 $x \in \{2,3,4,5,6\}$ 之後，我們可選擇停止（得分為 $x$）或繼續。每個得分都介乎 1 與 6 之間，所以每個策略的期望得分（expected score）都在 $[1,6]$ 之內。設

$$V = \text{任何策略所能達到的最大期望得分}$$

（嚴格來說，是對所有策略取的最小上界（least upper bound）；我們稍後會看到它確實可以達到）。

### 一個得分為 4 的策略

考慮以下策略：一擲出 4、5 或 6 就停止（當然，擲出 1 時也要停止）。設 $E$ 為其期望得分。以第一擲的結果為條件（condition）。擲出 $1,4,5,6$ 的概率（probability）各為 $\frac{1}{6}$，此時遊戲結束，得分就是該點數。擲出 2 或 3 的概率為 $\frac{2}{6}$；這時我們再擲，而由於骰子

沒有記憶，從那裏開始的期望得分仍然是 $E$。所以

$$E = \tfrac{1}{6}(1+4+5+6) + \tfrac{2}{6}E \quad\Longrightarrow\quad \tfrac{4}{6}E = \tfrac{16}{6} \quad\Longrightarrow\quad E = 4.$$

（遊戲以概率 1 結束，因為連擲 $n$ 次而仍未停止的機會是 $(\tfrac{1}{3})^n \to 0$。）因此 $V \geq 4$。

**沒有策略能勝過 4**

任取一個策略；它甚至可以取決於之前所有擲骰結果的整段歷史。看看它的第一擲 $x$。

- 若 $x = 1$，得分為 1。
- 若 $x \geq 2$ 而該策略停止，得分為 $x$。
- 若 $x \geq 2$ 而該策略繼續，那麼它此後的做法只不過是對一個全新的同樣遊戲所用的某個策略（先前的擲骰結果不能影響之後的結果）。所以它從這裏開始的期望得分至多為 $V$。

在每一種情況下，已知 $x \geq 2$ 時的期望得分都至多為 $\max(x, V)$。因此每個策略的期望得分至多為

$$g(V) = \tfrac{1}{6}\Big(1 + \sum_{x=2}^{6} \max(x, V)\Big),$$

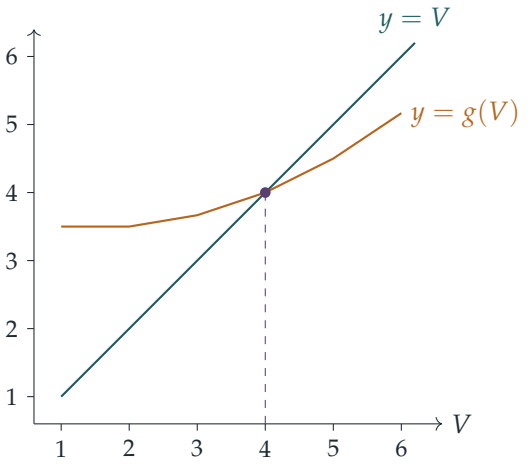
再對所有策略取最小上界，便得 $V \leq g(V)$。

? **想一想** 不等式 $V \leq g(V)$ 怎樣把 $V$ 確定下來？

考慮 $h(V) = g(V) - V$。當 $V$ 增加時，每一項 $\max(x, V)$ 的增長率是 0 或 1，而這樣的項只有五個，所以 $g$ 的增長率至多為 $\frac{5}{6}$。因此 $h$ 是嚴格遞減（strictly decreasing）的。在 $V = 4$ 時，

$$g(4) = \tfrac{1}{6}(1+4+4+4+4+5+6) = \tfrac{24}{6} = 4,$$

所以 $h(4) = 0$，而對每個 $V > 4$ 都有 $h(V) < 0$。由於 $h(V) \geq 0$，必定有 $V \leq 4$。




### 結論

綜合以上兩部分，$V = 4$：最佳的期望得分是 4，而做法是擲出 4、5 或 6 時停止，擲出 2 或 3 時再擲。這正是「若擲出的點數不小於 $V$ 就停止」這條規則：2 或 3 比繼續下去所期望的 4 差，而 5 或 6 則較好。

擲出 4 是打和的情況。擲出 4 時停止得 4 分，繼續下去的期望得分也是 4。所以「只在 5 或 6 時停止」同樣好：用同樣的條件論證，

$$E' = \tfrac{1}{6}(1+5+6) + \tfrac{3}{6}E' \quad \Longrightarrow \quad E' = 4.$$

其他任何規則都嚴格地較差；例如擲出 3 或以上就停止，得 $\frac{1+3+4+5+6}{5} = \frac{19}{5}$，而只在 6 時停止，得 $\frac{1+6}{2} = \frac{7}{2}$。

## 驗算

對於任何形如「擲出 1 及集合 $S$ 內的點數時停止」的規則，有一條巧妙的捷徑：每一擲令遊戲結束的概率都相同，而結束時的點數等可能地是 $\{1\} \cup S$ 中的任何一個，所以期望得分就是這些點數的平均值。對 $S = \{4,5,6\}$，它是 $\frac{16}{4} = 4$；對 $S = \{5,6\}$，它是 $\frac{12}{3} = 4$。我們精確地檢查了全部 32 條這樣的規則：只有這兩條得出 4，次佳的是 $\frac{19}{5}$。模擬 400 000 局遊戲，採用擲出 4 或以上就停止的規則，所得平均值為 3.998。

## 常見錯誤

- 把一擲的平均值 3.5 當作繼續下去的價值。繼續的價值高於 3.5，因為你保留了再擲的選擇權，所以「擲出 4 或以上就停止」不能靠與 3.5 比較來證明。

- 忘記了擲出 1 是強制的，於是把繼續的價值寫成彷彿可以毫無風險地一直擲下去。
- 只檢查閾值策略（threshold strategy），卻沒有解釋為何更巧妙、取決於歷史的規則不能做得更好。上面的無記憶性（memorylessness）論證涵蓋了所有策略。

## STEP 之外

這是最優停止（optimal stopping）的一個小例子，而方程 $V = g(V)$ 是一條貝爾曼方程（Bellman equation），它是動態規劃（dynamic programming）和馬可夫決策過程（Markov decision process）的基本工具。同樣的推理可用來為金融中的美式期權（American option）定價，也可解決著名的祕書問題（secretary problem）。Grimmett 與 Welsh 的 *Probability: An Introduction* 涵蓋了這裏所用的條件論證，而 Blitzstein 與 Hwang 的 *Introduction to Probability* 則有許多同類型的題目。

詳細解答

# STEP III 1999，第 1 題

純數學 • 三次方程的根

> **STEP III 1999，第 1 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/99-s3-q1

## 關鍵思路

這裏的一切都源自韋達定理（Vieta's formulas）：對於三次方程（cubic）$x^3 - px^2 + qx - r = 0$，設其根（roots）為 $\alpha, \beta, \gamma$，則

$$\alpha + \beta + \gamma = p, \qquad \alpha\beta + \beta\gamma + \gamma\alpha = q, \qquad \alpha\beta\gamma = r.$$

等比數列（geometric progression）是以乘法的方式圍繞中間項對稱的，所以根之積 $r$ 和比值 $q/p$ 能把中間的根找出來。等差數列（arithmetic progression）則是以加法的方式圍繞中間項對稱的，所以改由根之和 $p$ 把中間的根找出來。

## 解答

### 第 (i) 部分

設三個根為 $a/k$、$a$ 和 $ak$。要使 $a/k$ 有意義，必須有 $k \neq 0$；又因為根之積為 $r \neq 0$，所以亦有 $a \neq 0$。由韋達定理得

$$\begin{aligned} p &= \frac{a}{k} + a + ak = a\left(\frac{1}{k} + 1 + k\right), \\ q &= \frac{a}{k} \cdot a + a \cdot ak + \frac{a}{k} \cdot ak = a^2\left(\frac{1}{k} + 1 + k\right), \\ r &= \frac{a}{k} \cdot a \cdot ak = a^3. \end{aligned}$$

$p$ 和 $q$ 中出現了同一個括號。由於 $p \neq 0$，這個括號不等於零，所以可以相除：

$$\frac{q}{p} = a.$$

因此 $q/p$ 是其中一個根（即中間的那個）。於是 $q = ap$，從而

$$q^3 = a^3p^3 = rp^3, \qquad \text{即} \qquad q^3 - rp^3 = 0,$$

這正是所要證明的。

### 第 (ii) 部分

現在假設 $r = q^3/p^3$，並記 $a = q/p$，於是 $r = a^3$。把 $x = a$ 代入三次方程，並利用 $p\,a^2 = q^2/p = q\,a$，得

$$a^3 - pa^2 + qa - r = a^3 - qa + qa - a^3 = 0,$$

所以 $q/p$ 是一個根。

由於 $r \neq 0$，可知 $q \neq 0$，因此 $a \neq 0$。把另外兩個根稱為 $\beta$ 和 $\gamma$。三個根之積為 $r$，所以

$$a\beta\gamma = r = a^3 \quad \Longrightarrow \quad \beta\gamma = a^2 = \left(\frac{q}{p}\right)^2,$$

這裏兩邊除以了 $a \neq 0$。

> **想一想** 怎樣把 $\beta\gamma = a^2$ 化成 $a/k$、$a$、$ak$ 的形式？

我們需要選定 $k$。由於 $\beta\gamma = a^2 \neq 0$，$\beta$ 和 $\gamma$ 都不是零。定義 $k = \gamma/a$，於是 $\gamma = ak$ 且 $k \neq 0$。那麼

$$\beta = \frac{a^2}{\gamma} = \frac{a^2}{ak} = \frac{a}{k}.$$

所以三個根是 $a/k$、$a$ 和 $ak$，每一個都是由前一個乘以 $k$ 而得。它們成等比數列。（當 $\beta$ 和 $\gamma$ 不是實數時，公比 $k$ 可能是複數。這個論證並不要求它們是實數。）

綜合第 (i) 和第 (ii) 部分可知：當 $p \neq 0$ 且 $r \neq 0$ 時，三個根成等比數列，當且僅當（if and only if）$q^3 = rp^3$。

### 第 (iii) 部分

*必要性*。設三個根為 $b - d$、$b$ 和 $b + d$。它們的和是 $3b = p$，所以中間的根是 $b = p/3$。把 $x = p/3$ 代入三次方程，再乘以 27：

$$p^3 - 3p^3 + 9pq - 27r = 0 \quad \Longleftrightarrow \quad 2p^3 - 9pq + 27r = 0.$$

*充分性*。反過來，假設 $2p^3 - 9pq + 27r = 0$。把上面的步驟倒過來，可知 $p/3$ 是一個根。於是另外兩個根之和為 $p - p/3 = 2p/3$，所以可以把它們寫成 $p/3 - d$ 和 $p/3 + d$，其中 $d$ 是某個數（可能是複數）。三個根 $p/3 - d$、$p/3$、$p/3 + d$ 成等差數列。

因此，充分必要條件（necessary and sufficient condition）是

$$\boxed{2p^3 - 9pq + 27r = 0}.$$

## 驗算

- 根為 $1, 2, 4$（等比數列）時，$p = 7$，$q = 14$，$r = 8$。這時 $q/p = 2$ 正是中間的根，而且 $q^3 = 2744 = 8 \cdot 343 = rp^3$。
- 根為 $1, 2, 3$（等差數列）時，$p = 6$，$q = 11$，$r = 6$，而 $2 \cdot 216 - 9 \cdot 66 + 27 \cdot 6 = 432 - 594 + 162 = 0$。等比數列 $1, 2, 4$ 則給出 $686 - 882 + 216 = 20 \neq 0$，這也合乎預期。
- 我們亦用數千個隨機三次方程（包括有複數根的）對全部三個部分作了數值檢驗。所有結果都在 $10^{-11}$ 的誤差範圍內吻合。

## 常見錯誤

- 除以 $1/k + 1 + k$ 卻沒有說明它為何不等於零。它不等於零，是因為它等於 $p/a$，而 $p \neq 0$。
- 在第 (ii) 部分，只證明了 $q/p$ 是一個根以及 $\beta\gamma = (q/p)^2$，然後沒有具體寫出公比（common ratio），就斷言三個根成等比數列。
- 在第 (iii) 部分，只證明了一個方向。「充分必要」需要兩個方向都證明。

## STEP 之外

韋達定理說明，多項式的係數就是其根的**基本對稱函數**（elementary symmetric functions）。有一個經典定理指出，根的每一個對稱多項式（symmetric polynomial）都可以用這些基本對稱函數表示出來。這正是伽羅瓦理論（Galois theory）的起點，該理論解釋了多項式的根在甚麼情況下可以用根式（radicals）求出。順帶一提，代換 $x = y + p/3$ 可以消去三次方程中的 $x^2$ 項，而所得的缺項三次方程（depressed cubic）的常數項是 $2p^3 - 9pq + 27r$ 的倍數。想知道這些想法會通往何處，Allenby 的 *Rings, Fields and Groups* 是個很好的入門讀物。

詳細解答

# STEP III 1999，第 2 題

純數學 • 實根的計數

**STEP III 1999，第 2 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/99-s3-q2

## 關鍵思路

在第 (i) 部分，導數（derivative）可以分解為一個完全平方（perfect square）乘以 $\mathrm{e}^x$，所以函數永不遞減，而事實上它是嚴格遞增（strictly increasing）的。在第 (ii) 部分，新函數的導數恰好在 $(1+x^2)\mathrm{e}^x = k$ 時為零，而這正是第 (i) 部分的方程。由此可知，新函數只有一個轉向點（turning point），而且是極小點（minimum）。之後，一切都取決於 $x \to -\infty$ 時的情況，此時 $\tan^{-1} x \to -\frac{\pi}{2}$。$2/\pi$ 這個數就是由此而來。

## 解答

### 第 (i) 部分

由積法則（product rule），

$$\mathrm{f}'(x) = 2x\,\mathrm{e}^x + (1+x^2)\mathrm{e}^x = \mathrm{e}^x\left(x^2+2x+1\right) = \mathrm{e}^x(1+x)^2.$$

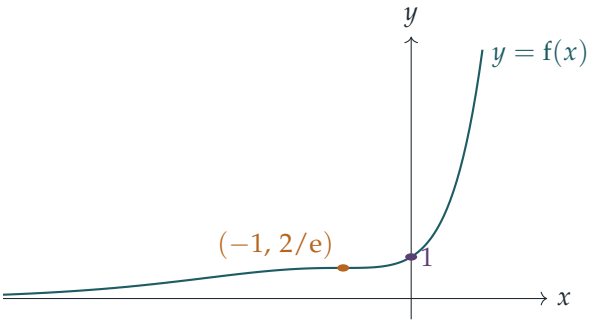
由於 $\mathrm{e}^x > 0$ 且 $(1+x)^2 \geq 0$，對所有 $x$ 都有 $\mathrm{f}'(x) \geq 0$，正如所求。等號只在 $x = -1$ 時成立。

? **想一想** $\mathrm{f}'(-1) = 0$，那麼 f 真的是嚴格遞增嗎？

是的。若 $a < b$，則 $\mathrm{f}(b) - \mathrm{f}(a) = \int_a^b \mathrm{f}'(x)\,\mathrm{d}x$，而被積函數（integrand）連續（continuous）、非負，並且在 $[a,b]$ 上最多只有一點為零，所以這個積分嚴格為正。單單一個駐點（stationary point）並不能阻止函數遞增。這裏 $x = -1$ 是一個駐拐點（stationary point of inflection）：$\mathrm{f}'$ 在它的兩側都是正的。

要畫出略圖，我們需要：

- 對所有 $x$ 都有 $\mathrm{f}(x) > 0$，因為兩個因子都是正的；
- $\mathrm{f}(0) = 1$，而拐點是 $(-1, 2/\mathrm{e}) \approx (-1, 0.74)$；
- 當 $x \to \infty$ 時，$\mathrm{f}(x) \to \infty$；
- 當 $x \to -\infty$ 時，$\mathrm{f}(x) \to 0$，因為 $\mathrm{e}^x$ 衰減的速度比任何多項式增長的速度都快（$x^2\mathrm{e}^x \to 0$）。所以負 $x$ 軸是一條漸近線（asymptote）。




現在 f 連續且嚴格遞增，並且當 $x \to -\infty$ 時 $\mathrm{f}(x) \to 0$，當 $x \to \infty$ 時 $\mathrm{f}(x) \to \infty$。所以它取遍 $(0, \infty)$ 內的每一個值（介值定理（intermediate value theorem））而且每個值恰好取一次（嚴格遞增）。它永遠不會取 $\leq 0$ 的值，因為 $\mathrm{f} > 0$。因此 $(1 + x^2)\mathrm{e}^x = k$ 在 $k > 0$ 時恰有一個實根（real root），在 $k \leq 0$ 時沒有實根。 ■

**第 (ii) 部分**

設

$$\mathrm{g}(x) = \mathrm{e}^x - 1 - k\tan^{-1} x.$$

立即可見 $\mathrm{g}(0) = 1 - 1 - 0 = 0$，所以 $x = 0$ 永遠是一個根。其次，

$$\mathrm{g}'(x) = \mathrm{e}^x - \frac{k}{1+x^2} = \frac{(1+x^2)\mathrm{e}^x - k}{1+x^2} = \frac{\mathrm{f}(x) - k}{1+x^2}.$$

兩種情況下都有 $k > 0$，所以由第 (i) 部分，恰有一個數 $a$ 滿足 $\mathrm{f}(a) = k$。因為 f 嚴格遞增，當 $x < a$ 時 $\mathrm{f}(x) < k$，當 $x > a$ 時 $\mathrm{f}(x) > k$。因此

$$\mathrm{g}'(x) < 0 \text{ 對於} x < a, \qquad \mathrm{g}'(x) > 0 \text{ 對於} x > a.$$

所以 g 在 $(-\infty, a]$ 上嚴格遞減（strictly decreasing），在 $[a, \infty)$ 上嚴格遞增，並在 $x = a$ 處有唯一的極小點。特別地，g 在 $a$ 的每一側最多只有一個根，所以總共最多有兩個根。

**$a$ 在哪裏？** 兩種情況下都有 $k < 1 = \mathrm{f}(0)$，所以 $\mathrm{f}(a) = k < \mathrm{f}(0)$，由此得 $a < 0$。於是 $x = 0$ 位於遞增的一段 $[a, \infty)$ 內。它是該處唯一的根，並且

$$\mathrm{g}(a) < \mathrm{g}(0) = 0.$$

餘下要判斷的，是 g 在 $(-\infty, a)$ 內是否還有一個根。

**$x \to -\infty$ 時的表現**。 由於 $\mathrm{e}^x \to 0$ 且 $\tan^{-1} x \to -\frac{\pi}{2}$，

$$\mathrm{g}(x) \to -1 + \frac{k\pi}{2} = \frac{\pi}{2}\left(k - \frac{2}{\pi}\right) \quad \text{當} x \to -\infty.$$

把這個極限（limit）記作 $L$。因為 g 在 $(-\infty, a]$ 上嚴格遞減，g 在該處的每一個值都嚴格小於 $L$。要看出這一點，取 $x \le a$ 和任意 $x' < x$。每一個 $y < x'$ 都滿足 $\mathrm{g}(y) > \mathrm{g}(x')$，所以令 $y \to -\infty$ 可得 $L \ge \mathrm{g}(x')$。因此 $\mathrm{g}(x) < \mathrm{g}(x') \le L$。

**(a)** $0 < k \le 2/\pi$。 這時 $L \le 0$，所以對每一個 $x \le a$ 都有 $\mathrm{g}(x) < L \le 0$。在 $a$ 的左方沒有根，故方程**恰有一個實根**，就是 $x = 0$。

**(b)** $2/\pi < k < 1$。 這時 $L > 0$，所以對所有足夠負的 $x$ 都有 $\mathrm{g}(x) > 0$，而 $\mathrm{g}(a) < 0$。由介值定理，在 $(-\infty, a)$ 內有一個根，而且因為 g 嚴格遞減，它是該處唯一的根。連同 $x = 0$，方程**恰有兩個實根**：一個負根（小於 $a$）和 $x = 0$。

> **想一想** 為甚麼邊界情況 $k = 2/\pi$ 歸入 (a)？

當 $k = 2/\pi$ 時，左方的極限恰好是 $L = 0$。遞減的一支從下方趨近 0，但永遠不會到達，所以不會出現第二個根。這一點正是靠嚴格單調性（strict monotonicity）來決定的。

## 驗算

- 在隨機選取的點上以數值方法對 f 求導，結果與 $\mathrm{e}^x(1+x)^2$ 吻合，而且 $\mathrm{f}(-1) = 2/\mathrm{e} \approx 0.736$。
- 我們在 $[-200, 10]$ 上用細密的格點數算 g 的變號次數。當 $k = 0.05, 0.3, 0.5, 0.6$ 和 $2/\pi \approx 0.6366$ 時有一個根，當 $k = 0.65, 0.7, 0.8, 0.9, 0.99$ 時有兩個根。當 $k = 0.65$ 時，負根離得很遠（在 $x \approx -31$ 附近），因為 $L = 0.65 \cdot \frac{\pi}{2} - 1 \approx 0.021$ 非常小。
- 當 $k = 0.7$ 時，極小點是 $a \approx -1.81$，而當 $k \to 1^-$ 時 $a \to 0^-$。這是吻合的：在 $k = 1$ 時，$\mathrm{g}'(0) = 0$，$x = 0$ 成為二重根（double root）。

## 常見錯誤

- 忘記了 $x = 0$ 永遠是一個根，於是只數算由 $-\infty$ 處變號而來的根。
- 以為當 $x \to -\infty$ 時 $\tan^{-1} x \to -\infty$。其實它是有界（bounded）的，趨向 $-\frac{\pi}{2}$。

► 由 $\mathrm{f}'(-1) = 0$ 斷定 f 在 $x = -1$ 處有轉向點。

## STEP 之外

這裏的論證是實分析（real analysis）中一種標準技巧的典範：在某區間上嚴格單調的連續函數在該區間內最多只有一個零點（zero），而介值定理（配合無窮遠處的極限使用）則告訴我們它何時確有一個零點。先找出函數的轉向點，再檢查函數在轉向點和兩端的正負號，從而數算根的數目，這正是羅爾定理（Rolle's theorem）及其推論背後的想法。Abbott 的 *Understanding Analysis* 嚴謹地闡述了連續性、介值定理和均值定理（mean value theorem）。

詳細解答

# STEP III 1999，第 3 題

純數學 • 黎曼和與極限

**STEP III 1999，第 3 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/99-s3-q3

## 關鍵思路

算式 $\frac{1}{n}\sum f(1+m/n)$ 是豎立在 $[1,2]$ 上的 $n$ 個闊度為 $\frac{1}{n}$ 的幼長矩形的總面積；矩形愈來愈幼時，它們便恰好填滿曲線下的面積。另外兩個極限（limit）其實是同一幅圖畫的變相：要做的就是抽出因子 $\frac{1}{n}$，再把每一項寫成只含 $m/n$ 的函數。

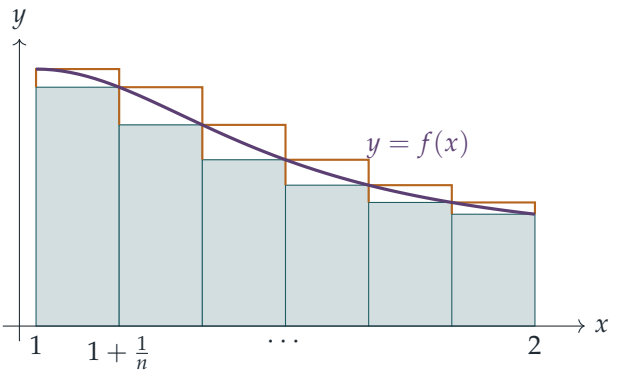
## 解答

### 極限公式

把 $[1,2]$ 分成 $n$ 條闊度為 $\frac{1}{n}$ 的長條；第 $m$ 條是 $\left[1+\frac{m-1}{n},\,1+\frac{m}{n}\right]$。在這長條上豎立一個矩形，其高度取 $f$ 在長條右端的值 $f(1+m/n)$。它的面積是 $\frac{1}{n}f(1+m/n)$，所以

$$S_n = \frac{1}{n}\sum_{m=1}^{n} f(1+m/n)$$

就是這些矩形的總面積。




草圖所示的是一個遞減的 $f$。着色（藍綠色）的矩形以右端的函數值為高，因此位於曲線之下；頂上的琥珀色梯級則顯示改用左端的函數值為高時所得的矩形，它們位於曲線之上。因此

$$S_n \;\le\; \int_1^2 f(x)\,\mathrm{d}x \;\le\; S_n + (\text{琥珀色部分的總面積}).$$

現在把所有琥珀色部分橫向平移，疊成闊度為 $\frac{1}{n}$ 的單獨一列：由於各段高度逐項相消（telescope），它們疊起來的總高度是 $f(1)-f(2)$。所以

$$0 \;\le\; \int_1^2 f(x)\,\mathrm{d}x - S_n \;\le\; \frac{f(1)-f(2)}{n} \;\longrightarrow\; 0 \quad \text{當} n\to\infty,$$

於是 $S_n \to \int_1^2 f(x)\,\mathrm{d}x$。遞增的 $f$ 可用同樣方法處理，只是不等號的方向相反；而任何合理的函數（例如連續而且分段單調（piecewise monotonic）的函數）都可以切成有限多個單調的部分，每部分所帶來的誤差都是 $\frac{1}{n}$ 的量級。這就說明了

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n} f(1+m/n) = \int_1^2 f(x)\,\mathrm{d}x.$$

**極限** $\ln 2$

**想一想** 怎樣才能令 $\dfrac{1}{n+m}$ 看起來像 $\dfrac{1}{n}f(1+m/n)$？

從分母抽出因子 $n$：

$$\frac{1}{n+m} = \frac{1}{n}\cdot\frac{1}{1+m/n} = \frac{1}{n}f(1+m/n) \qquad \text{其中} f(x)=\frac{1}{x}.$$

所以

$$\begin{aligned}\lim_{n\to\infty}\left(\frac{1}{n+1}+\cdots+\frac{1}{n+n}\right) &= \lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}\frac{1}{1+m/n}\\ &= \int_1^2 \frac{1}{x}\,\mathrm{d}x = \ln 2-\ln 1 = \ln 2,\end{aligned}$$

正如所求。（這裏 $f$ 是遞減的，所以草圖的論證完全適用；它甚至告訴我們，這個和介乎 $\ln 2 - \frac{1}{2n}$ 與 $\ln 2$ 之間。）

### 最後一個極限

第 $m$ 項是 $\dfrac{n}{n^2+m^2}$，其中 $m = 1, \ldots, n$。把分子和分母同時除以 $n^2$：

$$\frac{n}{n^2+m^2} = \frac{1}{n} \cdot \frac{1}{1+(m/n)^2}.$$

公式是在 $x = 1 + m/n$ 處計算 $f$ 的值，要與它吻合，便需要 $f(x) = \dfrac{1}{1+(x-1)^2}$，因為這樣就有 $f(1+m/n) = \dfrac{1}{1+(m/n)^2}$。這個 $f$ 在 $[1,2]$ 上是遞減的。因此

$$\lim_{n\to\infty} \sum_{m=1}^{n} \frac{n}{n^2+m^2} = \int_1^2 \frac{\mathrm{d}x}{1+(x-1)^2} = \Big[\tan^{-1}(x-1)\Big]_1^2$$
$$= \tan^{-1} 1 - \tan^{-1} 0 = \frac{\pi}{4}.$$

（另一個等價的做法是代入 $u = x - 1$，得出 $\int_0^1 \frac{\mathrm{d}u}{1+u^2}$。）

## 驗算

這些和可以直接計算。取 $n = 1000$，第二個和是 0.69290，而 $\ln 2 \approx 0.69315$；第三個和是 0.78515，而 $\pi/4 \approx 0.78540$。兩者的差距都約為 $\frac{1}{4000}$，在草圖所給出的界 $\frac{f(1)-f(2)}{n} = \frac{1}{2000}$ 之內。兩個和也都略低於極限；對於遞減的 $f$ 並以右端的函數值為高，結果正應如此。

## 常見錯誤

- 把第三個極限寫成 $\int_1^2 \frac{\mathrm{d}x}{1+x^2}$：公式是在 $1 + m/n$ 處而不是在 $m/n$ 處計算 $f$ 的值，所以那個 1 的平移必須放進 $f$ 裏面。
- 忘記最後一項（$m = n$）是 $\frac{n}{2n^2}$，它對應 $\frac{1}{n}f(2)$：求和的範圍是全部 $n$ 個右端點，而不是 $n+1$ 個點。

## STEP 之外

這些都是黎曼和（Riemann sum），而草圖的論證正是黎曼積分（Riemann integral）定義的雛形：當上和（upper sum）與下和（lower sum）可以夾迫到一起時，函數便是可

積（integrable）的，而閉區間上的每個連續函數都具有這個性質。同一個想法也可以倒過來用，把和化為積分來估算，例如用於證明 $1+\frac{1}{2}+\cdots+\frac{1}{n}-\ln n$ 收斂。Abbott 的 *Understanding Analysis*（一本數學分析入門書）中有一章講黎曼積分，寫得十分易讀。

詳細解答

# STEP III 1999，第 4 題

純數學 • 歐拉公式與多面體

**STEP III 1999，第 4 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/99-s3-q4

本題在 *Beyond STEP* 第 7.9 節只有五種柏拉圖立體中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch07/platonic-solids.html

## 關鍵思路

用兩種不同的方法數棱（edge）：一次逐個面（face）去數，一次逐個頂點（vertex）去數。每種數法都把 $E$ 用 $F$（或 $V$）表示出來，代入歐拉公式（Euler's formula）後，便只剩下一條關於 $F$ 的方程。之後整個分類只取決於一項觀察：$F$ 是正數，所以 $h$ 必須是正數，於是只剩下寥寥幾組整數對 $(m, n)$。

## 解答

### 推導 $F$ 的公式

每個面都是 $m$ 邊形，所以有 $m$ 條棱。逐個面去數它的棱，總數是 $mF$。但每條棱恰好是兩個面之間的邊界，所以這樣數會把每條棱數了兩次：

$$mF = 2E.$$

同樣地，每個頂點有 $n$ 條棱相交，所以逐個頂點去數棱，總數是 $nV$。每條棱恰好有兩個端點，所以這樣數也把每條棱數了兩次：

$$nV = 2E.$$

因此 $E = \frac{1}{2}mF$，而 $V = \frac{2E}{n} = \frac{mF}{n}$。代入 $V - E + F = 2$ 得

$$\frac{mF}{n} - \frac{mF}{2} + F = 2$$
$$F(2m - mn + 2n) = 4n \qquad \text{（兩邊乘以 } 2n\text{）.}$$

現在把題目給出的 $h$ 展開：

$$4 - (n-2)(m-2) = 4 - (mn - 2m - 2n + 4) = 2m + 2n - mn,$$

這正好就是括號內的式子。所以 $Fh = 4n$。這裏 $F$ 和 $n$ 都是正數，所以 $h \neq 0$，可以相除：

$$F = \frac{4n}{h}, \qquad h = 4 - (n-2)(m-2),$$

這就是所要證明的結果。

### 只有五種正多面體

**想一想** 有哪些對 $m$ 和 $n$ 的限制是來自幾何，而不是來自公式？

多邊形最少有三條邊，所以 $m \geq 3$。在立體的一個頂點處，最少要有三個面相交：只有兩個面在一個角上相交，得到的只是一塊扁平的摺片，而不是立體的一個角。所以同樣有 $n \geq 3$。於是 $m - 2 \geq 1$ 及 $n - 2 \geq 1$。

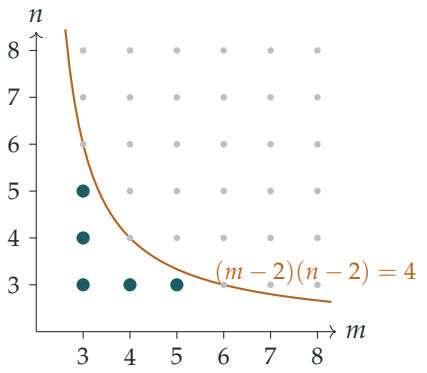
由於 $F > 0$ 及 $4n > 0$，公式迫使 $h > 0$，即

$$(m-2)(n-2) < 4.$$

因此兩個正整數 $m-2$ 與 $n-2$ 的積是 1、2 或 3，所以 $h = 3$、2 或 1。僅有的因數分解（factorisation）是

- $(m-2)(n-2) = 1$：$m = n = 3$，所以 $h = 3$；
- $(m-2)(n-2) = 2$：$(m,n) = (3,4)$ 或 $(4,3)$，所以 $h = 2$；
- $(m-2)(n-2) = 3$：$(m,n) = (3,5)$ 或 $(5,3)$，所以 $h = 1$。

下圖顯示的是同一回事：滿足 $m, n \geq 3$ 並嚴格位於雙曲線（hyperbola）$(m-2)(n-2) = 4$ 以內的格點（lattice point）。




對每一組數對，$F = 4n/h$、$E = \frac{1}{2}mF$ 及 $V = mF/n$：

| | $m$ | $n$ | $h$ | $V$ | $E$ | $F$ |
|---|---|---|---|---|---|---|
| 正四面體（tetrahedron） | 3 | 3 | 3 | 4 | 6 | 4 |
| 正八面體（octahedron） | 3 | 4 | 2 | 6 | 12 | 8 |
| 正方體（cube） | 4 | 3 | 2 | 8 | 12 | 6 |
| 正二十面體（icosahedron） | 3 | 5 | 1 | 12 | 30 | 20 |
| 正十二面體（dodecahedron） | 5 | 3 | 1 | 20 | 30 | 12 |

例如，對正二十面體，$F = 4 \cdot 5/1 = 20$、$E = \frac{1}{2} \cdot 3 \cdot 20 = 30$ 及 $V = 3 \cdot 20/5 = 12$。

所以 $(m, n)$ 最多只有五種可能的組合。每一種都確實存在，表中列出的五種為人熟悉的柏拉圖立體（Platonic solid）就是例證。所以正多面體（regular polyhedron）恰好有五種。

## 驗算

- 每一行都滿足 $V - E + F = 2$。對正十二面體，$20 - 30 + 12 = 2$。
- 正方體給出 $V = 8$、$E = 12$、$F = 6$，你可以拿一粒骰子來數一數。
- 把 $m$ 和 $n$ 對調，$V$ 和 $F$ 便會對調，而 $E$ 保持不變（正方體與正八面體、正十二面體與正二十面體）。這就是對偶性（duality）：在每個面的中心放一個頂點。
- 一項獨立的驗算：每個頂點的角虧（angle deficit）$2\pi - n\pi(m-2)/m$ 必須是正數，而全部 $V$ 個頂點的角虧總和是 $4\pi$（笛卡兒定理，Descartes' theorem）。五種立體全都符合，而用電腦搜尋 $3 \le m, n < 200$ 的範圍，也找不到其他滿足 $h > 0$ 的數對。

## 常見錯誤

- 數棱時忘記了因子 2：每條棱屬於兩個面，並有兩個端點。
- 沒有說明為甚麼 $n \geq 3$。少了這個條件，$n = 2$ 時對每個 $m$ 都有 $h = 4$，於是會出現無限多個「解」。
- 單憑數算就下結論說「恰好五種」。這個論證只說明最多有五種；你還需要實際的立體，才能說明每種情況都存在。

## STEP 之外

歐拉公式是拓撲學（topology）的起點。無論用哪種方法把一個曲面分割成面，$V - E + F$ 這個數（即歐拉示性數（Euler characteristic））都是一樣的，所以球面是 2，環面（torus）是 0，如此類推。在環面上重複這個論證，會得到無限多種「正則」的鋪砌（tiling），例如正方形網格。這五種立體的對稱羣（symmetry group），也是羣論（group theory）入門課程中的核心例子。想讀一本生動有趣的介紹，可看 Richeson 的 *Euler's Gem*；Armstrong 的 *Basic Topology* 則把這些想法再推進一步。

詳細解答

# STEP III 1999，第 5 題

純數學 • 斐波那契恆等式與數學歸納法

**STEP III 1999，第 5 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/99-s3-q5

## 關鍵思路

這其實是斐波那契數列（Fibonacci sequence），只是下標移了一位，使得 $u_n = F_{n+1}$。要證第一個恆等式（identity），可把 $u_{n+2}$ 和 $u_{n-1}$ 都用中間的一對 $u_{n+1}$、$u_n$ 表示：交叉項隨即互相抵消。至於倍增公式（doubling formulae），應以歸納法（induction）把兩個命題一併證明，因為兩者透過遞推關係（recurrence）互相支援。第一個恆等式正是完成奇數情況所需的工具。

## 解答

### 第一個恆等式

固定 $n \geq 1$，使得 $u_{n-1}$、$u_n$、$u_{n+1}$、$u_{n+2}$ 全部有定義。由遞推關係可得

$$u_{n+2} = u_{n+1} + u_n \qquad 及 \qquad u_{n-1} = u_{n+1} - u_n,$$

其中第二式來自 $u_{n+1} = u_n + u_{n-1}$。兩式分別平方後相加，

$$\begin{aligned} u_{n+2}^2 + u_{n-1}^2 &= \left(u_{n+1}^2 + 2u_{n+1}u_n + u_n^2\right) + \left(u_{n+1}^2 - 2u_{n+1}u_n + u_n^2\right) \\ &= 2\left(u_{n+1}^2 + u_n^2\right), \end{aligned}$$

這就是所求的恆等式。

### 倍增公式

以 $P(n)$ 表示以下兩式同時成立的命題：

$$u_{2n} = u_n^2 + u_{n-1}^2 \qquad 及 \qquad u_{2n+1} = u_{n+1}^2 - u_{n-1}^2.$$

數列的首幾項為 $u_0, u_1, \cdots = 1, 1, 2, 3, 5, 8, \ldots$。

**基礎情況**。當 $n = 1$ 時：$u_2 = 2 = 1^2 + 1^2 = u_1^2 + u_0^2$，而 $u_3 = 3 = 2^2 - 1^2 = u_2^2 - u_0^2$。所以 $P(1)$ 成立。

> **想一想** 為甚麼要把兩條公式一併證明，而不是逐條證明？

要由遞推關係求得 $u_{2n+2}$，我們需要 $u_{2n+1}$ 和 $u_{2n}$，所以 $n+1$ 時的偶數項公式需要用到 $n$ 時的兩條公式。其後 $n+1$ 時的奇數項公式又需要用到 $n+1$ 時的偶數項公式。兩個命題都無法單獨向前推進。

**歸納步驟**。假設 $P(n)$ 對某個 $n \geq 1$ 成立。

偶數項。先用遞推關係，再用 $P(n)$，

$$\begin{aligned} u_{2n+2} &= u_{2n+1} + u_{2n} \\ &= (u_{n+1}^2 - u_{n-1}^2) + (u_n^2 + u_{n-1}^2) \\ &= u_{n+1}^2 + u_n^2, \end{aligned}$$

這就是把 $n$ 換成 $n+1$ 後的偶數項公式。

奇數項。利用遞推關係、剛證得的一式以及 $P(n)$，

$$\begin{aligned} u_{2n+3} &= u_{2n+2} + u_{2n+1} \\ &= (u_{n+1}^2 + u_n^2) + (u_{n+1}^2 - u_{n-1}^2) \\ &= 2u_{n+1}^2 + u_n^2 - u_{n-1}^2. \end{aligned}$$

我們希望這等於 $u_{n+2}^2 - u_n^2$，而第一個恆等式正好在這裏派上用場。由於 $n \geq 1$，該恆等式給出 $u_{n+2}^2 = 2u_{n+1}^2 + 2u_n^2 - u_{n-1}^2$，所以

$$u_{n+2}^2 - u_n^2 = 2u_{n+1}^2 + u_n^2 - u_{n-1}^2 = u_{2n+3}.$$

這就是把 $n$ 換成 $n+1$ 後的奇數項公式，因為 $2(n+1)+1 = 2n+3$，而 $(n+1) \pm 1 = n+2,\ n$。

因此 $P(n)$ 蘊涵 $P(n+1)$。連同基礎情況，由歸納法可知 $P(n)$ 對所有正整數 $n$ 成立。

## 驗算

試取 $n = 5$：$u_{10} = 89$，而 $u_5^2 + u_4^2 = 64 + 25 = 89$；$u_{11} = 144$，而 $u_6^2 - u_4^2 = 169 - 25 = 144$。我們以精確的整數運算，對所有 $n \leq 299$ 驗證了第一個恆等式，並對所有 $n \leq 189$ 驗證了兩條倍增公式。另外留意 $u_{n+1}^2 - u_{n-1}^2 = (u_{n+1} - u_{n-1})(u_{n+1} + u_{n-1}) = u_n(u_{n+1} + u_{n-1})$，這是奇數項另一個方便好用的形式。

## 常見錯誤

- 只對偶數項公式作歸納：歸納步驟需要 $u_{2n+1}$，於是無法繼續。
- 在 $n = 0$ 時使用第一個恆等式，但此時 $u_{-1}$ 並無定義。在歸納步驟中，我們只在 $n \geq 1$ 時使用它。
- 忘記這裏的下標與慣常的斐波那契數相比移了一位（$u_0 = u_1 = 1$），以致用錯誤的數值驗證基礎情況。

## STEP 之外

倍增公式讓你只需大約 $\log_2 N$ 步便可算出 $u_N$，方法是把 $N$ 反覆減半：這就是「快速倍增」(fast doubling) 法，與快速冪 (fast exponentiation) 算法一脈相承。它背後是矩陣恆等式 $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{pmatrix}$，把這個矩陣平方，便可一次過得出兩條公式。Strang 的 *Introduction to Linear Algebra*（線性代數入門）透過這個矩陣的冪和特徵值（eigenvalues）來處理斐波那契數；而在 Davenport 的 *The Higher Arithmetic*（數論入門）中，講連分數（continued fractions）的一章會再次遇上相鄰斐波那契數之比，它們正是黃金比（golden ratio）的漸近分數（convergents）。

詳細解答

# STEP III 1999，第 6 題

純數學 • 參數曲線與面積

**STEP III 1999，第 6 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/99-s3-q6

## 關鍵思路

方程其實是說 $\left(x^{1/n}\right)^2 + \left(y^{1/n}\right)^2 = \left(a^{1/n}\right)^2$，這是一個喬裝了的圓。所以，像處理圓那樣，設 $x^{1/n} = a^{1/n}\cos t$ 及 $y^{1/n} = a^{1/n}\sin t$，然後取 $n$ 次方。正因為 $n$ 是奇數，我們才可以取負數的實 $n$ 次方根，所以曲線遍佈全部四個象限。

## 解答

### 參數化

因為 $n$ 是奇數，每個實數 $x$ 都恰好有一個實 $n$ 次方根 $x^{1/n}$，其正負號與 $x$ 相同，而 $x^{2/n}$ 的意思就是 $\left(x^{1/n}\right)^2$。設 $X = x^{1/n}$、$Y = y^{1/n}$ 及 $R = a^{1/n} > 0$，方程 $(*)$ 變成

$$X^2 + Y^2 = R^2,$$

這是一個半徑為 $R$ 的圓，我們以 $X = R\cos t$、$Y = R\sin t$ 把它參數化（parametrise）。取 $n$ 次方（由於 $n$ 是奇數，這是 $\mathbb{R}$ 上的一個雙射（bijection））便得

$$\boxed{x = a\cos^n t, \qquad y = a\sin^n t, \qquad 0 \le t \le 2\pi.}$$

曲線上每一點都可以這樣得出，因為圓上每一點 $(X, Y)$ 都可以這樣得出。

? **想一想** 為甚麼這是逆時針方向，而不僅僅是「某種」繞行曲線的方式？

映射 $(X, Y) \mapsto (X^n, Y^n)$ 保持每個坐標的正負號不變，所以它把每個象限映到自身。當 $t$ 由 0 增至 $2\pi$ 時，點 $(R\cos t, R\sin t)$ 沿逆時針方向繞圓一周，依次經過正 $x$ 軸、第一象限、正 $y$ 軸、第二象限，如此類推。它的像 $(x, y)$ 以相同的次序經過相同的軸和象限，

所以也是沿逆時針方向繞原點一周。（例如在第一象限內，$x = a\cos^n t$ 由 $a$ 遞減至 0，而 $y = a\sin^n t$ 則由 0 遞增至 $a$。）當 $n = 1$ 時，這就是我們熟悉的圓 $x = a\cos t$、$y = a\sin t$。

### $n = 3$ 時的草圖

現在 $x = a\cos^3 t$、$y = a\sin^3 t$，而曲線是 $x^{2/3} + y^{2/3} = a^{2/3}$。

對稱性。方程在 $x \mapsto -x$、$y \mapsto -y$ 以及 $x$ 與 $y$ 互換之下都保持不變。所以曲線對兩條坐標軸以及直線 $y = \pm x$ 都對稱，我們只需研究第一象限，即 $0 \le t \le \frac{\pi}{2}$。

截距。當 $t = 0$ 時，我們位於 $(a, 0)$；當 $t = \frac{\pi}{2}$ 時，位於 $(0, a)$。根據對稱性，曲線只在 $(\pm a, 0)$ 和 $(0, \pm a)$ 與坐標軸相交。在直線 $y = x$ 上（即 $t = \frac{\pi}{4}$ 時），我們位於 $\left(\frac{a}{2\sqrt{2}}, \frac{a}{2\sqrt{2}}\right)$，沿每條軸的方向只走了約 $0.35a$。

斜率。我們有

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -3a\cos^2 t\sin t, \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = 3a\sin^2 t\cos t,$$

所以當 $0 < t < \frac{\pi}{2}$ 時

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{3a\sin^2 t\cos t}{-3a\cos^2 t\sin t} = -\tan t.$$

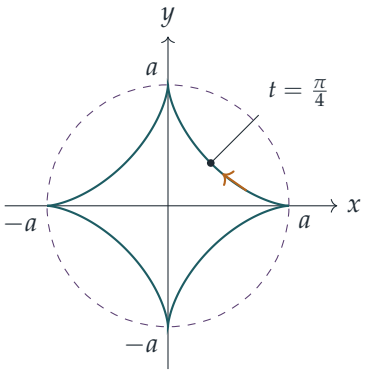
當 $t \to 0$ 時，這個值趨向 0；當 $t \to \frac{\pi}{2}$ 時，它趨向 $-\infty$。所以曲線到達 $(a, 0)$ 時與 $x$ 軸相切，到達 $(0, a)$ 時與 $y$ 軸相切。$\mathrm{d}x/\mathrm{d}t$ 和 $\mathrm{d}y/\mathrm{d}t$ 在 $t = 0$ 和 $t = \frac{\pi}{2}$ 時都等於零。第一象限的弧與由第四象限反射而來的弧在 $(a, 0)$ 相遇，而且兩者在該處都與 $x$ 軸相切，所以兩段弧會合成一個尖點。曲線在 $(\pm a, 0)$、$(0, \pm a)$ 這四點有尖點（cusp）。

凹凸性。把 $-\tan t$ 對 $x$ 求導，

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \frac{-\sec^2 t}{-3a\cos^2 t\sin t} = \frac{1}{3a\cos^4 t\sin t} > 0 \quad \text{當} 0 < t < \frac{\pi}{2}.$$

所以第一象限的弧是凸（convex）的：它向原點彎入，而不像圓那樣向外彎。它位於其弦 $x + y = a$ 的下方。事實上 $\cos^3 t + \sin^3 t \le \cos^2 t + \sin^2 t = 1$，等號只在尖點成立。

綜合以上各點，便得到一個四角星形（星形線（astroid））。下圖畫出這條曲線，並以虛線畫出半徑為 $a$ 的圓作比較。



### $n = 3$ 時的面積

利用上面的導數，

$$\begin{aligned} x\frac{\mathrm{d}y}{\mathrm{d}t} - y\frac{\mathrm{d}x}{\mathrm{d}t} &= a\cos^3 t \cdot 3a\sin^2 t\cos t + a\sin^3 t \cdot 3a\cos^2 t\sin t \\ &= 3a^2\sin^2 t\cos^2 t\,(\cos^2 t + \sin^2 t) \\ &= 3a^2\sin^2 t\cos^2 t = \tfrac{3}{4}a^2\sin^2 2t. \end{aligned}$$

由於 $\sin^2 2t = \frac{1}{2}(1-\cos 4t)$，而 $\cos 4t$ 在 $[0, 2\pi]$ 上的積分為零，

$$\int_0^{2\pi} \sin^2 2t\,\mathrm{d}t = \pi.$$

因此

$$A = \frac{1}{2}\cdot\frac{3a^2}{4}\cdot\pi = \boxed{\frac{3\pi a^2}{8}}.$$

公式給出正的面積，是因為曲線是沿逆時針方向繞行的，這正是第一部分中方向之所以重要的原因。

## 驗算

- 星形線位於菱形 $|x| + |y| \le a$ 之內，而這個菱形的面積是 $2a^2$，所以答案必定小於 $2a^2$。果然 $\frac{3\pi}{8}a^2 \approx 1.178a^2$。
- 作為一個不用該公式的驗算，取 $a = 1$，以數值方法計算 $4\int_0^a \left(a^{2/3} - x^{2/3}\right)^{3/2}\mathrm{d}x$，得 $1.178097\ldots$，與 $3\pi/8 = 1.1780972\ldots$ 吻合。
- 當 $n = 1$ 時，同樣的方法給出 $x\dot{y} - y\dot{x} = a^2$ 及 $A = \pi a^2$，即圓的面積，正如所料。

## 常見錯誤

- 寫下 $x = a\cos^n t$ 卻沒有解釋為甚麼 $x$ 可以是負數。正是因為 $n$ 是奇數，$x^{2/n}$ 在 $x < 0$ 時才有意義（參數化也才是完整的）。
- 把曲線畫成像壓扁了的圓那樣向外鼓出。二階導數顯示它是向內彎的，並在坐標軸上有尖點。
- 只在 $[0, \frac{\pi}{2}]$ 上積分而忘記乘以 4，或在 $-y\,\mathrm{d}x/\mathrm{d}t$ 一項弄錯正負號（兩項最終都是正的）。

## STEP 之外

這條面積公式是格林定理（Green's theorem）的特殊情況；格林定理把一個區域上的積分轉化為沿其邊界的積分。它也是求積儀（planimeter）背後的原理，求積儀是一種沿邊界描畫便能量度面積的機械儀器。像 $|x|^p + |y|^p = 1$ 這類曲線，也給出 $\ell^p$ 範數（norm）的「單位圓」，你會在分析和線性代數中遇到它們；星形線是 $p = \frac{2}{3}$ 的情況，它不是凸的，所以並不給出一個範數。星形線也是一個圓在半徑為其四倍的圓內滾動時，圓上一點所描出的軌跡。想看更多同類型的題目，可參閱 Siklos 的 *Advanced Problems in Mathematics*。

詳細解答

# STEP III 1999，第 7 題

純數學 • 群與同構

> **STEP III 1999，第 7 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/99-s3-q7

## 關鍵思路

表達式 $x+y+axy$ 幾乎就是一個乘積。乘以 $a$ 再加上 1，可得

$$1+a(x*y) = 1+ax+ay+a^2xy = (1+ax)(1+ay),$$

所以在變數代換（change of variable）$X = 1+ax$ 之下，運算 $*$ 其實只是喬裝了的普通乘法。羣（group）的每一項性質，都可由乘法的相應性質得出，而我們必須剔除的那個數，正是對應於 $X=0$ 的那一個。

## 解答

以下一律記 $\phi(x) = 1+ax$。由於 $a \neq 0$，$\phi$ 是由 $\mathbb{R}$ 到 $\mathbb{R}$ 的雙射（bijection），所以兩個實數 $u, v$ 相等，當且僅當 $\phi(u) = \phi(v)$。上面的恆等式可寫成

$$\phi(x*y) = \phi(x)\,\phi(y) \qquad \text{對所有實數} x, y. \tag{†}$$

### 結合律

在等式兩邊各用 (†) 兩次，

$$\phi\big((x*y)*z\big) = \phi(x*y)\,\phi(z) = \phi(x)\,\phi(y)\,\phi(z),$$
$$\phi\big(x*(y*z)\big) = \phi(x)\,\phi(y*z) = \phi(x)\,\phi(y)\,\phi(z).$$

由於實數乘法符合結合律（associativity），這兩式相等；又因為 $\phi$ 是一對一（one-to-one）的，所以 $(x*y)*z = x*(y*z)$。因此 $*$ 符合結合律。

（如果你想直接驗證：$(x*y)*z$ 和 $x*(y*z)$ 展開後都是對稱的表達式 $x+y+z+a(xy+yz+zx)+a^2xyz$。）

## 羣

**想一想** 要剔除的是哪一個數？為甚麼不能保留整個 $\mathbb{R}$？

在 $\phi$ 之下，$x = -\frac{1}{a}$ 這個數對應於 $\phi(x) = 0$，而 0 沒有乘法逆元(multiplicative inverse)。具體來說，對任何實數 $y$，

$$\left(-\tfrac{1}{a}\right) * y = -\tfrac{1}{a} + y - y = -\tfrac{1}{a},$$

所以 $-\frac{1}{a}$ 會把一切都吸收掉，無論與甚麼元素結合，都不可能得出單位元（identity）。因此我們取

$$G = \mathbb{R} \setminus \left\{-\tfrac{1}{a}\right\},$$

這正是所有滿足 $\phi(x) \neq 0$ 的 $x$ 所組成的集合。我們逐一驗證四條公理。

封閉性（*closure*）。若 $x, y \in G$，則 $\phi(x) \neq 0$ 且 $\phi(y) \neq 0$，所以由 (†) 得 $\phi(x * y) = \phi(x)\phi(y) \neq 0$。故此 $x * y \neq -\frac{1}{a}$，而 $x * y$ 當然是實數，所以 $x * y \in G$。

結合律。上面已對所有實數證明了這一點，所以對 $G$ 的元素自然也成立。

單位元。對每個 $x$，都有 $x * 0 = x + 0 + 0 = x$ 及 $0 * x = x$，而且因為 $\phi(0) = 1 \neq 0$，所以 $0 \in G$。因此 0 是單位元。(這與 $\phi(0) = 1$ 吻合，即對應乘法的單位元。)

逆元（*inverse*）。給定 $x \in G$，我們要找 $y$ 使得 $x * y = 0$，即 $x + y(1 + ax) = 0$。由於 $1 + ax \neq 0$，可以相除，得出

$$x^{-1} = \frac{-x}{1 + ax}.$$

這個數屬於 $G$，因為

$$\phi\left(\frac{-x}{1+ax}\right) = 1 - \frac{ax}{1+ax} = \frac{1}{1+ax} \neq 0,$$

而且由於 $*$ 是可交換（commutative）的（從公式明顯可見 $x * y = y * x$），所以同樣有 $x^{-1} * x = 0$。

四條公理全部成立，所以 $(G, *)$ 是一個羣（事實上是阿貝爾羣（abelian group）），其中被剔除的數是 $-\frac{1}{a}$。

## 2 階子羣

非零實數的乘法羣有子羣 (subgroup) $\{1, -1\}$。把它經 $\phi$ 拉回來：$\phi(x) = 1$ 給出 $x = 0$，而 $\phi(x) = -1$ 給出 $x = -\frac{2}{a}$。所以我們嘗試

$$H = \left\{0, -\tfrac{2}{a}\right\}.$$

這是 $G$ 中兩個不同的元素（因為 $a \neq 0$，$-\frac{2}{a} \neq 0$ 且 $-\frac{2}{a} \neq -\frac{1}{a}$）。直接計算，

$$\left(-\tfrac{2}{a}\right) * \left(-\tfrac{2}{a}\right) = -\tfrac{4}{a} + a \cdot \tfrac{4}{a^2} = 0,$$

以及 $0*0=0$，$0*\left(-\frac{2}{a}\right)=-\frac{2}{a}$。所以 $H$ 是封閉的，包含單位元，而且每個元素都是自己的逆元。因此 $H$ 是 $(G,*)$ 的子羣，恰好有兩個元素。

事實上這是唯一的一個：階（order）為 2 的羣必是 $\{e,h\}$，其中 $h*h=e$ 且 $h\neq e$；而 $h*h=0$ 即 $h(2+ah)=0$，所以 $h=-\frac{2}{a}$。

## 驗算

取 $a=1$。這時 $x*y=x+y+xy=(1+x)(1+y)-1$，被剔除的數是 $-1$，而 2 階子羣是 $\{0,-2\}$：的確 $(-2)*(-2)=-4+4=0$。此外 $2*3=11$，而 $(1+2)(1+3)=12=1+11$，與 (†) 所預期的一致。我們也用電腦以精確的有理數運算，對約 5000 組隨機選取的 $a,x,y,z$ 驗證了恆等式 (†)、結合律、$G$ 的封閉性、單位元 0、逆元公式，以及 $\{0,-\frac{2}{a}\}$ 的封閉性。

## 常見錯誤

- 忘記驗證逆元 $\frac{-x}{1+ax}$ 本身屬於 $G$，或忘記驗證 $x*y$ 永遠不會等於被剔除的那個數。正是封閉性這條公理，說明了為何要剔除的恰好是一個點。
- 以 $\{0\}$ 或 $\{0,-\frac{1}{a}\}$ 作為所求的子羣：前者只有一個元素，而 $-\frac{1}{a}$ 根本不在 $G$ 之中。

## STEP 之外

映射 $\phi(x)=1+ax$ 是由 $(G,*)$ 到羣 $(\mathbb{R}\setminus\{0\},\times)$ 的同構（*isomorphism*）：即一個把其中一種運算變成另一種運算的雙射。一旦找到同構，所有結構上的事實都可以不費吹灰之力地搬過去，這就是為甚麼 $(G,*)$ 恰好只有一個 2 階元素，正如 $\times$ 之下只有 $-1$ 一樣。同構，以及兩個羣可以外表不同卻「相同」這個想法，是抽象代數（abstract algebra）入門課程的核心；Allenby 的 *Rings, Fields and Groups* 是閱讀這方面內容的好選擇。

詳細解答

# STEP III 1999，第 8 題

純數學 • 分段微分方程

**STEP III 1999，第 8 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/99-s3-q8

## 關鍵思路

在每個區間上，方程都是頻率固定的簡諧運動（simple harmonic motion），所以該區間上的通解（general solution）是 $\cos nx$ 與 $\sin nx$ 的組合。唯一要做的工夫是銜接：連續性條件（continuity conditions）把上一個區間末端的 $y$ 和 $y'$ 的值，交給下一個區間作為起始值。各區間的長度都是 $2\pi$，結果每次的起始值都相同，於是出現一個規律，我們可以用歸納法（induction）加以證明。

## 解答

### 第 (i) 部分

當 $0 < x < 2\pi$ 時，$n = 1$，所以 $y'' + y = 0$，其通解為

$$y = A\cos x + B\sin x.$$

由於 $y$ 和 $y'$ 在 $x = 0$ 連續，條件 $y(0) = 0$ 和 $y'(0) = 1$ 適用於這個解：$y(0) = A = 0$ 及 $y'(0) = B = 1$。因此

$$y = \sin x \qquad (0 \le x \le 2\pi).$$

在右端，$y(2\pi) = \sin 2\pi = 0$ 及 $y'(2\pi) = \cos 2\pi = 1$。

### 第 (ii) 部分

當 $2\pi < x < 4\pi$ 時，$n = 2$，所以 $y'' + 4y = 0$，且

$$y = A\cos 2x + B\sin 2x.$$

由 $y$ 和 $y'$ 在 $x = 2\pi$ 的連續性，這個解在該處必須滿足 $y = 0$ 和 $y' = 1$，即第 (i) 部分末尾求得的值。由於 $\cos 4\pi = 1$ 及 $\sin 4\pi = 0$，

$$y(2\pi) = A = 0, \qquad y'(2\pi) = 2B\cos 4\pi = 2B = 1,$$

所以 $B = \frac{1}{2}$，而當 $2\pi \le x \le 4\pi$ 時 $y = \frac{1}{2}\sin 2x$，正如題目所求。

留意這裏同樣有 $y(4\pi) = \frac{1}{2}\sin 8\pi = 0$ 及 $y'(4\pi) = \cos 8\pi = 1$：下一個區間的起始值完全一樣。由此可提出以下斷言。

**斷言**。對每個正整數 $n$，

$$y = \frac{1}{n}\sin nx \qquad \text{對於} 2(n-1)\pi \le x \le 2n\pi.$$

用歸納法證明。$n = 1$ 的情況就是第 (i) 部分。假設斷言對 $n = k$ 成立。在 $x = 2k\pi$ 處，前一段給出

$$y = \frac{1}{k}\sin(2k^2\pi) = 0, \qquad y' = \cos(2k^2\pi) = 1,$$

因為 $2k^2\pi$ 是 $2\pi$ 的整數倍。在 $2k\pi < x < 2(k+1)\pi$ 上，方程為 $y'' + (k+1)^2 y = 0$，所以

$$y = A\cos(k+1)x + B\sin(k+1)x.$$

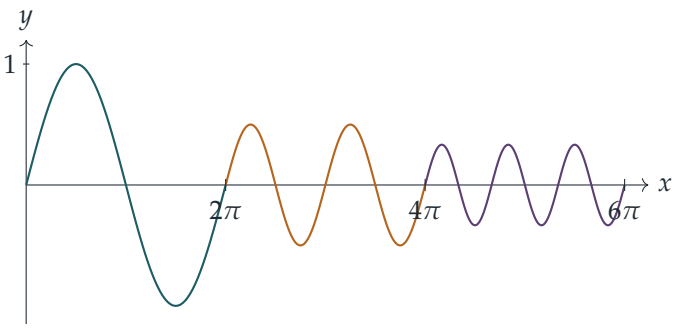
? **想一想** 在左端 $x = 2k\pi$ 處，$\cos(k+1)x$ 和 $\sin(k+1)x$ 的值是多少？

在該處 $(k+1)x = 2k(k+1)\pi$，同樣是 $2\pi$ 的整數倍，所以餘弦值是 1，正弦值是 0。因此由 $y$ 和 $y'$ 在 $x = 2k\pi$ 的連續性可得

$$A = 0, \qquad (k+1)B = 1,$$

所以在 $2k\pi \le x \le 2(k+1)\pi$ 上 $y = \frac{1}{k+1}\sin(k+1)x$。這正是 $n = k+1$ 時的斷言，故由歸納法，斷言對所有 $n \ge 1$ 成立。 □

所以只要 $2(n-1)\pi \le x \le 2n\pi$，就有 $y(x) = \frac{1}{n}\sin nx$。(在每個接合點 $x = 2n\pi$，相鄰兩條公式都給出 $y = 0$，所以並無歧義。)



**第 (iii) 部分**

利用第 (ii) 部分，在第 $n$ 個區間上

$$\int_{2(n-1)\pi}^{2n\pi} y^2\,\mathrm{d}x = \frac{1}{n^2}\int_{2(n-1)\pi}^{2n\pi} \sin^2 nx\,\mathrm{d}x = \frac{1}{2n^2}\int_{2(n-1)\pi}^{2n\pi} (1-\cos 2nx)\,\mathrm{d}x = \frac{1}{2n^2}\cdot 2\pi = \frac{\pi}{n^2},$$

因為 $\sin 2nx$ 在上下限處都等於零（每個積分限都是 $\pi$ 的倍數，而 $2n$ 是整數），所以餘弦項的積分為零。

把首 $N$ 個區間加起來，

$$\int_0^{2N\pi} y^2\,\mathrm{d}x = \pi\sum_{n=1}^{N}\frac{1}{n^2}.$$

這個無窮積分（infinite integral）就是 $\int_0^X y^2\,\mathrm{d}x$ 當 $X\to\infty$ 時的極限。由於 $y^2\ge 0$，它是 $X$ 的遞增函數；而當 $2(N-1)\pi \le X \le 2N\pi$ 時，它介乎在 $2(N-1)\pi$ 與 $2N\pi$ 處的值之間。由於級數收斂（converges），當 $N\to\infty$ 時這兩個值都趨向 $\pi\sum_{n=1}^{\infty}\frac{1}{n^2}$。因此

$$\int_0^{\infty} y^2\,\mathrm{d}x = \pi\sum_{n=1}^{\infty}\frac{1}{n^2},$$

正如題目所求。

## 驗算

- 每一段 $\frac{1}{n}\sin nx$ 都滿足 $y''=-n^2y$，起點處 $y=0$、$y'=1$，終點處亦是 $y=0$、$y'=1$，所以各段在函數值和斜率上都平滑地接合。
- 以數值方法（四階龍格－庫塔法，fourth-order Runge–Kutta）逐步求解這條微分方程（differential equation），走過首六個區間，所得結果與 $\frac{1}{n}\sin nx$ 相差不超過 $10^{-10}$，而 $y^2$ 由起點至 $12\pi$ 的積分算得 4.685336，與 $\pi(1+\frac{1}{4}+\cdots+\frac{1}{36})$ 吻合。
- 利用歐拉（Euler）的結果 $\sum 1/n^2=\pi^2/6$，這個積分等於 $\pi^3/6\approx 5.168$。

## 常見錯誤

- 在每個區間上都把初始條件（initial conditions）用於 $x=0$，而不是用於該區間的左端 $x=2(n-1)\pi$。這樣做在本題碰巧得出正確答案，只因為起點是 $2\pi$ 的倍數，而這一點必須寫明。
- 忘記 $y'=nB\cos nx$：正是因子 $n$ 產生了振幅（amplitude）$\frac{1}{n}$。
- 把積分寫成無窮級數之和，卻沒有說明極限為何存在；$y^2$ 非負，令這一步很容易說明。

## STEP 之外

這是係數為分段常數（piecewise-constant）的微分方程的一個簡單例子，這類方程可用來描述一個勁度在固定時刻切換的彈簧；在每次切換時，即使 $y''$ 出現跳躍，解仍然連續且可微。最後的答案 $\pi^3/6$ 把一條微分方程與巴塞爾和（Basel sum）$\sum 1/n^2 = \pi^2/6$ 聯繫起來；在大學本科的傅立葉級數（Fourier series）課程中，這個和通常最先是用帕塞瓦爾定理（Parseval's theorem）證明的。Boyce 與 DiPrima 的 *Elementary Differential Equations*（初等微分方程教科書）同時涵蓋二階線性方程和傅立葉級數。

詳細解答

# STEP III 1999，第 9 題

力學 • 牛頓殼層定理

**STEP III 1999，第 9 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/99-s3-q9

本題在 *Beyond STEP* 第 10.7 節球殼內的引力中有完整的引導式講解：https://beyondstep.keithmcnulty.org/zh-hant/ch10/shell-theorem.html

## 關鍵思路

一個有大小的物體所產生的引力（gravity），是它各個小部分的引力的向量和（vector sum）。對圓環而言，對稱性（symmetry）已經完成了大部分工作：圓環上位置相對的兩小段，其橫向的拉力互相抵銷，而且每一小段與質點（particle）的距離都相同，所以最後只剩下一個簡單的分量（component）。對球殼（shell）而言，我們把它切成許多圓環，使這些圓環的軸都通過質點，再對每一片應用圓環的結果，然後用積分把各片加起來。只要選對代換（substitution），這個積分便會化為零。

## 解答

### 圓環

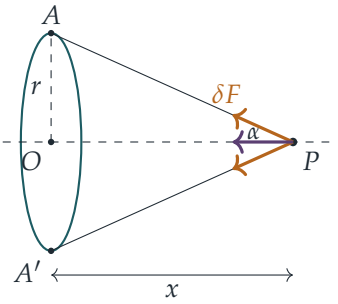
把圓環看成一個質量為 $m$ 的均勻細圓，並把質點看成位於 $P$ 的單位質量的點，它在軸上，與圓心 $O$ 的距離為 $x$。圓環上每一點與 $P$ 的距離都相同，同為

$$d = \sqrt{x^2 + r^2}$$

這可由 $P$、$O$ 和該點所組成的直角三角形，用畢氏定理（Pythagoras' theorem）得出。

*方向*。把圓環上位於 $A$ 的每一小段，與位於直徑另一端 $A'$ 的相等小段配成一對。它們與 $P$ 的距離同為 $d$，所以它們分別沿 $PA$ 和 $PA'$ 施加大小相等的拉力。這兩條直線關於軸互為反射，所以垂直於軸的分量互相抵銷，而沿軸的分量則相加，兩者都由 $P$ 指向

$O$。整個圓環正是由這樣的一對對小段組成，所以質點所受的合力沿着軸，指向圓環的中心。




大小。質量為 $\delta m$ 的一小段，以大小為 $G\,\delta m/d^2$ 的力沿 $PA$ 拉動質點。設 $\alpha$ 為 $PA$ 與軸 $PO$ 之間的夾角，則 $\cos\alpha = x/d$，所以這個力指向 $O$ 的分量為

$$\frac{G\,\delta m}{d^2}\cdot\frac{x}{d} = \frac{Gx\,\delta m}{(x^2+r^2)^{3/2}}.$$

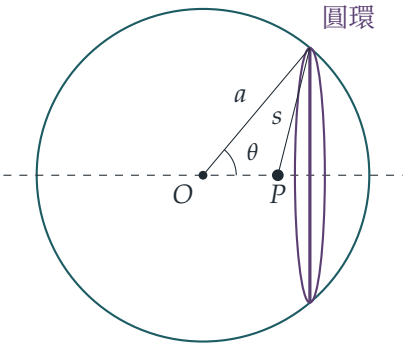
與 $\delta m$ 相乘的因子對每一小段都相同，所以對整個圓環求和，只需把 $\delta m$ 換成 $m$：

$$F = \frac{Gmx}{(x^2+r^2)^{3/2}},$$

方向指向圓心，正如題目所求。如果 $P$ 在圓環的另一側，只要把 $x$ 換成 $|x|$，結果仍然成立；在下一部分，我們會把 $x$ 當作有向距離（signed distance）使用，這時代入帶正負號的 $x$，公式便正確地給出力的有向分量。

**球殼**

設球殼的中心為 $O$，半徑為 $a$，質量為 $M$，則它的面密度（surface density）為 $M/(4\pi a^2)$。把質點放在 $P$，其中 $OP = b < a$。若 $b = 0$，從 $O$ 出發的每個方向都沒有分別，球殼上位置相對的點的拉力互相抵銷，所以力為零。以下取 $0 < b < a$，並以直線 $OP$ 為軸，由 $O$ 起沿指向 $P$ 的方向量度距離。



用垂直於 $OP$ 的平面把球殼切成許多薄圓環。位於極角（polar angle）$\theta$ 與 $\theta+\delta\theta$ 之間的圓環（極角在 $O$ 處由 $OP$ 方向起量度），半徑為 $a\sin\theta$，圓心位於軸上 $a\cos\theta$ 的位置，寬度為 $a\,\delta\theta$。它的質量是

$$\delta M = \frac{M}{4\pi a^2}\cdot 2\pi a\sin\theta\cdot a\,\delta\theta = \tfrac{1}{2}M\sin\theta\,\delta\theta.$$

這個圓環的圓心與 $P$ 的有向距離為 $a\cos\theta-b$。根據圓環的結果，它對 $P$ 的拉力沿着軸，而這個拉力在正方向（背離 $O$、穿過 $P$ 的方向）上的分量為

$$\frac{G\,\delta M\,(a\cos\theta-b)}{\big((a\cos\theta-b)^2+a^2\sin^2\theta\big)^{3/2}} = \frac{G\,\delta M\,(a\cos\theta-b)}{(a^2+b^2-2ab\cos\theta)^{3/2}}.$$

這個正負號的約定在兩種情況下都正確：當 $a\cos\theta>b$ 時，圓環在 $P$ 的外側，拉力指向正方向；當 $a\cos\theta<b$ 時，拉力指向相反方向。每個圓環的拉力都沿着同一條軸，所以總力沿着 $OP$，其分量為

$$F = \frac{GM}{2}\int_0^{\pi}\frac{(a\cos\theta-b)\sin\theta}{(a^2+b^2-2ab\cos\theta)^{3/2}}\,\mathrm{d}\theta.$$

**想一想** 用甚麼代換可以馴服這個分母？

根據餘弦定理（cosine rule），$a^2+b^2-2ab\cos\theta$ 這個量正是 $s$ 的平方，即 $P$ 到圓環的距離的平方（圖中標示為 $s$ 的線段）。因此我們作代換

$$s^2 = a^2+b^2-2ab\cos\theta, \qquad 2s\,\mathrm{d}s = 2ab\sin\theta\,\mathrm{d}\theta.$$

當 $\theta$ 由 0 變到 $\pi$ 時，$s$ 由 $a-b$ 遞增至 $a+b$（由於 $b<a$，前者為正數）。此外，

$$a\cos\theta-b = \frac{a^2+b^2-s^2}{2b}-b = \frac{a^2-b^2-s^2}{2b}.$$

利用 $\sin\theta\,\mathrm{d}\theta = \dfrac{s\,\mathrm{d}s}{ab}$，

$$\begin{aligned} F &= \frac{GM}{2}\int_{a-b}^{a+b}\frac{a^2-b^2-s^2}{2b}\cdot\frac{1}{s^3}\cdot\frac{s}{ab}\,\mathrm{d}s = \frac{GM}{4ab^2}\int_{a-b}^{a+b}\left(\frac{a^2-b^2}{s^2}-1\right)\mathrm{d}s \\ &= \frac{GM}{4ab^2}\left[-\frac{a^2-b^2}{s}-s\right]_{a-b}^{a+b}. \end{aligned}$$

在 $s=a+b$ 時，方括號內的值是 $-(a-b)-(a+b)=-2a$；在 $s=a-b$ 時，它是 $-(a+b)-(a-b)=-2a$。兩者相等，所以

$$F = \frac{GM}{4ab^2}\big(-2a-(-2a)\big) = 0.$$

力沿着 $OP$，而它沿 $OP$ 的分量為零，所以質點不受球殼的任何引力，正如題目所求。

## 驗算

- 圓環：在 $x = 0$ 時力為零（質點位於圓心，各方的拉力互相平衡）；而當 $x \gg r$ 時，力趨向 $Gm/x^2$，就像全部質量都集中在 $O$ 一樣。
- 球殼，質點在外：如果改為 $b > a$，積分下限變成 $s = b - a$，方括號在該處的值是 $(a + b) - (b - a) = 2a$。於是 $F = \frac{GM}{4ab^2}(-4a) = -GM/b^2$：這是一個指向 $O$ 的吸引力，就像全部質量都集中在中心一樣。這正是牛頓（Newton）的另一條球殼定理（shell theorem），可見積分應該沒有算錯。
- 數值驗證：把 2000 個點質量沿圓環排列再求和，結果與公式吻合至小數點後十位。對 $\theta$ 的積分使用辛普森法則（Simpson's rule），在隨機選取的幾個 $b < a$ 下，所得的值約為 $10^{-15}$；另外用蒙地卡羅方法（Monte Carlo）在球殼上隨機取 $2 \times 10^5$ 個點，取 $a = 2$、$b = 1.3$、$GM = 1$，所得的力小於 $10^{-3}$，而典型的量級是 $GM/a^2 = 0.25$。

## 常見錯誤

- 把圓環的質量元素取為 $M\,\delta\theta/\pi$。靠近 $\theta = \pi/2$ 的圓環較大，所以質量與 $\sin\theta\,\delta\theta$ 成正比，而不是與 $\delta\theta$ 成正比。
- 在圓環公式中使用 $|a\cos\theta - b|$。位於 $P$ 兩側的圓環，拉力方向相反，而這種抵銷正是整個問題的關鍵。
- 弄錯 $s$ 的上下限：在 $\theta = 0$ 時，距離是 $a - b$ 而不是 $b - a$，因為 $P$ 在球殼之內。

## STEP 之外

這就是牛頓的球殼定理，而利用引力的高斯定律（Gauss's law），它有一個更簡潔的證明：引力場穿過任何閉曲面的通量（flux），與曲面內的質量成正比，所以根據對稱性，空心球殼內部的引力場必定為零。同樣的數學也可以給出勢（potential），而球殼內部的勢是常數這一事實，正是調和函數（harmonic function）的一個早期例子；調和函數是大學一年級向量微積分（vector calculus）的研習內容。Kleppner 與 Kolenkow 所著的 *An Introduction to Mechanics*（力學入門教科書）對球殼定理和引力勢有細緻的論述。

詳細解答

# STEP III 1999，第 10 題

力學 • 下落鏈條的衝量

**STEP III 1999，第 10 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/99-s3-q10

## 關鍵思路

鏈條塌落期間，仍然懸在空中的部分處於自由落體（free fall）狀態：除了重力之外，沒有任何力拉着它。因此每個鏈環只是從原來的位置直接落下，各鏈環之間唯一的分別就是下落的高度。桌面對每個鏈環各施加一個衝量（impulse），使它停下。把這些衝量加起來，得到一個平方根之和；當 $n$ 很大時，這就是 $\int_0^1 \sqrt{x}\,\mathrm{d}x$ 的黎曼和（Riemann sum）。

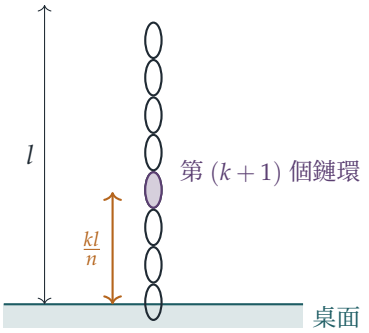
## 解答

### 模型

每個鏈環的質量為 $\frac{m}{n}$，長度為 $\frac{l}{n}$。鏈環是光滑（smooth）的，而鏈條完全柔軟，所以它不能傳遞推力。我們把每個鏈環看作一個質點（particle），它到達桌面時會即時靜止（碰撞屬非彈性（inelastic））。

? **想一想** 下方的鏈環正被桌面截停，為甚麼鏈條懸空的部分仍然是自由落體？

鏈條釋放後，若沒有張力（tension），每個鏈環的加速度都是 $g$；而張力為零的下落鏈條並無矛盾：所有鏈環一起運動，彼此保持接觸而不被拉長。桌面把鏈環向上推，使它停下，但柔軟的鏈條無法把這個推力傳給上方的鏈環。由於鏈環光滑，它們之間也沒有摩擦力。因此，仍在空中的鏈環除了重力之外不受任何其他力的作用，每個鏈環都自由下落，直至着地為止。




**第 $(k+1)$ 個鏈環**

第 $(k+1)$ 個鏈環下方有 $k$ 個鏈環，總長度為 $\frac{kl}{n}$。由於鏈條底端剛好觸及桌面，第 $(k+1)$ 個鏈環要下落 $\frac{kl}{n}$ 的距離才到達桌面。它由靜止開始自由下落，到達時的速率由 $v^2 = 2gs$ 給出：

$$v_k = \sqrt{\frac{2gkl}{n}}.$$

其後它被截停，所以它的動量（momentum）變化的大小為

$$\frac{m}{n}\sqrt{\frac{2gkl}{n}} = \frac{m}{n}\sqrt{2gl}\,\sqrt{\frac{k}{n}},$$

方向向上（鏈環失去了向下的動量 $\frac{m}{n}v_k$）。當 $k = 0$ 時，這個值為零：最底的鏈環本來就已經接觸桌面。

**總衝量**

根據牛頓第三定律（Newton's third law），桌面從第 $(k+1)$ 個鏈環所受的衝量，大小等於該鏈環的動量變化，方向向下。鏈環的編號為 $k+1 = 1, 2, \ldots, n$，所以桌面因碰撞而承受的總衝量為

$$J = \sum_{k=0}^{n-1} \frac{m}{n}\sqrt{\frac{2gkl}{n}} = m\sqrt{2gl}\ \cdot\ \frac{1}{n}\sum_{k=0}^{n-1}\sqrt{\frac{k}{n}}.$$

? **想一想** $\frac{1}{n}\sum_{k=0}^{n-1}\sqrt{k/n}$ 近似於哪一個積分？

把 $[0,1]$ 分成多個長條，這個和就是 $n$ 個闊度為 $\frac{1}{n}$ 的長方形的面積之和，各長方形的高度是 $\sqrt{x}$ 在長條左端點 $x = 0, \frac{1}{n}, \ldots, \frac{n-1}{n}$ 的值。所以當 $n$ 很大時，

$$\frac{1}{n}\sum_{k=0}^{n-1}\sqrt{\frac{k}{n}} \approx \int_0^1 \sqrt{x}\,\mathrm{d}x = \left[\tfrac{2}{3}x^{3/2}\right]_0^1 = \frac{2}{3},$$

因此

$$J \approx \tfrac{2}{3}\,m\sqrt{2gl},$$

這正是所求的結果。

再看看這個近似有多準確：$\sqrt{x}$ 是遞增函數，所以在長條 $[\frac{k}{n}, \frac{k+1}{n}]$ 上，左端點的高度 $\sqrt{k/n}$ 位於曲線之下，而在 $[\frac{k-1}{n}, \frac{k}{n}]$ 上則位於曲線之上。對所有長條求和，得

$$\int_0^{1-\frac{1}{n}} \sqrt{x}\,\mathrm{d}x \;\le\; \frac{1}{n}\sum_{k=0}^{n-1}\sqrt{\frac{k}{n}} \;\le\; \int_0^1 \sqrt{x}\,\mathrm{d}x, \quad \text{即} \quad \tfrac{2}{3}\Big(1-\tfrac{1}{n}\Big)^{3/2} \le \frac{J}{m\sqrt{2gl}} \le \tfrac{2}{3}.$$

上下界都趨向 $\frac{2}{3}$，所以相對誤差（relative error）最多約為 $\frac{3}{2n}$。

## 驗算

- 單位：$m\sqrt{gl}$ 是質量 × 速率，即動量，衝量的單位正應如此。
- 數值上，取 $m = 2$、$l = 3$、$g = 9.8$：$\frac{2}{3}m\sqrt{2gl} \approx 10.224$，而精確的和在 $n = 10$ 時為 9.363，$n = 100$ 時為 10.144，$n = 1000$ 時為 10.216。兩者之比趨近 1，而且始終落在上述的上下界之內。
- 最後着地的鏈環幾乎下落了整個長度 $l$，速率接近 $\sqrt{2gl}$。假如全部質量都以這個速率到達，衝量便是 $m\sqrt{2gl}$。大部分鏈環下落的距離都較短，所以答案是這個值的某個分數，十分合理。

## 常見錯誤

- 以為第 $(k+1)$ 個鏈環下落的距離是 $\frac{(k+1)l}{n}$。它下方其實只有 $k$ 個鏈環。
- 把已經躺在桌面上的那段鏈條的穩定重量也計算在內。題目只問碰撞所產生的衝量。

## STEP 之外

由於所有鏈環在 $\sqrt{2l/g}$ 的時間內全部着地，單是碰撞便對桌面施加了 $\frac{2}{3}m\sqrt{2gl}\big/\sqrt{2l/g} = \frac{2}{3}mg$ 的平均力。事實上，在每一刻，撞擊力都是已落在桌面上那段鏈條的重量的兩倍，所以放在桌下的磅秤，讀數會在最後一個鏈環着地的一刻升至 $3mg$。這類質量連續增加或減少的問題（下落的鏈條、火箭、雨點），要用牛頓第二定律（Newton's second law）的變質量（variable-mass）形式來處理。詳細而嚴謹的論述可參閱 Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教材）。

詳細解答

# STEP III 1999，第 11 題

力學 • 剛體轉動

> **STEP III 1999，第 11 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/99-s3-q11

## 關鍵思路

圓環一旦到達桌邊，便是一個繞固定點轉動的剛體（rigid body），因此兩條方程已足以解決整個問題：能量守恆（conservation of energy）給出任意角度時的角速率（angular speed）；對圓心應用牛頓第二定律（Newton's second law），並沿指向支點的方向分解，則給出法向反作用力（normal reaction）。當這個反作用力降至零時，圓環便失去接觸。第一部分求得的轉動慣量（moment of inertia），正是能量方程所需要的。

## 解答

### 轉動慣量

薄圓環上每一個質點與圓心 $C$ 的距離都是 $a$，所以對於通過 $C$ 並垂直於圓環所在平面的軸，轉動慣量為

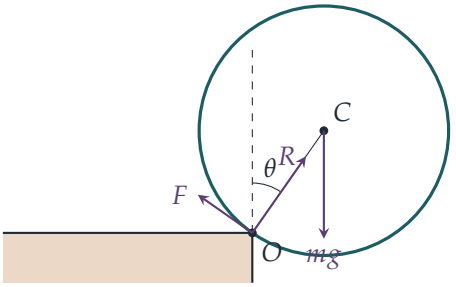
$$I_C = \sum \delta m\, a^2 = ma^2.$$

圓周上的一點 $O$ 與 $C$ 的距離為 $a$，而新的軸與原來的軸平行，故由平行軸定理（parallel axis theorem）得

$$I_O = I_C + ma^2 = 2ma^2.$$

### 滾過桌邊

把圓環看成均勻的剛性圓環，並把桌邊看成一個固定點 $O$。兩個表面都是粗糙的，我們假設摩擦力（friction）足夠大，使圓環始終不打滑：在桌面上它作滾動，越過桌邊時則繞 $O$ 轉動而不滑動。設 $\theta$ 為 $OC$ 與向上鉛垂線所成的角，因此圓環到達桌邊的一刻 $\theta = 0$，並記 $\omega = \dot{\theta}$。




**想一想** 圓環到達桌邊的一瞬間，它繞 $O$ 的角速率是多少？

以速率 $v$ 滾動而不打滑的圓環，角速率為 $v/a$，而它在每一瞬間的運動都是繞接觸點的轉動。在到達桌邊的一瞬間，接觸點正是 $O$，圓環只是繼續繞着它轉動：由於圓環沒有失去接觸，所以不會發生碰撞。因此當 $\theta = 0$ 時 $\omega = v/a$。

**能量**。只要圓環仍與 $O$ 保持接觸，作用於 $O$ 的力便不作功（施力點靜止不動），所以機械能（mechanical energy）守恆。圓心下降了 $a - a\cos\theta$，故

$$\tfrac{1}{2}(2ma^2)\,\omega^2 = \tfrac{1}{2}(2ma^2)\left(\frac{v}{a}\right)^2 + mga(1-\cos\theta), \qquad \text{即} \qquad a\omega^2 = \frac{v^2}{a} + g(1-\cos\theta).$$

**指向支點的方向**。圓心 $C$ 在以 $O$ 為中心、半徑為 $a$ 的圓上運動，所以它的加速度有一個由 $C$ 指向 $O$ 的分量 $a\omega^2$。作用於圓環的力有：重量 $mg$、沿 $OC$ 的法向反作用力 $R$（方向背離 $O$），以及垂直於 $OC$ 的摩擦力 $F$，後者沿 $OC$ 沒有分量。沿 $CO$ 方向分解：

$$mg\cos\theta - R = ma\omega^2.$$

代入能量方程，

$$R = mg\cos\theta - \frac{mv^2}{a} - mg(1-\cos\theta) = mg(2\cos\theta - 1) - \frac{mv^2}{a}.$$

這個量隨 $\theta$ 增加而減少，而當 $R$ 首次降至零時，圓環便與桌邊失去接觸，此時

$$2g\cos\theta = g + \frac{v^2}{a}, \qquad \text{即} \qquad \cos\theta = \frac{1}{2} + \frac{v^2}{2ag},$$

這正是所要證明的結果。

**想一想** 這個結果對任何 $v$ 都合理嗎？

當 $\theta = 0$ 時，$R = m(g - v^2/a)$，只有在 $v^2 < ag$ 時才是正數。這就是圓環滾過桌邊而不會立即失去接觸的條件，也是題目所假設的；它同時正是右方小於 1 的條件，所以存在一個有效的角 $\theta$。

為完整起見，沿垂直於 $OC$ 的方向分解，並利用 $I_O\ddot{\theta} = mga\sin\theta$，可得 $\ddot{\theta} = g\sin\theta/(2a)$ 及 $F = \tfrac{1}{2}mg\sin\theta$。這就是桌邊為防止打滑而必須提供的摩擦力，也說明了為何兩個表面都必須是粗糙的。

## 驗算

- 若 $v \to 0$，圓環只是從桌邊翻倒下去，並在 $\cos\theta = \frac{1}{2}$ 時失去接觸，即 $\theta = 60°$。當 $v^2 \to ag$ 時，$\theta \to 0$：圓環幾乎沒有碰到桌邊。兩個極限情況都合乎情理。
- 數值上，取 $a = 0.6\,\mathrm{m}$、$g = 9.8\,\mathrm{m\,s^{-2}}$ 及 $v = 1.5\,\mathrm{m\,s^{-1}}$，由 $\dot\theta = v/a$ 開始對 $\ddot\theta = g\sin\theta/(2a)$ 逐步計算，直至 $g\cos\theta = a\dot\theta^2$ 為止，得 $\cos\theta \approx 0.69132$，與 $\frac{1}{2} + v^2/(2ag) = 0.69133$ 吻合。

## 常見錯誤

- 在能量方程中使用 $I_C = ma^2$，或在 $\frac{1}{2}I_O\omega^2$ 之上另加一項平移動能（translational kinetic energy）。繞 $O$ 的轉動已經包含了圓心的運動。
- 忘記了初始動能，結果對任何 $v$ 都得出 $\cos\theta = \frac{1}{2}$。

## STEP 之外

從桌邊滾落是典型的*單側約束*（unilateral constraint）：桌邊只能推而不能拉，所以運動方程只在它們所預測的反作用力非負時才適用。質點從球面滑落，以及汽車駛過拱橋，都受同一個想法支配。在大學一年級的課程中，你會在學習剛體動力學（rigid-body dynamics）以及平行軸定理和垂直軸定理（perpendicular axis theorem）時遇到它。Kleppner 與 Kolenkow 的 *An Introduction to Mechanics*（力學入門教科書）對這些課題都有細緻的論述。

詳細解答

# STEP III 1999，第 12 題

概率與統計 • 離散聯合分佈

> **STEP III 1999，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/99-s3-q12

## 關鍵思路

聯合概率（joint probability）對 $j$ 和 $k$ 的依賴，主要是透過 $j+k$ 而來，所以應按總和 $n=j+k$ 把各項分組。餘下的是 $\sum_j \frac{1}{j!(n-j)!}$，它其實是一個喬裝了的二項式和：乘以 $n!$ 之後就變成 $\sum_j \binom{n}{j} = 2^n$。此後每一部分都是指數級數（exponential series）。

## 解答

### 第 (i) 部分

設 $n \geq 1$。事件 $\{X+Y=n\}$ 是事件 $\{X=j,\, Y=n-j\}$（$j=0,1,\ldots,n$）的不相交併集（disjoint union）。在其中每一個事件上都有 $j+k=n$，所以

$$\begin{aligned} \mathrm{P}(X+Y=n) &= \sum_{j=0}^{n} \mathrm{e}^{-1}\frac{n\lambda^n}{j!\,(n-j)!} = \mathrm{e}^{-1}\frac{n\lambda^n}{n!}\sum_{j=0}^{n}\binom{n}{j} \\ &= \mathrm{e}^{-1}\frac{n\lambda^n 2^n}{n!} = \frac{\mathrm{e}^{-1}(2\lambda)^n}{(n-1)!}, \end{aligned}$$

其中用到由二項式定理（binomial theorem）得出的 $\sum_j \binom{n}{j} = (1+1)^n = 2^n$，以及當 $n \geq 1$ 時 $n/n! = 1/(n-1)!$。

為完整起見，由於有因子 $j+k$，$\mathrm{P}(X+Y=0) = \mathrm{P}(X=0,\, Y=0) = 0$。

### 第 (ii) 部分

概率 $\mathrm{P}(X+Y=n)$ 對所有 $n \geq 0$ 求和，結果必須等於 1。$n=0$ 的一項是零，所以

$$1 = \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-1}(2\lambda)^n}{(n-1)!} = \mathrm{e}^{-1}(2\lambda) \sum_{m=0}^{\infty} \frac{(2\lambda)^m}{m!} = \mathrm{e}^{-1}\, 2\lambda\, \mathrm{e}^{2\lambda},$$

這裏我們設 $m = n-1$，並用了指數級數 $\sum_m t^m/m! = \mathrm{e}^t$。因此 $2\lambda\, \mathrm{e}^{2\lambda-1} = 1$，正是所要證明的結果。

### 第 (iii) 部分

設 $g(x) = 2x\, \mathrm{e}^{2x-1}$。由積法則（product rule），

$$g'(x) = 2\mathrm{e}^{2x-1} + 4x\, \mathrm{e}^{2x-1} = 2(1+2x)\, \mathrm{e}^{2x-1}.$$

當 $x > 0$ 時，$1+2x$ 和 $\mathrm{e}^{2x-1}$ 都是正數，所以 $g'(x) > 0$，而 $g$ 在 $x > 0$ 上嚴格遞增（strictly increasing）。

> ? **想一想** 為甚麼「遞增」可以推出「最多只有一個解」？

若 $0 < x_1 < x_2$，則 $g(x_1) < g(x_2)$，所以 $g$ 不可能在兩個不同的點上都取值 1。由於 $\lambda > 0$，第 (ii) 部分得出的方程 $g(\lambda) = 1$ 因此最多只有一個解。

現在試一試那個令指數變成零的明顯候選值 $x = \frac{1}{2}$：$g(\frac{1}{2}) = 1 \cdot \mathrm{e}^0 = 1$。所以 $\frac{1}{2}$ 是一個解，而根據唯一性（uniqueness），它就是唯一的解：

$$\lambda = \tfrac{1}{2}.$$

代入這個值，第 (i) 部分的結果變成：當 $n \geq 1$ 時，$\mathrm{P}(X+Y=n) = \dfrac{\mathrm{e}^{-1}}{(n-1)!}$。

### 第 (iv) 部分

由於 $X+Y$ 只取 $n \geq 1$ 的值，而其概率剛剛已經求出，

$$\mathrm{E}\left(2^{X+Y}\right) = \sum_{n=1}^{\infty} 2^n \frac{\mathrm{e}^{-1}}{(n-1)!} = 2\mathrm{e}^{-1} \sum_{m=0}^{\infty} \frac{2^m}{m!} = 2\mathrm{e}^{-1}\mathrm{e}^2 = 2\mathrm{e},$$

這裏同樣設 $m = n-1$ 並使用指數級數（該級數收斂，所以期望值（expectation）是有限的）。

## 驗算

- 當 $\lambda = \frac{1}{2}$ 時，第 (i) 部分的公式表明 $X+Y-1$ 服從平均值為 1 的泊松分佈（Poisson distribution）。它的概率 $\mathrm{e}^{-1}/m!$ 之和當然是 1，與第 (ii) 部分一致。

- 以數值方法把 $j+k<60$ 範圍內的聯合概率加起來，總和是 1.000000；把 $2^{j+k}\,\mathrm{P}(X=j, Y=k)$ 加起來則得 5.436564，與 $2\mathrm{e}\approx 5.436564$ 吻合。
- 對第 (iv) 部分作一個簡單的合理性檢查：$2^{X+Y}=2\cdot 2^{N}$，其中 $N\sim\mathrm{Po}(1)$，而對泊松變量有 $\mathrm{E}(s^N)=\mathrm{e}^{s-1}$，所以答案是 $2\mathrm{e}^1$。

## 常見錯誤

- 忘記了 $n/n!=1/(n-1)!$，結果在第 (ii) 部分的級數裏留下一個難以處理的因子 $n$。
- 把 $n=0$ 的一項當作非零：因子 $j+k$ 令這一項消失，這正是各個求和都由 $n=1$ 開始的原因。
- 在第 (iii) 部分只證明了 $\lambda=\frac{1}{2}$ 可行。你還需要用單調性（monotonicity）排除其他解。

## STEP 之外

$\mathrm{E}(s^{X+Y})$ 這個量是 $X+Y$ 的*概率母函數*（*probability generating function*），這裏是在 $s=2$ 處取值。母函數把概率的求和化為冪級數（power series），是大學第一門概率課程中用來辨認分佈和求矩（moments）的標準工具。Grimmett and Welsh, *Probability: An Introduction*（一本概率論入門書）對母函數和泊松分佈有清晰的講解。

詳細解答

# STEP III 1999，第 13 題

概率與統計 • 連續先驗分佈的貝葉斯定理

**STEP III 1999，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/99-s3-q13

## 關鍵思路

這裏有兩層隨機性：首先是我那一份的大小 $X$，然後是提子乾落在哪裏。一旦以 $X = x$ 爲條件（condition），每粒提子乾落在我那一份的概率都是 $x$，而且各粒之間互相獨立，所以提子乾的數目服從二項分佈（binomial distribution）。之後，第 (i) 部分用的是全期望定律（law of total expectation），第 (ii) 部分用的是帶有連續先驗分佈（continuous prior）的貝葉斯定理（Bayes' theorem）。

## 解答

首先求 $A$。概率密度函數（density）的積分必須等於 1：

$$\int_0^1 Ax\,\mathrm{d}x = \frac{A}{2} = 1 \quad \Longrightarrow \quad A = 2,$$

所以 $X$ 在 $[0,1]$ 上的密度為 $\mathrm{f}(x) = 2x$。

設 $N$ 為我那一份中提子乾的數目。我們以最自然的方式為「隨機放置」建立模型：每粒提子乾的位置在整個蛋糕內均勻分佈（uniformly distributed），並且與其他提子乾的位置以及我取多少蛋糕都互相獨立。因此，若我取了蛋糕的比例為 $x$，每粒提子乾落在我那一份的概率就是 $x$，各粒之間互相獨立，所以

$$N \mid (X = x) \sim \mathrm{B}(4, x).$$

### 第 (i) 部分

給定 $X = x$，二項分佈的平均值為 $\mathrm{E}(N \mid X = x) = 4x$。對 $X$ 取平均（即全期望定律），

$$\begin{aligned}\mathrm{E}(N) &= \int_0^1 \mathrm{E}(N \mid X = x)\,\mathrm{f}(x)\,\mathrm{d}x = \int_0^1 4x \cdot 2x\,\mathrm{d}x \\ &= 8\int_0^1 x^2\,\mathrm{d}x = \frac{8}{3}.\end{aligned}$$

等價地，$\mathrm{E}(N) = 4\,\mathrm{E}(X)$，其中 $\mathrm{E}(X) = \frac{2}{3}$。我那一份中提子乾數目的期望值是 $\frac{8}{3}$。

### 第 (ii) 部分

我們要求 $\mathrm{P}(X > \frac{1}{2} \mid N = 4)$。給定 $X = x$，四粒提子乾全部落在我那一份的概率是 $x^4$。

> **想一想** 當 $X$ 是連續的，而 $\mathrm{P}(X = x) = 0$ 時，我們怎樣運用貝葉斯定理？

我們把貝葉斯定理中的求和換成對密度的積分。在各種可能的份量大小之中，一小段 $[x, x + \mathrm{d}x]$ 的概率是 $2x\,\mathrm{d}x$，而在這一小段之下，四粒提子乾全中的機會是 $x^4$。把各小段加起來，

$$\mathrm{P}(N = 4) = \int_0^1 x^4 \cdot 2x\,\mathrm{d}x = 2\int_0^1 x^5\,\mathrm{d}x = \frac{2}{6} = \frac{1}{3}.$$

同樣地，只保留 $x > \frac{1}{2}$ 的那些小段，

$$\mathrm{P}\left(X > \tfrac{1}{2} \text{ 及 } N = 4\right) = \int_{1/2}^1 2x^5\,\mathrm{d}x = \frac{1}{3}\left(1 - \frac{1}{64}\right) = \frac{1}{3} \cdot \frac{63}{64}.$$

因此

$$\mathrm{P}\left(X > \tfrac{1}{2} \mid N = 4\right) = \frac{\frac{1}{3} \cdot \frac{63}{64}}{\frac{1}{3}} = \frac{63}{64}.$$

另一種看法：給定 $N = 4$，$X$ 的條件密度與 $x^4 \cdot 2x$ 成正比，所以它在 $[0, 1]$ 上是 $6x^5$。它的分佈函數（distribution function）是 $x^6$，而 $1 - (\frac{1}{2})^6 = \frac{63}{64}$。

## 驗算

- (ii) 的答案非常接近 1，這是合理的：找到全部提子乾，正是我取了一大塊的有力證據。先驗概率 $\mathrm{P}(X > \frac{1}{2}) = 1 - \frac{1}{4} = \frac{3}{4}$ 被推高至 $\frac{63}{64} \approx 0.984$。
- $\frac{8}{3}$ 介乎 0 與 4 之間，而且大於 2，這正是應有的結果：密度 $2x$ 偏向較大的份量。
- 以 $10^6$ 次試驗進行蒙地卡羅模擬（Monte Carlo simulation）（取 $X = \sqrt{U}$，其中 $U$ 服從均勻分佈，這樣得出的密度就是 $2x$，然後放置四粒提子乾），得到 $\mathrm{E}(N) \approx 2.667$、$\mathrm{P}(N = 4) \approx 0.333$，以及條件概率 $\approx 0.9846$，對比 $\frac{63}{64} = 0.984375$。

## 常見錯誤

- 忘記求 $A$，或者取 $A = 1$。這樣 f 就不是密度函數，(i) 的期望值也會算錯（雖然 $A$ 在 (ii) 中會約去）。
- 計算了 $\mathrm{P}(N = 4 \mid X > \frac{1}{2})$，而不是 $\mathrm{P}(X > \frac{1}{2} \mid N = 4)$。條件的方向弄反了。
- 以 $(\frac{1}{2})^4$ 代替積分：份量的大小是隨機的，所以我們必須把 $x^4$ 對密度取平均，而不是代入單一個數值。

## STEP 之外

這是貝葉斯推論（Bayesian inference）的縮影：密度 $2x$ 是未知比例的先驗分佈（prior），二項分佈是似然函數（likelihood），而 $6x^5$ 是後驗分佈（posterior）。先驗分佈與後驗分佈都是貝塔分佈（beta distribution），這是二項數據的共軛先驗（conjugate prior）的一個例子，而拉普拉斯（Laplace）的繼起法則（rule of succession）背後也是同一個計算。Blitzstein 與 Hwang 所著的 *Introduction to Probability*（概率論入門）把貝塔–二項分佈的來龍去脈，以及以連續隨機變量為條件的處理方法，講解得十分清楚。

詳細解答

# STEP III 1999，第 14 題

概率與統計 • 以條件化求期望值

**STEP III 1999，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/99-s3-q14

## 關鍵思路

在 H1 中，獎金只取決於首次成功在何時出現，因此期望值（expectation）是一個有限的等差等比級數（arithmetico-geometric series）之和 $\sum r\,p\,q^{r-1}$。處理這類和式的標準技巧，是把和式減去它自身的 $q$ 倍。在 H2 中，我們以回合數 $N$ 為條件，再把 H1 的答案按泊松分佈（Poisson distribution）取平均。這只需用到泊松和式 $\sum \lambda^n x^n/n!$ 與 $\sum n\lambda^n x^n/n!$，兩者都可由指數級數（exponential series）得出。

## 解答

### 遊戲 H1

設 $W$ 為她的獎金（以英鎊計）。各回合互相獨立，每回合成功的概率為 $p$。她首次成功出現在第 $r$ 回合，當且僅當首 $r-1$ 回合均失敗而第 $r$ 回合成功，所以

$$\mathrm{P}(W=r)=q^{r-1}p \qquad (1\le r\le n),$$

其餘情況下 $W=0$，對期望值沒有貢獻。因此

$$\mathrm{E}(W)=p\,S, \qquad \text{其中} S=\sum_{r=1}^{n} r\,q^{r-1}=1+2q+3q^2+\cdots+nq^{n-1}.$$

乘以 $q$ 後相減，並把 $q$ 的同次冪對齊：

$$\begin{aligned} S-qS &= 1+q+q^2+\cdots+q^{n-1}-nq^n \\ &= \frac{1-q^n}{1-q}-nq^n, \end{aligned}$$

這裏用了等比級數（geometric series）的求和公式；由於 $q \neq 1$，公式成立。因為 $1-q=p$，由此得 $S=\dfrac{1-q^n}{p^2}-\dfrac{nq^n}{p}$，於是

$$\begin{aligned}\mathrm{E}(W)=pS&=\frac{1-q^n-npq^n}{p}=\frac{1-q^n-n(1-q)q^n}{p}\\&=p^{-1}\left[1+nq^{n+1}-(n+1)q^n\right],\end{aligned}$$

這正是所要求的結果。

**遊戲 H2**

設 $N\sim\mathrm{Po}(\lambda)$ 為她獲准進行的回合數，則 $\mathrm{P}(N=n)=\mathrm{e}^{-\lambda}\lambda^n/n!$。記

$$g(n)=p^{-1}\left[1+nq^{n+1}-(n+1)q^n\right].$$

已知 $N=n\geq 1$ 時，她玩的是有 $n$ 個回合的 H1，所以獎金的期望值為 $g(n)$。已知 $N=0$ 時，她一無所獲。根據全期望定律（law of total expectation，即以 $N$ 為條件），

$$\mathrm{E}(W)=\sum_{n=1}^{\infty}\mathrm{P}(N=n)\,g(n).$$

**想一想** $g(n)$ 的三個部分若寫成由 $n=0$ 開始的完整泊松和式，會較易處理。我們可以讓和式由 $n=0$ 開始嗎？

可以：$g(0)=p^{-1}[1+0-1]=0$，與 $N=0$ 時獎金為零相符。所以

$$\mathrm{E}(W)=\frac{1}{p}\sum_{n=0}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}\left[1+nq^{n+1}-nq^n-q^n\right].$$

第一項之和為 1，因為各概率之和為 1。至於其餘各項，我們需要兩個級數，兩者對所有實數 $x$ 都成立：

$$\begin{aligned}\sum_{n=0}^{\infty}\mathrm{e}^{-\lambda}\frac{(\lambda x)^n}{n!}&=\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda x},\\\sum_{n=0}^{\infty}n\,\mathrm{e}^{-\lambda}\frac{(\lambda x)^n}{n!}&=\lambda x\sum_{n=1}^{\infty}\mathrm{e}^{-\lambda}\frac{(\lambda x)^{n-1}}{(n-1)!}=\lambda x\,\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda x}.\end{aligned}$$

取 $x=q$，並留意 $\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda q}=\mathrm{e}^{-\lambda(1-q)}=\mathrm{e}^{-\lambda p}$。於是

$$\sum_{n=0}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}\,q^n=\mathrm{e}^{-\lambda p},\qquad\sum_{n=0}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}\,nq^n=\lambda q\,\mathrm{e}^{-\lambda p}.$$

$nq^{n+1}$ 這一項是 $nq^n$ 的 $q$ 倍，所以它的貢獻是 $\lambda q^2 \mathrm{e}^{-\lambda p}$。綜合起來，

$$\begin{aligned}\mathrm{E}(W) &= \frac{1}{p}\left[1 + \lambda q^2 \mathrm{e}^{-\lambda p} - \lambda q\,\mathrm{e}^{-\lambda p} - \mathrm{e}^{-\lambda p}\right] \\ &= \frac{1}{p}\left(1 - \mathrm{e}^{-\lambda p}\right) + \frac{\lambda q(q-1)}{p}\,\mathrm{e}^{-\lambda p} \\ &= \frac{1}{p}\left(1 - \mathrm{e}^{-\lambda p}\right) - \lambda q\,\mathrm{e}^{-\lambda p},\end{aligned}$$

最後一步用了 $q - 1 = -p$。這就是所要求的表達式，以英鎊計。

## 驗算

- H1 取 $n = 1$：公式給出 $p^{-1}[1 + q^2 - 2q] = p^{-1}(1-q)^2 = p$。這是正確的，因為她以概率 $p$ 贏得 £1。
- 在 H1 中當 $n \to \infty$ 時，$q^n \to 0$ 且 $nq^n \to 0$，所以期望值趨向 $1/p$，即幾何分佈（geometric distribution）的平均值。同樣地，當 $\lambda \to \infty$ 時，H2 的結果也趨向 $1/p$。
- 當 $\lambda$ 很小時，把指數函數展開，得 $\lambda - \tfrac{1}{2}\lambda^2 p - \lambda q + O(\lambda^2) = \lambda p + O(\lambda^2)$。這合乎情理：她幾乎總是沒有回合或只有一個回合，而 $\mathrm{P}(N = 1) \approx \lambda$。
- 數值方面，對 $n \le 14$，H1 的公式與直接求和的結果完全吻合（以分數計算）。H2 的公式與截斷後的泊松和式吻合；另外，取 $\lambda = 2$、$p = 0.4$ 模擬 $4 \times 10^5$ 局遊戲，得 0.836，而公式給出的是 0.837。

## 常見錯誤

- 弄錯 $S - qS$ 的最後一項：和式 $S$ 止於 $nq^{n-1}$，所以 $qS$ 會剩下 $-nq^n$。
- 忘記 $\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda q}$ 可化簡為 $\mathrm{e}^{-\lambda p}$，結果在代數運算中迷失方向。

## STEP 之外

H2 中用到的和式 $\sum \mathrm{P}(N = n)\,x^n$，其實是**概率生成函數**（probability generating function）$G_N(x) = \mathrm{E}(x^N)$ 的值，而 $\sum n\,\mathrm{P}(N = n)\,x^n = xG_N'(x)$。對泊松變量而言，$G_N(x) = \mathrm{e}^{\lambda(x-1)}$。生成函數，加上以隨機的試驗次數為條件這一手法，是大學一年級概率論的核心工具，例如用於研究隨機和（random sums）與分支過程（branching processes）。Grimmett 與 Welsh 合著的 *Probability: An Introduction*（概率論入門）對兩者都有清晰的講解。

# 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

**1999 年詳細解答.** 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。